

Baseline Study of the PReSERVE Resilience Food Security Activity in Ethiopia



January 2023 | Volume II – Annex H

IMPEL | Implementer-Led Evaluation & Learning Associate Award



USAID
FROM THE AMERICAN PEOPLE

CAUSAL
DESIGN



Save the Children.

ABOUT IMPEL

The Implementer-Led Evaluation & Learning (IMPEL) Associate Award works to improve the design and implementation of Bureau for Humanitarian Assistance-(BHA)-funded resilience food security activities (RFSAs) through implementer-led evaluations and knowledge sharing. Funded by the United States Agency for International Development (USAID) BHA, IMPEL will gather information and knowledge in order to measure performance of RFSAs, strengthen accountability, and improve guidance and policy. This information will help the food security community of practice and USAID to design projects and modify existing projects to bolster performance, efficiency, and effectiveness. IMPEL is a seven-year activity (2019—2026) implemented by Save the Children (lead), TANGO International, Tulane University, Causal Design, Innovations for Poverty Action, and International Food Policy Research Institute (IFPRI).

RECOMMENDED CITATION

IMPEL. (2022). *Baseline Study of the PReSERVE Resilience Food Security Activity in Ethiopia (Vol. II)*. Washington, DC: The Implementer-Led Evaluation & Learning Associate Award.

PHOTO CREDITS

Save the Children / Charlie Forgham-Bailey

DISCLAIMER

This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the Implementer-Led Evaluation & Learning (IMPEL) award and do not necessarily reflect the views of USAID or the United States Government.

CONTACT INFORMATION

IMPEL Associate Award

c/o Save the Children

899 North Capitol Street NE, Suite #900

Washington, DC 20002

www.fsnnetwork.org/IMPEL

IMPEL@savechildren.org

PREPARED BY:



ACKNOWLEDGMENTS

The Causal Design research team would like to recognize several individuals who were instrumental in creating and refining this document. This includes Abiy Yadeta, Getahun Shibeshi, and Girma Deressa with Food for the Hungry, as well as Mohamoud Ahmed, Emily Marshall, Corena Sharp, and Tom Spangler with IMPEL. Special thanks also to Mara Mordini, Hilary Cook, Takawira Kapikinyu, Benita O'Colmain, Arif Rashid, and Adam Trowbridge with USAID for their ongoing technical support.

TABLE OF CONTENTS

Acknowledgments	i
List of Tables	iii
List of Figures	iii
Acronyms	iv
Annex H: Pre-Analysis Plan	5
1. Description of Study	5
1.1. Impact Evaluation Overview.....	5
2. Evaluation Approach	6
2.1. Research Objective	6
2.2. Evaluation Design	7
2.3. Power Calculations and Sampling Strategy	14
2.4. Cost-Effectiveness Analysis.....	18
3. Data Collection	21
3.1. Survey Design	21
3.2. Outcome Indicators	22
3.3. Enumerator Training.....	22
3.4. Data Management.....	22
3.5. Tentative Data Collection Timeline	23
4. Analysis.....	23
4.1. Descriptive Analysis	24
4.2. Impact Analysis.....	24
4.3. Supplementary Analysis on Resilience Indices	25
4.4. Cost-Effectiveness Analysis.....	26
5. Appendix	28
5.1. Data Management.....	28
5.2. Data Quality	28
5.3. Data Security & Research Ethics.....	30
5.4. Knowledge Management & Transparency	31
5.5. Data Storing/Sharing Guidelines	31
5.6. Role of the Implementing Partner.....	35
5.7. Intervention Packages	36
5.8. Indicators List.....	44

VOLUME I

Annex A: Intervention Packages

Annex B: List of Kebeles

Annex C: Summary Tables

Annex D: Treatment and Control Balance Tables

Annex E: Impact Evaluation and Non-Impact
Evaluation Balance Tables

Annex F: Summary Tables by Woreda

Annex G: Main Indicators with Standard
Errors and Confidence Intervals

LIST OF TABLES

Table 1. List of interventions across the three purposes.....	8
Table 2. Descriptive statistics of selected outcome indicators.....	14
Table 3. Minimum required number of households for selected outcomes in Amhara assuming an MDE of 0.1 standard deviations.	15
Table 4. Sampling strategy.....	18
Table 5. Baseline data collection timeline	23
Table 6. Endline data collection timeline (preliminary).....	23
Table 7. Data quality assurance activities.....	33
Table 8. Intervention packages.....	36
Table 9. Food for the Hungry PReSERVE indicators list	44

LIST OF FIGURES

Figure 1. Steps to select households receiving LD interventions	13
Figure 2. MDE (in SD) for different total sample sizes for the outcome “Children 6–23 months of age who meet minimum dietary diversity”	16

ACRONYMS

ANC	Antenatal Care
BHA	Bureau of Humanitarian Assistance
BL	Baseline
CEA	Cost-Effectiveness Analysis
DA	Development Agent
EL	Endline
FCS	Food Consumption Score
FIES	Food Insecurity Experience Scale
HH	Head of Household
IE	Impact Evaluation
IMPEL	Implementer-Led Evaluation and Learning Associate Award
LD	Livelihood Deepening
MAD	Minimum Acceptable Diet
MDD	Minimum Dietary Diversity
MDD-W	Minimum Diet Diversity-Women
MDE	Minimum Detectable Effect
ORT	Oral Rehydration Therapy
PCA	Principal Components Analysis
PMT	Proxy Means Test
PSNP	Productive Safety Net Program
R4D	Research for Development
RCT	Randomized Controlled Trial
RFSA	Resilience Food Security Activity
SD	Standard Deviation
TIGER – OR	Transition into Graduation through Enhanced Resilience – Operations Research
USAID	United States Agency for International Development
VESA	Village Economic and Social Associations
WASH	Water, Sanitation, and Hygiene
YES	Youth Economic Strengthening

ANNEX H: PRE-ANALYSIS PLAN

1. DESCRIPTION OF STUDY

Causal Design is part of the Implementer-Led Evaluation and Learning (IMPEL) consortium, the Bureau of Humanitarian Assistance's (BHA) primary mechanism for carrying out the evaluations of the resilience food security activities (RFSA) in Amhara, Ethiopia. Causal Design's support for this project will include a survey and evaluation design of the Food for the Hungry PRoSERVE RFSA. Throughout the implementation period, Causal Design will conduct an Impact Evaluation study using an experimental evaluation (randomized control trial (RCT)). The key evaluation deliverables include the following:

- Evaluability Assessment that will lay out the research design and approach to inform the Pre-Analysis Plan.
- A Pre-Analysis Plan that will outline the sampling strategy, survey design, outcomes for the analysis, estimation strategy, and additional methods (as appropriate).
- A Baseline Report that will summarize and analyze baseline (BL) survey data.
- An Impact Evaluation Report that will use the BL and endline (EL) data to estimate the impact of the RFSA.
- A Cost-Effectiveness Analysis (CEA) at EL, tied to the impact evaluation data.

In addition to these activities, Causal Design will also perform a process monitoring evaluation, which is meant to assess the implementation process itself. This evaluation is not part of the experimental evaluation and will not be discussed in this Pre-Analysis plan. A separate Pre-Analysis plan for the process monitoring will be delivered at the end of the year.

1.1. Impact Evaluation Overview

Causal Design will conduct an Impact Evaluation using BL and EL survey data in the intervention areas of the Food for the Hungry PRoSERVE RFSA in Amhara, Ethiopia. A BL survey for the RFSA will be carried out at the end of the second quarter of 2022, and a corresponding EL survey will be conducted in the second quarter of 2025. Additionally, a cost-effectiveness analysis, designed below, will be carried out in conjunction with the EL report.

Overview of BL survey: The BL survey will use quantitative methods to measure BL indicators collected in the RFSA target area in 2022. Causal Design will work closely with BHA and relevant stakeholders to identify other key learning objectives and ensure that the BL survey is able to contribute to this learning where possible.

Overview of EL survey: The survey will follow up with the same communities and households in the BL survey to estimate the ability of the RFSA intervention to directly impact household food security and well-being indicators as listed in BHA's standard indicators. The EL study will be carried out in 2025. The

same, or slightly modified,¹ survey will be administered to the same households as in the BL activity to ensure comparability across the two time periods.

Overview of Cost-Effectiveness Analysis: This will focus on the cost-effectiveness of PReSERVE’s livelihood interventions compared to Productive Safety Net Program (PSNP) 5 and other components of the PReSERVE RFSA interventions over the course of its entire implementation. The full results of this analysis will rely on data from the impact evaluation following the EL studies, along with relevant program and cost data. The results of the CEAs will be presented in the EL report.

2. EVALUATION APPROACH

2.1. Research Objective

The PReSERVE RFSA interventions are a package of interventions aimed at **improving food security** of vulnerable households in targeted PSNP communities in the Amhara Region **and contributing to a sustained reduction in rural poverty**. The primary objective of the impact evaluation will be to measure the marginal impact of the livelihood-deepening (LD) interventions on improving livelihoods and food outcomes in the targeted communities. The evaluation seeks to inform the larger knowledge base around the efficacy of LD activities among vulnerable populations and how benefits to vulnerable households can be further maximized.

In addition, the proposed evaluation contributes to a growing literature analyzing the effect of multifaceted “graduation models” implemented to target the reduction of poverty and broader enhancement of welfare in low-income countries (Banerjee et al. 2015, Bedoya et al. 2021, Brune et al. 2021, Bossuroy et al. 2022). Evidence has suggested that intensive multi-sectoral programs, often entailing asset transfers valued at \$500 or more, as well as consumption support, training and coaching visits, and other supplemental services, can have sustained positive effects on consumption, assets, and other household-level outcomes (Bandiera et al. 2017, Banerjee et al. 2021a, Banerjee et al. 2021b). These interventions generally entail a cost per recipient household of \$1,000 or more, of which at least 75% is received by the household in the form of direct cash or asset transfers.

However, the evidence base around lighter-touch and lower-cost interventions is more limited. We identify LD as a lighter touch intervention based on the intervention services received by households: they are receiving credit, not an asset outright, and while they receive transfers through the PSNP, the value of those transfers based on overall PSNP programmatic guidelines is lower than what is generally observed in other graduation models. This raises the important question as to whether this lighter-touch and lower-cost model can still be effective in generating transformational livelihoods impacts—a finding that has important implications for overall program design and cost-effectiveness in the graduation model space. This project will contribute to the literature by providing new evidence about the effects of a livelihoods-focused set of interventions for an extremely poor set of households (PSNP beneficiaries) in poor communities in rural Ethiopia.

¹ Due to the COVID-19 pandemic, anthropometric information will not be collected in the BL survey. This information will be collected in the EL survey.

2.1.1. Research Question

Research Question: *What is the marginal impact of the Livelihood Deepening package of interventions for relatively more vulnerable PSNP5-RFSA households on improving food security and other related well-being outcomes?*

2.2. Evaluation Design

The evaluation team will implement a **randomization-controlled trial using randomization at the household level** designed to estimate the effect of Food for the Hungry PReSERVE LD interventions. In a set of eligible kebeles, the research team will select a set of eligible households that will be part of the IE study. Half of those households will be allocated to the treatment group and will receive the LD activities and support in addition to the other components of the PReSERVE RFSA, while the other half will be allocated to the control group and will only receive the other components of the PReSERVE RFSA.

Using baseline and endline data, we will compare the households assigned to control and treatment in order to identify the direct impact of participating in LD, in addition to the other components of the PReSERVE RFSA, using BHA food security and nutrition and poverty indicators.²³ The RCT design will maximize the ability of the research to measure direct and attributional impacts. The CEA, combined with the impact evaluation findings, will allow the research team to explore a value-for-money dimension that assesses program effectiveness.

The following subsection gives a brief description of the PReSERVE interventions and discusses in more detail the LD interventions, whose additional impact will be studied in the IE. The other subsection outline identification, randomization, and sampling strategies for the impact evaluation as well as the CEA methods and strategy.

2.2.1. PReSERVE Interventions

To create sustainable change in targeted woredas, PReSERVE will implement a prioritized portfolio of evidence-based, catalytic interventions with the potential to increase the graduation of ultra-poor PSNP households from poverty. PReSERVE's interventions are organized around three main purposes:

1. Vulnerable HHs and Individuals Have Sufficient Quantity, Quality, and Diversity of Food at All Times
2. Vulnerable Community Members' Livelihoods Transformed
3. PSNP Systems Deliver Accountable, Effective, and Shock-Responsive Services

Food for the Hungry hopes to achieve the first purpose by focusing on interventions that address consumption for vulnerable households, availability of quality nutritious foods for men, women, and children, and improving children feeding behaviors that are practiced. Purpose two will focus on interventions that will increase and protect household assets, enhance individual and household resilience capacities, and help individuals and households sustainably engage in diversified livelihood.

² <https://www.usaid.gov/food-assistance/partner-with-us/implementation-and-reporting>

³ Even though we will collect information to compute a wide range of indicators, the analysis will focus on those more directly affected by the LD interventions.

Lastly, Food for the Hungry will address purpose three by centering on three sub-purposes: PSNP systems for planning construction and management of community assets improved, referrals and linkages to essential services and institutions improved, and PSNP services delivery improved. A more extensive list of interventions for purposes 1, 2, and 3 can be found in section 5.7.

Given both implementation realities and USAID learning goals and objectives, the evaluation will focus on the LD interventions implemented at the individual or household level. Table 1 presents a set of these interventions, their level of implementation, and the purpose they belong to. While there are LD interventions in Purposes 1 and 3, most of the LD interventions performed by Food for the Hungry are in Purpose 2. One of the main objectives of these interventions is to diversify viable livelihood opportunities within and outside of agriculture. To reduce the possibility of spillovers or contamination (i.e., control households also receiving the interventions that are part of the IE), the evaluation will focus on those interventions provided either at the individual or the household level.

The focus on LD is motivated by three factors. First, it did not seem that there was an opportunity to design an evaluation of the “whole of RFSA” programming. Such a strategy was deemed challenging to implement because the rollout of interventions across kebeles is not finalized yet and will be refined in the coming years. Second, LD is one of the most intensive components of Food for the Hungry programming and is expected to have the largest effects. Third, the use of the new TIGER-OR (Transition into Graduation through Enhanced Resilience – Operations Research)⁴ tool to define eligibility criteria to identify households for LD seemed to provide a particularly appropriate opportunity for an oversubscription design. For other program components, appropriate evaluation designs may be more complex to identify.

Table 1. List of interventions across the three purposes⁵

Purpose/ Intervention	List of LD interventions or Supportive services	Types of Service	Level of Implementation		
			Ind	HH	Kebele
Purpose 1: Vulnerable households and individuals have sufficient quantity, quality, and diversity of food at all times					
Irrigation	Construct family hand dug well for irrigation	LD intervention		X	
	Train farmers on irrigation based improved agronomic practices	LD intervention		X	
	Link irrigation user households with seed suppliers (agro-dealers/others) to buy seed for planting	LD intervention		X	

⁴ Research for Development (R4D) has created a tool, the Transition into Graduation through Enhanced Resilience – Operations Research (TIGER – OR). The goal of this tool is to support USAID RFSA activities to target PSNP-eligible households to receive livelihoods interventions. The research team thanks R4D for providing documentation about the TIGER-OR tool. The TIGER-OR tool is composed of three different tools. The first tool assigns kebeles a score, which allows the selection of kebeles to be targeted. The other two tools aid in the selection of households. The first household-level tool uses a proxy means test (PMT) to identify the poorest 20% of PSNP Households, while the other tool assigns to each household a capability score based on a survey. This capability score can be used to select the households that would benefit more from the LD interventions.

⁵ This is not an exhaustive list of interventions. All the activities can be found in section 5.7.

Purpose/ Intervention	List of LD interventions or Supportive services	Types of Service	Level of Implementation		
			Ind	HH	Kebele
Conservation Agriculture	Organize training and exposure visit to DAs and promoters on conservation agriculture techniques and improved agronomic practice	LD intervention			X
	Organize training and exposure visit to selected households on conservation agriculture	LD intervention		X	
	Support farmers to practice conservation agriculture techniques on their farm fields on selected crops (haricot bean, potato, vegetables, fruits)	LD intervention		X	
Purpose 2: Vulnerable Community Members' Livelihoods Transformed					
GRANT	Facilitate livelihood transfer to ultra-poor PSNP clients	LD intervention	X		
	Train LH transfer recipients on business management	LD intervention	X		
FRUIT	Handover nurseries to youth entrepreneurs	LD intervention	X		
	Link PSNP HHS with High value tree (fruit & others) nurseries to access fruit seedlings	LD intervention	X		
CREDIT & SAVING	Organize VESA groups	LD intervention	X		
	Link VESAs with MFIs and Unions for Informal Apprenticeship and credit access	LD intervention	X		
	Support initial materials for VESAs	LD intervention	X		
RuSACCOS	Channel guarantee loan fund to RuSACCOS	LD intervention			X
	Link individuals in the VESA groups with RuSACCOS	LD intervention	X		
YOUTH	Establish Youth Economic Strengthening (YES) Centers	LD intervention			X
	Provide loan fund for grantees selected male and female youth	LD intervention	X		
WAGE-BASED EMPLOYMENT	Provide Behavioral (soft) Skills Training for male and female youth	LD intervention	X		
	Provide Vocational Skill Training for male and female youth	LD intervention	X		

Purpose/ Intervention	List of LD interventions or Supportive services	Types of Service	Level of Implementation		
			Ind	HH	Kebele
OFF-FARM	Provide Behavioral (Soft) Skill Trainings for male and female youth	LD intervention	X		
	Provide Technical Skill Training for male and female youth	LD intervention	X		
NON-FARM	Provide Behavioral (Soft) Skill Trainings for male and female youth	LD intervention	X		
	Provide Vocational Skill Training for male and female youth	LD intervention	X		
	Provide Business Development Training for male and female youth	LD intervention	X		
Value chain	Facilitate improved access to market information	LD intervention		X	
Purpose 3: PSNP Systems Deliver Accountable, Effective, and Shock-Responsive Service					
	Train youths and women on nursery management	LD activity	X	X	
	Facilitate livelihood development in potential watersheds	LD activity	X		X
WATERSHED	Train user groups on technical areas and supportive skills (financial, saving and credit, and life skills)	LD intervention	X		
	Link mature watersheds for livelihoods activities by user groups	LD intervention	X		X
	Develop integrated, climate-smart, gender and nutrition sensitive annual public works plans that contribute to livelihood productivity	LD intervention			X
PRIVATE NURSERY	Establish/Strengthen private nurseries	LD intervention	X	X	
	Seedlings produced by private individuals	LD intervention	X		
	Train youths and women on nursery management	LD intervention	X		

Note: Ind stands for Individual.

2.2.2. Identification Strategy

To measure the marginal impact of the LD interventions, the evaluation team proposes to use a household-level randomized controlled trial. Within a set of kebeles selected to receive the LD

interventions,⁶ the evaluation team will survey a sample of the relatively most vulnerable eligible households.⁷ Half of them will be allocated to the treatment group and will receive the LD interventions specified in Table 1 and the appendix, while the other half will be allocated to the control group and will not receive these interventions. The evaluation will compare the outcomes of households and individuals in the treatment group against the same outcomes of households and individuals in the control group. Note that households in both the treatment and control arms will be eligible to receive interventions linked to purpose one and purpose three, though the precise services they receive will depend on the eligibility criteria for other interventions and whether households wish to participate in them.

Due to the reduced number of kebeles receiving the LD interventions (55), it was deemed infeasible by the research and program teams to conduct randomization at the kebele level (i.e., every eligible household in a treated kebele receiving the LD interventions). We are aware of the potential implementation challenges that are raised by excluding households from the LD interventions. Nonetheless, the evaluation team has noted that due to limited resources, Food for the Hungry PReSERVE is unable to provide the LD interventions to all the eligible households, independent of the evaluation design. Consequently, the IE design will have no effect on the total number of households receiving the LD interventions but will instead provide a nonbiased approach to uniformly determining the delivery of the LD program within the target sample.

The analysis can also explore the differential effects of LD when delivered in conjunction with other intervention components linked to purposes 1 and 3. However, there are two potential complexities to note here. First, the variation in participation in purposes 1 and 3 will be non-random and thus potentially correlated with other household characteristics. Second, the sample of households may or may not include large numbers of households participating in purposes 1 or 3. The sampling plan could potentially be adjusted to ensure that target households for another purpose are also included here. Given that the TIGER-OR tool is a new tool, simply executing the current sampling design with fidelity may already be a significant commitment, and we might want to be cautious about making the sampling design even more complex.

2.2.3. Randomization Strategy

The evaluation team will work with Food for the Hungry to randomize the rollout of the LD interventions at the household level.

Selection criteria of kebeles: Food for the Hungry used the TIGER-OR kebele tool to select a set of 55 kebeles in 4 Woredas (Tach Gayint, Lay Gayint, Sahila, and Simada) that will receive LD interventions. With the help of the tool, Food for the Hungry identified kebeles with enabling conditions that would increase a household's ability to improve livelihoods.

It is important to notice that the IE study is restricted to those 55 kebeles because those are the ones that will effectively be receiving LD interventions. Given the characteristics of these kebeles, the IE study

⁶ The next section explains in more detail the way these kebeles were elected and the implications of this strategy

⁷ The next section explains in more detail the set of households that will be sampled.

will be able to speak about the impact of LD interventions on kebeles, where households are more likely to benefit from LD activities.

Selection criteria of households: There are more than 20,000 PSNP households in the 55 selected kebeles. The selection of the households that will be part of the IE study involves three steps, which are represented in Figure 1.

In the first step, Food for the Hungry will use the TIGER-OR proxy mean test to identify the 20% poorest households. Those households will be part of the grants track and will not receive LD interventions. The other 80% of households will be part of the credit track and are eligible to receive the LD interventions. It is important to notice that the identification of the eligible households is done within kebele—i.e., for each one of the 55 kebeles, the 20% poorest PSNP households are not eligible to receive LD interventions.

The next step of the process involves assigning each one of the eligible households a capability score provided by the TIGER-OR tool. There are around 16,500 eligible households in the 55 kebeles. Due to capacity constraints, Food for the Hungry will only be providing the LD interventions to a subset of eligible households. The share of eligible households not receiving the LD interventions varies slightly by kebele because it is driven by the targets defined by Food for the Hungry at the Woreda level.

Within each kebele, the research team will select 60% of the households with the highest capability scores, which will automatically receive LD interventions and will not be part of the IE study. The remaining 40%⁸ will be part of the IE study and will be randomly allocated to either the treatment or control group.⁹ Even though this selection reduces the scope of the evaluation, this was considered appropriate for two reasons. First, all eligible households are PSNP beneficiaries and, therefore, are identified as among the poorest households in Ethiopia. Accordingly, the potential for learning is substantial, even in evaluating the effects of the interventions for the poorest households. Second, the proposed strategy combines the selection of households more likely to graduate (those with higher capability scores) with the randomization of households that will be included in the IE study. This enables the partner to balance the goals of the TIGER-OR selection process with the goals of the evaluation.

In the last step of the selection process, a subset of the lower 40% eligible households will be randomly allocated to receive the LD interventions (treatment group), while the remaining households will be allocated to the control arm. This allocation process will be implemented in Stata by the research team. The exact number of households allocated to the treatment and control group varies by kebele and is determined by the targets defined at the woreda level. Even though the number of households in the treatment and control groups is not the same, we will sample an equal number of households in both groups. This will maximize the power of the study and is possible because the number of households to be sampled is less than the bottom 40% of households.

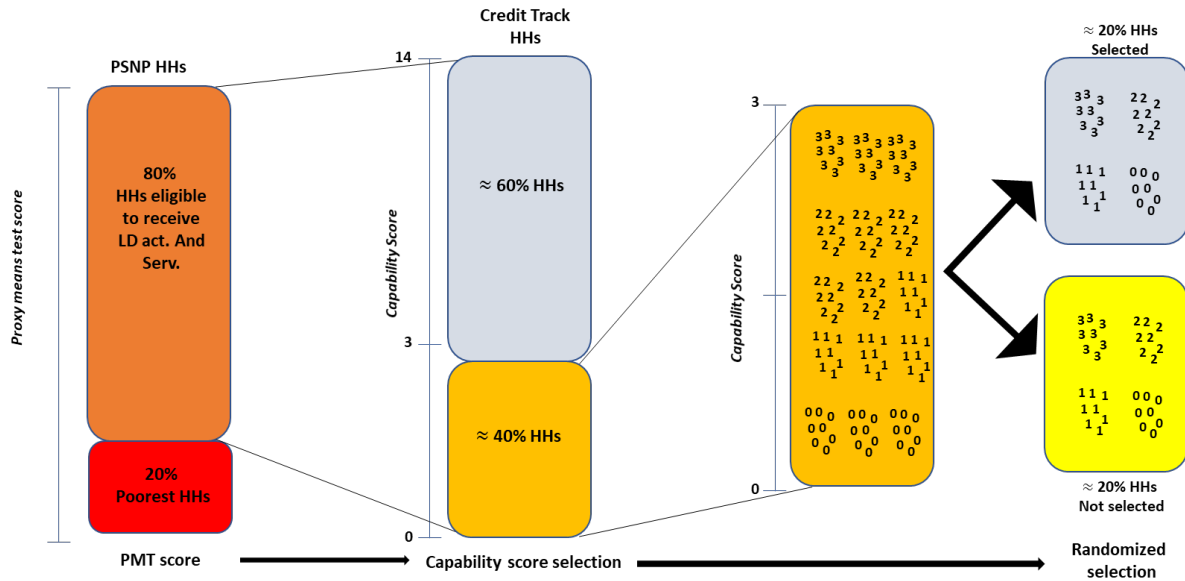
⁸ This number corresponds to approximately twice the share of eligible households not receiving LD interventions in the 55 kebeles. This share varies slightly by Woreda: kebeles in Lay Gayint have a share of approximately 22%, kebeles in Tach Gayint a share of approximately 21%, kebeles in Sahela a share of approximately 20% and kebeles in Simada a share of approximately 19%.

⁹ For ease of reading, we will refer to both groups as the upper 60% and the lower 40% in terms of the capability score.

The evaluation team will use a stratified randomization approach to guarantee a better balance. This technique involves dividing the sample of households into groups sharing similar characteristics. Based on the information provided by Food for the Hungry, the evaluation team will use two strata, namely kebeles and capability score ranges.¹⁰ The next step is to allocate households to treatment and control within each kebele. For example, let’s assume we have a kebele with 200 households in the lower 40%, and half will be allocated to treatment based on the kebele targets. Given the capability score ranges of 0–2.5, 3–5, and 5.5–8, the allocation of the treatment status will be done within each strata range. If 40 households have capability scores ranging from 0 to 2.5, then half of them (20) will be allocated to the treatment group and the other half to the control group. A similar process is done for households in the other two strata ranges. This approach ensures that treatment and control groups are balanced by the strata used.

As discussed in the next section, in addition to sampling households that will be included in the IE study (the bottom 40% based on their capability score), the research team will also sample some households in the upper 60%. This will allow us to create some basic comparison tables and descriptive statistics, which will be informative about the difference between the two groups of households. It is important to be aware that this supplementary exercise is not part of the experimental analysis, and therefore, no causal implications should be extracted from it.

Figure 1.¹¹ Steps to select households receiving LD interventions



¹⁰Households in the bottom 40% have capability scores in the range of 0–8. To reduce the number of strata we will work with three capability score ranges: 0 to 2.5, 3 to 5 and 5.5 to 8.

¹¹ The capability score numbers in the figure are just illustrative. The cut-off score might be slightly different by kebele, since it is determined by the distribution of capability scores within each kebele.

2.3. Power Calculations and Sampling Strategy

2.3.1. Power Calculations

To calculate the required sample size, we considered a set of focal outcome indicators that satisfied the following criteria: were relevant to the research question, available, easy to collect, and covered the population groups of interest. Table 2 contains descriptive statistics for seven focal indicators for Amhara and Oromia, computed using data collected as part of the SPIR program endline evaluation.¹² These indicators are either directly related to food insecurity for the relevant populations or are indicators most likely to be affected by the LD interventions. There are four types of indicators: one indicator for food insecurity at the household level, one indicator for women, one indicator for children, and three indicators most likely to be affected by LD interventions. As will be shown below, the proposed sample size would allow the evaluation team to identify a reasonable minimum detectable effect for all of them. We are confident that the statistical power will be enough to identify changes in other outcomes of interest.

Table 2. Descriptive statistics of selected outcome indicators

Variable	Mean	SD	N HH
Household is moderately or severely food insecure based on FIES score	45.3%	50%	3775
Met Minimum Dietary Diversity for Women	7.9%	27%	3704
Children 6-23 months of age who meet minimum dietary diversity	1.24%	11.10%	720
Consumption expenditure per adult equivalent in the past month (Birr)	591	535	3313
Household owns any poultry (as reported by men)	66%	47.8%	3700
Household has any savings (with RUSACCO, VESA, VSLA, MFI, at home, with a relative)	44.7%	49.71%	3788
Household has taken out a loan for productive purposes in the last 12 months	26.1%	43.91%	3804

Note: SD stands for standard deviation. N. HH stands for number of households that participated in the SPIR endline evaluation survey.

The evaluation team ran a set of power calculations for each one of the seven indicators. In these power calculations, we assumed a minimum detectable effect (MDE) of 0.1 standard deviations and explored

¹² Outcome Data from SPIR evaluation endline; See Alderman, Harold; Gilligan, Daniel O.; Hidrobo, Melissa; Leight, Jessica; Ramani, Gayathri V.; Taffesse, Alemayehu Seyoum; and Tabet, Heleene. 2020. Impact evaluation of the strengthen PSNP4 institutions and resilience (SPIR) development food security activity (DFSA): Midline report

the minimum sample size needed to obtain such MDE. Table 3 presents those calculations. The following assumptions were used in their computation:

- Sample size will be based on a household level randomized design
- Base level (mean) and standard deviation (SD): values specified in Table 2
- Power level: 80%
- Confidence level: 95%
- MDE: 0.1 standard deviations

Table 3.¹³ Minimum required number of households for selected outcomes in Amhara assuming an MDE of 0.1 standard deviations.

Variable	Mean	MDE	Required number of households
Household is moderately or severely food insecure based on FIES score	45.3%	5pp	2,442
Met Minimum Dietary Diversity for Women	7.9%	2.6pp	2,850
Children 6–23 months of age who meet minimum meal frequency	67.91%	4.67pp	11,840
Consumption expenditure per adult equivalent in the past month (Birr)	591	53.5	2,476
Household owns any poultry (as reported by men)	66%	4.73pp	2,382
Household has any savings (with RUSACCO, VESA, VSLA, MFI, nak, at home, with a relative)	44.7%	4.97pp	2,492
Household has taken out a loan for productive purposes in the last 12 months	26.1%	4.39pp	2,600

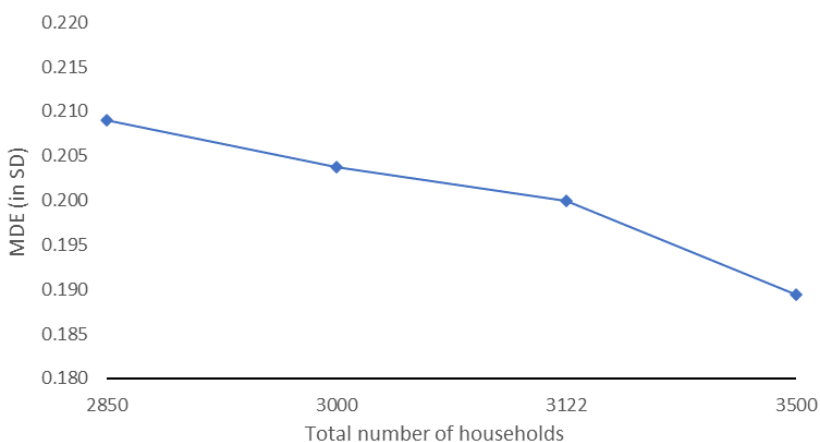
The last column of the table computes the minimum required sample size to be able to obtain an MDE of 0.1 standard deviations. We can see that for all the indicators, except the one concerning children, we would need a sample size not larger than 3000 to obtain the desired MDE. The high sample size for the children’s outcome is related to the fact that not each household has a child in the age bracket of 6 to 23 months. Based on data from SPIR endline evaluation and the 2016 Ethiopia DHS survey we assumed that around 19% of households have a child 6–23 months old.

Since obtaining an MDE of 0.1 standard deviations for children’s outcomes would require a sample size outside of the allocated budget, we explored the implications of using sample sizes of around 3,000

¹³ The MDE was computed using the command `power` in STATA. For continuous variables the command used was `power twomeans `var_mean', diff(`diff5' `diff10' `diff20' `diff30') power(0.8) sd(`sd') onesided`. For binary variables the command used was `power twoprop `var_mean', diff(`diff5' `diff10' `diff20' `diff30') power(0.8) onesided`. The variables ``var_mean'` and ``sd'` were directly obtained from the SPIR data.

households. Figure 2 shows the MDE (in standard deviations) for four different sample sizes. All the sample sizes would allow us to obtain an MDE of around 0.2 standard deviations.

Figure 2.¹⁴ MDE (in SD) for different total sample sizes for the outcome “Children 6–23 months of age who meet minimum dietary diversity”



Based on the different power calculations, we recommend a sample size of 3123 households. This sample size would allow us to estimate MDEs around 0.2 standard deviations for outcomes associated with children 0–23 months old. For the other outcomes, the MDEs are less than 0.1 standard deviations.

The 3,123 households refer to households that are part of the IE study (bottom 40% based on the capability score). In addition to these households, we will randomly sample 700 from the group of households with the highest capability scores (the upper 60%). This number was informed by budgetary concerns and is enough to allow the computation of basic comparisons and descriptive statistics.

The final step to determine the proposed sample size considers the possibility of attrition (non-response). Assuming an attrition rate of 10%, the study suggests sampling 4,235 households.¹⁵ The exact number of households sampled per kebele will depend on the number of eligible households per kebele.

2.3.2. Sampling Strategy

Based on the discussion in the previous section, the evaluation team will be sampling 4235 households in the 55 kebeles. This section will discuss the strategy used to sample households and individuals within the households.

¹⁴ All the power calculations portrayed in this graph were computed using the Stata command `twoprop `var_mean', n(2850 3000 3122 3500) power(0.8) one-sided (upper)`.

¹⁵ For households in the bottom 40%, $3,123 / 0.90 = 3,470$ households total, or 63.09 household per kebele if we were to sample the same number of households per kebele. We rounded the number of households to 3,465, corresponding to 63 households per kebele. Out of the 3,465 households, 1,732 correspond to control households and 1,733 correspond to treatment households.

For households in the upper 60%: $700 / 0.9 = 770$ households total, or 14 household per kebele if we were to sample the same number of households per kebele. Adding up to the two sets of households surveyed (3,465 and 770), we get 4,235 total households surveyed.

Households will be randomly sampled from the set of households that are part of the IE study (the bottom 40%). We will sample the same number of households from the control and the treatment group in each kebele to maximize the power of the design.

Household inclusion criteria:¹⁶ To be included in the sample, households must be a PSNP beneficiary, as these are the households targeted for the Food for the Hungry PReSERVE interventions. We will also limit the sample of households to those with women of reproductive age (15–49 years old), as these are target women for many indicators. Additionally, this strategy will increase the likelihood that the household will have a child under the age of 5.

This strategy will primarily exclude elderly households and is justified by the following reasons: (i) the majority of PSNP households have at least one woman of reproductive age¹⁷ and (ii) because of the nature of the LD interventions, elderly households are not likely to be substantially affected by those interventions or to be eligible to participate in them. The majority of the interventions, such as those related with nutrition, youth, and livelihoods, target households with children, youth, or working-age members. In addition, many nutrition and health outcomes are specific to young children or women of reproductive age. Accordingly, the sample selected will be representative of the households predominantly targeted by the PSNP+RFSA interventions and, more particularly, the LD interventions.

Within individual household selection:¹⁸ The evaluation team proposes to randomly select or purposively select one target individual for each outcome,¹⁹ as opposed to interviewing every eligible individual. The primary reason for this choice is that interviewing every eligible individual in the household is very time-consuming and costly not only for the household but also for data collection. For power reasons, we do not need more than one individual per household, and in general, outcomes would be highly correlated within households. Thus, the additional information provided is limited. Typically, interviewing multiple individuals per household is useful if the objective is to compare outcomes across individuals in the same household: for example, in polygamous households comparing outcomes for first versus second wives. However, this is not part of the IE design and thus, it is not worth the additional costs. Given the previous considerations, Table 4 provides more detail about the sampling strategy.

¹⁶ This sampling strategy was discussed with BHA and approved.

¹⁷ Based on authors calculations from a PSNP4 dataset, 82.5 percent of PSNP households had a woman of reproductive age.

¹⁸ This sampling strategy was discussed with BHA and approved. The sampling strategy will be applied to every surveyed household.

¹⁹ The within-household selection processes will be embedded in the survey tool. When a random member of the household needs to be selected, the survey tool will do the randomization using the household roster.

Table 4. Sampling strategy

Module	Sampling choice
D: Children’s Nutritional Status and Feeding Practices ²⁰	Prevalence of exclusive breastfeeding (children 0-5 months): randomly sample one child in this age range. Children’s feeding practices and diets (children 6-23 months): randomly sample one child in this age range. Children’s diarrhea (children 0-59 months): randomly sample one child in this age range
E: Women’s Health, Nutritional Status, Dietary Diversity, and Family Planning	Randomly sample one woman 15–49 years old
G: Agriculture	Select the person most informed about agriculture production in the household ²¹
J: Gender (Cash)	Select the adult most knowledgeable about household affairs and spouse ²²
K: Gender Access to Credit and Group Participation	Select the adult most knowledgeable about household affairs and spouse ²³

To sample the households, we will use the sample frame constructed by Food for the Hungry to select the eligible households. This dataset contains information on households’ members and their ages. For each kebele, households with women of reproductive age in the treatment and control group will be randomly ranked, and the first set²⁴ will be surveyed in each group.

2.4. Cost-Effectiveness Analysis

The CEA will focus on the cost-effectiveness of the LD implementation over the course of its entire implementation. This analysis will be viewed from the following dimensions:

- Impact results data measuring key outcomes of the LD interventions;
- Detailed cost per outcome calculations for selected and relevant units of key outcomes; and
- Perceptions of effectiveness from the implementers.

The conclusions will contribute to building a body of knowledge towards understanding the cost-effectiveness of resilience, food security, and emergency interventions.

²⁰ Causal Design will randomly select children 0-5 months, 0-23 months, and 0-59 months. This selection will be done independently, so that the same child could be selected twice (e.g., a child 8 months old could be selected for age bracket 6-23 months and age bracket 0-59 months).

²¹ This selection will be done by asking the household head to address issues around ownership and control. The survey will contain follow-up questions on specific individuals involved in different activities.

²² In the case that the household head is not married or in a union, the questions related to a couple will not be asked. In the case of polygamous households, we will randomly select from available wives.

²³ In the case that the household head is not married or in a union, the questions related to a couple will not be asked. In the case of polygamous households, we will randomly select from available wives.

²⁴ The number of households selected will be specific to each kebele and will depend on the relative size of the eligible households in the kebele.

2.4.1. CEA Perspective

The CEA will use the donor’s perspective of costs and outputs: often called the **program perspective**. While this perspective does not capture the true cost of providing economic development interventions to society (such as the cost of the farmer’s time or materials), it is a useful perspective for understanding the cost-effectiveness of programs. This perspective is often used to understand where cost savings can be achieved for the implementing organizations (IPs). It is also done to compare alternative development approaches for achieving the same outcome or output. From the program perspective, expenditures by external stakeholders (e.g., households) will *not* be included in the final cost-effectiveness analysis, but these will be considered and discussed as leveraged contributions from other parties.

2.4.2. Activity-Based Costing

We will aim to use activity-based accounting of project delivery costs.²⁵ This approach allows for a detailed, disaggregated understanding of the implementation costs by the interventions that make up an intervention. This approach will be facilitated by the total non-salary costs accounting done by Food for the Hungry PReSERVE and its implementing partners. This should provide more information on interventions where economies of scale or efficiencies can be achieved, which may be useful for future program design and management decisions and may also be helpful for understanding the cost of scaling any interventions.

Detailed discussions with Food for the Hungry (forthcoming) will determine how they plan to set up their financial system to ensure that we will be able to track costs according to interventions at either the purpose level or ideally the intervention level. To conduct the CEA appropriately, we will need to track costs at the intervention-level identified under each purpose (listed in the Appendix), though few implementers track their costs at the intervention level. At a minimum, we hope to disaggregate Activity Costs into at least the following categories:

- Purpose 1: Vulnerable HHs and Individuals Have Sufficient Quantity, Quality, and Diversity of Food at All Times
- Purpose 2: Vulnerable Community Members’ Livelihoods Transformed
- Purpose 3: PSNP Systems Deliver Accountable, Effective, and Shock-Responsive Services

Most of the LD activity and service costs will fall under Purpose 2; however, there are LD costs associated with interventions in Purpose 1 and Purpose 3. We will work with Food for the Hungry to try to separate out LD costs and particularly only those LD costs at the household and individual levels (as opposed to the kebele-level LD). We anticipate that this level of disaggregation (by interventions under each purpose and even by the level of implementation, such as household vs. kebele) is unlikely, and the analysis will focus on costs at the lowest level of activity disaggregation possible. This will result in several limitations that will be fully documented in the EL report.

We will collect administration costs separately (e.g., project staff salaries, equipment, supplies, fringe benefits). We will also try to include indirect expenses from headquarter staff or those who are not

²⁵ Sometimes referred to as bottom-up costing.

billable entirely to PReSERVE. However, these costs will not vary by activity and will be treated separately (see the analysis section below). Lastly, we will separate the refinement period and implementation period costs in the final analysis.

2.4.3. Period of Analysis and Real Figures for all Cases

The timeline for both the cost and the measure of outputs will be from the beginning of the BL evaluation data collection (2022) until EL data collection (2025). We will use 2022 as a base year (using real figures by removing inflation from the analysis).²⁶

2.4.4. Cost-Effectiveness Ratio

The analysis will evaluate the cost-effectiveness ratio (CER) of the incremental impact resulting from economic development funding. The CER will reflect the cost on a per-unit basis, with different measures of effectiveness for the units. These measures of effectiveness are incremental outcome measures that will come directly from the impact evaluation. The ratio takes the form below:

$$CER = \frac{Fixed\ Costs + Activity\ Costs}{Measure\ of\ Effectiveness}$$

A discussion on fixed and activity costs and which measures of effectiveness might be used are included in the CEA Analysis section below.

2.4.5. Qualitative Context for Interpretation CEA Results

To complement the CER results, we will also interview key stakeholders to better understand where they perceive any cost efficiencies to exist and what may have been driving those results. Once the CER results are available, interviews with project managers and other key stakeholders will explore qualitative aspects of implementation. This line of inquiry will help the evaluation team understand what aspects of implementation may have been rather expensive and any strategies for making the implementation more affordable. Given the likely limitations of isolating the costs for LD interventions from Food for the Hungry’s financial system, we will likely rely on interviews with Food for the Hungry to help us understand by how much our proposed CEA approach under- or overestimates the cost-effectiveness of the LD interventions.

Finally, we will leverage these interviews to understand whether the CEA results might be capturing non-RFSA or non-LD interventions and if that might lead to interventions appearing more or less cost-efficient than they would be in the absence of the actions of external stakeholders.

Key questions for the qualitative CEA inquiry include:

- Have the planned outcomes been achieved, and if not, why not? Was this due to implementation challenges or to other factors, independent of the project’s ability to deliver?

²⁶ Base year means the analysis will be done in 2022 USD. This implies that inflation in years after 2022 will be removed from the cost calculations, allowing for a comparison on real costs alone.

- Was the project able to leverage resources from other parties? What other costs were there (beyond implementing partners) that were incurred for PReSERVE’s achievements?
- What other interventions were operational in the same areas as PReSERVE that may have helped contribute to PReSERVE’s achievements?
- From your perspective, what aspects of this intervention were done cost-effectively? Why? Which aspects were not done cost-effectively? Why not?
- What were unexpected costs or relatively expensive costs in the PReSERVE program?

Additionally, we will thoroughly review project documentation (e.g., quarterly, and annual reports and the midterm evaluation) for possible cost variations and their causes.

3. DATA COLLECTION

All questionnaires will be drafted initially in English using Open Data Kit (ODK) software. After receiving BHA, IMPEL, and IPs approval, Causal Design will translate surveys into Amharic utilizing local partner staff.

3.1. Survey Design

Impact Evaluation: Causal Design intends to implement the same household questionnaire at both the BL and EL periods, with the exception of anthropometry, which will only be collected at the endline. The foundation of the survey will draw from selected BHA indicators from the BHA baseline/endline indicators handbook,²⁷ as confirmed by BHA, IMPEL, and IPs. Where required, surveys will be adapted to local context and adjustments will be made between survey periods. The questionnaires include a combination of the following modules:

- Module A: Household identification and informed consent
- Module B: Household roster
- Module C: Food access (e.g., FCS and FIES)
- Module D: Children’s nutrition and health
- Module E: Women’s nutrition, breastfeeding, and antenatal care
- Module F: Household water, sanitation, and hygiene
- Module G: Agriculture
- Module H: Household Poverty
- Module K: Gender Access to Credit and Group Participation
- Module J: Gender–Cash
- Module R: Resilience Module.

Community-level data: In addition to the household surveys, a short community-level survey will be conducted at the kebele level to gather information available at the local level, such as development and aid projects operating in the area and local public services.

²⁷ <https://www.usaid.gov/food-assistance/documents/ffp-indicators-handbook-part-i-indicators-baseline-and-endline-surveys-RFSA>

3.2. Outcome Indicators

The list of outcome indicators for the RSFA are in the appendix. These indicators reflect discussions with USAID-BHA and intervention programming that Food for the Hungry will implement in target areas. These indicators are also listed in the BHA baseline/endline indicators handbook referenced in section 3.1 above.

Anthropometric data: Due to the COVID-19 pandemic and the ability to evaluate using only EL data within the evaluation design, anthropometric data will not be collected as part of the baseline survey.

We are also defining primary and secondary outcomes in order to structure the analysis around a subset of outcomes that are of particular importance in assessing program impacts. The definition of the primary outcomes will particularly guide the analysis reported in any subsequent academic paper.

Primary outcomes include:

- The prevalence of moderate and severe food insecurity, based on the FIES
- The percent of households with poor, borderline, and adequate food consumption score (FCS)
- The percent of farmers who used financial services in the past 12 months
- Daily per capita expenditure
- Percent of people living on less than \$1.90/day
- The percent of women and men in a union who earned cash over the past 12 months.

These outcomes are identified as primary given that they relate to livelihoods outcomes, and this evaluation is designed to identify the effects of LD activities. Other indicators, enumerated in Table 8 (below), will be secondary outcomes. They will be reported in BHA deliverables and in the paper, but will not be primary.

3.3. Enumerator Training

Causal Design will engage a local data collection partner to recruit and train enumerators. Survey manuals and other training materials will be developed prior to the training period. The Causal Design team will monitor practice surveys during the training to verify comprehension and functionality of the instrument and the performance of the enumerators. Currently, Causal Design has scheduled (1) a pre-test focused on the survey tool, followed by 2) testing during enumerator training, and then 3) field piloting²⁸ before starting the survey process.

3.4. Data Management

Data protocols: Questionnaires will be drafted using ODK and all household survey data will be collected with electronic tablets utilizing SurveyCTO, a standard data collection application that allows for secure data storage and options for monitoring data quality. Causal Design staff will monitor incoming survey data to flag potential enumeration errors early in the data collection process. In accordance with best practices and regulation around human subject testing and data privacy, access to personally

²⁸ The pilots will take place in 3 kebeles outside of the area of study. These kebeles will be chosen so as to have similar characteristics to the area of study.

identifiable data will be limited and anonymized data will be utilized for analysis. Data management protocols will also be approved by a U.S certified Internal Review Board (IRB) and by the Ethiopian Society of Sociologists, Social Workers and Anthropologists (ESSWA) prior to any data collection, storage, and analysis.

Initial validation: Causal Design will work closely with IPs and BHA to review preliminary findings in accordance with the contextual validation activity outlined previously.

Quality assurance: Causal Design’s internal Data Management Protocol (see section 5) outlines the interventions and strategies that the research team implements to ensure that all data collection efforts meet industry and sector standards and expectations. This includes assurance that the data reflect high levels of USAID’s five data quality standards²⁹: validity, reliability, precision, integrity, and timeliness. These efforts are then mapped onto the wider phases of the project to demonstrate when they are being implemented and at what level. A full version of this protocol is available upon request.

3.5. Tentative Data Collection Timeline

The following table contains a list of all the activities related to data collection, as well as a tentative timeline. Note these timelines are preliminary and might be subject to change.

Table 5. Baseline data collection timeline

Activity	Completion Time
IRB Approval (Local and US)	May 27, 2022
Survey training and pilot	May 23, 2022 – June 7, 2022
Quantitative data collection	June 9, 2022 – July 27, 2022
Preliminary Indicator Tables	August 3, 2022

Table 6. Endline data collection timeline (preliminary)

Activity	Completion Time
Survey training and pilot	January 13 – January 31, 2025
Quantitative data collection	February 3 – February 28, 2025
Preliminary Indicator Tables	March 14, 2025

4. ANALYSIS

To assess the impact of the LD interventions over the course of the program, the Causal Design team will (1) present preliminary descriptive analysis utilizing baseline data, (2) conduct regression analysis to estimate the impact of the LD interventions in the PReSERVE RFSA utilizing both baseline (BL) and endline (EL) rounds of data, and (3) estimate the cost-effectiveness of the package of intervention.

²⁹ Conducting Data Quality Assessments | Program Cycle | Project Starter

4.1. Descriptive Analysis

The preliminary analysis using the BL data will show the extent to which the randomization was successful in achieving balance at BL. The survey will provide BL data on the status of communities and households across BHA standard indicators. The regression analysis can then correct for any imbalances by adding additional controls as needed.

4.2. Impact Analysis

The evaluation interventions will use Ordinary Least Squares or OLS³⁰ for continuous outcomes and linear probability models for binary outcomes. As needed, the research team will incorporate additional specifications that are meant to enhance statistical power, increase the validity of constructed comparison groups, or both. Examples include ANCOVA regressions that control for baseline levels of outcomes of interest.

The IE model will estimate the average effect of the LD interventions on relatively more vulnerable eligible households in 55 kebeles in Tay Gayint, Lach Gayint, Sahila, and Simada. The analysis will compare eligible households that received the LD interventions with eligible households that did not receive those interventions.³¹ The basic ANCOVA model:

$$(1) \quad y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 y_{ij0} + \varepsilon_{ij}$$

where y_{ij} refers to the outcome variable for individual or household i in kebele j at EL; T is the treatment dummy, which takes a value of 1 if the household was treated or 0 if it was not treated; and ε_{ij} is the error term. In this instance, y_{ij0} refers to BL values for the same outcome indicator.

$$(2) \quad y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 y_{ij0} + M_j + \varepsilon_{ij}$$

To account for the stratification, equation (2) adds a set of block dummies, M_j . Although the random selection of households the regressors in (1) and (2) are exogenous, we include a third specification controlling for various household-level X_{ij} and village-level X_j covariates in equation (3). This is because randomization will be conducted prior to the BL with limited information. Therefore, we cannot rule out the possibility of imbalance in some dimensions.

$$(3) \quad y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 y_{ij0} + \beta_3 X_{ij} + \beta_4 X_j + M_j + \varepsilon_{ij}$$

Sampling weights: In order for the evaluation team to include sample weights, we need to have a complete list of PSNP households with women of reproductive age for each study kebele. If we are able to obtain such a list, then the evaluation will include results of both weighted and unweighted estimations. Sampling weights will be calculated as the inverse of the probability of selection of the household in a given kebele. This will give us a representative sample of PSNP households with women of reproductive age in the 55 target kebeles; however, these 55 target kebeles are selected based on

³⁰ In statistics, ordinary least squares (OLS) and linear probability models (LPM) are methods for estimating the unknown parameters in a linear regression model. They are standard econometric methods used to establish and estimate empirical relationships between outcomes and a range of explanatory factors

³¹ As was specified in the section “Randomization strategy”, we will only be looking at eligible households in the bottom 44% according to their capability scores.

specific criteria mentioned above and not representative of all PReSERVE kebeles. Separate weights will also be calculated for indicators and adjusted to compensate for household and individual non-response. Given that sample weights are not needed to measure the causal impacts of the LD interventions, we will also conduct unweighted estimations.³²

The evaluation will include results of both weighted and unweighted estimations. Sampling weights will be calculated as the inverse of the product of the probabilities of selection from each of the stages of selection (e.g., woreda, kebele, household, eligible individual selection). Separate weights will be calculated for indicators and adjusted to compensate for household and individual non-response.

Standard errors and p-values: Standard errors will be clustered at the kebele level. Given the large number of outcome variables, it is important to consider that as the number of outcomes tested increases, the likelihood of finding a statistically significant effect when there is no true effect (Type I error) increases. To account for this, we will report both the standard p-values and the p-values corrected for Family-Wise Error Rate and the sharpened q-values corrected for the False Discovery Rate. To generate q-values, outcomes will be organized into outcome ‘families’ according to sector (i.e., food security, child nutrition and health, women’s health, WASH, agriculture, poverty, gender dynamics, and resilience).

Attrition and missing data: In the case of significant levels of attrition, BL data on originally selected households will be compared with BL data of households that are present at EL. The research team will be able to test if attrition (or non-response) was imbalanced (by regressing the attrition dummy on treatment status) and/or non-random (by regressing the attrition dummy on various outcome indicators measured at BL). If attrition was found to be non-random and imbalanced, we can construct Lee Bounds—a conservative measure of the upper and lower bounds based on the most extreme sample selection—or conduct Inverse Probability Weighting.

The analysis will not attempt to impute missing data points and responses will be ignored for the purposes of impact analysis. This will apply to questions where respondents refused to answer, stated an inability to answer, or are otherwise unable to respond. Cases of implausible data will be shared with the enumeration team to verify the validity of the response or understand the root of the error.

4.3. Supplementary Analysis on Resilience Indices

In addition to the impact analysis described above, we will conduct additional descriptive analyses on the BL 8 (Adaptive Capacity Index), BL 9 (Absorptive Capacity Index), and BL 25 (Transformative Capacity Index). This analysis aims to provide additional insights on which elements of adaptive, absorptive, and transformative capacities are driving the overall index scores to provide useful programming insights for Catholic Relief Services. The analysis will be conducted at endline for both baseline and endline adaptive, absorptive, and transformative indices in order to provide insights for Food for the Hungry on which capacities were relatively strongest and weakest at both points in time and which have seen the most growth over the evaluation period. We will report the weights on the index subcomponents generated from the principal components analysis (PCA) procedure to assess this. We will also assess

³² For a review of when sample weights are needed for causal estimates see Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. "What are we weighting for?." *Journal of Human resources* 50.2 (2015): 301-316.

how these weights compare to the overall distributions of each subcomponent to determine which subcomponents are most driving the overarching index scores.

4.4. Cost-Effectiveness Analysis

Measures of effectiveness for the PReSERVE evaluation are proposed below. These measures were identified based on how closely these results can be attributed to Food for the Hungry programming. At the same time, we have also tried to identify measures that can best capture the full extent of the interventions and its results, or at least reflect most of the interventions. These overall outcome measures (after endline data collection) could include costs per:

- Reduced incidence of people living on less than \$1.90 per day
- Household with reduced poor or borderline food consumption score (FCS).

All measures of effectiveness will come from the impact evaluation specified above and will reflect the incremental benefit attributable to LD funding (to the extent possible).

4.4.1. Cost Data

As indicated in the CER explained above, cost data will be captured for fixed costs and for activity costs in the year in which the expense occurs. Definitions and collection plans for each type of cost are outlined below.

Fixed costs: Fixed costs are those costs that do not change based on the implementation of interventions. Fixed costs include the salaries of the senior project management personnel (e.g., the Chief of Party), financial, contract, and monitoring and evaluation (M&E) staff. Additional fixed costs, such as rent, security, and utilities, that were incurred in Ethiopia will also be considered. We will not include non-Ethiopia fixed costs, which might slightly underestimate the costs of project implementation, but we believe the burden of collecting the data will be high. This data will need to be collected in cooperation with Food for the Hungry and its partners.

Activity Costs: Activity costs are operational costs. Data for these costs (and the associated activity-based costing) will be pulled from the PReSERVE financial database. Our team is confirming with Food for the Hungry that their financial data can be disaggregated according to the interventions identified in section 2.4.2 (above). To the extent possible, we will provide detailed cost data for the materials that were required for implementation of each activity (e.g., labor, transportation, infrastructure). We will also perform the analysis for all implementation costs, and only those after the refinement period.

Costs will be collected from Food for the Hungry as well as its partners.

4.4.2. Cost-Effectiveness Analysis Results

As the Impact Evaluation is focused on measuring the impact of the LD interventions compared to a control group, or population without access to LD interventions (at the household or individual level), the associated CEA will only be able to produce one cost-effectiveness ratio per outcome measure

identified above.³³ For example, our results may suggest that it costs \$15 per household with an improved FCS score as a result of PReSERVE LD interventions. Without another treatment arm, there is no immediate relevant comparison group to compare this figure to. As a result, the question becomes whether \$15 per household is relatively cost-effective.

To answer that question, we would need to compare this cost-effectiveness ratio to another cost-effectiveness ratio for a similar program with the same outcome measure. Given the impact evaluation's narrow focus on only LD interventions, this may limit the comparability of this CEA with those being constructed for other RFSA impact evaluations. Nonetheless, CEAs of other external programs will be reviewed to determine if these results may help characterize or provide some insights about the relative effectiveness of PReSERVE's LD intervention:

- CEAs from other RFSA impact evaluations in Madagascar which will have similar cost accounting and outcome measures. Other RFSA impact evaluations and accompanying CEAs may also be comparable (e.g., in Malawi and Uganda).
- We anticipate that we can compare the PReSERVE CEA results to the CEA results for the second RFSA that Causal Design is evaluating in Ethiopia, which will use the same outcome measures and cost accounting, though the programming that is measured under the impact evaluation is different.
- Finally, we will review literature to identify other similar programming, such as Feed the Future activities, that were subject to an impact evaluation and measured cost-effectiveness using the same outcome measures. We believe this specific criterion will limit the amount of evidence that may be available for comparison. However, it may be possible to impute cost-effective ratios with available information (e.g., we may be able to create our own cost-effectiveness ratios if impact evaluations provide overall program costs and the same outcome measures).

There will be limitations to comparing LD CEA results directly to any of these other programs listed above, specifically related to the similarity of the programming and the context in which each activity operated. All limitations will be discussed in the final report.

4.4.3. Qualitative Data

We intend to complement our understanding of these results in interviews with key stakeholders (in the form of key informant interviews). These interviews will take place after the cost-effectiveness analysis is complete and contingent on the availability of the stakeholders for interviews. Selected interviewees for key informant interviews will be individuals with relevant experience and who are knowledgeable about project implementation and the associated costs to provide rich insight. Currently, we anticipate these key informant interviews to be held with project management and possibly USAID personnel who are very familiar with the implementation of the interventions, but we may also include external stakeholders or direct participants. These interviews will be semi-structured interviews, driven by the methodology questions identified earlier in this document, as well as the results of the CEA analysis.

³³ This may be possible if multiple treatment arms are examined for different IWM+ and if it is possible to disaggregate the costs by these approaches. Whether this is possible is still being explored.

5. APPENDIX

5.1. Data Management

The objective of this section is to provide detailed guidance towards Causal Design’s policy and protocols when storing, coding, and reporting data collected or shared with any staff member. All staff members including permanent salaried staff, permanent/part-time consultants, and previous staff are bound to uphold these agreements as part of their employment agreement with Causal Design, as indicated in the employee handbook.

If any violations or accidental sharing of information that is not encrypted is mistakenly shared outside of Causal Design. The staff member shall immediately notify the Chief Privacy Officer (Keith Ives, also CEO) and the appropriate notifications will be sent to the IRB, clients, and any study participant whose data has been compromised.

This handbook drawn from an array of resources around data management and data quality assurance mechanisms including:

Handbook for Safeguarding Sensitive Personally Identifiable Information. Department of Homeland Security, March 2012

Callahan, Mary Ellen. Handbook for Safeguarding Sensitive Personally Identifiable Information, Handbook for Safeguarding Sensitive Personally Identifiable Information § (2012).

“Research Protocols.” Innovations for Poverty Action, August 23, 2018. <https://www.poverty-action.org/researchers/research-resources/research-protocols>.

Chuang, Erica, Harrison Diamond Pollock, and Stephanie Wylstra. “Reproducible Research: Best Practices for Data and Code Management.” Innovations for Poverty Action., November 2015.

5.2. Data Quality

- Create inception plan before launching survey operations: The inception plan is an operational plan that covers timelines, staffing needs, logistics, and procurement for your survey, at all stages, including questionnaire development, training, piloting, tracking, interviews, and quality assurance. Your inception plan must be in line with your budget(s); for example, you cannot survey more respondents in the baseline than your budget estimated — without overspending during your endline.
- Create data quality assurance plan and materials before launch: The data quality assurance plan lays out in detail the requirements for backchecks, high frequency checks, accompaniments, spot checks, and any other data quality assurance activities. The scope of the data quality assurance plan should not only include technical products, but also data flow, roles and responsibilities, reporting schedules, actionable items based on output, and incentive programs for the field team. It also includes your staffing needs, which may change over the course of the survey.

- Bench test survey (ideally at least 2 weeks in advance): Bench testing means testing your survey in the office with a minimum of three different testers. You will save time and money by making sure your survey works well BEFORE launching field data collection. Bench testing is an iterative process wherein testers run the survey in different scenarios and provide feedback, while the programmer(s) make changes; note that even small changes to a survey must go through the bench testing process again, as it is easy to make mistakes that affect other parts of the survey. This process works best if the "paper" survey is considered mostly complete and has already been reviewed by central decision-makers on the project.
- Pilot survey (ideally at least 1 week in advance): Every survey must be piloted prior to the beginning of the survey in communities outside your study sample. Your pilot should look as close to actual surveying as possible — you may even decide not to tell your field team it is a pilot. Ideally, every question that is included in the final survey should be piloted prior to launch. For surveys using Digital Data Collection, a pilot should include field testing of both the survey program and devices. Remember to leave time to make corrections of errors you identified during piloting.
- Accompany surveyors in first week of survey: Field supervisors must accompany a subset of field officers' interviews to monitor field officer performance and to check for survey issues. All field officers must be personally accompanied at least once during the first week of the survey. Accompaniments can be scaled down as the survey progresses, especially by leveraging digital supplements like audio recordings and meta-data.
- Implement and act on high frequency checks: High frequency checks provide insight into ongoing field team and data quality concerns before they become too entrenched or too late to manage. By running HFCs, you can regularly analyze (comparative) field officer performance, compliance with ethics requirements, response frequencies and outliers, duplicates, and other project-specific data quality issues. HFCs are meant to provide the evidence needed to successfully guide and manage a field team on a daily basis, and thus, must be accompanied by strict guidance on roles and responsibilities, reporting schedules, and triggered actions (e.g., what outliers would trigger re-interviewing a household).
- Implement and act on backchecks: A backcheck (also known as a field audit or re-interview) refers to when a highly qualified field officer (also known as a back-checker) visits a respondent a second time to re-administer a selection of questions from the original questionnaire. Those backcheck responses are then compared to the original responses. An IPA generated code `bcstats` program can be used to identify discrepancies between answers, and thus to identify problems with the questionnaire, field team, or both. Your quality assurance plan should have included a backcheck randomization plan, as well as an action plan for what to do when you encounter discrepancies.
- Double enter & reconcile paper surveys: Although paper surveying is now uncommon, there are strict protocols for data entry from paper surveys. Each survey must be entered by two separate data entry operators who cannot compare responses. When there are discrepancies between their entries, they must be reconciled by a third data entry operator who looks at the original survey closely. In-house data entry can be replaced by online firms, which also provide double entry and allow for you to review discrepancies against the original survey responses.

5.3. Data Security & Research Ethics

If the IRB is used on your project:

- The Principal Investigator (PI) is responsible for maintaining Institutional Review Board (IRB) approval throughout project lifecycle (e.g., submissions, renewals, amendments, human subjects certificates): Any study conducting human subjects research must have the approval of at least one IRB; note that each project is different, so you should consult with your PIs and IRB Coordinator about how best to get IRB coverage for your project. A typical lifecycle includes approval of the initial research protocol, annual renewals, and amendments when critical items change, such as the questionnaire, staffing, research protocol, or risk level. All project staff, partners, and investigators who can see encrypted personally identifying information (PII) must have up-to-date human subjects' certificates. Any deviation from the protocol or any unexpected risk to respondents must be reported as unexpected events to the IRB. Use Salesforce to keep track of all IRB approvals and upcoming renewal dates.
- Retire your project with all IRBs once the project is complete: Once your study is complete, you should retire or otherwise officially close out your IRB with all the reviewing IRBs. For the Causal Design IRB, you should retire your study when (A) all study interventions and activities are complete and (B) you are no longer actively, regularly working with identified data. Other IRB(s) may have slightly different standards or procedures, so you should check with your reviewing IRB administrator(s) where relevant as well.

Whether the IRB is used on your project or not:

- Create data security plan and set up encryption (using Whisp.ly to transfer between partners <https://whisp.ly/en?>) before launch: Respondents' confidential data should be encrypted **at all stages**, starting at the moment of data collection. This includes while it is on the data collection device, during wireless transmission, while on an external server (e.g., Kobotoolbox, Commcare, SurveyCTO, etc), when it is on a cloud storage system (e.g., GoogleDrive or Dropbox), and while on laptops and removable media (hard drives, flash drives). Any time the data is stored on a server that is not controlled by Causal Design, it must be separately encrypted so that the company that controls the server cannot access the data. You must plan beforehand how you will ensure encryption at each of these steps and how it will be maintained after your project has been officially closed if you are retaining any PII. **If you are using any IRB any un-encrypted data is uploaded to the cloud or e-mailed, you must file an unexpected event report to your IRB(s) and comply with any ruling they make. If you are not using an IRB, you should report this to the Chief Privacy Office of Causal Design, Keith Ives.**
- Maintain data security plan (especially encryption) throughout project lifecycle: At every stage of the project lifecycle, data should be properly protected. Among other things, this means personally identifying information (PII) should remain encrypted during storage and transmission, and passwords should be restricted to the critical members of your research staff.
- Use new UID in deidentified dataset: When you share or publish un-encrypted data, it must be deidentified, i.e., there must be no identifying information in the dataset, such as name or address, or a combination of variables that can be used to identify a respondent. You should also replace your original unique identifier (UID) with a new unique identifier. You should do this

at the end stage of your project, when you have finished matching across waves or different data collection activities.

5.4. Knowledge Management & Transparency

- Back up data in at least two locations: There must be at least two copies of the data available at all times. During data collection, this will likely mean on a KoboToolBox/SurveyCTO/CommCare server, as well as on a laptop and synced to Google Drive. Do not delete server data until it has fully synced to Google Drive as a protection from laptop theft. Post-data collection, this could mean backing up your data on an external hard drive on the extremely rare chance that a major cloud service like Google Drive fails.
- Save ALL project files to and ONLY to Google Drive: Causal Design project files must be stored in the My Drive\3_CD_Projects superstructure on Google Drive. This includes in particular: raw data files, final versions of questionnaires, back check questionnaires, survey manual, project log and survey notes, high frequency check files, analysis do-files, IRB documentation, and replication code.

5.5. Data Storing/Sharing Guidelines

The following bullets are intended for projects which are completed and are going to be stored long term on the Google Drive or any other survey.

5.5.1. Detailed Steps for Preparing Data and Code:

Remove personally identifiable information (PII): Check thoroughly for PII, and make sure to remove before sharing with the data repository team.

- All direct identifiers, such as unique IDs (social security numbers, bank account numbers, and so on), should be removed before storing. Indirectly identifying data such as combinations of variables which could uniquely identify participants should also be considered carefully before storing or sharing data.

Include clear variable labels and code value labels:

- Make sure that variables are clearly labeled.
- If it is a variable collected directly from the questionnaires, indicate this with a question number. If it is constructed, either include the construction in the name or label or, if complex/lengthy, include additional information in notes.
- Ensure that value code labels are provided, as they are needed for interpreting the data.

Include code file(s) with headers/comments:

- **Headers:** Include header with name of person who last wrote/edited the code, date, and software used (package and version).
- **Comments:** Use comments in the code to indicate which tables are produced.

Prepare Readme files:

- Please indicate: 1) which files are included in what is shared and 2) how data and code files relate (i.e., what code runs on which data, to produce which outputs). We have a template for readme files that we are happy to share and is located on every project folder.

Include survey instruments:

- Ensure that you are sharing the final version used to collect the data.

5.5.2. Data Curation Steps That Data Repository Staff Will Complete

As the data repository team works on the dataset submitted, we will conduct the following three steps to ensure the quality of the materials that we share in our repository.

Confirming there is no personally identifiable information (PII) shared in data or code files

- It is the responsibility of the original researcher (s) to ensure that PII is removed. IRB protocols do not permit sharing PII with the data repository team. However, the DR Unit will double-check that PII is removed before sharing because of the high level of importance of maintaining confidentiality of research participant’s information.

Examining data and code for usability:

- The data repository team will examine variable names and labels, value codes, and the statistical code. As a part of sharing high-quality data, we will attempt to fill in variable labels and/or notes in the dataset where we are able to glean further information from published tables or communication with researchers. Where there are a large number of unclear variables, we may ask the researcher(s) to improve the dataset before publishing.
- We will run the statistical code to ensure that it produces the published tables.

Checking and sharing related materials:

- Supplementary readme file: As we conduct our data curation steps, we will track and share information that will help site users understand the steps that we took and what we found. We will confirm with the original researcher before sharing this file along with the data.
- Study-level metadata: We have created a custom template with fields that we will fill in from all studies.

5.5.3. Project Language for Quality Assurance and Control

The following plan outlines the activities and strategies that the research team intends to put in place to help ensure that the data collection for the IMPEL meets industry and sector standards and expectations. This includes assurance that the data reflect high levels of the following dimensions: validity, reliability, precision, integrity, and timeliness (USAID 2016). These efforts are then mapped onto the wider phases of the project to demonstrate when they are being implemented and at what level.

Table 7. Data quality assurance activities

Project Phases	Data Quality Assurance Activity	Quality Dimensions
Phase 1: Discovery and Design	Literature Review and Sector Assessment	Validity
	Indicator Mapping	Validity
	Questionnaire Designed to Promote Proper Response Coding	Integrity
Phase II: Collection	Integration of Data Collection Activities with Existing IMPEL staff capacity	Reliability/Integrity
	Develop Data Collection Protocols and enumerator training	Reliability
Phase III: Analysis	Preliminary Data Spot Checks	Integrity
	Enumeration team review and Feedback	Validity
Phase IV: Reporting	Scheduled Analysis and Reporting	Timeliness

Validity

The research team will work closely with IMPEL and BHA project staff to ensure that the indicators and research design are valid measures. This is primarily addressed through efforts leading up to the proposed design of research activities. In this case, the research team combines the following to ensure that the proposed indicators and methods are valid for the scope of the research:

- **Literature Review and Sector Assessment** of current thinking and practice focused both on wider academic and implementation-based publications and on IMPEL specific reports and projects related to measurement; and
- **Sector Experts** feedback and consultation is included into all phases of the baseline.

Analysis created as a result of research efforts will undergo stakeholder review to further ensure that findings are interpreted correctly and account for contextual realities.

Reliability

The research team will also ensure that protocols are put in place to ensure consistency in data collection efforts. This includes the creation and implementation of training (if necessary), sampling, and data collection protocols, which undergo internal peer-review.

Precision

At the outset, the project will build on efforts to ensure data *validity* and utilize the literature review, desk research, and project documentation to comment on and revise our *analysis plan* that connects theories of change pathways to research objectives.

Integrity

Data integrity within the IMPEL program is delivered through specific systems and processes that manage data entry and safeguards to ensure proper data input.

The data input will rely on the following ecosystem:

- The questionnaire will be designed to provide clear instruction on *proper response coding*;
- Daily updates to our server will ensure proper data input by centralizing data input across data collectors; and
- The research team will conduct *spot checks on data* taken during the population survey.

Timeliness

In order to ensure data timeliness, the research team has created a project work plan to ensure that scheduled analysis quickly follows data collection in order to guarantee that relevant findings can inform project implementation decisions and strategy.

Security

In order to ensure that the data is secure, Causal Design will use SurveyCTO’s encryption system to create a private key. Only the evaluation team and the data collection field manager will have access to this key. Only data that has all PII removed will be shared to ensure respondent privacy.

5.5.4. Quality Assurance

During the implementation of this research, our Team Lead, Reimar Macaranas, supported by Causal Design’s Project Management Office, will use state-of-the-art enterprise resource planning software to manage the project timeline, budget, and resources to ensure high-quality, on-time delivery of all work products. Causal Design uses Intuit’s suite of programs, which integrates timesheets, accounting, staff availability, budgeting, and project management functions to provide integrated access to all information needed to effectively manage projects. Mr. Macaranas will adhere to Causal Design’s policy that any changes to implementation plans, or timelines are immediately updated in this system, to ensure we can always provide a real-time estimate of the expected resources necessary to complete a task or project, including both staff time and budget.

Causal Design also understands the paramount importance of QA/QC on all work products and technical deliverables, and of effective and frequent communication between the Team Lead and IMPEL. Causal Design’s “no surprises” policy requires all project managers to keep clients regularly informed about progress, challenges, solutions, and concerns. IMPEL will therefore always be fully informed of all relevant activities and immediately consulted when guidance is needed. This policy ensures that the Causal Team and KWSH will be partners in critical decision-making on and problem-resolution in all matters.

5.5.5. Quality Control

Quality Control for all products will be managed by Mr. Reimar Macaranas. His academic training, years of leadership in research and evaluation, and role as Chief Operations Officer will be utilized to ensure all the Team’s products meet or exceed the expectations of IMPEL. Mr. Macaranas will also provide executive-level oversight and senior technical review of all project tasks and deliverables. He will ensure IMPEL has access to the Causal Team’s key technical personnel that can answer questions at any time. He will verify that Causal Design’s rigorous QC procedures are implemented and ensure that all deliverables submitted to IMPEL meet the highest quality standards and require minimal rounds of

revision. These quality control systems will ensure that the Causal Team provides the highest possible quality services to IMPEL with minimal service disruption.

5.5.6. Data Processing and Procedures

Quantitative data will be collected using tablets and stored in a secure cloud-based server; analysis will be done using STATA. Causal Design will manage team for doing the data clean up, data entry, data analysis, and reporting.

5.5.7. Ethical Considerations

We will ensure that our team, including all enumerators and contractors working on the project, adhere to the ethical guidelines outlined in the American Evaluation Association’s Guiding Principles for Evaluators. The Causal team has experience in preparing IRB protocols for evaluations. For many evaluations, we have successfully worked with Solutions IRB to obtain IRB clearances on domestic and international studies.

After recruiting household survey participants, we will obtain informed consent for each person to be interviewed. We will explain the purpose of the study, the topics of the interview/focus group, the person’s rights as a participant, including that their responses will remain confidential, and that participation is voluntary. We will provide contact information for the study investigators and appropriate IRB(s) (if used). The data collectors will review the information to be collected. We will use plain language and translate into Khmer. Participants will provide oral consent. The Causal Team will work with IMPEL to obtain any necessary national or local IRB clearances as appropriate.

5.6. Role of the Implementing Partner

Causal Design will work closely with Food for the Hungry to ensure the feasibility and success of the IE study. Food for the Hungry will provide support during the different stages of the study. Some of the roles of the IP are listed below:

- Provide a list of the different interventions.
- Provide a list of the activity’s actual implementation communities.
- Provide a survey sampling frame and study sites.
- Review the baseline and endline study protocol.
- Brief the research team on the political, social, and cultural norms and context in which the baseline and endline data collection will take place. This will guarantee that the data collection approach is informed by “do no harm.”
- Introduce the research team to the communities prior to data collection.
- Observe enumerator and supervisor trainings.
- Participate in periodic conference calls to receive updates on data collection.
- Review and comment on draft reports.

5.7. Intervention Packages

Table 8. Intervention packages³⁴

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
Purpose 1: Vulnerable households and individuals have sufficient quantity, quality and diversity of food at all times					
Commodity	Distribute full package food commodities to PSNP	Supportive service		X	
Irrigation	Construct family hand dug well for irrigation	LD intervention		X	
	Train farmers on irrigation based improved agronomic practices	LD intervention		X	
	Link irrigation user households with seed suppliers (agro-dealers/others) to buy seed for planting	LD intervention		X	
	Construct water harvesting structures	LD intervention		X	
	Provision of fruit and fodder seedlings from central nurseries	LD intervention		X	
	Provision of package of pumpkin, watermelon, kale, etc.	LD intervention		X	
Conservation Agriculture	Organize training and exposure visit to DAs and promoters on conservation agriculture techniques and improved agronomic practice	LD intervention			X
	Organize training and exposure visit to selected households on conservation agriculture	LD intervention		X	
	Support farmers to practice conservation agriculture techniques on their farm fields on selected crops (haricot bean, potato, vegetables, fruits)	LD intervention		X	
	Integrate natural resources conservation with no-tillage agriculture (high-value perennial crops, Apiary)	LD intervention		X	

³⁴ The interventions that are targeted for individuals or households are indicated as LD interventions. There are some of these interventions whose level of implementation is the kebele level. Those interventions can be provided to treatment or control group, or neither of the two (i.e. provided to non-PSNP households) and will not be evaluated as part of the IE study.

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
	Train Food for the Hungry promoters and DAs on forage/fodder production strategies, utilization and marketing	LD intervention			X
	Form and train group members on fodder seed and feed production, utilization and marketing	LD intervention		X	
	Support farmers to practice improved forage production and utilization technologies (improved variety seeds and techniques)	LD intervention		x	
	Support farmers practice improved poultry management technologies (poultry housing, feed, etc. for local poultry breeds) to increase ASF production	LD intervention		X	
	Establish poultry production groups to provide for their members	LD intervention	X		X
	Train farmers on HH level poultry farming and feed preparation	LD intervention	X		X
	Train DAs and promoters on general animal care, improved livestock housing, medical conditions, and treatments				X
	Train farmers on general animal care, improved livestock housing (for poultry and ruminants), medical conditions, and treatments		X		
Nutrition	Train WHDAs on preparation, preservation, and use of local foods	Supportive service			X
	Conduct regular NCG sessions	Supportive service	X		
	Conduct home visits for tailored counseling and identification of SAM/MAM cases	Supportive service	X		
	Conduct referral of SAM/MAM cases	Supportive service	X		
Gender	Facilitate discussion sessions on HTP using gender club's comic book for schoolboys and girls		X		X
	Promote positive social norm development through GOG discussion session for community members	Supportive service		X	
	Transitioning of GOG to GMF	Supportive service		X	

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
	Train male advocates from GMF to promote gender equality	Supportive service			X
	Conduct gender club discussion session using comic book	Supportive service			X
	Cascade male involvement training for male advocates	Supportive service			X
	Broadcast gender equality audio, video messages and public demonstration community members	Supportive service			X
WASH	Train volunteers on production of age-appropriate play objects	Supportive service			X
	Establish community-based playground	Supportive service			X
	Train WHDAs on baby WaSH module	Supportive service			X
	Conduct community conversation on Baby WaSH module	Supportive service		X	
	Facilitate construction and use of handwashing stations at model households	Supportive service		X	
	Facilitate construction and use of improved latrine with handwashing facilities for all LD households	Supportive service		X	
	Train members of traditional groups (CBOs) on latrine quality improvements	Supportive service			X
	Construct and rehabilitate drinking water sources for rural communities (springs and wells)	Supportive service	X		X
	Establish rural piping system for human consumption	Supportive service	X		X
Purpose 2: Vulnerable Community Members' Livelihoods Transformed					
GRANT	Facilitate livelihood transfer to ultra-poor PSNP clients	LD intervention	X		
	Train LH transfer recipients on business management	LD intervention	X		
FRUIT	Handover nurseries to youth entrepreneurs	LD intervention	X		

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
	Link PSNP HHS with High value tree (fruit & others) nurseries to access fruit seedlings	LD intervention	X		
CREDIT & SAVING	Organize VESA groups	LD intervention	X		
	Link VESAs with MFIs and Unions for Informal Apprenticeship and credit access	LD intervention	X		
	Support initial materials for VESAs	LD intervention	X		
	Train VESA leaders on saving and financial management	LD intervention	X		
	Facilitate regular VESA discussion	LD intervention	X		
RuSACCOs	Channel guarantee loan fund to RuSACCOs	LD intervention			X
	Link individuals in the VESA groups with RuSACCOs	LD intervention	X		
	Link RuSACCOs with MFI to access additional lending capital	LD intervention			X
YOUTH	Establish Youth Economic Strengthening (YES) Centers	LD intervention			X
	Provide loan fund for grantees selected male and female youth	LD intervention	X		
WAGE-BASED EMPLOYMENT	Provide Behavioral (soft) Skills Training for male and female youth	LD intervention	X		
	Provide Vocational Skill Training for male and female youth	LD intervention	X		
	Mentor and graduate for male and female youth	LD intervention	X		
OFF-FARM	Provide Behavioral (Soft) Skill Trainings for male and female youth	LD intervention	X		
	Provide Technical Skill Training for male and female youth	LD intervention	X		
	Provide Business Development Training for male and female youth	LD intervention	X		
NON-FARM	Provide Behavioral (Soft) Skill Trainings for male and female youth	LD intervention	X		

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
	Provide Vocational Skill Training for male and female youth	LD intervention	X		
	Provide Business Development Training for male and female youth	LD intervention	X		
	Facilitate 'Work Based Learning' (WBL) for male and female youth (OJT, On Job Training (OJT), Job Shadowing, Apprenticeship, etc.)	LD intervention	X		
	Provide coaching and mentoring services for male and female youth	LD intervention	X		
Value chain	Facilitate improved access to market information	LD intervention		X	
	Train producers' associations on improved marketing system	Supportive service			X
	Form marketing groups of producers	LD intervention			x
	Facilitate engagement in selected value chains	LD intervention	X		
	Potato value chain	LD intervention	X		
	Haricot bean value chain	LD intervention	X		
	Fish value chain	LD intervention	X		
	Sheep value chain	LD intervention	X		
	Goat value chain	LD intervention	X		
	Organize localized trade events to connect producers and buyers	LD intervention			X
	Provide information to producers (quality requirements, pricing, timing, location, payment terms, etc.)	LD intervention	X		
	Facilitate the establishment of one stop shopping center a main woreda town	LD intervention			X
	Facilitate business linkages with output buyers	LD intervention			X

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
	Facilitate business linkages with input suppliers	LD intervention			X
	Identify local service providers (such as AWEA, Chamber of Commerce, Coops, etc.) for market linkage activities	LD intervention			X
	Co-create market linkage services with local providers (that respond to gaps identified in VCAs)	LD intervention			X
	Market actors supported to align quality of production to market needs	LD intervention			X
	Train individuals in institutions on gender and value chain	LD intervention			X
	Train producers' associations on improved marketing system	LD intervention	X		X
	Upgrade local market centers	LD intervention			X
	Assess the capacity of local groups, or local service providers, to implement the farmer group market development strategy	LD intervention			X
	Design warehouse receipt pilots and train stakeholders on management of the various roles in a warehouse receipt activity	LD intervention			X
Other IGAs	Facilitate engagement of LG HHs into various on-farm, non-farm and off-farm	LD intervention	X		
	Fattening	LD intervention	X		
	Honey production	LD intervention	X		
	Production of wheat, teff, chickpea, etc	LD intervention	X		
DEMONSTRATION	Demonstrate appropriate technologies in the FTCs	LD intervention			X
	Demonstrate appropriate technologies in schools	LD intervention			X
	Create a model ecosystem in watersheds with integrated technologies (crop, natural resource management, and livelihood technologies)	LD intervention			X

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
EXTENTION SERVICE	Train educated youth female as extension promoter to provide extension service for women	LD intervention			X
	Support farmers with technologies adoption	LD intervention	X		
MARKETING	Form marketing groups of producers	LD intervention	X		
	Train producers and marketing groups in group marketing, financial management, bookkeeping, etc.	LD intervention	X		
	Train women producers on assertiveness and negotiation skill	LD intervention	X		
	Link marketing groups with markets	LD intervention	X		
POST-HARVEST	Train producers on improved post-harvest handling	LD intervention	X		
	Train youths on sorting, grading, packaging, and storage	LD intervention	X		
	Link youths with national and international markets	LD intervention	X		
Purpose 3: PSNP Systems Deliver Accountable, Effective, and Shock-Responsive Service					
	Train youths and women on nursery management	LD activity	X	X	
WATERSHED	Treat degraded areas with physical SWC measures	Supportive services	X	X	X
	Treat degraded areas with biological SWC measures	Supportive services	X	X	X
	Facilitate livelihood development in potential watersheds	LD intervention			X
	Train user groups on technical areas and supportive skills (financial, saving and credit and life skills)	LD intervention	X		
	Link mature watersheds for livelihoods activities by user groups	LD intervention	X		X
	Develop integrated, climate-smart, gender, and nutrition-sensitive annual public works plans that contribute to livelihood productivity	LD intervention			X

Purpose/ Intervention	List of LD interventions or Supportive services	Types of service	Level of Implementation		
			Ind.	HH	Kebele
	Construct small-scale reservoirs such as community ponds for perennial horticultural production	LD intervention			X
PRIVATE NURSERY	Establish/Strengthen private nurseries	LD intervention	X	X	
	Seedlings produced by private individuals	LD intervention	X		
	Train youths and women on nursery management	LD intervention	X		
	Establish youth groups as suppliers of SWC tools and seeds	LD intervention	X		
	Train youths on constructing SWC tools and seeds	LD intervention	X		
LINKAGE	Conduct BCC sessions during Public Work activities	Supportive services	X		
	Train model PSNP clients to work as nutrition champions	Supportive services	X		
	Facilitate awareness creation sessions on existing essential services	Supportive services	X		
DRR	Train DRR committees and CBO leaders on linkages, information exchange and timely response	Supportive services			X
	Update/develop community DRR plans	Supportive services			X
	Conduct soft skill trainings to selected CBOs leaders on their roles in Development and DRR responses	Supportive services			X
	Establish Community Food Contingency Reserves (CFCR) to respond to localized shocks	Supportive services			X
Participatory Monitoring	Market Information Analysis	Supportive services			X
	Resilience monitoring	Supportive services			X

5.8. Indicators List

Table 9. Food for the Hungry PRESERVE indicators list

BL #	RFSA Indicators	Relevant Modules
BL 6	Prevalence of moderate and severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)	C
BL 10	Percent of households with poor, borderline, and adequate food consumption score (FCS)	C
BL 12	Prevalence of children 6–23 months receiving a minimum acceptable diet (MAD)	D
BL 13	Prevalence of exclusive breastfeeding of children under 6 months of age	D
BL 14	Percent of children under age 5 who had diarrhea in the prior 2 weeks	D
BL 15	Percent of children under 5 years old with diarrhea treated with Oral Rehydration Therapy	D
BL 39	Prevalence of children 6–23 months consuming a diet of minimum diversity (MDD-C)	D
BL 11	Percent of women of reproductive age consuming a diet of minimum diversity (RiA)	E
BL 26	Percent of births receiving at least four antenatal care (ANC) visits during pregnancy	E
BL 36	Percent of women in a union who have knowledge of modern family planning methods that can be used to delay or avoid pregnancy	E
BL 37	Percent of women in a union who made decisions about modern family planning methods in the past 12 months	E
BL 16	Percent of households using basic drinking water services	F
BL 17	Percent of households with soap and water at a handwashing station on premises	F
BL 18	Percent of households in target areas practicing correct use of recommended household water treatment technologies	F
BL 19	Percent of households in target areas practicing open defecation	F
BL 27	Percent of households with access to a basic sanitation service	F
BL 21	Percent of producers who have applied improved management practices or technologies	G

BL #	RFSA Indicators	Relevant Modules
BL 22	Yield of targeted agricultural commodities within target areas ³⁵	G
BL 29	Percent of farmers who used financial services (savings, agricultural credit, and/or agricultural insurance) in the past 12 months	G
BL 30	Percent of farmers who practiced the value chain interventions promoted by the activity in the past 12 months	G
BL 1	Prevalence of Poverty: Percent of people living on less than \$1.90/day	H
BL 2	Depth of Poverty of the Poor: Mean Percent shortfall of the poor relative to the \$1.90/day	H
BL 40	Daily per capita expenditures (as a proxy for income) in USG-assisted areas	H
BL 32	Percent of women and men in a union who earned cash in the past 12 months	J
BL 33	Percent of women in union and earning cash who report participation in decisions about the use of self-earned cash	J
BL 34	Percent of women in union and earning cash who report participation in decisions about the use of spouse/partner's self-earned cash	J
BL 35	Percent of men in union and earning cash who report spouse/partner participation in decisions about the use of self-earned cash	J
BL 41	Percent of women/men in a union who are members of a community group	K
BL 42	Percent of women/men in a union with access to credit	K
BL 43	Percent of women/men in a union who make decisions about credit	K
BL 8	Adaptive capacity index	R
BL 9	Absorptive capacity index	R
BL 23	Ability to recover from shocks and stresses index	R
BL 24	Percent of households that believe local government will respond effectively to future shocks and stresses	R
BL 25	Transformative capacity index	R
BL 38	Index of social capital at the household level	R
BL 31	Percent of households participating in group-based savings, micro-finance or lending programs	R/K

³⁵ Per BHA specification only yield information from the production of livestock will be computed. The measurement of crop yield required a subjective estimation of the weight of the crop, which was considered inaccurate. To obtain the weight of the livestock units, average weight will be obtained from secondary sources such as the International Livestock Research (ILRI) and Food and Agriculture Organization.