MONITORING & EVALUATION (M&E) PLAN FOR NURU NIGERIA

RAY MARSHALL CENTER FOR THE STUDY OF HUMAN RESOURCES

Photo credits: Cover page photo by maxime niyomwungeri on Unsplash

MONITORING & EVALUATION (M&E) PLAN FOR NURU NIGERIA

NOVEMBER 2018

RMC Ray Marshall Center for the Study of Human Resources
TEXAS LBJ School
The University of Texas at Austin
Lyndon B. Johnson School of Public Affairs

3001 Lake Austin Blvd., Suite 3.200
Austin, TX 78703 (512) 471-7891
www.raymarshallcenter.org

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

1. OVERVIEW

Nuru International's mission is to eradicate extreme poverty in fragile rural areas to build communities resilient to violent extremism. Nuru International considers fragile states to continue to be a source of instability and relative deprivation in the world. Vulnerabilities in marginalized communities are ripe for exploitation by violent extremist groups and ideologies. Nuru International envisions a world in which all people live in an enabled environment with lasting, meaningful choices. Free of the burdens of vulnerabilities that threaten the stability and resilience of households and communities, people will be able to thrive and to exercise their agency.

Nuru International believes that building resilient communities is best done by local leaders who produce and manage programs locally and nationally, combining and leveraging their knowledge, innovation and experience with international and expatriate expertise, support and resources. Nuru International is committed to eliminating extreme poverty through investments in self-sustaining and scalable interventions.

Nuru's integrated development model takes a holistic approach to address the most prevalent and fundamental challenges of extreme poverty through an intentional and co-creative design process. By addressing vulnerabilities in an integrated way, Nuru enables individuals not just to improve their own well-being and climb out of the poverty trap, but also to avoid falling back into poverty. Without a holistic approach to development, well-intentioned reforms and investment in one sector risk being squandered because they are not supported by measures in other sectors.

The Ray Marshall Center (RMC), an organized research unit in the LBJ School of Public Affairs at The University of Texas, is providing technical assistance to support Nuru's monitoring and evaluation (M&E) efforts. The RMC's experience and expertise supports Nuru's work by demonstrating the effectiveness and robustness of its integrated approach to addressing poverty.

The RMC has been at the forefront of pro-poor education, workforce, and social policy research since its creation in 1970. The RMC identifies and fosters creative solutions to challenging problems through a variety of applied research activities, including policy research analysis, state-level consulting on capacity development, program monitoring and evaluation, and impact evaluations, using both quantitative and qualitative methods. The RMC research staff is composed of a core of PhD and Master's level researchers with highly competitive quantitative and qualitative research and evaluation skills, from a wide range of disciplines including international development, economics, education, social policy, public affairs, public health, sociology and geography.

Context

The next phase of Nuru's work will take place in Michika LGA, Adamawa State, Nigeria. Michika is a Local Government Area of Adamawa State, Nigeria, located in eastern Nigeria and bordered on the east by the Republic of Cameroon. Michika is made of sixteen wards. The principal tribe and language in Michika is the Kamwe. The Boko Haram terrorist group seized Michika in September 2014 in its efforts to create an Islamic state. Michika was recaptured by the Nigerian military in January 2015. However, locals who fled the area returned to find burned-out houses, blown-up bridges, ransacked banks, ruined schools and hospitals, and desecrated Christian churches and cemeteries (Freeman 2015). Michika is also struggling with the challenge of Boko Haram fighters or sympathizers in the local population.

With a mission based in eradicating extreme poverty, Nuru is approaching the Nigeria country project with an open mind towards the types of outcomes that will be needed in this new context. Nuru's monitoring and evaluation work has involved the application of customized poverty measurement systems, implementation of the M-PAT and the MPI, as well as variations on measuring program outcomes and local leadership sustainability. In this next phase, Nuru is addressing the necessities that arise in a fragile state context, particularly with regard to creating and fostering resilience to shocks in conflict-affected communities.

In any M&E plan, monitoring refers to the systematic collection of data on specified indicators to assess the extent of progress and achievement of objectives (OECD-DAC 2010). Monitoring also provides critical information that informs decisions about a program as it is being implemented (Corlazzoli and White 2013). Evaluation is the systematic and objective assessment of a program, including its design, implementation and results to determine the program's efficiency, effectiveness, impact and sustainability (OECD-DAC 2010). Both components contribute to improve decision making and management by keeping the program on track towards achieving its goals and by integrating lessons learnt into planning.

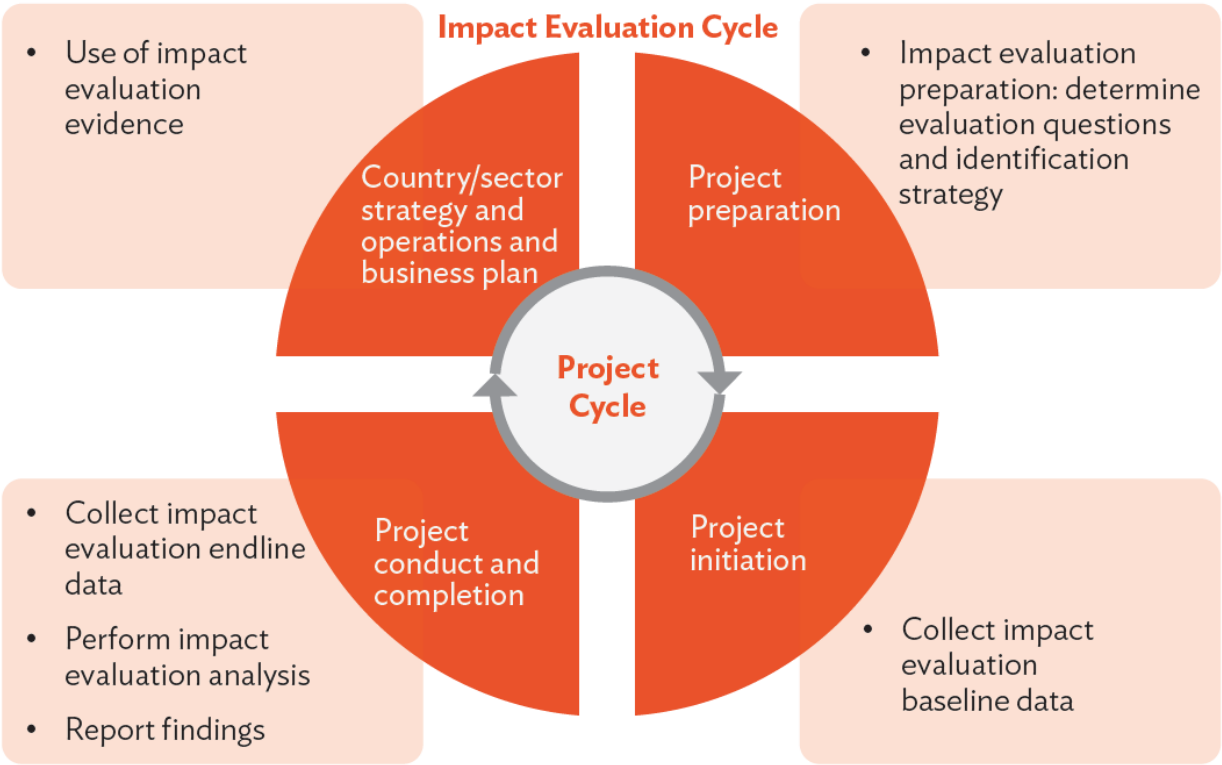
Figure 1 below illustrates the desired relationship between evaluation and project design and implementation, and is one to which the RMC and Nuru will adhere.

Organization of the M&E Plan

The remainder of this document proceeds as follows: Section 2 provides an introduction to the evaluation including the conceptual framework for the program design as well as research questions. Section 3 describes the program design guiding Nuru's programming in Nigeria, including the theory of change. Section 4 provides a brief overview of NN's programming as it relates to its livelihoods, social services, and leadership sustainability interventions. Section 4 describes the evaluation design in detail, including the qualitative and quantitative data collection, sampling design, data collection instruments, data analysis, timeline, work plan and

deliverables. Appendix 1 contains a list of the indicators while Appendix 2 describes the methods guiding the quantitative evaluation component.

Figure 1. Impact Evaluation and the Project Cycle



Source: Asian Development Bank, "Impact Evaluation Development Interventions Guide." Program Strategies and Activities

2. INTRODUCTION

In August 2018, Nuru revised its mission statement for NN to read:

> "*Eradicate extreme poverty in fragile rural areas to build communities resilient to violent extremism*."

This revision makes clear the assumption of a causal link between extreme poverty, as interpreted in terms of vulnerabilities, to building resilience to violent extremism, and, as such, establishes parameters within which NN's program design should operate. And, this assumption builds on Nuru's prior and ongoing work in Kenya and Ethiopia, during which program feasibility, as it relates to poverty reduction, has been proven. As noted above, however, applying tested poverty reduction strategies to reduce vulnerability in VE contexts is uncharted territory for Nuru. Given this, and given that NN's outcomes have the potential to not only directly benefit the individuals it serves, but also to inform the larger, global debate about how best to combat VE, it is essential that this evaluation goes beyond more common M&E practices and focuses on measuring impact in a methodologically sound and convincing manner.

Conceptual framework for the program design

Target population

Extremely poor households within marginalized communities with a high level of advocacy for extreme change will be the target population for Nuru's intervention. Nuru will neither specifically target nor program for the actual supporters of VEOs or the perpetrators of violence for VEOs. The goal will be to address the vulnerabilities that allow VEOs to imbed themselves in communities rather than address the complexity associated with what triggers individual radicalization. Through stabilizing communities and insulating them from VEOs, Nuru aims to reduce the overall ability of VEOs to garner both local supporters and new recruits from local communities.

Nuru's Impact programs will be driven by locally identified needs and vulnerabilities of the target population. Programming will use farmer organizations as the vehicle in communities to organize the population, reduce transaction costs for service delivery, and establish feedback loops. Simplifying programming under the two themes of Livelihoods and Social Services will aim to more deeply and meaningfully integrate interventions while reducing complexity. Feedback loops will allow for accountability of progress by both parties (FOs and Nuru) and for greater ownership of impact.

Partnerships

Partnerships will be crucial in creating the enabling environment through which Nuru's work will be sustained. These partnerships will fall into two key areas: Scaling Partnerships and Stabilizing Partnerships. Government and large International NGOs will be the key Scaling Partners to replicate Nuru's model regionally. Stabilizing Partners will build on Nuru's foothold in communities to deepen and widen impact.

Women's empowerment and gender mainstreaming

In addition, women's empowerment and gender mainstreaming will underpin Nuru's programming in Nigeria. Nuru believes that by unlocking the potential of both women and men within households, poverty solutions will be more sustainable and equitable. To this end, Nuru is committed to:

- Fostering an understanding of how gender affects programming
- Building gender equitable programming
- Analyzing gender disaggregated data for equitable data driven decision making
- Understanding gender roles in conflict resolution, peacebuilding, and countering violent extremism
- Creating gender equitable policies within the organization
- Doing no harm in the communities Nuru serves

Rigorous monitoring of gender-related metrics will be a central part of this evaluation. Collection and analysis of data on gender outcomes is essential to understanding more about the relationship gender has to poverty, resilience, and violent extremism.[1]

Conceptual framework for the evaluation

We have conceptualized the evaluation as a mixed-methods evaluation, comprising a qualitative component to provide context for findings from the quantitative component. With this in mind, we propose the following analytical plan:

- RCT, with wards assigned to treatment or control, and HH randomly selected from within control group wards to be matched with self-selected participants from the treatment wards (see "Evaluation Design" section below for additional detail)
- Qualitative components, including focus groups (FGD), key informant interviews (KII), and direct observation at beginning, mid-, and end-points

[1] Women Deliver. Improve Data and Accountability for Girls and Women. http://womendeliver.org/wp-content/uploads/2017/09/D4G_Infographic_11.pdf

- Outcomes analysis, using difference-in-differences coupled with inverse probability weighting (a propensity-based method) to measure differences between treatment and comparison groups
- Panel data collection over the duration of the project in service of either fixed or random effects panel data analysis to determine the effects that the interventions taken individually and as integrated have on (1) vulnerabilities, and (2) capacities, with indices created for each dependent variable.

Rigor in any impact evaluation depends on how clearly the counterfactual is defined. For this reason, we devote significant space below to outlining our process for selecting a comparison group that will mostly closely match participant households. As with the interventions themselves, the approach we take is innovative in that it is an amalgam of methods that, taken together, will mostly clearly measure the near-, intermediate-, and long-term impacts from NN.

Research questions

The key research question motivating this evaluation, and reflective of the TOC, is:

> ***Can and to what extent does Nuru programming build community resilience, by alleviating the sense of injustice and fostering social inclusion?***

Key assumptions underlying the study are:

Assumption 1: Building community resilience to shocks also makes communities more insulated from shocks and stressors related to violent extremism. This removes several of the potential vulnerabilities in these communities (necessary but not sufficient to completely insulate them from VE threat).

Assumption 2: Alleviating the sense of injustice and fostering social inclusion better link households and communities to local government through various pathways (more active citizenry, bigger tax base, etc.), and local government therefore has a vested interest in protecting the people.

3. NN PROGRAM DESIGN

Theory of change

Nuru's theory of change accurately reflects both the scope and complexity of its Nigeria project, with the key discriminating factor from its Kenya and Ethiopia programs being the inclusion of an anticipated impact on violent extremism resulting from its poverty eradication projects. As it stands, Nuru's TOC is comprehensive, builds on its previous work and lessons from Kenya and Ethiopia, and is rooted in the relevant TOC literature. Nuru's end game is to make populations less vulnerable to violent extremist organizations by strengthening, economically and socially, the areas that those groups are exploiting. By eradicating extreme poverty and restoring agency, communities will no longer suffer these particular vulnerabilities to VEOs. The TOC informs the mechanics by which this will work, and the evaluation will, in turn, inform programming and ultimately success.

A useful exercise (for the planning phase of the project), therefore, would be for Nuru's program and M&E teams to consider the assumptions that connect activities to outputs to outcomes. Questions to consider in this exercise include:

- What is required in order for an anticipated output or outcome to follow from any given activity?
- Are there conditions, contexts, norms that are being taken for granted in linking particular activities to outcomes?
- Are there political, economic, or organizational contexts that either need to be in place or need to be avoided in order for the TOC to function as a solid schema?
- Is there a sound understanding of what motivates individuals to behave in way expected/hoped for?

For example, an objective of NN is to establish farmers' organizations that are self-financing and are able to address future vulnerabilities. However, there are several implicit assumptions between intervention activities and this particular outcome, including:

- capabilities (accounting, management, business development, etc.) exist among participants for investing FO profits into new business lines and poverty-reduction strategies;
- participants will naturally look to FOs, as opposed to other actors, for the type of leadership required to address future vulnerabilities;
- a plan exists for addressing potential competition for the FOs arising elsewhere in the wards; or
- FOs will be sufficient to provide the social cohesion required to reduce vulnerabilities and, ultimately, counter violent extremism.

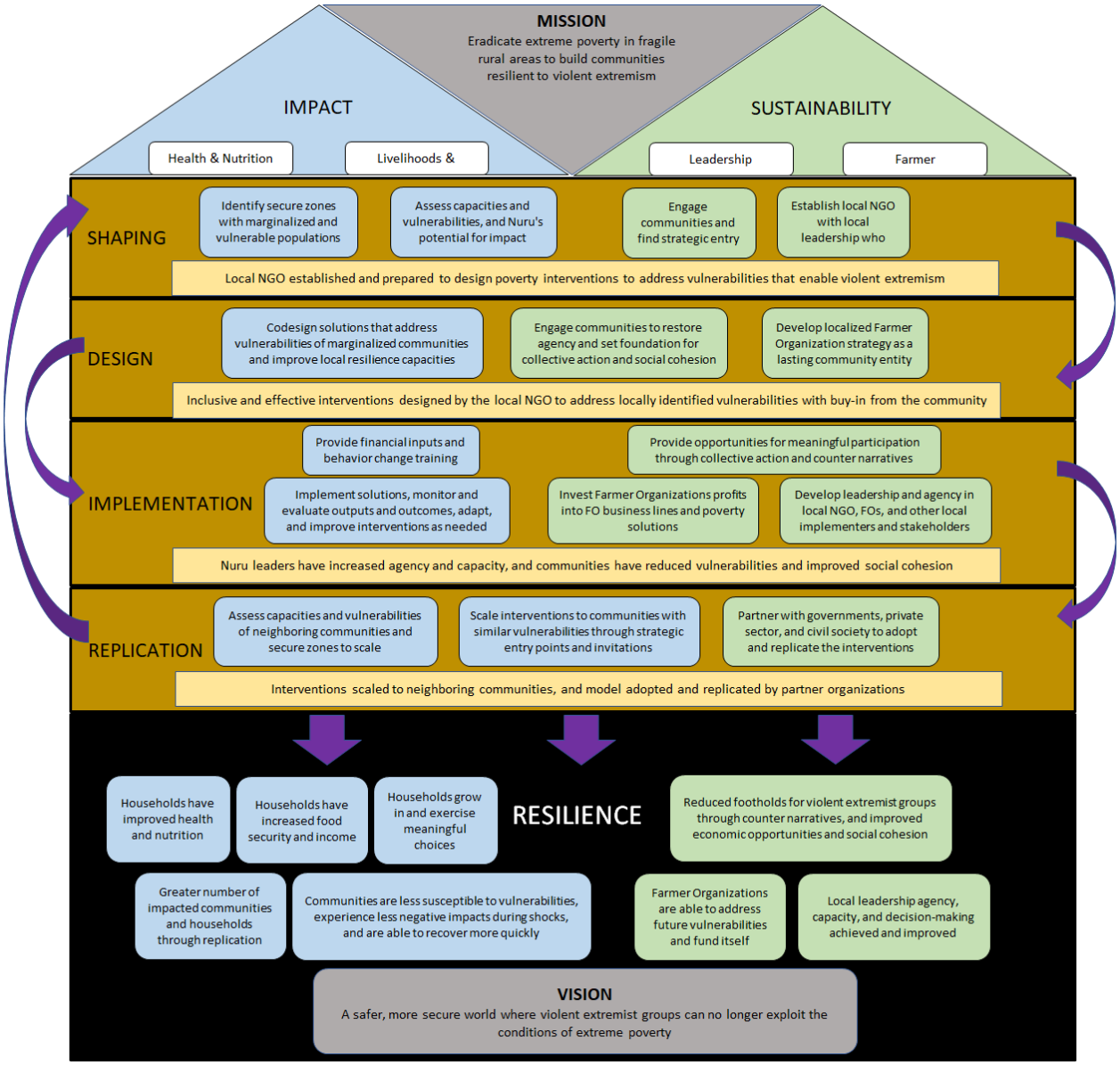
A similar exercise should be conducted related to each of the other seven objectives presented in the current TOC.

In addition, any number of exogenous factors, beyond the intervention, could influence CVE-related outcomes. It is a useful exercise to identify the exogenous factors at the outset of creating the TOC in order to delineate between, and so bear in mind, activities over which the program has some control and those over which it has less or no control. For the purposes of the NN evaluation, these exogenous factors would include, among others:

- political change at local or national level;
- armed conflict in the area; and
- drought

We are not making the argument that, unless each of these assumptions can be answered in the affirmative, the intervention will fail. Rather, we present these as examples of assumptions that will need to be examined and tested in order to help ensure that activities lead to expected outcomes and, therefore, a successful intervention.

Figure 2. Theory of change



Defining terms

An essential first step in the preparatory phase of the project is to agree to a definition of terms, "resilience" and "violent extremism" particularly. This is important for a number of reasons, not least of which is that how these concepts will be measured depends largely on how they are defined at the outset. Prior evaluations of Nuru's multidimensional poverty impacts suffered from a lack of alignment between Nuru's individual interventions, whether related to livelihoods or social services, and the metrics used to measure the integrated effect that these interventions had on individual households. This is provided as a cautionary tale that should drive toward early consensus on what, precisely, Nuru aims to achieve through its Nigeria program.

Defining "resilience"

Numerous definitions of resilience are present in the literature, each with slightly different emphases. Mercy Corps' Strategic Resilience Assessment (STRESS) offers useful guidance for settling on several key definitional questions:

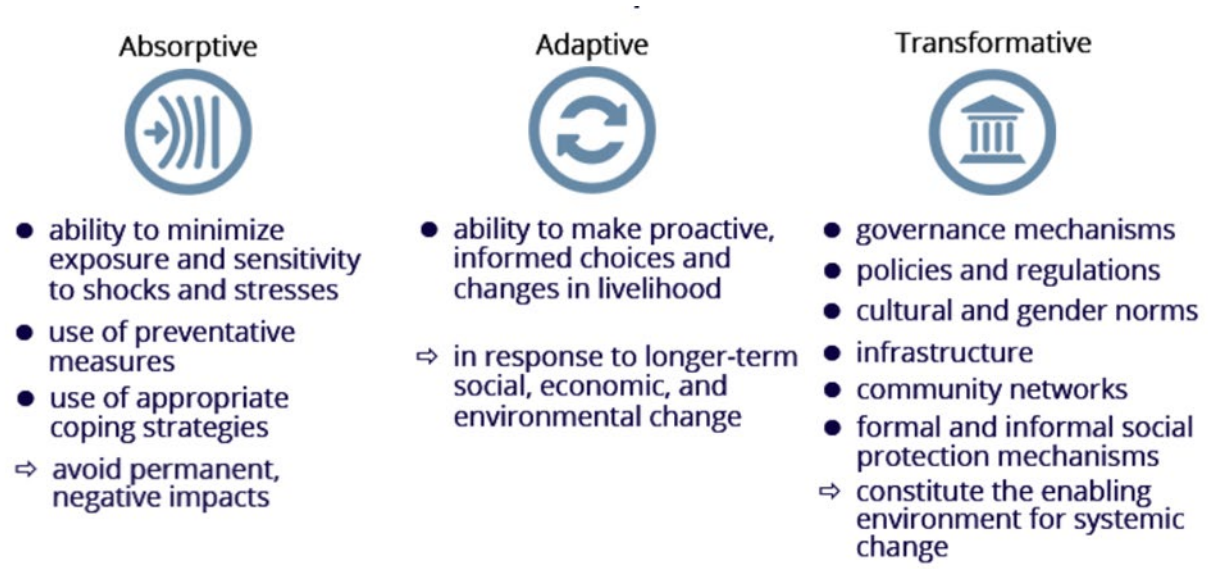
Resilience of what: refers to both the target geography and the elements of key systems within that geography that relate to the development trends of interest (in NN's case, social systems and economic systems);

Resilience for whom: refers to the need for clarity about not only the populations to be served, but also how differences within these populations may require differing interventions, and require bringing differing actors to bear if resilience is to be attained at a community level;

Resilience to what: refers to the need to prioritize among the several different types of shocks and stresses that need to be addressed in order to build resiliency, in addition to gaining a contextual understanding of the trends surrounding these shocks and stresses, and how they are currently managed (Mercy Corps also notes that teams should be able to distinguish between shocks and stresses on the one hand, and more common development constraints on the other hand); and

Resilience through what: refers to, generally, the capability to manage shocks and stresses over time. With regards to "resilience through what", Mercy Corps also recommends conceptualizing resilience subdivided into three dimensions—absorptive, adaptive, and transformative (**see Appendix 1 for resilience indicators).**

Figure 3. Resilience Capacities



Defining "violent extremism"

There is less disagreement in the literature as to how CVE is defined when compared to "resilience," but substantial variation in terms of identifying what works. Identifying what works is beyond the remit of this evaluation plan, but evaluation of outcomes will depend, as with resilience, on a clear definition (so, perhaps, in a sense defining the term does point toward certain strategies). Given our understanding of Nuru's objectives in Nigeria, we recommend the definition offered by USAID: "encouraging, condoning, justifying, or supporting the commission of a violent act to achieve political, ideological, religious, social, or economic goals." It follows, then, that countering these actions would involve a set of activities designed to weaken support for them.

Nuru's TOC identifies multiple anticipated strategies with regards to CVE, including building and implementing a counter-narrative to CVE, reducing marginalization, addressing economic drivers, and supporting change agents in the form of farmers cooperatives. Each of these strategies could, arguably require its own TOC, and will certainly involve different metrics needed to measure impact.

Given Nuru's approach to improving well-being, however, we understand that primus inter pares in this list would be addressing economic drivers of VE. In their toolkit, USAID identifies economic drivers of VE as one of several contributing factors, and it identifies several assumptions that likely underpin most attempts to CVE through addressing economic need (of which NN is one), including:

- Frustration with lack of economic opportunity and/or relative economic deprivation is

driving support for violent extremism;

- Violent extremist groups are using financial incentives to entice recruits;
- Improved economic conditions at the individual and/or community level will reduce frustration and/or limit attractiveness of VE groups' financial incentives; and
- It is possible for donor interventions to improve economic conditions in the short term.

Defining roles for farmer organizations

Farmers' organizations will play a central role in NN. By promoting member-owned, member-run and member-benefiting businesses, NN seeks to introduce a means for restoring agency to the community, providing information feedback loops to Nuru, and assisting in promoting counter-narratives to VEOs. SCOPEinsight will be used to assess the performance of farmer cooperatives and to benchmark their relative performance against regional, national and international standards.

4. EVALUATION DESIGN

In her classic guidance, Judy Baker (World Bank, 2000) outlines the essential steps in designing a poverty impact evaluation in a developing country setting. Baker notes that most successful impact evaluations in these settings are composed of ten common elements:

1. Determining whether or not to carry out an evaluation
2. Clarifying objectives of the evaluation
3. Exploring data availability
4. Designing the evaluation
5. Forming the evaluation team
6. If data will be collected:
    a. Sample design and selection
    b. Data collection instrument development
    c. Staffing and training fieldwork personnel
    d. Pilot testing
    e. Data collection
    f. Data management and access
7. Ongoing data collection
8. Analyzing the data
9. Writing up the findings and discussing them with policymakers and other stakeholders
10. Incorporating the findings in project design

Work to date on the M&E plan by Nuru has effectively addressed the first three items in this list, as outlined above, although additional work may be required regarding the second item (we assume that the objectives of the evaluation will be further defined in the upcoming project planning meetings[2]). A few questions that may help guide the discussion of the evaluation's use-value are:

- Will the information from an impact evaluation strengthen the program or policy being tested?
- Will the information from an impact evaluation identify activities for scale-up or transfer?

[2] Finalizing the evaluation design should occur, to the extent possible, concurrently with finalizing the project plan. However, it should not, and cannot, precede the finalized project plan as many of the critical components for the evaluation can only be determined once the project plan is complete. Given that the project plan is not yet finalized, this draft of the evaluation will not include certain elements (e.g., metrics specific to interventions that have yet to be designed).

- Will the information from an impact evaluation inform future funding decisions?[3]

The remaining seven items will serve as a rough outline for the remainder of this evaluation plan, with accommodation made for the particularities of NN and the remaining items to be defined in its implementation plan.

Evaluation framework

We propose a mixed methods evaluation framework that is designed to: (1) answer the key questions motivating Nuru's Nigeria work (i.e., can and to what extent does Nuru programming reduce or eliminate vulnerabilities by building capacities for resilience in households and communities? Does the resulting resilience make communities less vulnerable to violent extremist organizations?); (2) incorporate the state of the art with regards to evaluation methodologies focused on building resilience and CVE; and (3) build on Nuru Kenya and Nuru Ethiopia's evaluation strategies. This framework is modified from USAID's "Baseline-RMS-Endline" approach, and from Tincani and Poole's mixed-methods framework (2015).

No single data tool or evaluation approach can address the measurement challenges inherent in international development programs. Different methods meet specific purposes, from measuring outcomes to understanding context, and bring their own strengths and limitations. A mixed methods approach can overcome the limitations and enhance the strengths of individual methods.

The quantitative component of the evaluation is important for monitoring and evaluating program outcomes and impacts. However, the quantitative evaluation is constrained by the limitations inherent in the process by which quantitative data from closed-ended questions in surveys are collected.[4] Household surveys are limited in are limited in their ability to ask important questions about the social, cultural and political context. A qualitative approach will be essential for Nuru to understand complex multidimensional constructs like resilience and women's empowerment. Qualitative data collection methods can help Nuru understand the drivers of resilience as well as the processes and interrelationships relevant to household and community resilience. The qualitative component will enable a better understanding of the significance of changes that are measured quantitatively, as perceived by households. Qualitative data collection methods will also be better suited than quantitative data collection

---

[3] Adapted from Jetha, Qayam, Kanan, Harini, and Escueta, Maya. 2017. "Impact Evaluability Toolkit." J-PAL South Asia and CLEAR South Asia.

[4] Rao, Vijayendra and Michael Woolcock (2003) "Integrating Qualitative and Quantitative Approaches in Program Evaluation", in Francois J. Bourguignon and Luiz Pereira da Silva (eds.) *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools* New York: Oxford University Press, pp. 165-90

methods for collecting information on sensitive topics (such as violent extremism and gender-based violence). A qualitative approach is also important for monitoring a - qualitative data can help Nuru assess the process of program implementation and how this affected outcomes and impacts.

Figure 4. Evaluation dimensions

Evaluation Components	Research Questions	Methods	Link to Survey	Survey sample	Timing
Resilience to Vulnerabilities	Can and to what extent does Nuru programming reduce or eliminate vulnerabilities by building resilience capacities in households and communities?	• Household Questionnaire developed by REAL consortium		All 500 intervention households and 1000 randomly selected comparison households	Baseline., midpoint and endpoint
Perceptions of social cohesion and security	Can and to what extent does Nuru programming impact perceptions of social cohesion and of security?	• Household Questionnaire developed by REAL consortium • Mercy Corps VRAI tool		All 500 intervention households and 1000 randomly selected comparison households	Baseline., midpoint and endpoint
Response to shocks	To what extent do communities/households have the capacities to cope with irregular/unanticipated shocks?	• Recurrent Monitoring Survey		50 randomly selected intervention households	Every 2 months following an irregular/ unanticipated shock. For 1 year
Gender	How does Nuru's gender equality work contribute to more gender-equitable decision-making in households?	• Household Questionnaire developed by REAL consortium • MercyCorps BRIDGE tool		All 500 intervention households and 1000 randomly selected comparison households	Baseline., midpoint and endpoint

Evaluation team

Formation of the evaluation team has begun with the hire of an M&E Program Specialist (expat Field Team staffer), a Nigerian M&E manager and field officers, and the continuing

engagement of existing external evaluators and Nuru's in-house Impact and Analytics Director and its Monitoring and Evaluation Strategic Advisor. For the purposes of defining evaluation-related roles, we propose the following duties, subject to amendment and expansion as needed:

- Evaluation manager(s)—members of Nuru's core staff, responsible for general administration of the evaluation, including oversight of external evaluator and field manager;
- External evaluator—responsible for quantitative analysis, design/refinement of instruments and protocols, sampling design, qualitative analysis, impact evaluation, and interim and final reports;
- Fieldwork manager and staff—responsible for supervising data collection, including planning routes for data collection, scheduling fieldwork teams, hiring enumerators and supervisors (if any), and data reporting to external evaluator.

Data collection

Following Tincani and Poole's framework, the remainder of this section proceeds as follows:

- description of the initial participatory qualitative information gathering component of the evaluation, including tools and methods for conducting focus groups and KII
- description of the quantitative components, including sampling design, Recurrent Monitoring Survey, and data analysis method (survey instruments are included in the appendices);
- description of the qualitative data collection to occur over the course of the evaluation.

The baseline tools will also be used at the midpoint and endpoint assessment, and will be applied to the same households and individuals, to the extent feasible. The objective is to build panel data collection into the evaluation, both via the surveys, and via the RMS, in order to facilitate the analysis and impact evaluation.

Qualitative data collection

Baseline qualitative data collection

Participatory qualitative data collection will be the first step in data collection. The objective of this component of the baseline work is to understand better how resilience and CVE are perceived locally, as well as attitudes and behavior that reflect gender norms. Approaches to collect this qualitative data include focus groups and key informant interviews (KII)s. Baseline qualitative data will also be used to inform the quantitative baseline data collection.

- Focus group discussions—a minimum of four focus group discussions, one in each of the

two intervention and two comparison group wards, will be conducted prior to program implementation (early 2019). These FGDs will involve 8-12 participants and will be representative of the community population in terms of gender, age, and SES. The purpose will be to conduct a problem analysis (i.e., to gather information on how resiliency and CVE are perceived in each ward). We will use purposive sampling of participants, with the aim of selecting relatively homogenous groups in order to facilitate open discussion. The group will be presented with a guide to the discussion topics (see Appendix 5 and Appendix 6), and the FGD will be attended by a facilitator and a note-taker. Logistics will be determined on site by the evaluation manager.

- Key informant interviews (KII) with a non-randomly chosen sample of individuals who are representative of key cross-sections of the population to be served will be conducted for the same purposes as those that apply to the FGDs, namely to understand how vulnerabilities and violent extremism are defined and linked (see Appendix 7). The key differences between the FGDs and the KIIs are that, in this case, we aim for representation from the range of stakeholders involved, and that these interviews will be conducted on an individual basis, although, as with FGDs, attended by two evaluation staff members. We will rely on local leaders to identify no fewer than five interviewees in each of the five wards.

We will use the information collected from the FGDs and KIIs to inform and refine the baseline survey instruments for the quantitative data collection. *As such, this component necessarily precedes the quantitative data collection* component of the set of baseline stages, and it should be an integral part of the program planning phase of Nuru's work.

Midpoint and endpoint qualitative data collection

Qualitative data will also be collected at midpoint and endline through participatory approaches such as focus groups (FGs) and key informant interviews (KIIs). Well-organized and detailed focus group discussion guides and key informant interview guides will be developed for use by field staff. Field staff facilitating FGs or conducting KIIs will receive training from the evaluators. Focus group discussions and KIIs will be recorded, either through field notes or audio recordings. Information gathered will be analyzed to identify trends and patterns and derive meaning. The analysis will be systematic (using a planned approach in a consistent manner) and verifiable (with evidence that can be reviewed e.g. audio recordings, transcripts, content analysis, field notes etc.).

Quantitative data collection

Baseline quantitative data collection

The baseline household questionnaire will be administered to the full panel of 500 treatment

households and 1000 randomly selected comparison households in April 2019.

Annual quantitative data collection

A reduced form of the household questionnaire will be administered to treatment households only, in April 2020 and April 2022. This will allow the evaluation team to analyze program outcomes.

Midpoint and endpoint data collection

The full household questionnaire will be administered to the full panel of 500 treatment households and 1000 randomly selected comparison households in April 2021 and April 2023. This will allow the evaluation team to analyze program impacts.

Sampling design

NN will contend with many of the typical problems facing development interventions, not least of which are:

- the need to account for selection bias when program participants will self-select,
- the potential for sample contamination, as well as
- the need for precision in the representativeness of the sample in light of limited resources.

Given that communities are somewhat split along certain cleavages, e.g.:

- Rural, Christian, Higgi ethnic group, crop farmers, also aligns to their political affiliation
- Peri-urban or "urban", Muslim, livestock grazers / pastoralists / business people, favor Muslim candidates, Fulani.

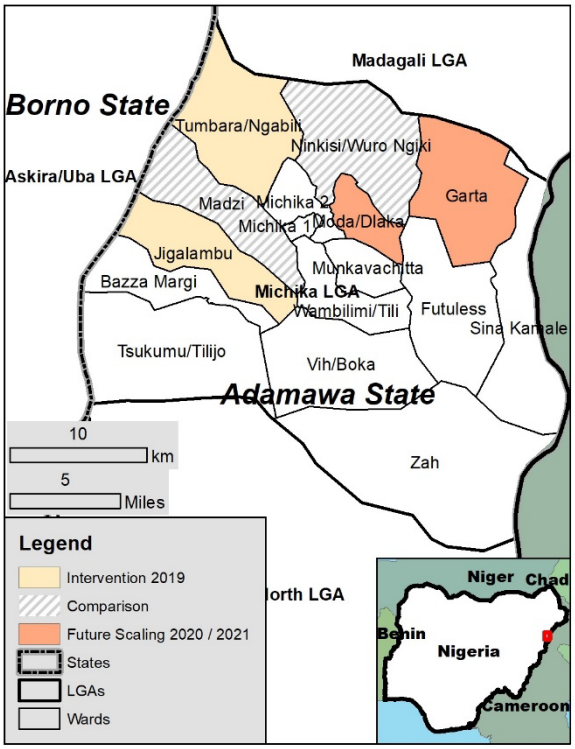
While these generalizations somewhat broad, they permit a useful stratification strategy, i.e.:

- Madzi, Garta and Tumbara/Ngabili are about 90% Christian / 10% Muslim, almost entirely "rural";
- Jigalambu, Ninkisi/Wuro Ngiki, and Moda/Dlaka are about 70% Christian / 30% Muslim, include portions of peri-urban around Michika town. These peri-urban fringes contain the Muslim population, and the Christian farmers are more dispersed.

Based on this stratification, we have randomly assigned one ward to 2019 intervention, one ward to comparison, and one ward to future scaling, for each of these groups, to the effect that the six wards are all assigned one of the three treatments or non-treatments.

Treatment	Tumbara/Ngabili	Jigalambu
Control	Madzi	Ninkisi/Wuro Ngiki
Future scaling	Garza	Moda/Dlaka

- All 500 treatment households in the two intervention wards (250 from Tumbara/Ngabili and 250 from Jigalambu) will be included in the study;
- 1000 households from the comparison wards (500 from Madzi and 500 from Ninkisi/Wuro Ngiki) will be randomly selected, prior to project implementation, to be included in the study.



This sampling design facilitates the use of an inverse probability weighting approach to balance covariates, and a difference-in-differences approach (described in Appendix 2) to measure NN's impact across several outcomes, including those related to resilience, vulnerabilities, and CVE.

Data collection instruments

Compensation

Cash incentives will be used for comparison group households to take surveys. Cash incentives will be used for comparison group individuals to participate in FGs and KIIs.

Household survey

A household questionnaire has been developed [5] (see Appendix 3) that adapts modules from

- existing Nuru household surveys;
- the resilience-focused household questionnaire developed by the Resilience Evaluation, Analysis and Learning (REAL) Consortium (funded by the USAID Center for Resilience and led by Save the Children with Food for the Hungry, Mercy Corps, and TANGO International as partners)[6]
- the women's empowerment focused BRIDGE household questionnaire, developed by Mercy Corps
- the CVE focused VRAI household questionnaire, developed by Mercy Corps

The instrument will allow the evaluation team to measure shock exposure, resilience capacities, responses, and recovery (see Appendix 1). Specifically, the instrument will allow for the measurement of resilience capacities as a set of indices, one for each of the three dimensions of resilience capacity—absorptive capacity, adaptive capacity, and transformative capacity—and one overall index combining these three indexes.

However, the findings from the participatory qualitative data collection carried out in weeks -4 through -2 will likely yield information that will be used to refine the survey instruments. We anticipate an additional two weeks between the completion of the qualitative data collection and the start of implementation will be needed to revise the instruments.

Quick Tap, a real-time, data collection tool, will be used by enumerators, and data will be made accessible to external evaluators, who will clean it and prepare for analysis, as it is collected.

Recurrent Monitoring Survey

USAID, working with TANGO International, developed the RMS as a means "to capture real-time household and community response to shocks and stresses as they occur" (USAID, Ethiopia PRIME vol. 1, 2015), and has found it useful in the context of resilience-related projects in Ethiopia, Niger, and Burkina Faso. In addition to measuring how material conditions

---

5 It may be the case, documented in the course of the qualitative baseline data collection. A third, and significantly less desirable option, would be to identify secondary data sources from which metrics of CVE could be drawn.however, that Nuru chooses not to include questions related to CVE in its survey instrument. If this is the case, however, attention will need to be paid to the qualitative instruments (described above) to ensure that critical metrics of CVE are documentable and

[6] TANGO International. (2018). Resilience and Resilience Capacities Measurement Options: Full Approach—Household Questionnaire. Produced by TANGO International as part of the Resilience Evaluation, Analysis and Learning (REAL) Associate Award.

change as the result of a given shock, the RMS also enables USAID to identify which components of their programming prove most effective in terms of mitigating the shock.

The RMS consists of a relatively short survey of 15-20 questions (see Appendix 4) that can be quickly completed, and it is designed to be fielded immediately after a shock is experienced. The RMS will be re-administered every two months for a year (the objective is to conduct a number of rounds sufficient to drawing defensible conclusions), and will be administered to a subsample of 50 households drawn from the baseline sample of 500 intervention households. USAID notes that the "Baseline—RMS—Endline" approach, a variation of which we have adopted, is best suited to projects, like NN, that incorporate adaptive management, and that seek to answer questions about resilience dynamics.

Fielding of the RMS is triggered by a particular shock relevant to the project's objectives. An important first step for the evaluation, therefore, would be defining a "shock threshold", past which it is recognized that a shock has occurred and, at which time, the RMS would be fielded.

Given the dual nature of NN's long-term objective, we believe the following shock thresholds related to livelihoods and to VE are appropriate:

- **Livelihoods:** in their examination of rainfall shocks on agricultural productivity in Nigeria, Amare et al (2018) found that a negative rainfall shock, **defined as a wet season with a rainfall measure at least one standard deviation (or approx. 520mm for Adamawa)[7] below a 30 year historic mean**, decreases agricultural productivity and hence decreases household consumption by 37%, and that they have a negative, significant impact for asset-poor and non-poor households. The RMS will be administered to households randomly selected from the region of the livelihoods shock event.
- **VE:** The Armed Conflict Location and Event Data Project (ACLED) identifies 9 event types that are included as units of observation for its data collection purposes. Its "violence against civilians" event type, defined as "a violent act upon civilians by an armed, organized, and violent group. By definition, civilians are unarmed and not engaged in political violence. Rebels, governments, militias, external forces, and rioters can all commit violence against civilians. Protesters are also civilians, and significant violence against protesters falls under this category" (ACLED Codebook, Version 8, 2017). ACLED events categorized as "violence against civilians" with a violent extremist organization named as the primary actor will be identified as a VE shock. The RMS will be administered to households randomly selected from the vicinity of the VE shock

[7] Estimated from Africa Rainfall Climatology Version 2 (ARC2) of the National Oceanic and Atmospheric Administration's Climate Prediction Center data

event.

Pilot testing

The next major step in the implementation of the evaluation will be to pilot test the baseline household questionnaire and the Recurrent Monitoring Surveys (see Appendix 3 and Appendix 4 for instruments). Prior to piloting the instruments, we will need to ensure that data entry programs and data collection procedures are in place (i.e., staff are assigned to data entry tasks, a schedule for surveying households is settled upon, guidelines for conducting surveys are established and understood, etc.).

Once these programs and procedures are in place, we will randomly select a test sample for piloting the instruments. This sample does not need to be a particular size or composition—the pilot test is meant to determine (1) how long it takes to conduct the survey; (2) whether or not the instruments collect the required data; and (3) whether or not any modifications need to be made (note: the survey instruments included in the appendices are only drafts, so should not be considered final until they are piloted in the field). If possible, we will avoid piloting the instrument among populations expected to be included in the evaluation. Finally, this pilot testing process can be used to winnow down the number of field staff assigned to data collection (e.g., the pilot test may indicate that fewer staff than anticipated are needed, or that some staff are better suited to other duties related to the evaluation).

Data analysis

As data is received, we will inspect it for any anomalies, missing data, and will complete any necessary data cleaning in order to prepare it for analysis. We will maintain data on a secure server, housed at the University of Texas, Austin, and access to the data will be limited to external evaluators and Nuru staff.

Using the intermediate approach and methodology shared by the REAL Consortium[8], we will calculate the individual components of resilience capacity as well as the three resilience capacity indices – absorptive, adaptive, and transformative capacities – and an overall resilience index. The indexes can then be compared from baseline to endline for both the treatment and the control groups to measure change in resilience.

Deliverables

1. Baseline report

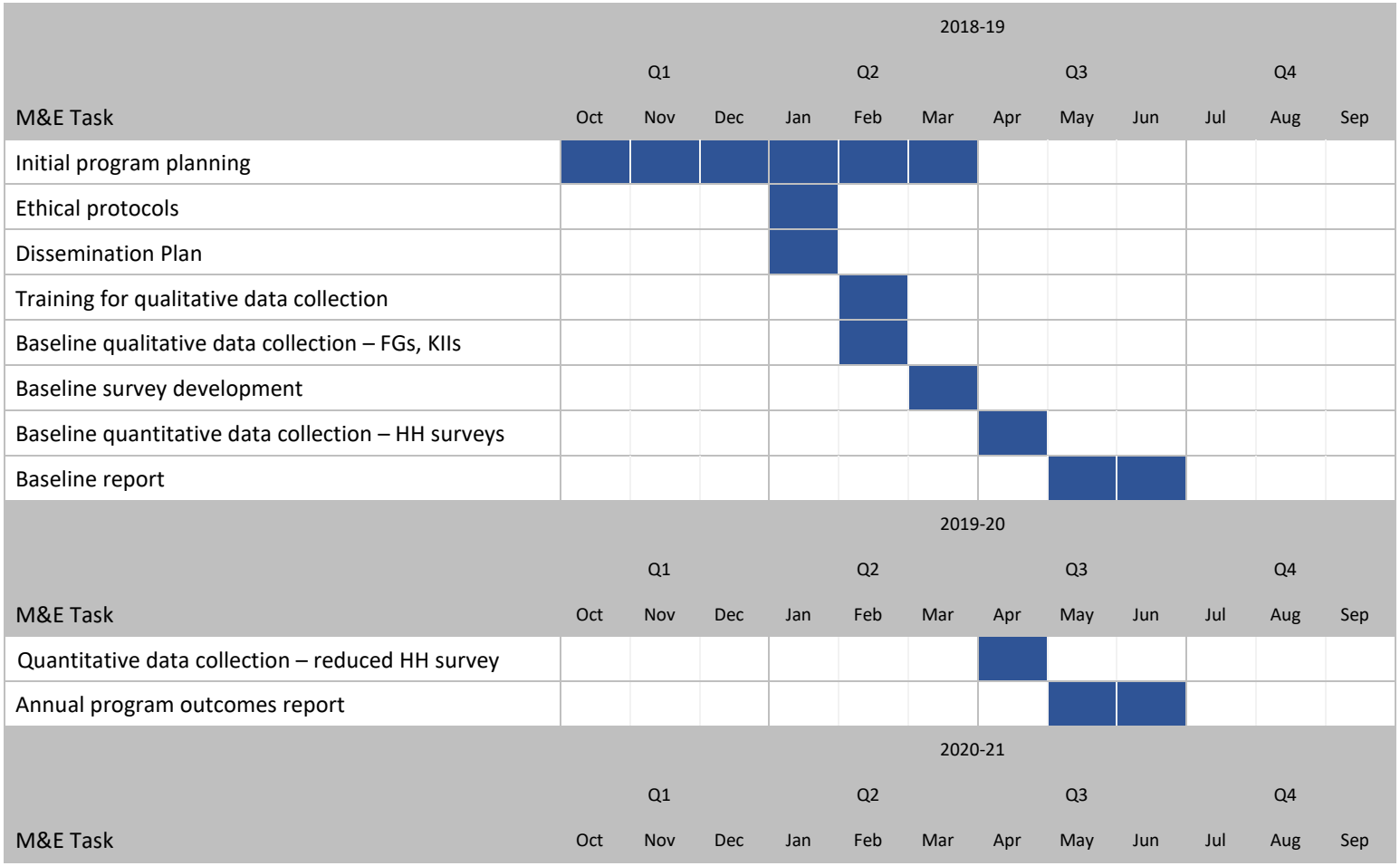
[8] TANGO International. (2018). Methodological Guide: A Guide for Calculating Resilience Capacity. Produced by TANGO International as part of the Resilience Evaluation, Analysis and Learning (REAL) Associate Award.

    - A baseline report will document the findings from the participatory qualitative and quantitative data collection activities at baseline described above. It will become the basis for measuring change resulting from Nuru's interventions, and will serve as the primary data resources for conducting impact evaluations at later points in the project.
2. Annual outcome evaluation reports
    - As with Nuru Ethiopia and Nuru Kenya, we anticipate analyzing annual household survey data and reporting program outcomes for monitoring purposes.
3. Midpoint and endpoint impact evaluation reports
    - As with Nuru Ethiopia and Nuru Kenya, we anticipate applying propensity score and difference-in-difference methods to measure programmatic impact. Impact evaluations will be conducted and reported at midpoint and endpoint. Additionally, focus groups and KIIS will be conducted at midpoint and endpoint and findings from the qualitative analysis will be incorporated into the midpoint and endpoint reports.
4. Policy briefs
    - We anticipate the interventions and evaluation to produce numerous policy-relevant findings, although likely not until later stages of the program. Policy briefs will be developed at endpoint that highlight these findings and suggest courses of action.
5. Fully documented data sets, design and analysis protocols
    - The evaluation will produce numerous, rich data sets, and will employ specific design and analysis protocols. These will be made available to Nuru, and will be presented in a fashion designed to permit replication.
6. LSI dashboard
    - We will work with the M&E team to design an LSI dashboard.
7. Dissemination plan
    - Depending on Nuru's program strategy, a dissemination plan will be developed that will direct what and how evaluation findings will be provided, and to whom.
8. Ethical protocols on protection of human subjects (to be completed), including ensuring informed consent, obtaining IRB approval
    - We will be responsible for designing and submitting an application for ethical approval of the evaluation.

Timeline and work plan

This evaluation study will span four years, beginning in 2019 and ending in 2023, with a final endpoint data collection in April 2023 and final impact reports published by the end of 2023.

Figure 5. Timeline and Work plan



Midpoint quantitative data collection – full HH questionnaire												
Midpoint qualitative data collection												
Midpoint Impact Report												
	2021-22											
		Q1			Q2			Q3			Q4	
M&E Task	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
1st follow-up quantitative data collection – reduced HH surveys												
Annual program outcomes report												
	2022-2023											
		Q1			Q2			Q3			Q4	
M&E Task	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Endpoint quantitative data collection– full HH questionnaire												
Endpoint qualitative data collection												
Endpoint Impact Report												
Policy briefs												
Data sets, design and analysis protocols												

APPENDIX 1. METRICS

Resilience

Absorptive Capacity Index	Adaptive Capacity Index	Transformative Capacity Index
• Availability of informal safety nets • Bonding social capital • Access to cash savings • Asset ownership • Shock preparedness and mitigation • Availability of/access to insurance • Availability of/access to humanitarian assistance	• Bridging social capital • Education/training • Livelihood diversification • Asset ownership • Availability of financial resources • Aspirations • Locus of control • Confidence to adapt	• Availability of/access to formal safety nets • Bridging social capital • Social cohesion • Gender equitable decision-making index • Local government responsiveness

Gender

Gender-equitable decision making in

- decisions on how household income is used
- household decisions over health care and nutrition
- decisions about major household purchases
- decisions about your children's education

Social cohesion

- Confidence in government
- Social relations
- Freedom of religion
- Perception of security

APPENDIX 2. QUANTITATIVE METHODS

Inverse probability weighting approach

Inverse probability weighting is one of the methods that take advantage of the identification of a propensity score. The propensity score is the estimated probability of being in the treatment group given the observable characteristics from a regression model of participation (Rosenbaum and Rubin, 1983). This probability is obtained from the "participation equation", a multivariate regression in which the dependent variable is dichotomous, taking the value of 1 for those in the treatment group, and 0 for those in the comparison group. The independent variables will include all observed variables that may affect participation in NN, but are not affected by the intervention. We will apply data collected from the baseline surveys to the participation model.

Approaches based on the propensity score rely on observable characteristics of program participants and non-participants in order to artificially create a comparison group. The key assumption underpinning the use of the propensity score is that, once pre-treatment characteristics are controlled for, whether an individual receives treatment or not is essentially randomly assigned and the treatment effects can be identified.

With inverse probability weighting, inverse probability weights are calculated from the participation regression model, creating balance between treatment and comparison groups. We will use either a logit or probit estimator to calculate average treatment effects on any key outcomes, controlling for pre-treatment characteristics (for NN, we will want to measure multiple outcomes related to vulnerabilities and CVE (will likely create indices—TBD)).

Impact analysis

Once we have created sufficiently balanced treatment and comparison groups, we employ a difference-in-differences methodology to measure impact[9], comprised of the following steps:

a. survey households in treatment and control districts at baseline
b. measure outcomes for each intermediate and final impact indicator, in both treatment and comparison groups
c. calculate differences between treatment and control across all impact indicators

[9] Additional methods may be considered, depending on ease of collecting data, and availability of longitudinal data (e.g., regression modeling or statistical matching methods). Other methods, while more rigorous than did, will not be considered due to their poor applicability to the program designed (e.g., regression discontinuity or randomization at the individual level).

d. conduct follow-up survey at year end
e. calculate mean differences between before and after across all indicators for both treatment and control
f. calculate difference between mean differences= program impact