



FEED THE FUTURE

The U.S. Government's Global Hunger & Food Security Initiative

Feed the Future Survey Implementation Document

Feed the Future Population-Based Survey Sampling Guide

Diana Maria Stukel, PhD

April 2018



USAID
FROM THE AMERICAN PEOPLE

FANTA III
FOOD AND NUTRITION
TECHNICAL ASSISTANCE

fhi360
THE SCIENCE OF IMPROVING LIVES

This guide is made possible by the generous support of the American people through the support of the Office of Health, Infectious Diseases, and Nutrition, Bureau for Global Health, U.S. Agency for International Development (USAID), USAID Bureau for Food Security, and USAID Office of Food for Peace, under terms of Cooperative Agreement No. AID-OAA-A-12-00005, through the Food and Nutrition Technical Assistance III Project (FANTA), managed by FHI 360.

The contents are the responsibility of FHI 360 and do not necessarily reflect the views of USAID or the United States Government.

April 2018

Recommended Citation

Diana Maria Stukel. 2018.
Feed the Future Population-Based Survey Sampling Guide.
Washington, DC: Food and Nutrition Technical Assistance Project, FHI 360.

Contact Information

Food and Nutrition Technical Assistance III Project
(FANTA)
FHI 360
1825 Connecticut Avenue, NW
Washington, DC 20009-5721
T 202-884-8000
F 202-884-8432
fantamail@fhi360.org
www.fantaproject.org

Acknowledgments

I am deeply indebted to a number of colleagues—Anne Swindale (USAID/BFS), Kiersten Johnson (USAID/BFS), Arif Rashid (USAID/FFP), Mario Chen (FHI 360), and Megan Deitchler (FHI 360)—for their insightful and thorough comments and suggestions on previous versions of this manuscript. They were instrumental in helping me shape the guide into its current form, and their feedback catalyzed a final version that was greatly improved in terms of technical accuracy, clarity, and readability. I am also greatly appreciative of the many thought-provoking and informative conversations and email exchanges I had with Anne Swindale, Kiersten Johnson, Peter Lance (UNC MEASURE), Keith Rust (Westat Inc.), Tom Krenzke (Westat Inc.), Sara Viviani (FAO), Carlo Cafiero (FAO), Mario Chen (FHI 360), Marga Eichleay (FHI 360), Zo Jariseta Rambeloson (FHI 360), Kathleen Ridgeway (FHI 360), Andres Martinez (FHI 360), and Zeina Maalouf-Manasseh (FHI 360), all of whom provided stellar advice and inputs on various technical topics. I wish to acknowledge Jeff Feldmesser, whose editorial expertise enormously improved the readability and flow of my earlier rough drafts. Finally, a debt of gratitude is owed to the wonderful FANTA Communications team, who were responsible for the layout, typesetting, and overall transformation of this document into a professional publication.

Table of Contents

Acknowledgments	i
Abbreviations and Acronyms	v
1. Introduction	1
2. Calculating the Sample Size for a PBS	7
2.1 Survey Purposes and Types of Indicators.....	7
2.2 Sample Size Calculations to Power Statistical Tests of Differences over Time for Indicators of Proportions or Means Using Comparative Analytical PBSs	10
2.2.1 Sample Size Calculations to Power Statistical Tests of Differences over Time for Indicators of Proportions.....	10
2.2.2 Sample Size Calculations to Power Statistical Tests of Differences over Time for Indicators of Means	18
2.2.3 Computing the Final Sample Size for the Survey.....	20
2.2.4 Adjustment 1: Inflation for the Number of Households to Contact.....	22
2.2.5 Adjustment 2: Inflation for Anticipated Household Non-Response	24
2.2.6 Computing the Final Sample Size for the Survey.....	24
2.2.7 Adjusting the Final Sample Size at the Second Time Point prior to the End-Line PBS.....	26
2.3 Sample Size Calculations to Ensure Adequate Precision for Estimates of Indicators of Proportions or Means Using Descriptive PBSs.....	29
2.3.1 Sample Size Calculations to Ensure Adequate Precision for Estimates of Indicators of Proportions.....	29
2.3.2 Sample Size Calculations to Ensure Adequate Precision for Estimates of Indicators of Means	30
2.3.3 Computing the Final Sample Size for the Survey.....	31
3. Development of a Sampling Frame	33
3.1 First Stage Sampling Frame of Clusters/Enumeration Areas.....	33
3.2 Second Stage Sampling Frame of Segments.....	35
3.3 Third Stage Sampling Frame of Households.....	36
3.4 Fourth Stage Sampling Frame of Individuals	37
4. Stratification and Allocation of the Sample	39
4.1 Stratification	39
4.2 Allocation of the Sample to Strata	39
4.2.1 Proportional Allocation	40
4.2.2 Equal Allocation	40
4.2.3 Power Allocation.....	40
4.3 Examples of Stratification and Allocation	41
4.4 A Special Application of Allocation: Joint Baseline and End-Line PBS.....	43

5. First Stage Sampling of EAs	46
5.1 Deciding on the Number of EAs to Sample.....	46
5.2 Randomly Selecting a Sample of EAs	48
5.3 Systematic PPS Sampling	49
5.3.1 Additional Issues to Consider for the First Stage Selection of EAs	53
6. Preparation for Second and Third Stage Sampling: Listing Exercise for Sampled EAs ..	56
7. Potential Second Stage Sampling of Segmented Enumeration Areas.....	58
8. Third Stage Sampling of Households within Sampled EAs or Segments	60
8.1 Fractional Interval Systematic Sampling.....	60
8.2 Considerations to Take into Account When Selecting Households within Sampled EAs	63
8.3 A Special Application of Fractional Interval Systematic Sampling: Joint Baseline and End-Line PBS.....	65
8.4 A Special Application of Fractional Interval Systematic Sampling: The Treatment of Inaccessible EAs	65
9. Fourth Stage Sampling of Individuals within Sampled Households.....	66
10. Sample Weighting.....	68
10.1 Calculating Probabilities of Selection and Sampling Weights	69
10.1.1 Overview of How to Calculate Probabilities of Selection	69
10.1.2 Calculating the Probability of Selection at the First Stage Sampling of EAs.....	71
10.1.3 Calculating the Probability of Selection at the Second Stage Sampling of Segments	73
10.1.4 Calculating the Probability of Selection at the Third Stage Sampling of Households	74
10.1.5 Calculating the Probability of Selection at the Fourth Stage Sampling of Individuals.....	75
10.1.6 Calculating the Overall Probability of Selection	75
10.1.7 Calculating the Sampling Weights to Account for Probabilities of Selection.....	78
10.2 Adjusting Survey Weights for Household and Individual Non-Response.....	79
10.2.1 Adjusting for Household Non-Response	79
10.2.2 Adjusting for Individual Non-Response	80
10.3 Calculating the Final Sampling Weights	81
11. Data Analysis for Descriptive Surveys: Producing Single-Point-in-Time Estimates of Indicators along with Their Standard Errors and Confidence Intervals	83
11.1 Producing Estimates of the Indicators	83
11.2 Producing Confidence Intervals and Standard Errors Associated with the Estimates of the Indicators	85
11.2.1 Calculating Confidence Intervals and Standard Errors Associated with Estimates of Indicators of Proportions	86
11.2.2 Calculating Confidence Intervals and Standard Errors Associated with Estimates of Indicators of Means.....	87

11.2.3 An Example of Calculating a Confidence Interval and a Standard Error for an Estimate of an Indicator of Proportions.....	88
11.2.4 Interpreting Standard Errors and Confidence Intervals	89
12. Data Analysis for Comparative Analytical Surveys: Statistical Tests of Differences	91
12.1 Statistical Tests of Differences for Indicators of Means	91
12.2 Statistical Tests of Differences for Indicators of Proportions.....	93
Appendix A. Derivation of Inflation Factor for Number of Households to Contact Given an Initial Sample Size of Individuals	95
Appendix B. Syntax for Statistical Software Packages SAS, SPSS, and STATA.....	101

Abbreviations and Acronyms

A-WEAI	abbreviated Women’s Empowerment in Agriculture Index
BBS	beneficiary-based survey
BFS	Bureau for Food Security (USAID)
BL/FE	baseline and final evaluation
BMI	body mass index
CAPI	computer-assisted personal interviewing
CI	confidence interval
DEFF	design effect
DF	degrees of freedom
DFSA	development food security activity
DHS	Demographic and Health Surveys
DP	depth of poverty
EA	enumeration area
EBF	exclusive breastfeeding
FANTA	Food and Nutrition Technical Assistance III Project
FAO	Food and Agriculture Organization of the United Nations
FFP	Office of Food for Peace (USAID)
FIES	Food Insecurity Experience Scale
GFSS	Global Food Security Strategy
GPS	global positioning system
HAZ	height-for-age z-score
HHS	Household Hunger Scale
ICC	intra-class correlation coefficient
IP	implementing partner
LSMS	Living Standards Measurement Studies
M&E	monitoring and evaluation
MAD	minimum acceptable diet
MDD-W	minimum dietary diversity for women
MICS	Multiple Indicator Cluster Surveys
MOE	margin of error
NRVCC-C	nutrient-rich value chain commodities (children)
NRVCC-W	nutrient-rich value chain commodities (women)
NSO	national statistics office

PBS	population-based survey
PCE	per capita expenditure
PIRS	Performance Indicator Reference Sheet
PP	prevalence of poverty
PPP	purchasing power parity
PPS	probability proportional to size
RCT	randomized control trial
RS	random start
SE	standard error
SRS	simple random sampling
U.S.	United States
USAID	U.S. Agency for International Development
USG	U.S. Government
WAZ	weight-for-age z-score
WDDS	Women's Dietary Diversity Score
WEAI	Women's Empowerment in Agriculture Index
WHZ	weight-for-height z-score
ZOI	Zone of Influence

I. Introduction

Feed the Future, a United States Government (USG) initiative led by the U.S. Agency for International Development (USAID), is the USG’s global hunger and food security initiative. Phase one of the initiative was launched in 2010. Phase two was launched in 2017 and is guided by the USG Global Food Security Strategy (GFSS) 2017–2021,¹ which presents an integrated whole-of-government strategy and agency-specific implementation plan, as required by the Global Food Security Act of 2016.

This sampling guide provides technical guidance on the design of population-based surveys (PBSs) to support the collection and analysis of data for Feed the Future Zone of Influence (ZOI)² PBS indicators, which include the suite of FFP baseline and final evaluation (BL/FE) indicators. The guide is intended for use mainly by Feed the Future monitoring and evaluation (M&E) specialists, M&E contractors, and USAID Office of Food for Peace (FFP) development food security activity (DFSAs) implementing partners (IPs).³

PBSs are conducted among a sample of the entire population living within a Feed the Future ZOI or an FFP DFSA implementation area. This is in contrast to beneficiary-based surveys (BBSs), which are conducted among a sample of a project’s direct beneficiary population.^{4,5} In general, baseline, monitoring,⁶ and end-line PBSs are used in the Feed the Future context in one of two ways: to monitor project progress (monitoring PBSs only) or to see if there has been change over time at the population level in key outcomes and impact indicators (baseline and end-line PBSs).⁷ In contrast, BBSs are typically used in the context of project monitoring to ensure that implementation is rolling out as expected and that interventions are on track for achieving their intended outcomes and targets in the direct

¹ The vision of the strategy is a world free from hunger, malnutrition, and extreme poverty, where thriving local economies generate increased income for all people; where people consume balanced and nutritious diets, and children grow up healthy and reach their full potential; and where resilient households and communities face fewer and less severe shocks, have less vulnerability to the shocks they do face, and are helping to accelerate inclusive, sustainable economic growth. For more information, see <https://www.usaid.gov/what-we-do/agriculture-and-food-security/us-government-global-food-security-strategy>.

² ZOIs are the geographic zones where Feed the Future programmatic interventions are concentrated within a country and where population-level impacts on poverty, hunger, and malnutrition are measured.

³ Although FFP is part of the Feed the Future initiative, this guide will make reference to the FFP and non-FFP parts of the Feed the Future initiative as separate entities when relevant.

⁴ This guide uses the term “project” to refer to FFP-funded DFSAs and to non-FFP-funded activities under the broad banner of Feed the Future projects. See USAID Automated Directives System glossary for the definitions of project and activity (<https://www.usaid.gov/who-we-are/agency-policy/glossary-ads-terms>).

⁵ Direct beneficiaries are those who come into direct contact with the set of interventions (goods or services) provided by the project in each technical area. Individuals who receive training or benefit from project-supported technical assistance or service provision are considered direct beneficiaries, as are those who receive a ration or other type of good. These should be distinguished from indirect beneficiaries, who benefit indirectly from the goods and services provided to the direct beneficiaries, e.g., members of the household of a beneficiary farmer who received technical assistance, seeds and tools, other inputs, credit, or livestock; or neighboring farmers who observe technologies being applied by direct beneficiaries and elect to apply the technologies themselves.

⁶ Feed the Future “monitoring PBSs” are typically conducted every 3 years. The first monitoring PBS is generally conducted 3 years after the initial baseline PBS and is used to monitor project progress, whereas subsequent monitoring PBSs, conducted 3 years after the preceding one, are used to see if there has been change since the baseline in key outcome and impact indicators. For the remainder of this guide, the second monitoring PBS, which is conducted 6 years after the baseline PBS (in the Feed the Future context) and which is meant to be compared to the baseline PBS, will always be called an “end-line PBS” to avoid confusion. Note that Feed the Future may phase out monitoring PBSs in the future.

⁷ In the FFP context, baseline and end-line PBSs are used to see if there has been change over time in key outcome and impact indicators. However, although FFP IPs sometimes conduct monitoring BBSs, they generally do not conduct monitoring PBSs.

beneficiary population. The rationale for conducting baseline, monitoring, and end-line PBSs at the population level relates to the expectation that the effects of a project should spread beyond direct beneficiaries to the general population within the Feed the Future ZOI (or FFP DFSA implementation area) over the life of the award.

For phase two, Feed the Future phase one indicators were revised, including the set of ZOI PBS indicators (a subset of all Feed the Future indicators). Each ZOI PBS indicator has an associated Performance Indicator Reference Sheet (PIRS) that provides the information needed to gather data and report on the indicator.⁸ Table 1 provides a list of the 20 Feed the Future phase two ZOI PBS indicators. Feed the Future target countries must establish baseline values for these indicators by 2020, and the data for these indicators are to be collected every 3 years thereafter through monitoring and end-line PBSs. As part of Feed the Future, FFP DFSAs report on many of these phase two indicators (as well as other indicators unique to FFP) in its set of BL/FE indicators.⁹

There is substantial overlap between the Feed the Future phase one and phase two ZOI PBS indicators; the Feed the Future phase two ZOI PBS indicators that were also Feed the Future phase one ZOI PBS indicators are indicated by an asterisk in Table 1. Some of the phase one ZOI PBS indicators were dropped in 2018; these are listed in Table 2. Feed the Future focus countries continuing past 2018 as target countries and existing FFP DFSAs are required to collect end-line data on the phase one indicators for which they also collected data at baseline, even if those indicators were dropped as phase two indicators. Feed the Future focus countries not continuing as target countries are required to report only end-line results for the prevalence of poverty and stunting indicators, and only when secondary data are available to do so.

Table 1. Feed the Future Phase Two ZOI PBS Indicators

Indicator	Type of Indicator	Sampling Group Implicated ^a	Disaggregation Required
Prevalence of Poverty (PP): Percent of people living on less than \$1.90/day using 2011 purchasing power parity (PPP) ^b	Proportion	Household	Gendered household type
Depth of Poverty (DP) of the Poor: Mean percent shortfall of the poor relative to the \$1.90/day poverty line using 2011 PPP ^c	Proportion	Household	Gendered household type
Prevalence of Moderate and Severe Food Insecurity in the population (based on the Food Insecurity Experience Scale [FIES]) ^d	Proportion	Household	Gendered household type
Percentage of Households below the Comparative Threshold for the Poorest	Proportion	Household	Gendered household type

⁸ The complete set of Phase 2 Feed the Future ZOI PBS indicators and their PIRSs can be found in the publication “Feed the Future Indicator Handbook: Definition Sheets,” which is located at https://feedthefuture.gov/sites/default/files/resource/files/Feed_the_Future_Indicator_Handbook_Sept2016.pdf.

⁹ The complete set of FFP BL/FE indicators and their PIRSs can be found in the publication “FFP Indicators Handbook – Part 1: Indicators for Baseline and Final Evaluation Surveys,” which is located at https://www.usaid.gov/sites/default/files/documents/1866/Part%20I_Baseline%20and%20Final%20Evaluation_04.13.2015.pdf. FFP will post an updated version of this document reflecting the Feed the Future Phase 2 indicators after the Feed the Future Phase 2 indicator handbook is finalized in March 2018.

Indicator	Type of Indicator	Sampling Group Implicated ^a	Disaggregation Required
Quintile of the Asset-Based Comparative Wealth Index			
Ability to Recover from Shocks and Stresses Index	Index	Household	Gendered household type
Index of Social Capital at the Household Level	Index	Household	Gendered household type Type of social capital
Proportion of Households That Believe Local Government Will Respond Effectively to Future Shocks and Stresses	Proportion	Household	Gendered household type
Proportion of Households Participating in Group-Based Savings, Microfinance, or Lending Programs	Proportion	Household	Gendered household type Product type
Percentage of Households with Access to a Basic Sanitation Service	Proportion	Household	Gendered household type Location
Percentage of Households with Soap and Water at a Handwashing Station Commonly Used by Family Members	Proportion	Household	Gendered household type Location
Abbreviated Women's Empowerment in Agriculture Index (A-WEAI) Score ^e	Index	Primary adult female and male decision makers in household	Age
Prevalence of Exclusive Breastfeeding (EBF) of Children under 6 Months of Age*	Proportion	Children age 0–5 months	Sex
Prevalence of Children 6–23 Months Receiving a Minimum Acceptable Diet (MAD)*	Proportion	Children age 6–23 months	Sex
Prevalence of Stunted (height-for-age z-score [HAZ] < - 2) Children under Five (0–59 Months)*	Proportion	Children age 0–59 months	Sex
Prevalence of Healthy Weight (weight-for-height z-score [WHZ] ≤ 2 and ≥ - 2) among Children under Five (0–59 Months)	Proportion	Children age 0–59 months	Sex
Prevalence of Wasted (weight-for-age z-score [WAZ] < - 2) Children under Five (0–59 Months)*	Proportion	Children age 0–59 months	Sex
Prevalence of Underweight (body mass index [BMI] < 18.5) Women of Reproductive Age*	Proportion	Nonpregnant women age 15–49 years	Age
Prevalence of Women of Reproductive Age Consuming a Diet of Minimum Diversity (MDD-W) ^f	Proportion	Women age 15–49 years	Age
Proportion of Producers Who Have Applied Targeted Improved Management Practices or Technologies	Proportion	Producers	Management practice or technology type Sex of producer Age of producer Commodity type

Indicator	Type of Indicator	Sampling Group Implicated ^a	Disaggregation Required
Yield of Targeted Agricultural Commodities within Target Areas	Mean	Producers	Sex of producer Age of producer Commodity type

^a Note that the concept of “sampling group” differs from that of “(proxy) respondent group.” For instance, for the indicator “Prevalence of Children 6–23 Months Receiving a Minimum Acceptable Diet (MAD),” the sampling group consists of children age 6–23 months, because this is the group for which information is required. However, the (proxy) respondent group consists of mothers or caregivers of children age 6–23 months, because these are the individuals who provide the information on behalf of the sampling group. Similarly, for a household-level indicator, the sampling group consists of households, but the respondent group consists of responsible adults residing within the households who can provide information on behalf of the households. When a proxy respondent is not needed, the sampling group and the respondent group are the same.

^b Feed the Future reported on the PP indicator in phase one and will continue to report on the PP indicator under phase two. However, because the international extreme poverty threshold and PPP rates used to compute the indicator have changed from \$1.25 using 2005 PPP (used in phase one) to \$1.90 using 2011 PPP (used in phase two), the phase two indicator is considered a different indicator from the phase one indicator. Computing the phase one indicator requires its own analysis that is different from that of the phase two indicator.

^c Feed the Future reported on the DP of the poor indicator under phase one. However, in addition to the changes in the international extreme poverty threshold and PPP rates used to compute the indicator, the phase two indicator differs from the phase one indicator in that it focuses only on DP of the poor. Computing the phase one indicator requires its own analysis that is different from that of the phase two indicator.

^d Feed the Future reported on the prevalence of households with hunger indicator under phase one. The phase two indicator uses a different measurement tool that captures the broader food insecurity experience, and it uses a longer time period (12 months versus 30 days). Computing the phase one indicator requires its own analysis that is different from that of the phase two indicator.

^e Feed the Future developed and reported on the Women’s Empowerment in Agriculture Index (WEAI) under phase one. Under phase two, a shorter, streamlined version of the original WEAI, the A-WEAI, is used. However, because there is no requirement to report on the full WEAI (phase one indicator) moving forward, the indicator is not included in Table 2.

^f Feed the Future reported on the women consuming a diet of minimum diversity under phase one using the Women’s Dietary Diversity Score (WDDS). This indicator reports the mean number of food groups consumed by women of reproductive age in the last 24 hours, based on nine food groups. The phase two indicator, MDD-W, is based on 10 food groups and reports the prevalence of women of reproductive age consuming at least 5 of the 10 food groups in the last 24 hours. Computing the phase one indicator requires its own analysis that is different from that of the phase two indicator.

Table 2. Feed the Future Phase One ZOI PBS Indicators Dropped in 2018

Indicator	Type of Indicator	Sampling Group Implicated	Disaggregation Required
Prevalence of Poverty (PP): Percent of people living on less than \$1.25/day using 2005 PPP	Proportion	Household	Gendered household type
Depth of Poverty (DP): Mean percent shortfall relative to the \$1.25/day poverty line using 2005 PPP	Proportion	Household	Gendered household type
Prevalence of Households with Hunger (Household Hunger Scale [HHS])	Proportion	Household	Gendered household type
Average Daily Per Capita Expenditures (PCE)	Mean	Household ^a	Gendered household type
Prevalence of Underweight Children	Proportion	Children age 0–59 months	Sex
Prevalence of Anemia among Children	Proportion	Children age 6–59 months	Sex
Prevalence of Anemia among Women	Proportion	Women age 15–49 years	Pregnant women/ nonpregnant women
Women’s Dietary Diversity Score (WDDS)	Mean	Women age 15–49 years	None
Prevalence of Women Consuming Nutrient-Rich Value Chain Commodities (NRVCC-W)	Proportion	Women age 15–49 years	Commodity type
Prevalence of Children Consuming Nutrient-Rich Value Chain Commodities (NRVCC-C)	Proportion	Children age 6–23 months	Commodity type Sex

^a The data for the PCE indicator are collected at the household level, but the indicator is reported at the individual level. The indicator is computed by summing the sample weighted expenditure at the household level across all households in the sample, and then dividing the sum by the sample weighted sum of household members in the sample. Similarly, the PP and DP indicators are also reported at the individual level.

Given the importance of PBSs for monitoring the performance of Feed the Future projects, there is a need for uniform and comprehensive guidance across countries and over time on how to design and implement these surveys. This guide is designed to help meet this need.

While there are a multitude of possible designs for quantitative surveys, this guide promotes the use of stratified multi-stage cluster sampling designs,¹⁰ where it is assumed that there are three or four stages of sampling: i) clusters or census enumeration areas (EAs)¹¹, ii) segments within sampled clusters (only if applicable), iii) households within sampled segments (or clusters if segmentation is not applicable), and iv) individuals within sampled households.¹²

¹⁰ This guide recommends the use of multi-stage cluster sampling designs over simple random sample (SRS) designs to ensure the geographic spread of the selected random sample and to facilitate logistical considerations of fieldwork. For those readers who want a greater understanding of when multi-stage cluster sampling designs versus SRS designs are appropriate, please see: Kalton, Graham. 1983. *Introduction to Survey Sampling*. Newbury Park, CA: Sage Publications, Inc.

¹¹ A census EA is a geographical statistical unit that is created to support the implementation of a census. In rural areas, an EA is usually a community, a part of a community, or a group of small communities, with its location and boundaries well defined and recorded on census maps.

¹² Theoretically, a PBS could also have more stages of sampling if the geography to be covered spans large areas, such as in national surveys. In such cases, there might be multiple stages of clustering that precede sampling at the household and individual levels. However, for the purposes of this guide, there is an assumption that the geographic coverage of the Feed the

This guide is structured as follows:

- Chapter 2 provides guidance on calculating sample sizes for Feed the Future PBSs.
- Chapter 3 discusses the development of sampling frames to be used as the foundation for sample selection.
- Chapter 4 addresses issues regarding stratification and allocation of the sample.
- Chapters 5–9 describe the four stages of sampling and include a discussion on listing exercises for sampled clusters.
- Finally, Chapters 10–12 detail the post-fieldwork analysis component, including the construction of sampling weights; the production of single-point-in-time estimates for indicators of interest, along with their standard errors (SEs) and confidence intervals (CIs); and the implementation of tests of differences over time for indicators of proportions and means.

Future ZOI or FFP implementation area is relatively compact, so that it is reasonable to limit the design to one stage of clustering only.

2. Calculating the Sample Size for a PBS

The first step in the survey design process is to calculate the sample size. This chapter starts by describing the different survey purposes and types of indicators, and the different sample size calculations associated with each survey purpose and indicator type. Focus then turns to providing formulas for determining the initial sample size for two types of surveys having different aims: to power statistical tests of differences over time for indicators of proportions and means and to ensure high-precision single-point-in-time estimates of indicators of proportions and means. In both cases, the input parameters to the initial sample size calculation are described in detail and recommendations on how to estimate them are provided. The various indicators that are candidates to drive the overall sample size for the surveys are introduced, as are the rules for choosing among the indicators. Two multiplicative adjustments to the initial sample size formula are given, to permit the computation of a final sample size of the required number of households to interview. Illustrative examples are provided throughout the section.

2.1 Survey Purposes and Types of Indicators

The formulas used to calculate the sample size for a survey depend on two factors: the **survey purpose** and the **type of indicator**.

Surveys generally have one of two purposes under Feed the Future: They are either **descriptive** or **comparative analytical**.

- The first survey purpose is to provide a snapshot of the situation at a single point in time. This requires a descriptive survey, where the intention is to provide a sample size to achieve a reasonable level of precision (i.e., a small SE) by specifying a “margin of error” (MOE) (described in more detail later in the guide) for indicator estimates. The first Feed the Future monitoring PBS, which is conducted 3 years following the baseline PBS, has this purpose. Estimating change is not advisable in cases where the two time points are spaced close together (e.g., 3 years after a baseline PBS) because little policy-relevant change is likely to have taken place for most priority Feed the Future ZOI PBS indicators.
- The second survey purpose is to conduct statistical tests of differences between indicators from different groups or at different time points. This requires a comparative analytical survey, where, in the Feed the Future context, the underlying data are collected at different points in time (e.g., at the start of the Feed the Future strategy¹³ and 6 years later, or at the start and end of a FFP DFSA) and typically for indicators of proportions or means. For these surveys, the intention is to provide a sample size that controls for the levels of inferential errors associated with the statistical tests of differences. The Feed the Future baseline and end-line PBSs are comparative analytical surveys. For the remainder of this guide, the second comparative analytical survey, which is conducted 6 years after the baseline PBS and which is compared to the baseline PBS, will always be called an end-line survey, to avoid confusion.

¹³ Feed the Future multi-year strategies outline the strategic planning for the USG’s global hunger and food security initiative. These documents represent coordinated, whole-of-government approaches to address food security that align in support of partner country priorities. The strategies reflect analysis and strategic choices made at the time of writing and, while interagency teams have formally approved these documents, they may be modified as appropriate.

The two types of surveys with differing purposes require different formulas to calculate the overall sample size for the surveys. The formulas for descriptive surveys are simpler and tend to result in smaller sample sizes than those for comparative analytical surveys, although this is not always the case. In the context of Feed the Future PBSs, there are several scenarios that warrant the use of either descriptive surveys or comparative analytical surveys. Two such scenarios are described in Box 1 and these are used as examples throughout the guide.

Box 1. Two Example Scenarios of Feed the Future Multi-Year Strategies and ZOIs

Scenario 1: A Feed the Future multi-year strategy commences in 2012 and there is no change in the definition of the ZOI over time. Under this scenario, a baseline PBS is conducted in 2012, where the aim of the PBS, as a comparative analytical survey, is to enable a statistical test of differences to detect changes in indicators of interest relative to a future survey. A monitoring PBS is conducted in the ZOI in 2015; the aim of this assessment, as a descriptive survey, is to produce single-point-in-time estimates of indicators, along with their SEs and CIs, and to monitor Feed the Future progress at the population level. An end-line PBS is conducted 3 years later, in 2018; the aim of this PBS, as a comparative analytical survey, is to enable a statistical test of differences to detect changes in indicators of interest relative to the baseline PBS conducted in 2012.

Scenario 2: A Feed the Future multi-year strategy commences in 2012 and there is a change in the definition of the ZOI in 2018: Some districts are dropped from original ZOI and some new districts are added. The dropped and new ZOI districts can now be divided into three strata: dropped (i.e., original 2012 ZOI) districts, common districts that are in both the original 2012 ZOI and the new 2018 ZOI, and new (i.e., 2018 ZOI) districts. Under this scenario, a baseline PBS is conducted in 2012, where the aim of the PBS, as a comparative analytical survey, is to enable statistical tests of differences to detect changes in indicators of interest relative to a future survey. In 2018, the strata with the dropped 2012 districts and the 2012/2018 common districts serve as the basis for an end-line PBS on the original 2012 ZOI. The aim of the 2018 end-line PBS, as a comparative analytical survey, is to enable a statistical test of differences to detect changes in indicators of interest relative to the baseline PBS conducted in 2012 in the original ZOI. In addition, in 2018, the strata with the new 2018 districts and the common 2012/2018 districts serve as the basis for a baseline PBS on the new 2018 ZOI. The aim of the 2018 baseline PBS, as a comparative analytical survey, is to enable a statistical test of differences to detect changes in indicators of interest relative to an end-line PBS in the future. In 2015 and in 2021, ZOI monitoring PBSs are conducted; the aim of these PBSs, as descriptive surveys, is to produce single-point-in-time estimates of indicators of interest along with their SEs and CIs, to monitor progress in measures of food security at the population level, relative to the original 2012 ZOI and the new 2018 ZOI, respectively.

It is clear that, under Scenario 2, it would be ideal to conduct **one** PBS in 2018 that serves as the data collection vehicle for both the baseline PBS on the new 2018 ZOI and the end-line PBS on the original 2012 ZOI. This is because, in all likelihood, there will be considerable overlap in the indicators for which data must be collected for the original ZOI and the new ZOI, and any differences in the set of indicators would not be substantial enough to justify the cost and burden of two separate surveys. A road map for addressing this challenge is discussed later in the guide, in Section 4.4, Section 5.1 (Example 2), Section 8.3, and Section 10.1.2.

As can be seen from the examples above, comparative analytical surveys imply two or more surveys. In the Feed the Future context, one typically focuses on comparative analytical surveys at two time points, commonly termed “pre” and “post” surveys (i.e., baseline and end-line PBSs), the results of which are compared through a statistical test of differences on the indicator(s) of interest. This guide describes the designs of comparative analytical surveys in the context of this pairing, using what are known as “adequacy evaluation” designs.¹⁴ In the FFP context, such designs are commonly used for performance evaluations. The particular designs described do **not** use “control” or “counterfactual” groups (which are groups that are not subject to project interventions), nor do they use “randomization” (i.e., the randomized assignment of project interventions to individuals or clusters, which are typically used to avoid “selection bias” or the bias induced by purposively targeting individuals or geographic areas for project interventions). Designs that are “pre-post” with randomization of interventions and the use of control groups are known as randomized control trials (RCTs); RCTs permit statements of attribution to project interventions, i.e., the degree to which the observed changes were caused by the project interventions. Given the constraints of using a simple pre-/post-design without control groups or randomization, a statistical test of differences permits an assessment of whether change has or has not occurred but does not permit attribution of any observed changes or lack thereof to project interventions. This is because any change that occurred may be attributable (at least in part) to external factors that have not been controlled for in the comparative analysis, such as government policies, government-funded infrastructure improvements, climatic anomalies, civil strife, economic shifts, changes in population composition, and related interventions by other organizations. Therefore, care must be exercised in interpreting the results of baseline/end-line PBSs, and any statements regarding attribution of observed changes to project interventions must be avoided (or at least provided in a context of appropriate caveats).

The second factor that influences the formula used to calculate sample size for a PBS is the type of indicator. There are several types of indicators for which data can be collected through sample surveys, for example, proportions (which are often expressed as prevalences, such as “Prevalence of Stunted Children under Five”), means (e.g., “Yield of Targeted Agricultural Commodities”), and totals (e.g., “Number of Hectares under Improved Technologies,” which is not typically collected through PBSs), as well as other less common types of indicators, such as ratios, percentiles, and medians.

Each type of indicator described above necessitates a different formula for calculating the associated sample size. The Feed the Future Phase One and Phase Two ZOI PBS indicators in Tables 1 and 2 are usually proportions or means, although some indicators take somewhat different forms, such as indexes

¹⁴ For a more in-depth discussion on adequacy evaluation designs, see page 144 of: Mason, J.B.; Habicht, J.-P.; Tabatabai, H.; and Valverde, V. 1984. *Nutritional Surveillance*. Geneva: World Health Organization.

that are complex composites of indicators (e.g., “Abbreviated Women’s Empowerment in Agriculture Index” or “Ability to Recover from Shocks and Stresses Index,” both in Table I). This guide is limited to the computation of sample sizes for indicators that are either proportions or means only. Although PBSs may collect data that will support the production of indicators that are indexes, such as those in Table I, the sample sizes underpinning the PBSs will not be based on these indexes.

2.2 Sample Size Calculations to Power Statistical Tests of Differences over Time for Indicators of Proportions or Means Using Comparative Analytical PBSs

This section provides a description of the sample size formulas that should be used for statistical tests of differences over time, using the scenarios described in Box I, in the context of Feed the Future baseline and end-line PBSs. But first, it is important to understand the concepts underlying such tests, and how they should be structured and interpreted.

2.2.1 Sample Size Calculations to Power Statistical Tests of Differences over Time for Indicators of Proportions

In general, any statistical test is underpinned by a hypothesis about a particular indicator of interest. The hypothesis is expressed in terms of both a “null hypothesis” (denoted H_0) and an “alternative hypothesis” (denoted H_A). The null hypothesis generally articulates the status quo or a worsening situation (e.g., the “Prevalence of Stunted Children” is the same or higher at the second time point [i.e., end-line] than it was at the first time point [i.e., baseline]), whereas the alternative hypothesis articulates an improved situation (e.g., the “Prevalence of Stunted Children” is lower at the end-line than it was at the baseline). For statistical tests of hypotheses, the burden of proof lies in demonstrating that an improved situation has occurred by rejecting the null hypothesis and accepting the alternative hypothesis.

For instance, suppose the aim of establishing a test of differences for the “Prevalence of Stunted Children under Five,” which is an indicator of proportions. Assume that P_1 represents the true prevalence (or proportion) of stunted children at baseline and that P_2 represents the true prevalence of stunted children at end-line. If the project is attempting to decrease the prevalence of stunting over time, the null hypothesis would be stated as:

$$H_0: P_1 - P_2 \leq \delta$$

and the alternative hypothesis as:

$$H_A: P_1 - P_2 > \delta$$

The null hypothesis states that there has been no change or an increase over time in the prevalence of stunted children (i.e., a deterioration in stunting). The alternative hypothesis states that there has been a decrease over time in the prevalence of stunted children (i.e., an improvement in stunting). In other words, the prevalence of stunted children at the baseline exceeds that at the end-line by a quantity that is greater than zero (i.e., δ , a positive number.) The quantity δ is called the “minimum meaningful effect size.” It is set by the researcher, team, or study lead, and represents the minimum difference that is deemed important to detect in the indicator between the two time points.

In the Feed the Future context, δ can be set to the change **expected** to be achieved in the indicator over the time period between the two time points (i.e., the indicator target). Alternatively, δ can be set to a value that is somewhat smaller than the target (e.g., 80% or 90% of the target). The rationale for doing the latter is that a test of differences will detect any change that is **at least** as large as δ . Therefore, setting δ to be somewhat smaller than the expected target allows IPs that come close to, but don't quite succeed in, achieving targets to demonstrate that indeed **some** meaningful change took place between the two survey occasions and, therefore, that at least some progress toward achieving the target was made. However, a disadvantage to setting δ to a value that is smaller than the target is that the sample size required to detect a smaller change will be larger—sometimes much larger—than that required to detect a change at least as large as the target. This clearly has cost implications for the associated PBS. It is up to Feed the Future teams to decide whether it is preferable to set δ equal to the target value or equal to some percentage (e.g., 80% or 90%) of the target value.

Additionally, when setting the value for δ , it is important to take into consideration the number of years anticipated between the two surveys between which values will be compared, in case it is necessary to prorate the 6-year target, for example, to fit the time frame of the surveys.

We illustrate the determination of δ , taking these two considerations into account. Suppose that the prevalence of stunted children in the ZOI at the time of the baseline PBS is known from external sources to be approximately 40% (or 0.40). The Feed the Future team in the country has set its target for the reduction in the prevalence of stunted children in the ZOI at 20% of the baseline value—an 8 percentage point drop or a decrease of 0.08 in the 0.40 prevalence (20% of 40% is 8%)—over 6 years. However, recognizing that a reduction of 20% of the baseline value is an ambitious target, and wanting to be able to demonstrate some meaningful change even if the results fall short of the target after 6 years, the Feed the Future team in the country decides that achieving 80% of the 20% reduction target (a 16% reduction) is a minimum meaningful change, and that δ should be set to this amount. That is, the team considers a 16% drop in the 40% baseline value, a 6.4 percentage point (0.064) decrease in the prevalence of stunting, to be a meaningful change.¹⁵

To further complicate the situation, the team intends to test for change by conducting an end-line PBS 5 years after the baseline PBS. That means that the team wants to detect a difference that could be achieved after 5 of the 6 years (i.e., a reduction of five-sixths of 6.4 percentage points or 5.3 percentage points). Therefore, given all the above constraints, δ would be equal to 0.053.¹⁶

As an additional cautionary note, Feed the Future teams and IPs should be aware that it is possible to set the value of δ so that it is too small to be considered relevant for policy or programming. For instance, if the minimum difference is set to $\delta = 0.02$ in the context of the stunting indicator (i.e., a

¹⁵ As mentioned earlier, if the team feels the target is not overly ambitious, it could also choose to set δ to the full 20% reduction of the baseline value (i.e., 8 percentage points). It will clearly be easier for the IPs to achieve a reduction in stunting of 6.4 percentage points or more than it will be to achieve a reduction of 8 percentage points or more. However, to statistically detect the former would require a larger sample size, which has cost implications.

¹⁶ This obviously assumes a constant rate of change over the period of implementation, which, in most cases, is an inaccurate assumption. However, for simplicity sake, this assumption is used here.

reduction of 2 percentage points), then, even if the stunting level does meet the threshold by dropping 2 percentage points over the life of the Feed the Future 6-year strategy, the change may not be significant enough to be considered relevant from a policy or programmatic perspective. In addition, such a small change would likely require a very large sample, greatly increasing cost for minimal benefit.

Notice that the above hypothesis test is relevant for indicators of proportions where the aim is to see a decrease over time, such as for “Prevalence of Stunted Children under Five,” “Prevalence of Poverty,” and many of the other indicators in Tables 1 and 2. For other indicators of proportions in Tables 1 and 2, such as “Prevalence of Exclusive Breastfeeding (EBF)” and “Prevalence of Children 6–23 Months Receiving a Minimum Acceptable Diet (MAD),” the aim is to see an increase over time, and therefore the appropriate null hypothesis should be reversed from before and stated as:

$$H_0: P_2 - P_1 \leq \delta$$

and the alternative hypothesis as:

$$H_A: P_2 - P_1 > \delta$$

In either case, these alternative hypotheses are “one-sided,” not “two-sided.” To explain what is meant by “one-sided” versus “two-sided,” suppose one were to use the null hypothesis:

$$H_0: P_2 - P_1 = 0$$

and the alternative hypothesis:

$$H_A: |P_2 - P_1| > \delta$$

which is equivalent to:

$$H_A: P_2 - P_1 > \delta \text{ or } H_A: P_2 - P_1 < -\delta$$

In the above case of a “two-sided” hypothesis, if one were to reject the null hypothesis and accept the alternative, it would mean that there has been either an increase or a decrease in the indicator in question or, in other words, that there has been either an improvement in the situation or a deterioration in the situation relating to the indicator. However, for all the indicators in Tables 1 and 2, there is a clear desired direction of change that the indicators are attempting to achieve, and the use of a one-sided hypothesis is preferable given that the main interest in conducting statistical tests of differences in this particular context lies in determining if indicators have come close to achieving their targets. In principle, a two-sided hypothesis could reveal either an improvement **or** a deterioration in the situation, and, although a deterioration is certainly possible for all the Feed the Future indicators under consideration in Tables 1 and 2, it is very unlikely that the implementation of the Feed the Future strategy would have caused such a deterioration. As such, the statistical tests of hypotheses in this context are less focused on determining if there has been a deterioration in the situation.¹⁷

¹⁷ In the biomedical literature, the use of a one-sided statistical hypothesis is controversial because of the concern that unearthing potential harmful effects stemming from treatment interventions can be masked, whereas the use of a two-sided hypothesis would reveal them. However, in the development setting, a marked deterioration in the situation is less likely due to the implementation of the Feed the Future strategy and more likely due to external factors (e.g., long-term drought or food

One of the advantages of a one-sided hypothesis is that it generally requires a smaller sample size than a two-sided hypothesis. However, one criticism of its use is that it is easier to reject the null hypothesis (and hence show significant improvement in the indicator in question) with a one-sided hypothesis than it is with a two-sided hypothesis.

In any statistical test based on an underlying set of hypotheses, it is important to control for two types of error. The first type of error (**type I error**) happens when the null hypothesis is incorrectly rejected (i.e., the alternative hypothesis is not true). In other words, one concludes that the desired level of change has occurred when in fact it has not. The probability of a type I error, denoted by α and called the **significance level**, is a prespecified value set by the user, typically $\alpha = 0.05$. The **confidence level** of the test, denoted by $1 - \alpha$, is the complement of α . It represents the probability of correctly concluding that the desired level of change has not occurred (or more accurately, that the results are inconclusive¹⁸). In other words, the confidence level is the probability of correctly not rejecting the null hypothesis when it should not be rejected (i.e., when the alternative hypothesis is not true). If the type I error is set at $\alpha = 0.05$, then the confidence level is $1 - \alpha = 0.95$.

The second type of error (**type II error**) happens when one concludes that the desired level of change has not occurred (or that the results are inconclusive) when in fact it has. In other words, the null hypothesis is incorrectly **not** rejected (i.e., the alternative hypothesis is true). The probability of a type II error, denoted by β , is a prespecified value set by the user, typically set to $\beta = 0.20$. The **power** of the test, denoted by $1 - \beta$, is the complement of β . It represents the probability of correctly rejecting the null hypothesis when it should be rejected (i.e., when the alternative hypothesis is true). In other words, it is the probability of correctly concluding that the desired level of change has occurred. If the type II error is set at $\beta = 0.20$, then the power is $1 - \beta = 0.80$. Table 3 summarizes these concepts.

Table 3. Probabilities of Study Decisions Under the Alternative Hypothesis

Study Decision	Alternative Hypothesis (change has occurred)	
	False	True
Do not reject null hypothesis (results inconclusive)	Correct Decision: confidence level ($1 - \alpha$)	Type II Error (β)
Reject null hypothesis/accept alternative hypothesis (desired level of change occurred)	Type I Error: significance level (α)	Correct Decision: power ($1 - \beta$)

price volatility). Therefore, unlike the biomedical context, there would be little reason to believe that a Feed the Future strategy (or a specific project) should be adjusted because it is the cause of the deterioration. As such, the use of one-sided statistical hypotheses is less of a concern in the current context. For a more in-depth discussion on the controversy regarding one-sided hypotheses, see <http://www.jerrydalla.com/LHSP/onesided.htm>.

¹⁸ It is more accurate to say that the results are inconclusive because, even if we fail to reject the null hypothesis, it does not mean the null hypothesis is true. That's because a hypothesis test does not determine **which** hypothesis (i.e., null or alternative) is true, or even which is most likely; it assesses **only** whether available evidence exists to reject the null hypothesis.

Note that the type I error is typically controlled more tightly (at $\alpha = 0.05$) than the type II error (at $\beta = 0.20$) because, in principle, there is a greater willingness to make a type II error than a type I error. By more tightly controlling the type I error, it is more difficult to mistakenly conclude that an indicator has improved when in fact it has not. In other words, greater emphasis is put on protecting the population being served by Feed the Future (e.g., children who are stunted) from false results, since such false results (concluding that stunting has decreased when in fact it has not) could cause a Feed the Future team to be unaware that project or strategy adjustments are needed to achieve targets, and could lead to less success in improving the well-being of the population served. Conversely, by more loosely controlling the type II error, it is easier to make the error of not concluding that an indicator has improved when in fact it has. While less damaging to the populations served, this could result in lost opportunities to demonstrate success of the initiative, and lead a team to believe adjustments are needed where they are not. If Feed the Future teams are concerned about this and if resources permit, they could consider controlling type II error more tightly (at $\beta = 0.10$).

For the above hypothesis test (i.e., whether a Feed the Future indicator has improved over time), a sample size is computed prior to the survey at the baseline. The same sample size is used for the survey at the end-line, and the statistical test of change is conducted after the data from the end-line survey has been collected. The implementation of the statistical test is discussed in more detail in Chapter 12.

For a one-sided test of hypothesis for an indicator of proportions (i.e., for most of the indicators in Tables 1 and 2), the initial sample size formula is given by¹⁹:

$$n_{initial} = D_{est} * \left[\frac{z_{1-\alpha} \sqrt{2 \bar{P} (1-\bar{P})} + z_{1-\beta} \sqrt{P_{1,est} (1-P_{1,est}) + P_{2,est} (1-P_{2,est})}}{\delta} \right]^2 \quad (1)$$

where:

$n_{initial}$ is the initial sample size required by the surveys for each of the two time points (i.e., for both the baseline and end-line PBSs).

δ represents the minimum meaningful effect size to be achieved over the time frame specified by the two surveys; note that $\delta \neq 0$ in order to compute formula (1).

$P_{1,est}$ represents a survey estimate of the true (but unknown) population proportion P_1 at baseline. A value can be obtained from a recent survey that collects data on the same indicator, conducted in the same country or region of the country.²⁰

¹⁹ For a derivation of this sample size formula see Chapter 3 in: Lance, P. and Hattori, A. 2016. *Sampling and Evaluation: A Guide to Sampling for Program Impact Evaluation*. Chapel Hill, North Carolina: MEASURE Evaluation, University of North Carolina. Available at: <https://www.measureevaluation.org/resources/publications/ms-16-112>.

²⁰ Such values can be obtained from any number of internationally sponsored surveys, such as the USAID-sponsored Demographic and Health Surveys (DHS) or the UNICEF-sponsored Multiple Indicator Cluster Surveys (MICS), both of which provide estimates for many maternal and child health and nutrition indicators, at national and subnational levels. In addition, the World Bank-sponsored Living Standards Measurement Studies (LSMS) provide estimates for many poverty-related indicators. It is recommended to use indicator estimates from these sources for the same geographic area in which Feed the Future is operating (typically subnational) if they exist, as national-level estimates may be quite different from subnational estimates. National-level estimates should be used only if subnational estimates do not exist.

$P_{2,est}$ represents a survey estimate of the true (but unknown) population proportion P_2 at end-line. Since $P_{2,est}$ represents a future value that is unknown at baseline when the initial sample size must be computed, it can be approximated by $P_{1,est} - \delta$ for indicators where a decrease is expected over time (such as “Prevalence of Stunted Children under Five”). Similarly, it can be approximated by $P_{1,est} + \delta$ for indicators where an increase is expected over time (such as “Prevalence of Exclusive Breastfeeding”).

$$\bar{p} = \frac{P_{1,est} + P_{2,est}}{2}$$

$Z_{1-\alpha}$ is the value from the Normal Probability Distribution corresponding to a confidence level $1 - \alpha$ (see Table 4a for various possible values of $Z_{1-\alpha}$). For $1 - \alpha = 0.95$, the corresponding value is $z_{0.95} = 1.64$.

$Z_{1-\beta}$ is the value from the Normal Probability Distribution corresponding to a power level of $1 - \beta$ (see Table 4b for various possible values of $Z_{1-\beta}$). For $1 - \beta = 0.80$, the corresponding value is $z_{0.80} = 0.84$.

D_{est} is the estimated design effect (DEFF) of the survey, which represents the ratio of the statistical variance (square of the SE) under the current multi-stage cluster sampling design to the statistical variance under a design using simple random sampling (SRS). Computing a DEFF can be complex, and the numerator of the ratio should include all contributions to the variance that reflect deviations from SRS, such as stratification, allocation (both discussed in Section 4), all stages of sampling that imply any type of “clustering” (e.g., households within sampled EAs or individuals within sampled households, both discussed in greater detail in later chapters) and potential unequal probabilities of selection of units at various stages. For instance, the contribution to the DEFF due to clustering of EAs depends on the size of the sampled EAs and the degree of homogeneity within EAs with respect to the indicator in question, the latter of which is normally measured using an intra-class correlation coefficient (ICC). Often, survey practitioners use estimates of the DEFF from previous surveys; however, this is appropriate only if the design of the previous survey is similar to that of the current survey, which is rarely the case. However, the ICC (which is denoted by ρ and has a different value for each indicator) is more “portable” across surveys because it is independent of the survey design. When the ICC is available, the contribution to the DEFF due to one level of clustering (say, at the EA level) is estimated by:

$$DEFF_{clustering} = 1 + (\rho * (\bar{n} - 1))$$

where \bar{n} is the average sample size of households per EA, across all sampled EAs in the survey.²¹ However, values of ρ are often hard to come by, and, in addition, this estimate of the DEFF represents only a partial contribution to the overall DEFF, because it reflects only the contribution

²¹ If there was more than one level of clustering in the design of the multi-stage PBS, the formula for DEFF due to clustering would be more complex than that given above. However, this guide assumes only one level of clustering in the PBS design and assumes any other contribution to the overall DEFF is negligible.

due to one level of clustering. Nevertheless, when ρ is available, using this formula is the preferred method of computing an estimate of the DEFF.

Sometimes, practitioners will instead use a default value of 2 for the overall DEFF, but this default value is likely an underestimate for many survey designs and for many indicators. For instance, the DEFF across 17 Feed the Future FFP and non-FFP surveys in 14 countries using multi-stage cluster sampling designs similar to the one suggested in this guide have median values 2.25 for the “Prevalence of Stunted Children under Five” indicator (with values ranging between 1.44 and 4.0) and 4.82 for the “Prevalence of Poverty” indicator (with values ranging between 1.25 and 10.89). Similarly, the DEFF across a variety of Gallup surveys, sponsored by the Food and Agriculture Organization of the United Nations (FAO) in 140 countries, using multi-stage cluster sampling designs, have a median value of 5.0 for the “Prevalence of Moderate and Severe Food Insecurity.” In this case, the standard deviation of the DEFF across all countries is approximately 1.0, indicating no excess of variability in the DEFF values across the countries.²²

It is evident that different indicators in the same survey can have very different DEFFs, and therefore it is infeasible to suggest a “one size fits all” rule of thumb DEFF value for all indicators. Furthermore, when ICCs for the various indicators are not available, Feed the Future teams are urged to use the rule-of-thumb values suggested in Table 5 for three key Feed the Future ZOI PBS indicators, where the values in the table are based on the median values from the 17 Feed the Future surveys in 14 countries described above.

Table 4a. $z_{1-\alpha}$ and $z_{1-\alpha/2}$ for Selected Values of α

α (type I error)	$1 - \alpha$ (confidence level)	$z_{1-\alpha}$	$1 - \alpha/2$	$z_{1-\alpha/2}$
0.10	0.90	1.28	0.95	1.64
0.05	0.95	1.64	0.975	1.96
0.025	0.975	1.96	0.9875	2.24
0.01	0.99	2.33	0.995	2.58

Table 4b. $z_{1-\beta}$ for Selected Values of β

β (type II error)	$1 - \beta$ (power)	$z_{1-\beta}$
0.30	0.70	0.53
0.20	0.80	0.84
0.15	0.85	1.03
0.10	0.90	1.28
0.05	0.95	1.64
0.025	0.975	1.96
0.01	0.99	2.33

²² The actual range of values across all countries was not available.

Table 5. Recommended DEFF Values for Select Feed the Future ZOI PBS Indicators

Feed the Future PBS Indicator	Recommended DEFF Value to Use for Sample Size Calculations
Prevalence of Stunted Children under Five	2.0
Prevalence of Moderate and Severe Food Insecurity (based on the FIES)	5.0
Prevalence of Poverty (PP)	5.0

To illustrate the initial sample size computation, consider the following example. A Feed the Future country is planning a baseline PBS in the ZOI, with an end-line PBS conducted after 6 years to test for statistical change in the indicators. As the baseline PBS is being planned, the Feed the Future team knows that one of the potential indicators on which the overall sample size for the survey will be based is the “Prevalence of Stunted Children under Five.” The stunting level for children under 5 years of age is 40%, based on the most recent DHS conducted a year ago in the same region of the country where the ZOI is located. The team sets a target of reducing stunting levels by 20% of the baseline value over the 6-year period. The team decides to set the minimum meaningful effect size to be 80% of the expected target change of 20% reduction in stunting over the 6-year period (i.e., 80% of 20% target reduction from 40% stunting, or 16% reduction from 40% stunting, which is equivalent to 6.4% reduction in stunting), and so they set $\delta = 0.064$.²³ A default value of $DEFF = 2$ (the recommended value given by Feed the Future [see Table 5]) is assumed. Finally, the statistical test of differences is to be conducted assuming a confidence level of $1 - \alpha = 0.95$ and a level of power of $1 - \beta = 0.80$.

In this example:

$$\begin{aligned}
 P_{1,est} &= 0.40 \\
 \delta &= 0.16 * P_{1,est} = 0.064 \\
 P_{2,est} &= P_{1,est} - \delta = 0.336 \\
 z_{1-\alpha} &= 1.64 \\
 z_{1-\beta} &= 0.84 \\
 D_{est} &= 2 \\
 \bar{P} &= (0.40 + 0.336) / 2 = 0.368
 \end{aligned}$$

Plugging these values into formula (1) provides the following:

$$n_{initial} = 2 \left[\frac{1.64\sqrt{2 * 0.368 (1 - 0.368)} + 0.84\sqrt{0.4 (1 - 0.4) + 0.336 (1 - 0.336)}}{0.064} \right]^2 = 1,403$$

²³ The team has decided to set the minimum meaningful effect size to be 80% of the expected target change over the 6-year period to be able to detect changes that come close to but don't quite achieve the full target value.

That means that both the baseline PBS and the end-line PBS must sample 1,403 children under 5 years of age. It is assumed that for each of the two surveys, an independent (i.e., new) sample is drawn.²⁴

2.2.2 Sample Size Calculations to Power Statistical Tests of Differences over Time for Indicators of Means

The sample formula to establish a test of differences for the value of an indicator that is a mean (such as the “Yield of Targeted Agricultural Commodities within Target Areas”) is different from that used for an indicator of proportions. Assume that \bar{X}_1 represents the population mean value of the indicator at baseline and \bar{X}_2 represents the population mean value of the indicator at end-line. If the project is attempting to influence an improvement in such an indicator, then one would expect to see an increase in the mean over time. The null hypothesis can be stated as:

$$H_0: \bar{X}_1 - \bar{X}_2 \geq \delta$$

and the alternative hypothesis as:

$$H_A: \bar{X}_1 - \bar{X}_2 < \delta$$

For a one-sided test of hypothesis for an indicator of means (i.e., for a few of the indicators in Tables 1 and 2), the initial sample size formula is given by²⁵:

$$n_{initial} = D_{est} * \left[\frac{(z_{1-\alpha} + z_{1-\beta})^2 * (\sigma_{X_{1,est}}^2 + \sigma_{X_{2,est}}^2)}{\delta^2} \right] \quad (2)$$

where:

δ represents the minimum meaningful effect size to be achieved over the time frame separated by the two surveys.

$\bar{X}_{1,est}$ represents a survey estimate of the true (but unknown) population mean value \bar{X}_1 at baseline. A value for this can be obtained from a recent survey that collects data on the same indicator, conducted in the same country or region of the country.

$\bar{X}_{2,est}$ represents a survey estimate of the true (but unknown) population mean value \bar{X}_2 at end-line. Since $\bar{X}_{2,est}$ represents a future value that is unknown at the time that the sample size computation is made at baseline, it can be approximated by $\bar{X}_{1,est} + \delta$ in cases where an increase is expected over time or by $\bar{X}_{1,est} - \delta$ in cases where a decrease is expected over time.

$\sigma_{X_{1,est}}$ is the standard deviation of the distribution of $X_{1,est}$, which is the distribution of individual values underpinning the indicator across all sampled individuals at baseline. An estimate of $\sigma_{X_{1,est}}$ can be obtained from a recent survey that collects data on the same indicator, conducted in the same

²⁴ This scenario is in contrast to one where the same random sample is used at both time points, i.e., from a longitudinal panel survey. In that scenario, a sample is randomly drawn at baseline and the sampled units are tracked and re-interviewed at end-line. Although the use of panel samples is discouraged in the Feed the Future context, a more in-depth discussion regarding the potential use of the same set of first stage sampled clusters or EAs over the two time points is provided in Section 5.2.2.

²⁵ For a derivation of this sample size formula see Chapter 3 in: Lance, P. and Hattori, A. 2016.

country or region of the country. If no such survey exists, an estimate can be obtained from the following approximation based on the normal distribution:

$$\sigma_{X_{1,est}} = \text{Standard Deviation } (X_{1,est})$$

$$= \frac{\text{maximum value of } X_{1,est} \text{ for any individual} - \text{minimum value of } X_{1,est} \text{ for any individual}}{6} \quad (3)$$

Plausible maximum and minimum values for an individual are estimated by the Feed the Future team, using experience and expert knowledge as guides.

$\sigma_{X_{2,est}}$ is the standard deviation of the distribution of $X_{2,est}$, which is the distribution of individual values underpinning the indicator across all sampled individuals at end-line. Since the sample size computation is initially undertaken at the time of the baseline, this end-line value is unknown.

Therefore, one can simply set the estimate for $\sigma_{X_{2,est}}$ to be the same as that for $\sigma_{X_{1,est}}$.²⁶

$Z_{1-\alpha}$, $Z_{1-\beta}$, and D_{est} are the same as in Section 2.2.1.

The following example illustrates the computation of the sample size for this type of indicator. The Feed the Future team plans a baseline PBS and an end-line PBS 6 years after the baseline. It has decided that one of the potential indicators on which the overall sample size for the survey will be based is the “Yield of Targeted Agricultural Commodities within Target Areas” indicator, where the commodity is maize. From the most recent agriculture survey conducted a year ago in the same country, it is known that the average yield of maize for a typical producer is 1.50 metric tons per hectare in the region where the ZOI is located.

The team determines that yield for maize needs to increase to 1.75 metric tons per hectare over the 6-year period to achieve the targeted increases in other indicators, such as income from agricultural sales. The team decides to set the minimum meaningful effect size to be 80% of the expected target change over the 6-year period (i.e., 80% of $[1.75 - 1.50 = 0.25] = 0.20$), so that an increase to 1.70 metric tons ($1.50 + 0.20$) per hectare of maize after 6 years will be considered meaningful. Therefore, they set $\delta = 1.70 - 1.50 = 0.20$.

No prior information exists for the standard deviations of this indicator, so formula (3) is used to obtain an estimate. The team estimates that the minimum value for the yield of maize for any individual producer in the ZOI is 0 metric tons per hectare and that the maximum yield of maize for any individual producer is 2.4 metric tons per hectare. Using formula (3), an estimate for the standard deviation is $(2.4 - 0.0) / 6 = 0.4$ metric tons per hectare. Given the absence of any other information, the same standard deviation is assumed for both baseline and end-line. A prior agriculture survey, which has a similar sampling design to the current one being planned, gives a DEFF value for the yield indicator as $D_{est} = 3$.

The same assumptions on $Z_{1-\alpha}$ and $Z_{1-\beta}$ are made as for the prior example for proportions.

²⁶ Alternatively, one could also assume a constant coefficient of variation over the two time points. For instance, one could take $\sigma_{X_{2,est}}$ to be a prorated estimate of $\sigma_{X_{1,est}}$. In other words, $\sigma_{X_{2,est}} = \left(\frac{\bar{X}_{2,est}}{\bar{X}_{1,est}}\right) * \sigma_{X_{1,est}}$.

In the above example:

$$\begin{aligned}\bar{X}_{1,est} &= 1.5 \\ \bar{X}_{2,est} &= 1.70 \\ \delta &= 0.2 \\ \sigma_{X_{1,est}} &= 0.4 \\ \sigma_{X_{2,est}} &= 0.4 \\ z_{1-\alpha} &= 1.64 \\ z_{1-\beta} &= 0.84 \\ D_{est} &= 3\end{aligned}$$

Plugging these values into formula (2) results in:

$$n_{initial} = 3 * \left[\frac{(1.64 + 0.84)^2 * (0.4^2 + 0.4^2)}{0.2^2} \right] = 149$$

That means that the surveys must collect data on 149 producers of maize. It is assumed that for both the baseline and end-line PBSs, an independent (i.e., new) sample of 149 producers of maize will be drawn.

2.2.3 Computing the Final Sample Size for the Survey

When calculating an overall sample size for a PBS, one should keep in mind that the survey will collect data in support of a substantial number of indicators, some of which may be proportions and others of which may be means (see Tables 1 and 2), but all of which will have their own sample size requirement. Even when considering indicators within the same sampling group (e.g., two indicators at the household level), different input parameters will likely result in different sample sizes. However, one indicator only, from among all indicators on which data are to be collected through the survey, can determine the overall sample size for the survey. The challenge lies in selecting that indicator.

The recommendation is that the sample size for all **key** indicators from among the indicators being collected in the survey should be calculated and that the **largest sample size** resulting from all candidate sample sizes computed should be chosen to be the overall sample size for the survey.

In the case of Feed the Future comparative analytical PBSs, the recommended key indicators on which to base the sample size calculation are²⁷:

1. Prevalence of Moderate and Severe Food Insecurity
2. Prevalence of Stunted Children under Five
3. Prevalence of Poverty (PP) at \$1.90 2011 Purchasing Power Parity (PPP)

Because these three high-level indicators are likely to have the smallest amount of targeted change among all Feed the Future ZOI PBS indicators, the required sample size will almost always be large

²⁷ For further guidance, see: "Feed the Future ZOI Survey Guidance for Target Countries." 2018. Available at: <https://agrilinks.org/post/feed-future-zoi-survey-methods>.

enough to capture statistically significant changes (if they occur) in lower-level Feed the Future ZOI PBS indicators, such as “Prevalence of Exclusive Breastfeeding (EBF)” and “Prevalence of Children Receiving a Minimum Acceptable Diet (MAD).” Although many households may need to be visited to find one child in the appropriate age range for these last two indicators (i.e., children age 0–5 months for EBF and children age 6–23 months for MAD), the amount of change targeted by Feed the Future IPs is typically much larger for EBF and MAD than it is for the three key indicators. Because the size of the targeted change typically overwhelms the rarity of the sample group in the sample size computations, the indicators needing to capture larger change will have smaller sample sizes associated with them, and therefore the sample size requirements for EBF and MAD will likely be smaller than those for the three key indicators.

Regardless, there will always be one indicator that ultimately drives the overall sample size of the survey (i.e., the one with the largest sample size needs), and this overall sample size will be applicable to all indicators collected through the survey. This effectively means that all other indicators will likely have more sample size than is required and a statistical test of differences that is undertaken at end-line for these indicators will likely have more power than necessary.

Table 6 illustrates the calculation of the initial sample size for the three key Feed the Future ZOI PBS indicators, assuming the input parameters given in the table.

Table 6. Calculation of Initial Sample Size for Three Key Feed the Future ZOI PBS Indicators

Indicator	Sampling Group	$P_{1, est}$	δ	$P_{2, est}$	\bar{P}	$z_{1-\alpha}$	$z_{1-\beta}$	D_{est}	$n_{initial}$
Prevalence of Moderate and Severe Food Insecurity	Household	0.50	0.080	0.42	0.46	1.64	0.84	5	2,395
Prevalence of Stunted Children under Five	Children 0–59 months	0.40	0.064	0.336	0.368	1.64	0.84	2	1,403
Prevalence of Poverty	Household	0.60	0.096	0.504	0.552	1.64	0.84	5	1,654

In Table 6, sample sizes based on the three key Feed the Future ZOI PBS indicators have been computed, assuming a 20% reduction from baseline values over 6 years for each of the indicators, and assuming a minimum meaningful effect size of 80% of the target, thus using $\delta = (0.80 * 0.2 * P_{1, est}) = (0.16 * P_{1, est})$. The estimated values for $P_{1, est}$ are obtained from external sources and it is assumed that $P_{2, est} = P_{1, est} - \delta$ for all three indicators, because successful implementation of the strategy translates to a decrease over time in the value of all three indicators. The values of D_{est} are taken from Table 5. The computation produces initial sample sizes $n_{initial} = 2,395$, $n_{initial} = 1,403$, and $n_{initial} = 1,654$ for “Prevalence of Moderate and Severe Food Insecurity,” “Prevalence of Stunted Children under Five,” and “Prevalence of Poverty,” respectively.

However, before the sample size for the survey can be finalized, two adjustments to these initial sample sizes must be made: inflation for the number of households to contact and inflation for anticipated household non-response.

2.2.4 Adjustment I: Inflation for the Number of Households to Contact

Although a basic sample size for each indicator under consideration should be computed, a final choice of sample size for the survey can be made only after the sample size computations for all indicators under consideration are comparable. As such, one must take into account the sampling groups involved in indicators under consideration when deciding on the overall sample size for the survey. For example, the data for the “Prevalence of Exclusive Breastfeeding (EBF) of Children under 6 Months of Age” indicator are collected for children 0–5 months, whereas the data for the “Prevalence of Stunted Children under Five” indicator are collected for children 59 months of age or less.

Furthermore, although sample size requirements are expressed in terms of the sampling groups associated with each of the indicators, the sample size must then be converted to reflect the number of households that must be contacted to encounter the targeted number of individuals, as households rather than individuals are the first point of contact in the field. Since the correspondence between households and eligible members of the household for each sampling group is not one-to-one, each indicator under consideration has a different conversion factor in terms of the number of households to contact. This conversion effectively puts all sample size computations on a “level playing field” in terms of the number of households that must be visited in each case, so that the sample size requirements for each indicator can be compared. As a result, the sample size requirements (in terms of the number of households to be contacted) for indicators related to more “rare” sampling groups, such as children under 6 months of age for the EBF indicator, may be greater than that for other indicators.

Although some households will have exactly one member from a given sampling group, other households will have more than one eligible member from the sampling group and some households will have no eligible members from the sampling group. For sampling purposes, it is essential therefore to have not only an estimate of the number of eligible members from the various sampling groups that must be sampled, but also an estimate of the number of households that need to be visited to obtain the required sample of eligible members from the associated sampling groups.

For these reasons, the initial sample size, $n_{initial}$, should be inflated to determine the number of households to contact to meet the required sample for individual-level sampling groups. The following formula should be used:

$$adj_1 = \left[A * \frac{1}{(1 - e^{-\lambda})} \right] + 0.5 \left[(1 - A) * \frac{1}{(1 - e^{-\lambda})} \right] \quad (4)$$

where:

$$A = (1 + \lambda) * e^{-\lambda}$$

The derivation of this formula is provided in Appendix A of this guide²⁸. In the above formula, e refers to the exponential function, found on any scientific hand calculator under the symbol exp or e^x . The parameter λ is the estimated **average** number of individuals in the sampling group per household. For example, consider the stunting indicator with the unit of analysis being children 0–59 months of age, and assume that in the geographic area being surveyed the average household size is 4.27 and the proportion

²⁸ This inflator was previously published in 2012 as an addendum to the FANTA Sampling Guide (1999) by Robert Magnani; the inflator is sometimes referred to as the “Stukel-Deitchler Inflator”.

of children in the population that are 0–59 months is roughly 0.16 (equivalent to 16%),²⁹ then the estimated average number of children 59 months of age or less per household is $\lambda = 4.27 * 0.16 = 0.6832$. The above formula is valid for all values of $\lambda \leq 1.5$. For values of $\lambda > 1.5$, it can be shown that the computation will result in an overall **deflation** from the original sample size n ; however, a value of $\lambda > 1.5$ rarely occurs in practice.

Finally, the inflation adjustment given in formula (4) assumes that all eligible members of the sampling group within a sampled household will be selected for interviewing, which is the procedure that should be followed for Feed the Future ZOI PBSs. More details on “take all” sampling of individuals within households is given in Chapter 9.³⁰

The sample size adjusted for the number of households to contact is then given by:

$$n_{adj_1} = n_{initial} * adj_1$$

where $n_{initial}$ is given by equation (1) or (2), depending on whether the indicator is a proportion or a mean, respectively.

Note that if the sample size calculation is based on a household-level indicator, such as the “Prevalence of Poverty” or the “Prevalence of Moderate and Severe Food Insecurity,” the inflation factor described above is not required because the household is the sample unit for the indicator, and one can set $adj_1 = 1$.

To illustrate the first adjustment, we continue the example given in Section 2.2.1 on the “Prevalence of Stunted Children under Five.” Assume $n_{initial} = 1,403$ and $\lambda = 0.6832$, as per the computation above. Also, assume that all eligible children age 0–59 months within a sampled household are selected for the survey. Therefore, using formula (4):

$$A = (1 + 0.6832) * e^{-0.6832} = (1.6832) * 0.5049 = .8500$$

and

$$adj_1 = \left[0.8500 * \frac{1}{(1 - 0.5050)} \right] + 0.5 \left[(1 - 0.8500) * \frac{1}{(1 - 0.5050)} \right] = 1.8688$$

Finally,

$$n_{adj_1} = 1,403 * 1.8688 = 2,622$$

Therefore, 2,622 households must be sampled to achieve the goal of collecting data on 1,403 children age 0–59 months, assuming all sampled households respond.

²⁹ Figures for both the average household size and the proportion of children in the target age group are typically obtained from the most recent national census or from some other national or internationally sponsored survey.

³⁰ If only one eligible individual from the sampling group is randomly selected in sampled households having multiple eligible individuals (not the recommended approach), the formula given in (4) simplifies considerably, although field logistics and sample weighting are more complicated. See Appendix A for details.

2.2.5 Adjustment 2: Inflation for Anticipated Household Non-Response

Another adjustment that needs to be made relates to the anticipated household non-response (denoted by adj_2). In all household surveys, it is expected that some proportion of households selected for the survey will be unreachable, unavailable, or unwilling to respond to any of the survey questions; this is called household non-response.

Despite the best efforts of interviewers, there is usually some residual non-response that remains, even after several attempts to complete an interview with the household selected for the sample. To ensure that the targeted number of households provide completed interviews despite household non-response, the sample size should be inflated by multiplying by the inverse of the expected response rate so that the resultant sample size after fieldwork is as close as possible to the targeted sample size.³¹

The expected response rate can be estimated using information from reports on PBSs conducted in the same geographic area and with the same (or similar) survey population, including reports on Feed the Future ZOI surveys, the Demographic and Health Surveys (DHS), the Living Standards Measurement Studies (LSMS), or the Multiple Indicator Cluster Surveys (MICS).

If no past information is available on non-response rates, a generally accepted rule of thumb is to assume an estimated response rate of 90%–95%. That is to say, if a response rate of 95% is assumed, then the sample size should be multiplied by $adj_2 = 1/0.95 = 1.052$. If there are reasons to believe that the response rate will be low (e.g., if the planned number of attempts to reach selected households is low or if the length of the survey questionnaire is long), then it is best to assume a response rate that is closer to 90%. However, based on response rates obtained in previous ZOI PBSs, Feed the Future recommends assuming an anticipated response rate of 95%.

2.2.6 Computing the Final Sample Size for the Survey

The final sample size (denoted by n_{final}), which is a product of the initial sample size and both adjustments, then becomes:

$$n_{final} = n_{initial} * adj_1 * adj_2 = n_{adj_1} * adj_2 \quad (5)$$

To illustrate the computation of the final sample size, the example of the “Prevalence of Stunted Children under Five” indicator continues, assuming $n_{adj_1} = 2,622$ households and $adj_2 = 1/0.95 = 1.052$. In this case, $n_{final} = 2,622 * 1.052 = 2,760$. This means that $n_{final} = 2,760$ households should be sampled to ensure that anthropometric data on $n_{initial} = 1,403$ children age 0–59 months are

³¹ Inflation of the sample size to account for anticipated non-response in individuals who are selected for interviewing within each sampled household is not made, given the differential patterns of missingness at the individual level across indicators within the same sampling group. For instance, the “Prevalence of Stunted Children under Five” indicator may experience one level of missingness (with respect to height/length data) but the “Prevalence of Wasted Children under Five” indicator may experience a different level of missingness (with respect to weight data)—although both indicators collect data on children under the age of 5 years. Therefore, it is infeasible to inflate for anticipated non-response for children under the age of 5 years (the sampling group in question), because the missingness pattern differs across different indicators for the same sampling group. Alternatively, inflating for anticipated individual non-response at the indicator level (rather than the sampling group level) is considered too unwieldy because it necessitates different sampling weights for each indicator rather than just for each sampling group, which is the more typical practice.

collected, after accounting for the number of households that need to be visited and household non-response.

Table 7 continues the prior example to illustrate the computation of the final sample size for three key Feed the Future ZOI PBS indicators.

Table 7. Calculation of Final Sample Size for Three Key Feed the Future ZOI PBS Indicators

Indicator	Sampling Group	$n_{initial}$	λ	adj_1	n_{adj_1}	adj_2	n_{final}
Prevalence of Moderate and Severe Food Insecurity	Household	2,395	N/A	1.0	2,395	1.052	2,521
Prevalence of Stunted Children under Five	Children 0–59 months	1,403	0.6832	1.8688	2,622	1.052	2,760
Prevalence of Poverty	Household	1,654	N/A	1.0	1,654	1.052	1,741

Recall that the recommendation for determining the overall sample size for the survey (which is denoted $n_{final-overall}$) is to compute the sample size for all **key** indicators from among the indicators being collected in the survey and then to choose the **largest sample size** resulting from all candidate sample sizes. In the above example, the “Prevalence of Stunted Children under Five” indicator has the greatest sample size requirement, namely, that 2,760 households be sampled to ensure that anthropometric data on 1,403 children 0–59 months old are collected. Therefore, the overall sample size for the survey is $n_{final-overall} = 2,760$ households. Because the largest sample size is chosen as the overall sample size for the survey, it meets and exceeds the needs of the other two indicators, “Prevalence of Moderate and Severe Food Insecurity” and “Prevalence of Poverty.”

Note that Feed the Future estimates are required to be produced for the indicator disaggregates specified in Tables 1 and 2. Depending on the indicator, there are various required disaggregates, for example, by sex (two categories: male, female), by age (two sets of categories: 15–29 or 18–29, ≥ 30), by gendered household type (four categories: adult male but no adult female, adult female but no adult male, adult male and adult female, child but no adult), and by commodity type (number of categories varies by project). This means that, for instance, because there are four categories of gendered household type, Feed the Future requires reporting on the “Prevalence of Poverty” for each of the four categories. While it would be ideal to conduct statistical tests of differences for the “Prevalence of Poverty” at the level of each of the four disaggregated categories with the same power and statistical significance as that of the overall level of the indicator, doing so would entail undertaking a separate sample size calculation for each of the four categories of gendered household type, and then summing the estimated sample sizes for each disaggregate to get the total sample size required for the indicator. This would result in very large and likely infeasible total sample sizes. In addition, such a calculation would necessitate taking into account the input parameters specific to the disaggregates implied by each gendered household type (e.g., the prevalence of poverty for each gendered household type). More importantly, ensuring a survey with a sample size that included the appropriate number of each gendered household type would entail having information on the sampling frame available relating to gendered household type, so that each gendered household type could be specifically targeted for

sampling (i.e., randomly selecting a specific sample for each gendered household type). However, it is very unlikely that such information on disaggregates would be available on the sampling frame and, therefore, in most cases, it would be infeasible to undertake statistical tests of differences for disaggregates of indicators at the same level of power as the overall level of the indicators. Therefore, sample size calculations based on disaggregates and associated statistical tests at the disaggregate level should be avoided for all indicators.

2.2.7 Adjusting the Final Sample Size at the Second Time Point prior to the End-Line PBS

There is a final issue worthy of consideration for those undertaking sample size calculations. As mentioned earlier, both the baseline and end-line PBSs must sample $n_{final-overall}$ households. That is to say, an independent (i.e., new) sample of $n_{final-overall}$ households is drawn at each of the two time points. As a result, the combined sample size across the two time points should be $2 * n_{final-overall}$.

However, it is not always the case that the sample sizes at baseline and end-line are identical. There are two scenarios in which a survey implementer may wish to re-compute the sample size for the indicator driving the sample size of the survey before the end-line PBS is undertaken. Appropriate methods for addressing the two scenarios are presented below.

The first scenario whereby a survey implementer may wish to re-compute the sample size occurs after conducting the baseline PBS. At that point, it is possible to compute the input parameters for the sample size computation based on actual values from the baseline and determine if the initial sample size computed for the baseline was “accurate.”³² If it is discovered that $n_{initial}$ for this indicator actually fell short of what it should have been at baseline based on the updated parameters, a remedial measure that is typically used, albeit sub-optimal, has two parts: i) revise the sample size for the end-line based on the updated input parameters based on the results of the baseline and ii) “top-up” the sample size for the end-line by the shortfall amount encountered at baseline based on the updated value of $n_{initial}$.

The second scenario whereby a survey implementer may wish to re-compute the sample size also occurs after conducting the baseline but is somewhat simpler than the first scenario. It may be found that the input parameters for the sample size computation were accurate (as verified by a re-computation after the baseline is conducted), but that, regardless, the final sample size resulting from the first survey did not meet the target sample size, $n_{initial}$. This could happen for a multitude of reasons, including unanticipated high non-response rates or suspension of data collection due to security issues partway through the baseline. In this case, the typical (but sub-optimal) remedial measure is simply to top-up the sample size for the end-line by the amount of the shortfall encountered at the baseline.

Regardless of which of the two scenarios occurs (or indeed even if both scenarios occur), the end result is that there is a shortfall in the sample size from the baseline that needs to be rectified prior to commencing fieldwork for the end-line. However, neither of the above remedial measures that are

³² For example, prior to the end-line PBS, it may be of value to investigate whether the DEFF on which the initial sample size was based is accurate, using the results of the baseline PBS. To that end, the intra-cluster correlation from the baseline PBS can be computed and used as input to re-compute an updated DEFF in advance of the end-line PBS. The formula to use is that for $DEFF_{clustering}$ given in Section 2.2.1.

typically used is recommended. The reasons why are illustrated through an example of the second scenario.

Continuing the example of the “Prevalence of Stunted Children under Five” indicator referenced earlier, assume that a sample size of $n_{initial} = 1,403$ children under 5 years of age has been used at baseline, and, after conducting that survey, it is discovered that the actual realized sample size was $n_{initial,actual} = 1,000$ and therefore there was a shortfall for the first survey of 403 children. A survey implementer may feel tempted to top up the sample size for the end-line by the shortfall encountered at baseline by using a sample size at end-line of $1,403 + 403 = 1,806$. The problem with this approach is that the power of the statistical test produced by the pair of sample sizes (1,000, 1,806) for the two time points will rarely, if ever, be at the same level as that using the ideal pair of sample sizes (1,403, 1,403) for the two time points. For instance, a simple computation shows that while the pair of sample sizes (1,403, 1,403) results in a power of 80%, the pair (1,000, 1,806) results in a power of 77%. If the sample size does not include the top-up in the amount of the shortfall, and the pair (1,000, 1,403) is used instead, the power is only 73%.

As can be seen from this example, topping up the sample for the end-line to compensate for any shortfalls in the intuitive way described above does not produce the power that was envisioned at the time of the baseline—and therefore is not recommended. Instead, regardless of which of the two scenarios above occurs, when there is a need to increase the sample size for the end-line to compensate for shortcomings at the baseline, it is best to find a pair of sample sizes that has the same power as the original sample size calculation, where the first element of the pair, the sample size at baseline, is the one actually realized at the baseline ($n_{initial,actual}$).³³ The appropriate value for the second element of the pair can be obtained by using a multiplicative inflation factor (K) to be applied to the actual realized sample size at baseline ($n_{initial,actual}$) and can be approximated using the input parameters estimated from the baseline by applying one of the following two formulas:

For indicators that are proportions³⁴:

$$K = \frac{D_{actual} * P_{2,target} * (1 - P_{2,target}) * \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(P_{1,actual} - P_{2,target})^2}}{n_{initial,actual} - [D_{actual} * P_{1,actual} * (1 - P_{1,actual}) * \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(P_{1,actual} - P_{2,target})^2}]} \quad (6)$$

where $P_{2,target} = P_{1,actual} \pm \delta$ depending on whether an increase or decrease is expected in the indicator over time. $P_{1,actual}$ and D_{actual} values are computed from the baseline PBS.

³³ If the sample size at baseline exceeded what was required, then no adjustment is needed to the sample size at end-line. This is because the pair of sample sizes will then provide power to the statistical test in excess of 80%, which is acceptable because the type II error is reduced.

³⁴ Formula (6) is based on a somewhat different statistical test of differences than the one given in formula (1) for proportions, but the resulting sample sizes for the two statistical tests of differences can be shown to be roughly the same. Formula (6) was derived by the author of this guide by drawing on the equation for unequal sample sizes at the two time points for indicators of proportions given in: Chow, S.; Shao, J.; and Wang, H. 2008. “Sample Size Calculations in Clinical Research.” Second Edition. Boca Raton, FL: Chapman & Hall/CRC, Taylor & Francis Group. Page 89.

For indicators that are means³⁵:

$$K = \frac{D_{actual} * \sigma_{X_{1,actual}}^2 * \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(\bar{X}_{1,actual} - \bar{X}_{2,target})^2}}{n_{initial,actual} - [D_{actual} * \sigma_{X_{1,actual}}^2 * \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(\bar{X}_{1,actual} - \bar{X}_{2,target})^2}]} \quad (7)$$

where $\bar{X}_{2,target} = \bar{X}_{1,actual} \pm \delta$ depending on whether an increase or decrease is expected in the indicator over time. $\bar{X}_{1,actual}$, D_{actual} , and $\sigma_{X_{1,actual}}$ values are computed from the baseline PBS.

Using the multiplicative factor, the sample size to be used at end-line is given by $n_{initial,adj} = n_{initial,actual} * K$. Finally, if one is using an individual-level indicator to drive the sample size for the survey, $n_{initial,adj}$ must also be adjusted to the appropriate number of households using formula (4), to arrive at a final sample size for the survey at end-line, $n_{final-overall,adj}$ (after inflating for anticipated household non-response as well).

The following numerical illustration continues the example used earlier where the final sample size for the survey was $n_{final-overall} = 2,760$ households and $n_{initial} = 1,403$ children under 5 years of age based on the “Prevalence of Stunted Children under Five” indicator. Suppose that after the baseline was conducted, anthropometric data on only $n_{initial,actual} = 1,000$ children were obtained, rather than the expected $n_{initial} = 1,403$ children. Assume that the input parameters were re-computed using the baseline data, and it was found that they remained unchanged from the estimates used during the initial sample size calculation (i.e., $D_{actual} = D_{est} = 2$, $P_{1,actual} = P_{1,est} = 0.40$, and $P_{2,target} = P_{2,est} = 0.336$). Then, using formula (6) for proportions, the sample size at end-line would be inflated by:

$$K = \frac{2 * 0.336 * (1 - 0.336) * \frac{(1.64+0.84)^2}{(0.4-0.336)^2}}{1,000 - [2 * 0.4 * (1 - 0.4) * \frac{(1.64+0.84)^2}{(0.4-0.34)^2}]} = 2.445$$

and the new initial sample size for the survey at end-line is $n_{initial,adj} = 1,000 * 2.445 = 2,445$. This means that at end-line, $n_{initial,adj} = 2,445$ children (rather than the initially computed $n_{initial} = 1,403$ children) should be sampled to maintain 80% power over the two survey time points. This example illustrates that while the pair (1,000, 1,806) results in a power of 77% (shown earlier), it takes substantially more than 1,806 children at end-line to achieve the desired power of 80%; as we can see, it is the pair (1,000, 2,445) that finally results in a power of 80%.

In cases such as these where there is a sample shortfall that needs to be compensated for at end-line, teams should weigh the costs of such a large increase in sample size for the relatively small increase in power when deciding whether or not to increase the sample size at end-line.

³⁵ Formula (7) was derived by the author of this guide by drawing on the equation for unequal sample sizes at the two time points for indicators of means given in: Chow, S.; Shao, J.; and Wang, H. 2008. Page 58.

As a final note, this number ($n_{initial,adj} = 2,445$ children) must be inflated using the adjustments from Section 2.2.4 and Section 2.2.5, to arrive at the final adjusted sample size for the survey at end-line, $n_{final-overall,adj}$.

2.3 Sample Size Calculations to Ensure Adequate Precision for Estimates of Indicators of Proportions or Means Using Descriptive PBSs

This section provides a description of the sample size formulas that should be used to ensure “high precision” single-point-in-time estimates of indicators of proportions or means, along with their SEs and CIs, in support of descriptive PBSs. For such purposes, in the Feed the Future context, monitoring PBSs are conducted as described in Box 1. Such monitoring PBSs are conducted 3 years after the baseline PBS.

2.3.1 Sample Size Calculations to Ensure Adequate Precision for Estimates of Indicators of Proportions

The formula for calculating the initial sample size for the estimation of indicators of proportions that ensures adequate precision is given by:

$$n_{initial} = \frac{D_{est} * z_{1-\alpha/2}^2 * P_{est} * (1 - P_{est})}{MOE^2} \quad (8)$$

where:

D_{est} is the estimated DEFF of the survey. The recommended values to use for key Feed the Future ZOI PBS indicators are given in Table 5, and should be used unless the monitoring PBS has a similar sample design as the baseline PBS, in which case DEFF computed from the baseline should be used instead.

$z_{1-\alpha/2}$ is the critical value from the Normal Probability Distribution. For Feed the Future, the significance level is typically set at $\alpha = 0.05$, giving a value of $z_{1-0.05/2} = z_{0.975} = 1.96$ (see Table 4a).

P_{est} represents the estimated prevalence (or proportion) at the time of the monitoring PBS. A value for this can be obtained from the baseline PBS after adjusting for the target increase or decrease to be achieved in the intervening 3 years.

MOE is the margin of error. This value is typically set between 5% and 10%, but it is recommended that Feed the Future teams set the margin of error to 5% or $MOE = 0.05$, because a relatively high level of precision is required for Feed the Future performance monitoring and reporting purposes.

To illustrate, suppose a Feed the Future team wants to conduct a monitoring PBS 3 years after the baseline PBS has been conducted. The team decides that one of the potential indicators on which the overall sample size for the survey is to be based is the “Prevalence of Stunted Children under Five” indicator. The baseline value for stunted children was 40%. Their targeted reduction over 6 years is 20%, so the team assumes the value for stunted children after 3 years would be $0.40 * (1 - ((3 / 6) * 0.20)) = 0.36$ or 36.0%. Using the value from Table 5, they assume $DEFF = 2$.

In the above example, $D_{est} = 2$, $z_{1-\alpha/2} = z_{0.975} = 1.96$, $P_{est} = 0.36$, and $MOE = 0.05$. Plugging these values into formula (8) results in the following value:

$$n_{initial} = \frac{2 * 1.96^2 * 0.36 * (1 - 0.36)}{0.05^2} = 709$$

This means that the monitoring PBS must sample 709 children under 5 years of age in support of the “Prevalence of Stunted Children under Five” indicator.

2.3.2 Sample Size Calculations to Ensure Adequate Precision for Estimates of Indicators of Means

The formula for calculating the initial sample size for the estimation of indicators of means that ensures adequate precision is given by:

$$n_{initial} = \frac{D_{est} * z_{1-\alpha/2}^2 * \sigma_{X_{est}}^2}{MOE^2} \quad (9)$$

where:

$\sigma_{X_{est}}$ is the standard deviation of the distribution of X_{est} , and is the distribution of values underpinning the indicator across all sampled individuals at the time of the monitoring PBS. A value for this can be obtained from a recent survey that collects data on the same indicator, conducted in the same country or region of the country. If such a survey does not exist, an estimate can be obtained from the approximation given in formula (3).

MOE is the margin of error. In the case of indicators of means, it is recommended that Feed the Future teams set the margin of error to 5% of the target value of the indicator for the year in which the monitoring PBS is taking place, or $MOE = 0.05 * \bar{X}_{est}$.

\bar{X}_{est} represents the target value of the indicator value (a mean) at the time of the monitoring survey. A value for this can be obtained from the baseline PBS after adjusting for the target increase or decrease to be achieved in the intervening 3 years.

$z_{1-\alpha/2}$ and D_{est} are the same as in Section 2.3.1.

To illustrate, suppose the Feed the Future team decides to conduct a monitoring PBS 3 years after the baseline PBS has been conducted, and one of the potential indicators on which the overall sample size for the survey is to be based is the “Yield of Targeted Agricultural Commodities within Target Areas” indicator for the commodity of maize. The baseline value for this indicator was 1.50 metric tons per hectare of maize. The target over 6 years is to increase this to 1.75 metric tons per hectare of maize or to increase the indicator by 0.25 metric tons per hectare. So, the assumed value for “Yield of Targeted Agricultural Commodities within Target Areas” after 3 years would be $1.5 + ((3 / 6) * 0.25) = 1.625$ or 1.625 metric tons per hectare of maize. Because no prior information exists for the standard deviations for this indicator, formula (3) is used to obtain an estimate. It is known that the minimum value for any individual producer in the ZOI in which the project is working is 0 metric tons per hectare of maize and the maximum value for any individual producer in the same area is 2.4 metric tons per hectare of maize. Using formula (3), an estimate for the standard deviation is $(2.4 - 0) / 6 = 0.4$. Using values from the

baseline PBS, a value for DEFF is assumed as $D_{est} = 3$. The same assumption on $z_{1-\alpha/2}$ as for the example above is made.

In the above example, $\bar{X}_{est} = 1.625$, $\sigma_{\bar{X}_{est}} = 0.4$, $z_{1-\alpha/2} = z_{0.975} = 1.96$, $MOE = 0.05 * 1.625 = 0.08125$, and $D_{est} = 3$. Plugging these values into formula (9) results in the following value:

$$n_{initial} = \frac{3 * 1.96^2 * 0.4^2}{0.08125^2} = 280$$

Because data on yield are only meaningful at the commodity level, this means that the monitoring PBS must collect data from 280 producers of maize. Separate sample size computations will be needed for the “Yield of Targeted Agricultural Commodities within Target Areas” indicator in relation to other applicable commodities (if there are any).

2.3.3 Computing the Final Sample Size for the Survey

Similar to the case of comparative analytical PBSs, before the sample size can be finalized, the same two adjustments from Section 2.2.4 and Section 2.2.5 must be made to the initial sample size: adjustment 1 to inflate the number of households to be contacted (if the indicator is at the individual level only) and adjustment 2 to inflate for the anticipated household non-response. Note that for indicators such as per capita expenditure (PCE) in Table 2, adjustment 1 is not needed since the sampling group for the indicator is the household itself. In summary, the final sample size, n_{final} , is given by formula (5), and $n_{initial}$ is given by formula (8) or formula (9), depending on whether the indicator is a proportion or mean, respectively.

As with the case of comparative analytical surveys in Section 2.2.3, the same general recommendation is made: that the sample size (n_{final}) for all **key** indicators from among the indicators being collected in the survey be calculated and that the **largest sample size** resulting from all candidate sample sizes computed be chosen to be the overall sample size for the survey, denoted $n_{final-overall}$.

In the case of descriptive PBSs in support of Feed the Future, the recommended key indicators on which to base the sample size calculation are the same as those used to determine the overall sample size for comparative analytical PBSs:

1. Prevalence of Moderate and Severe Food Insecurity
2. Prevalence of Stunted Children under Five
3. Prevalence of Poverty at \$1.90 2011 PPP

With descriptive PBSs, it is possible that the overall sample size for the survey can be small,³⁶ and therefore Feed the Future is recommending that survey implementers adopt a minimum overall sample size for the survey of 1,050 households. That is to say $n_{final-overall}$ should be 1,050 or more, after taking into account the two adjustments to $n_{initial}$, the second adjustment of which assumes an

³⁶ While it is often the case for descriptive PBSs that sample sizes resulting from formula (8) or (9) may be small, such sample sizes are more often reasonably large for comparative analytical PBSs using formula (1) or (2). For this reason, we do not include a discussion on recommended minimum sample size for comparative analytical PBSs.

anticipated household response rate of 95%. If the actual household response rate encountered in the field is 95%, then there will be completed interviews for 1,000 sampled households.

A minimum overall sample size for the survey of 1,050 households (or 1,000 households after household non-response) has been established based on the need to:

- Ensure reasonable precision for required disaggregates (see Tables 1 and 2), given that each category of disaggregate will have a sample size of considerably fewer than 1,050 households (or 1,000 households after household non-response)
- Ensure reasonable precision for district or other subproject-level geographic areas, should Feed the Future teams wish to produce these for their own internal monitoring needs
- Justify the basic cost investment in the PBS implementation

3. Development of a Sampling Frame

As mentioned earlier, this guide focuses on the use of stratified multi-stage cluster sampling designs, where it is assumed that there are three or four stages of sampling: i) clusters or EAs, ii) segments within sampled clusters (if applicable), iii) households within sampled segments (or within sampled clusters, if segmentation is not applicable), and iv) “take-all” of individuals within sampled households. This chapter discusses the sampling frames at each stage of sampling, and their critical importance in the implementation of PBS surveys. A sampling frame is the essential backbone of all survey implementation. It comprises complete lists of the units (i.e., clusters/EAs, segments, households, or individuals) from which a representative sample can randomly be drawn at each stage of the survey. Without such frames, it is impossible to undertake a representative survey.

A high-quality survey frame should be comprehensive, complete, and up to date. “Comprehensiveness” refers to the **type** of information that is included on the frame (i.e., clusters/EAs, segments, households, or individuals), while “completeness” refers to the **extent** to which information on **all** relevant units is reflected in the frame. With regard to frames being up to date, it is important that Feed the Future teams keep track of the geographic composition of the Feed the Future ZOI or FFP DFSA implementation areas, and any related changes that occur over time. Information on the geographic composition is used to update the first stage sample frame of clusters/EAs provided by the national statistics office (NSO) and to make adjustments as needed at the time of the next survey to reflect any changes in geographic coverage of the Feed the Future ZOI or FFP DFSA implementation area.

It is also important that once sampling frames for EAs are finalized, copies be retained and stored to allow for retrospective and prospective comparisons of the list of EAs at each survey occasion. The three qualities defining high-quality frames (i.e., comprehensive, complete, and up-to-date) will be elaborated on later in the chapter.

3.1 First Stage Sampling Frame of Clusters/Enumeration Areas

In a typical PBS, a first stage “cluster” refers to a geographic area for which random selection occurs at the first stage of sampling. It is recommended that PBSs use EAs (rather than communities) defined by the national census whenever possible at the first stage of sampling, because: i) the population and/or household counts for each EA that are needed for sampling at that stage are readily available from the census; ii) census EAs are usually of roughly equal size, which helps with lister/enumerator workload distributions; and iii) the use of EAs helps minimize the need for segmentation, which more often occurs when communities of large or uneven size are used.³⁷

In general, for the **first stage EA frame** to be considered “complete,” it should consist of an exclusive (i.e., with no duplicates) and exhaustive (i.e., with none missing) set of all EAs in the Feed the Future ZOI or FFP DFSA implementation area. For the EA frame to be considered “comprehensive,” it should include, at a minimum, the following information:

- A unique ID number for the EA

³⁷ Note that segmentation implies an additional stage of sampling. Segmentation is described in more detail in Chapter 7.

- The name of the EA (if one exists)
- The location of the EA (e.g., census geographic code or global positioning system [GPS] coordinates)
- Information on all appropriate higher-level geographic areas (e.g., provinces or districts) in which the EA is contained
- The number of households in the EA (obtained via census files)

Information on the first stage sampling frame of EAs is usually obtained from the NSO or the national census office. Alternatively, it can be obtained from a prior survey vehicle that used the required frame in the recent past. It is important to use a source of information that is as recent as possible (e.g., the most recent census or a prior survey of recent vintage) to ensure that the set of EAs is not out of date, given that the definitions of EAs can change and EAs can be reshaped over time by national authorities. The use of up-to-date sources is also important to ensure that the information on each EA (e.g., a count of the number of households in the EA) is as accurate as possible.

Note that a distinction should be made between Feed the Future ZOI and FFP DFSA implementation area sampling frame coverage, because historically PBSs undertaken in each have used somewhat different definitions of their respective coverage areas. For Feed the Future ZOIs, the first stage sampling frame typically includes all EAs within the entire ZOI, which may span a larger geography than the combined area covered by all the projects in the given country. On the other hand, for FFP DFSA implementation areas, the first stage sampling frame is typically limited to include EAs within the DFSA implementation areas only, and so has a much more restricted definition.³⁸

When conducting a comparative analytical PBS (where the main aim is testing change between two time points), there are special issues to consider regarding the first stage sampling frame. The issues relate to how to reconcile the fact that the relevant geography in the Feed the Future ZOI (e.g., districts) or FFP DFSA implementation area (e.g., communities) may be different at the beginning of a project when the baseline is conducted compared to when a subsequent survey is conducted. This means the associated EAs that form the first stage sampling frame may also be different at the time points. There are two scenarios of how this can happen in the Feed the Future context.

- **Scenario A:** This scenario is relevant in the FFP DFSA context only. For this scenario, a FFP IP plans to swap in and/or swap out communities over the life of the DFSA and these swaps **are known** at the time of the design of the baseline (e.g., the project adopts a phased approach to programming where the DFSA commences implementation in a few communities, phases in additional communities during the first few years of the DFSA, and tapers off implementation in

³⁸ In the FFP context, one of the difficulties in using EAs is that FFP IPs typically define their implementation areas in relation to communities, and the correspondence between EAs and the communities in which IPs work is not always straightforward. In such cases, a “cross-walk” must be built between EAs and DFSA implementation communities. If the DFSA implementation communities don’t follow EA boundaries, then the recommendation is to develop a cross-walk that defines the first stage sampling frame in such a way that only those EAs that overlap geographically by 50% or more with the DFSA implementation communities are included in the frame.

communities near the end of the DFSA).³⁹ Because the coverage area for a PBS is typically limited to include the set of EAs that correspond to the FFP DFSA implementation communities only, this issue potentially adds some complexity to PBS implementation, since the coverage area may be different at the beginning of the DFSA (when the baseline PBS is conducted) than it is at the end of the DFSA (when the end-line PBS is conducted). Feed the Future recommends that the first stage sampling frame contain all the EAs to be covered over the life of the project, even if some of the included EAs have no project activities taking place at some period during the life of the project. The rationale for this recommendation is that projects are ultimately responsible for achieving results for the entire project area as defined over the life of the project.

- **Scenario B:** This scenario is relevant in both Feed the Future ZOI and FFP DFSA contexts alike. In this scenario, the Feed the Future team or the FFP IP needs to add or remove districts or communities between the baseline PBS and the end-line PBS, and these changes **are not known** at the time of the design of the baseline. This can happen, for instance, if the Feed the Future ZOI or FFP DFSA implementation area undergoes civil strife and some of the areas targeted for implementation are dropped after the baseline PBS is conducted. In this case, if programming is not implemented in the conflict-affected areas, the end-line PBS should include only EAs relating to the reduced geography that excludes the conflict-affected areas. The analytical comparison to the baseline PBS should include only that part of the frame from the baseline PBS that corresponds to the set of EAs used in the sample frame of the end-line PBS. Since some of the EAs from the baseline PBS will not be used in the comparison at end-line, there will be a shortfall of sample size from the first survey, similar to the situation described in Section 2.2.7. Once again, formulas (6) and (7) can be used to adjust the end-line PBS sample size to compensate. Finally, although data from some of the EAs from the baseline PBS will not be used in the analytical comparison, no weighting adjustment of any kind needs to be made to compensate for the fact that they are not being used; they are simply dropped from the analysis. This is because at the time of the end-line PBS, these EAs are considered “out of scope.” The EAs that are kept for the purposes of analysis are the true “survey domains” of interest and these already have predetermined probabilities of selection associated with them.⁴⁰

3.2 Second Stage Sampling Frame of Segments

Occasionally, there is rapid change (growth or reduction) in the size of some of the EAs’ populations between the time of the last census and the PBS, as ascertained by the listing operation (described in detail in Chapter 6). In general, if a sampled EA has grown too large by the time of field operations,⁴¹ field teams will typically divide the EA into “segments” and subsample one of the segments, a process called “segmentation.” Segmentation entails an additional stage of sampling. (The process is described in more detail in Chapter 7.) EAs are divided into segments on the basis of the number of households in the EA, and segmentation is implemented by grouping all households within the EA into different

³⁹ In the case of Feed the Future ZOIs, non-FFP IPs may also choose to swap in and/or swap out communities within the ZOI. Regardless, the ZOI definition remains static over time because the ZOI is typically larger than the combined set of project implementation communities, and PBSs are conducted at the ZOI level rather than at the project implementation level.

⁴⁰ One can think of this situation as similar to that of producing “survey domain” estimates for disaggregates by sex (male versus female). When a disaggregated estimate for “females” is produced, for example, the sampling weights are not adjusted to account for the fact that the data for “males” is not used in the construction of the disaggregated estimate.

⁴¹ See Chapter 7 for more information on segmentation and a discussion of what constitutes “too large” in this context.

segments. Therefore, the sampling frame used at the second stage of sampling of segments is the same as the sampling frame used at the third stage of sampling of households, since the household listing exercise within each sampled EA serves as a frame for both. (See Section 3.3 for more details on the third stage sampling frame.) Since segmentation occurs only in those few EAs where the EA has grown too large, at the third stage of sampling, either households are sampled within segments (when segmentation occurs) or households are sampled within EAs (when segmentation does not occur).

3.3 Third Stage Sampling Frame of Households

In a typical Feed the Future PBS, the third stage sampling frame consists of a complete and comprehensive list of all households within the selected EAs (or within segments sampled at the second stage, if segmentation is required). If a sufficiently recent listing of households (from within the last few months) does not exist prior to the main survey fieldwork, Feed the Future requires that one be generated through a listing field operation. This listing should be conducted within 8–10 weeks of the main data collection.⁴²

Note that there is a distinction between listing dwelling units (physical structures) and listing households⁴³ within sampled EAs. For Feed the Future PBSs, the sampling unit is the household; hence, households rather than dwelling units, should be listed. Although most conventional surveys list dwelling units within sampled EAs,⁴⁴ listing households rather than dwelling units has the advantage that it eliminates from the main survey fieldwork the additional step of sampling households within dwelling units. Additionally, the adjustment given by formula (4) in Section 2.2.4 does not adjust for the number of dwelling units to sample, but rather the number of households to sample, so listing households is in conformity with the overall sample size computation. To ensure that proper and comprehensive listing takes place, contact must be made with all households (assuming there is more than one) within each dwelling unit encountered, to obtain basic information on the residents within.⁴⁵ This information on residents is needed to determine if the dwelling unit comprises one or more households. If the dwelling unit comprises more than one household, each household is listed separately during the listing exercise.

⁴² For more information on the listing process, see: “Feed the Future Household Listing Manual.” 2018. Available at: <https://agrilinks.org/post/feed-future-zoi-survey-methods>.

⁴³ A “dwelling unit” is a room or a group of rooms normally intended as a residence for one or more households. There are various possible definitions of a “household,” but this guide uses the following Feed the Future definition: A household consists of all people, including adults and children, who live together (i.e., sleep) under the same roof, share cooking or housekeeping arrangements, and recognize the same lead male or female decision makers in the household. Household members can include servants, lodgers and agricultural laborers, and other non-family members, as well as family members. In some cases, one may find a group of people living together in the same dwelling unit, but each person has separate eating arrangements; they should be counted as separate one-person households.

⁴⁴ This is because listing dwelling units is a relatively inexpensive exercise that requires only a simple walk through a sampled EA to identify dwelling structures from the exterior, whereas listing households necessitates the additional step of ascertaining the identification of household(s) within each dwelling structure. Most conventional surveys in developed countries deem the additional cost of listing households prohibitive, but in developing countries, field operations are often less costly, and so the listing of households rather than dwellings is considered a viable option.

⁴⁵ Basic information on the composition of the household only is collected at the listing stage only. A more complete gathering of household information is obtained at the time of interviewing sampled households, through the completion of a “household roster.”

For the **third stage household frame** to be considered “complete,” it should consist of an exclusive (i.e., with no duplicates) and exhaustive (i.e., with none missing) listing of all households within the set of sampled EAs or segments. For the household frame to be considered “comprehensive,” it should include, at a minimum, the following information for each household:

- A unique household identification number
- A unique dwelling identification number (corresponding to the dwelling containing the household)
- The household location (e.g., address, relative location or GPS coordinates, if available)
- The community name to which the household belongs
- The location of the community in which the household is located (e.g., census geographic code or GPS coordinates, if available)
- The EA number corresponding to the EA in which the household is contained
- All appropriate higher geographic levels (e.g., province or district) in which the household is contained
- The complete name of a responsible male or female adult⁴⁶ who can provide information on the composition of the household

3.4 Fourth Stage Sampling Frame of Individuals

In a Feed the Future PBS, the fourth stage sampling frame consists of all individuals categorized as household members within sampled households from the third stage. This frame is established through a household roster, which is a listing of all household members along with associated information; it is generated by interviewers during the main data collection, after sampled households have been located and contact has been established with a responsible adult within these households. In general, all household members are included on the roster, where “household members” are defined as “adults or children who live together and eat from the ‘same pot.’ ”⁴⁷ To construct the household roster, basic information on each household member is obtained from a responsible adult in the household. After the household roster has been established, individuals who are eligible to respond to the various questionnaire modules corresponding to the different indicators can be interviewed.

⁴⁶ The case where the household is headed by a child and there is no adult residing within is addressed in: “Feed the Future Household Listing Manual.” 2018. Available at: <https://agrilinks.org/post/feed-future-zoi-survey-methods>.

⁴⁷ For the purposes of a Feed the Future PBS, “household members” are defined more completely as “usual residents of a household who have spent the night in the household within the past 6 months. A household consists of one or more persons (adults or children) who live in the same dwelling unit. They can be related or unrelated (including family members, but also including servants, lodgers, agricultural laborers, friends, or other non-family members), but they should acknowledge the same person or persons as lead decision makers for the household, share the same housekeeping and cooking arrangements, and share the same contiguous roof. In some cases, one may find a group of people living together in the same dwelling unit, but each person has separate eating arrangements; they should be counted as separate one-person households. Note that dwellings intentionally designed to shelter unrelated groups of people, such as army camps, school dormitories, refugee camps, or prisons are considered households.” See: Feed the Future. 2018. *Zone of Influence Survey Country Report Template*. Washington, DC: USAID. Available at: <https://agrilinks.org/post/feed-future-zoi-survey-methods>.

The basic information to be collected on each member of the selected household might include:

- The name of household member
- The relationship of the household member to the responsible adult
- The age of the household member
- The sex of the household member
- Any information needed to ascertain if the individual is eligible for the various questionnaire modules
- The education level of the household member (suggested auxiliary information to be used in analysis)
- The marital status of the household member (suggested auxiliary information to be used in analysis)
- Ideally, the cellular telephone number (if feasible) of at least one household member (to help locate and contact one member of the household, preferably the responsible adult)⁴⁸

⁴⁸ Such contact information may be needed in the case where all required interviews within a sampled household cannot be completed during one visit to the household. In the event that multiple visits are required, it may be necessary to ascertain best times to visit for follow-up interviews or to determine if a previously absent member of the household with whom an interview was desired has returned and is available for interviewing.

4. Stratification and Allocation of the Sample

4.1 Stratification

“Stratification” is the process by which a target population is divided into subgroups (called strata) that have similar characteristics. For instance, in a Feed the Future ZOI PBS, teams may wish to stratify their ZOI by a geographic area, such as province, district, or region, and by urban and rural areas within them. Strata are often geographic in nature, but they don’t have to be.

The principal objective of stratification is to reduce the SEs of survey estimates. SEs are a measure of variability of survey estimates and offer a sense of how “precise” the estimates are; the smaller the SE, the more precise the estimate. Because typically part of the overall sample is allocated to each stratum and samples are selected independently within each stratum, the SEs of estimates across strata depend on the variability existing within the strata but not between the strata. This has the effect of reducing the overall variability across the sample.⁴⁹

It can also be useful to stratify the sample if separate estimates are desired at the stratum level. In the case of Feed the Future PBSs, estimates are not required at the stratum level, but rather are required only at the overall ZOI level (for Feed the Future non-FFP) or at the DFSA level (for FFP). To obtain stratum-level estimates with the same precision as at the ZOI level, the overall sample size would need to be replicated for each stratum, which could be very costly.

Another noteworthy distinction to make is that strata are not the same as survey domains. A survey domain is a population subgroup for which separate survey estimates are desired, but which were not planned for in advance (by allocating sufficient sample size to each stratum prior to fieldwork). An example of survey domains is any of the required Feed the Future PBS indicator disaggregates in Tables 1 and 2 (such as male versus female).

It is important to understand that stratification should **not** be considered a stage of sampling because strata are purposively determined, whereas the units at each stage of sampling are randomly selected. Therefore, it is not necessary to build sampling weights to reflect stratification, although sampling weights are typically computed separately within each stratum. For Feed the Future PBSs, it is suggested that survey implementers use stratification whenever possible (to increase precision of estimates), employing whichever geographic levels within the country of interest make the most sense.

4.2 Allocation of the Sample to Strata

As mentioned earlier, once the total sample size, $n_{final-overall}$, has been determined, the overall sample should be appropriately allocated to (i.e., divided among) the different strata. Different allocation schemes are available, depending on the situation. Three allocation schemes are described below, followed by a description of a few Feed the Future-specific scenarios where the various allocation schemes might be appropriate.

⁴⁹ Another benefit of stratification is that it conveniently allows for a flexible sample design that can be different for each stratum.

4.2.1 Proportional Allocation

For a multipurpose survey where there are many indicators of interest that span a variety of sampling groups and where the interest lies in producing estimates at the overall Feed the Future ZOI or FFP DFSA level only, proportional allocation is best. It is best in the sense that it produces “optimal” estimates with the lowest possible variance or highest possible precision for a fixed sample size. Proportional allocation allocates sample to each stratum proportional to the stratum size, using a size measure, such as number of households in each stratum. The formula for allocating the sample size $n_{final-overall}$ to the various strata using proportional allocation is given by:

$$n_{final-overall}_h(\text{proportional}) = n_{final-overall} * \frac{N_h}{\sum N_h}$$

where:

$n_{final-overall}_h(\text{proportional})$ is the portion of the sample size (from $n_{final-overall}$) to be allocated to stratum h

N_h is the number of households in stratum h

$\sum N_h$ is the number of households across all strata in the ZOI or FFP DFSA implementation area

4.2.2 Equal Allocation

If optimal estimates (i.e., estimates with the lowest possible variance/highest possible precision) are required at the stratum level, rather than at the overall level, and sufficient sample size is available for allocation to each stratum to ensure the precision, then the best allocation scheme to use is equal allocation.⁵⁰ The formula for allocating the sample size $n_{final-overall}$ to the various strata using equal allocation is given by:

$$n_{final-overall}_h(\text{equal}) = \frac{n_{final-overall}}{H}$$

where H is the overall number of strata into which the population is stratified.

4.2.3 Power Allocation

If optimal estimates are required at both the overall level and the stratum level, then neither of the above two allocation schemes is best—assuming a fixed sample size at the overall level. Proportional allocation is optimal for estimates at the overall level, but will not result in estimates that are optimal at the stratum level. On the other hand, equal allocation is optimal for estimates at the stratum level, but will not result in estimates that are optimal at overall level. What this means is that, for a fixed sample size at the overall level, estimates at the overall level under a proportional allocation scheme are more precise than estimates at the overall level under an equal allocation scheme. Similarly, for a fixed sample size at the overall level, estimates at the stratum level under an equal allocation scheme are more precise than estimates at the stratum level under a proportional allocation scheme.

⁵⁰ Under equal allocation, estimates at the overall level will still have high precision, but they will not be optimal. This means that estimates at the overall level would have higher precision under proportional allocation than they would under equal allocation, for a fixed sample size.

A power allocation is a compromise allocation scheme when optimality is desired at both the overall level **and** the stratum level. The formula for allocating the sample size $n_{final-overall}$ to the various strata using power allocation is given by:

$$n_{final-overall}_h(power) = n_{final-overall} * \frac{N_h^\alpha}{\sum N_h^\alpha}$$

where α is a fraction such that $0 \leq \alpha \leq 1$.⁵¹

A power allocation is an allocation proportional to the size measure N_h raised to the power α . At the extremes, using $\alpha = 1$ results in proportional allocation while using $\alpha = 0$ results in equal allocation. A power value between 0 and 1 provides a compromise allocation between proportional and equal allocation. Feed the Future recommends using the power of $\alpha = 0.5$ when using power allocation, because this is a value that is often used in practice

4.3 Examples of Stratification and Allocation

There are a number of situations in the Feed the Future context that warrant the use of stratification and the application of allocation schemes. A few examples are given below.

Example 1: For Feed the Future ZOI PBSs, the main interest is in producing optimal (highest possible precision) indicator estimates at the overall ZOI level. However, geographically diverse areas, such as districts or urban/rural zones within the ZOI, are often designated as strata. This is because it is of interest to ensure representation of each geographic area in the sample to capture the variation across the geographic areas. This is achieved through stratification and allocation of some of the overall sample to each stratum—although the interest is not necessarily in producing estimates at the stratum level. In this case, because producing high-precision estimates at the overall ZOI level is the chief aim, the use of proportional allocation to the strata is preferred.

Table 8. Example of Various Allocation Schemes for a Sample Size of $n_{final-overall} = 2,760$ Households

Strata (Districts)	Number of Households in Strata (N_h)	Proportional Allocation of Sample	Equal Allocation of Sample	Power Allocation of Sample ($\alpha = 0.5$)
Ciril	2,289	333	460	396
Greyling	2,241	327	460	391
Morthand	4,804	701	460	573
Rhun	3,769	550	460	508
Buckland	3,012	439	460	454
Udun	2,808	410	460	438
Total	18,923	2,760	2,760	2,760

⁵¹ For more information on power allocation, see: Bankier, M.D. 1988. "Power Allocations: Determining Sample Sizes for Subnational Areas." *The American Statistician*. 42:3.

Consider the example in Table 8 that illustrates a proportional allocation scheme, using a fictitious ZOI that consists of the six districts that are considered strata and an overall sample size of $n_{final-overall} = 2,760$ households to be allocated to the six strata. The example is continued from Section 2.2.6 where the “Prevalence of Stunted Children under Five” indicator drives the overall sample size for the survey. Here, the size measure that is used is “the number of households” in each stratum, the values for which are available from the most recent census. To see how the computation is done, consider the first stratum entry (“Ciril”) in the column for proportional allocation. This is computed as:

$$n_{final-overall,h}(proportional) = n_{final-overall} * \frac{N_h}{\sum N_h} = 2,760 * \frac{2,289}{18,923} = 333$$

This means that because Ciril contains 12% of the population, it will be assigned 12% of the sample under proportional allocation.

Example 2 (Special Case for FFP DFSAs only): FFP is typically interested in producing estimates for each separate DFSA within the overall FFP implementation area. In this case, a sample size calculation is undertaken to support high-precision estimates at the level of each DFSA. The sample size is identical (i.e., the input parameters are the same) for all DFSAs, which means that the sample size for the overall DFSA implementation area is equal to the sample size for each DFSA multiplied by the number of DFSAs. This is equivalent to using equal allocation, where each DFSA implementation area is considered a stratum. However, in this example, a bottom-up approach to sample size allocation is used instead of a top-down approach. That is to say, the sample size is determined at the stratum level—for one stratum—and then the same sample size is replicated for all strata. Although there is also an interest in estimates at the overall FFP implementation area, equal allocation is used here, rather than power allocation, which might at first appear to be more appropriate. The reason is that a bottom-up approach will provide a sample size at the overall combined FFP implementation area that is large enough so that estimates at that level are not likely to suffer any substantial issues with diminished precision.

Table 8 also provides an example to illustrate an equal allocation scheme, but using a top-down approach, that is to say, where the overall sample size is determined first and then each stratum is allocated an equal portion of the overall sample size. A top-down approach is more typically used when resource limitations dictate that the overall sample size be fixed in advance of any allocation at lower levels, but the primary interest is producing precise estimates for each stratum. This is the more typical scenario that is used for Feed the Future PBSs (as compared to PBSs undertaken in support of FFP DFSAs).

An example of power allocation is also provided in Table 8. Notice that, compared to proportional allocation, power allocation narrows the range of the sample sizes across the various strata. That is to say, the smallest sample size across all strata under the proportional allocation is smaller than the smallest sample size under the power allocation. Similarly, the largest sample size across all strata under the proportional allocation is larger than the largest sample size under the power allocation.

4.4 A Special Application of Allocation: Joint Baseline and End-Line PBS

Now consider a special application of allocation related to the case described in Scenario 2 in Box 1 of Section 2.1. In that scenario, there is a need for two surveys—an end-line PBS for the original ZOI and a baseline PBS for the new ZOI—and there is a desire to conduct both PBSs using the same survey vehicle. This is akin to one survey consisting of two “sub-surveys.” Recall that three sets of strata are used (i.e., dropped districts, common districts, and new districts). Assuming that the main interest is in producing high-precision estimates at the overall ZOI level, an overall sample size can be computed for the baseline PBS (sub-survey 1), and proportional allocation can be used to allocate the overall sample to the sets of strata consisting of common districts and new districts. Similarly, an overall sample size can be computed for the end-line PBS (sub-survey 2), and proportional allocation can be used to allocate the overall sample to the sets of strata consisting of common districts and dropped districts. Since both sub-surveys will be administered together, one issue is that the sample size allocated to the strata of common districts may differ for the two sub-surveys. For instance, suppose that sub-survey 1 has the larger allocation to the set of strata consisting of common districts in comparison to that for sub-survey 2. In this case, it is recommended first to draw a sample within the set of common districts using the larger of the two allocated sample sizes to be used for the purposes of sub-survey 1. Then, a second phase of sampling can be introduced where a subsample corresponding to the smaller of the two allocated sample sizes can be drawn from the larger sample for the purposes of sub-survey 2.⁵² Alternatively, if sub-survey 2 has the larger allocation to the set of strata consisting of common districts (in comparison to the allocation for sub-survey 1), then the roles of the two surveys are reversed.

To illustrate the concept, see the example given in Table 9 where Argonath and Edoras constitute the “dropped” set of strata from the original ZOI, Hornburg and Westfold constitute the “common” set of strata between the original ZOI and the new ZOI, and Bree and Eryn constitute the “new” set of strata from the new ZOI. A baseline PBS must be conducted in support of the new ZOI and an end-line PBS in support of the original ZOI. Suppose a sample size calculation in support of the end-line PBS gives $n_{final-overall} = 2,760$ households (which has been computed at the time of the baseline PBS for the original ZOI several years earlier), and a proportional allocation results in 1,806 households to the set of strata with common districts and 954 households to the set of strata with dropped districts. Suppose further that a sample size calculation in support of the baseline PBS also gives $n_{final-overall} = 2,760$ households,⁵³ and a proportional allocation results in 1,644 households to the set of strata with common districts and 1,116 households to the set of strata with new districts. In this case, the two separate proportional allocations require 1,806 and 1,644 households, respectively, allocated to the set

⁵² The issue of the timing of the draw for the second phase sample depends on which of two scenarios is in play. In the first scenario, if the questionnaire used for each of the two surveys is identical, then the second phase sample can be drawn **after** data collection and just prior to data analysis, to facilitate the production of estimates for the survey with the smaller allocation of sample to the strata with the common districts. However, in the second scenario, if additional questionnaire modules are required for the second phase sample, then the second phase sample must be drawn **before** fieldwork, to facilitate the administration of different questionnaire modules to the first phase sample units versus the second phase sample units within the common districts. Generally, this occurs in the context where there is a common core questionnaire applied to the entire first phase sample, and additional questionnaire modules (above and beyond the common core questionnaire) applied to the second phase sample. However, for ease of training and for assurance of quality control, it is usually preferable to develop a single, comprehensive questionnaire and administer it at all phases—and revert back to the first scenario.

⁵³ In this example, the sample sizes for the baseline and end-line PBSs are the same. But in practice, they usually will not be because different input parameters will be used in the sample size formula for the two surveys.

of strata with common districts. The recommendation is to randomly select 1,806 households (the larger number) through a “first phase”⁵⁴ sample for the purposes of administering the end-line PBS. Then, in a “second phase” sample, 1,644 households are subsampled from the 1,806 households sampled at the first phase, for the purposes of administering the baseline PBS.⁵⁵ More detail on the various phases of sampling, and how to conduct the subsampling and weighting for this particular scenario is discussed in Section 5.1 (Example 2), Section 8.3, and Section 10.1.2.⁵⁶

The last column of Table 9 provides the required sample size for the jointly administered baseline/end-line PBS. Here, the set of strata with dropped districts require 954 households as per the end-line sample size requirement, while the set of strata with new districts require 1,116 households as per the baseline sample size requirement. For the set of strata with the common districts, the maximum sample size between the baseline and end-line sample size requirements (1,806 households) is used. Therefore, the total sample size for the jointly administered baseline/end-line PBS is 3,876. If the baseline PBS and the end-line PBS were administered separately, the overall sample size requirement would have been $2,760 + 2,760 = 5,520$ households. Since the joint baseline/end-line requires only 3,876 households, this means that there is an overall sample size savings of $5,520 - 3,876 = 1,644$ households through the joint administration of the two surveys.

⁵⁴ There is a need to distinguish “phases” of sampling from “stages” of sampling. In a survey with multiple stages of sampling, different entities (e.g., clusters, households, individuals) are sampled at the different stages. For instance, sampled individuals are selected within sampled households, which in turn have been selected within sampled clusters. In a sample with multiple phases, the same entities (e.g., households) are sampled at the different phases. For instance, a set of households are subsampled at a second phase from an original set of households sampled at the first phase.

⁵⁵ Note that in Table 8 there are really two levels of stratification: i) the first level that designates the type of strata as “dropped,” “common,” and “new,” and ii) the second level that delineates the districts within the first level of stratification. For instance, the first-level “common” strata consist of the districts of Hornburg and Westfold. Therefore, although we select 1,806 and 1,644 households, respectively, from the set of “common” strata, in reality, there should be an additional step whereby a proportional allocation of the (1,806 and 1,644) households to the two districts of Hornburg and Westfold is undertaken, and then random selection within each of the districts is performed separately. In this case, the 1,806 households would be allocated as Hornburg (1,012) and Westfold (794) for the end-line, whereas the 1,644 households would be allocated as Hornburg (921) and Westfold (723) for the baseline.

⁵⁶ An alternative strategy for dealing with this situation would be to use the full sample of 1,806 households for both the baseline and end-line PBSs, instead of subsampling 1,644 households from the 1,806 households at a second phase of sampling for the baseline. This means that there would be an excess of $1,806 - 1,644 = 162$ households above and beyond what is required for the baseline PBS sample size in the common strata. While an excess of sample size generally increases the overall precision of estimates, it could also have the opposite effect of reducing the overall precision of estimates by destabilizing the optimal proportionality established between the common and new strata (where proportional allocation was used) for the baseline. The extent to which the increase in precision is offset by the decrease in precision is difficult to ascertain in advance. Therefore, the strategy in the main text is the one that is recommended to Feed the Future survey implementers.

Table 9. Example of Allocation for Joint Baseline/End-Line PBS with a Baseline Sample Size of $n_{final-overall} = 2,760$ Households and an End-Line Sample Size of $n_{final-overall} = 2,760$ Households

First-Level Strata	Second-Level Strata (Districts)	Number of Households in Strata (N_i)	Number of Households in Baseline PBS Strata	Proportional Allocation for Baseline PBS Sample	Number of Households in End-Line PBS Strata	Proportional Allocation for End-Line PBS Sample	Sample Size for Joint Baseline/ End-Line PBS (Households)
Dropped (Original ZOI)	Argonath	2,289			4,530	954	954
	Edoras	2,241					
Common (Original & New ZOI)	Hornburg	4,804	8,573	1,644	8,573	1,806	1,806
	Westfold	3,769					
New (New ZOI)	Bree	3,012	5,820	1,116			1,116
	Emyn	2,808					
Total		18,923	14,393	2,760	13,103	2,760	3,876

5. First Stage Sampling of EAs

5.1 Deciding on the Number of EAs to Sample

After the overall sample size, $n_{final-overall}$ (with respect to number of households to visit) is determined, and after stratification and allocation of sample to the strata has taken place, each stratum, denoted by h , has its own sample size, $n_{final-overall_h}$, according to the allocation scheme used. At that point, sampling takes place within each stratum separately and independently. Recall that there are at least two stages of sampling to arrive at that point: the sampling of EAs followed by the sampling of households within selected EAs.⁵⁷ That means that a decision needs to be made about how to convert $n_{final-overall_h}$ into the number of EAs to sample in stratum h (denoted by m_h). From there, one can compute the number of households to sample in stratum h across all EAs (denoted by n_h) by dividing $n_{final-overall_h}$ by m_h .

In a multi-stage sampling design with a given sample size, there is no prescriptive formula for determining the combination of the number of EAs and the number of households within each EA to sample. There are competing interests in terms of what is most operationally expedient versus what is most statistically efficient.

On one hand, for a fixed sample size, for operational expediency, it is clearly optimal to select the smallest number of EAs possible, with a greater number of households per EAs. When a smaller number of EAs is selected, the time and cost of transportation to, from, and between the EAs are decreased—and potentially the number of interviewers can also be decreased. However, the statistical efficiency decreases, as measured by an increase in the DEFF.

On the other hand, for statistical efficiency, it is recommended to select the smallest number of households possible from each EA, and therefore to select the largest number of EAs. This is because each additional household within the same EA adds a decreasing amount of new information, assuming that EAs tend to be homogeneous in terms of the characteristics of their households. In this case, the operational efficiency is diminished because travel time to, from, and between the EAs is increased given the larger number of EAs. However, a greater number of EAs with fewer households reduces the DEFF, which in turn reduces the sample size required and potentially lowers the cost of the survey. Therefore, it is not always clear which of the two scenarios will play a greater role in reducing overall costs of the survey.

Given these opposing considerations, a compromise must be struck. To make an appropriate decision, available budget and resources must be assessed, including the number of available interviewers, the ease of access to and between sampled EAs, and time constraints, among other considerations. However, each survey will potentially face a different set of constraints, and it is not possible to provide a definitive recommended number of EAs to select in each instance.

⁵⁷ For the time being, there is an assumption of no segmentation, and therefore, the stage of sampling related to the sampling segments within selected EAs is dropped.

It is possible, however, to provide a rule of thumb concerning the number of sampled households to allocate to each sampled EA: For Feed the Future ZOI PBSs, a range of 20–30 households for each selected EA is recommended because, in most cases, this represents a logistically feasible number of households per EA in which to conduct interviews without compromising statistical efficiency by inducing a DEFF that is too large.

Note that it is always best to sample the identical number of households in each sampled EA, rather than varying the number of households sampled in different EAs (e.g., sampling a proportional number of households relative to the EA size). There are two reasons for doing so. First, interviewing a constant number of households in each sampled EA leads to roughly equal interviewer workload assignments of households (assuming that the number of eligible individuals within households is roughly constant). Second, a carefully constructed design that couples a constant “sample take” of households in each EA coupled with the use of probability proportional to size (PPS) sampling of EAs at the first stage of sampling leads to what is called a “self-weighting” design (which is discussed in greater detail in Section 10.1.5). Self-weighting has the main advantage of increasing precision by diminishing the overall DEFF of the survey (which is a component of the SE) associated with the estimates of indicators—in particular, the contribution to the DEFF related to unequal weighting.

Example 1: The following example illustrates the application of this rule of thumb of using 20–30 households for each selected EA. Suppose that a PBS is being designed where there is a need to sample $n_{final-overall} = 2,760$ households, and stratification has been undertaken within the ZOI using the districts in Table 8 in Section 4.3. Suppose further that optimal estimates are required at the overall ZOI level, and therefore a proportional allocation scheme corresponding to the third column of Table 8 is used. With the given resources, a decision has been made to select 25 households in each sampled EA. Sampling is undertaken independently in each stratum. In the first stratum, Ciril, 333 households are allocated. To determine the number of EAs to sample in Ciril, 333 is divided by 25 to obtain 13.3. Since it is not possible to sample a fractional number of EAs, this number is rounded up to 14. Therefore, in Ciril (stratum h), $m_h = 14$ EAs are randomly selected, and then 25 households are randomly selected within each of the 14 EAs. Rounding the number of clusters from 13.3 to 14 means that in effect $n_h = 14 * 25 = 350$ households will be interviewed rather than 333 households in Ciril—an excess of 17 households. The alternative would be to round 13.3 down to 13, and in doing so to interview $13 * 25 = 325$ households. In this case, there would be a shortfall of 8 households from the required 333. It is always preferable to include a little excess sample over what is required than it is to be short of sample. However, it is acceptable to offset any excess by making small reductions in other strata. For instance, in Morthand, 701 households have been allocated. It is acceptable to round this down to 700 so that it will divide evenly by 25. The computation to determine the number households in each EA should be repeated for all six strata.

Example 2: Continuing with the example described in Section 2.1 (Box 1) and Section 4.4, where both a baseline PBS and an end-line PBS are to be conducted using one survey vehicle and each sub-survey requires $n_{final-overall} = 2,760$ households, the original ZOI and the new ZOI are stratified using the districts in Table 9 in Section 4.4. In this case, two separate proportional allocations of 1,644 and 1,806 households are allocated to the set of strata with common districts for the baseline PBS and end-line PBS, respectively. Again, it has been decided that, with the given resources, 25 households will be

selected in each sampled EA. For the baseline PBS, 1,644 / 25 or approximately 66 EAs are to be sampled in the set of strata with common districts. For the end-line PBS, 1,806 / 25 or approximately 72 EAs are to be sampled in the set of strata with common districts. As discussed before, it is appropriate to adopt a “two-phase” approach at the first stage of sampling, whereby 72 EAs are sampled at the first phase of the first stage of sampling and then 66 of the 72 EAs are subsampled at the second phase of the first stage of sampling, in the set of strata with common districts. The 72 sampled EAs are used in determining results for the end-line PBS while the 66 subsampled EAs are used in determining results for the baseline PBS. The method for randomly selecting 66 of the 72 EAs is described in Section 8.3. At the next stage of sampling, 25 households are selected from each of the 72 sampled EAs. For the 66 subsampled EAs, the 25 sampled households within these EAs will be in common for the baseline and end-line PBSs. Deciding on the number of EAs to sample within the set of strata with dropped districts and new districts from Table 9 is done in a manner similar to Example 1, so is not elaborated upon here.

5.2 Randomly Selecting a Sample of EAs

After the number of EAs to be randomly selected has been determined, the next step in the survey design process is to randomly select the sample of EAs from the sampling frame, independently within each stratum. In most instances, the method used to randomly select a sample of EAs at the first stage of sampling is PPS sampling. In general, PPS sampling selects EAs according to a “size measure” that is related to the key indicator of interest. The “total number of households” in each EA is most frequently used as the size measure. Such size measures can be obtained from most recent national census files.

In general, using PPS sampling ensures that EAs with a greater number of households have a greater chance of being selected from the frame, while EAs with fewer households have a smaller chance of being selected from the frame. It is an efficient way of sampling if the number of households per EA varies greatly across all the EAs on the sampling frame. However, in most cases, national census offices attempt to create EAs in such a way that they are roughly of equal size, so that EAs have roughly the same number of households (although in some countries EA sizes can vary considerably⁵⁸). In countries where EA sizes are roughly the same, it may be tempting to use systematic sampling at the first stage of sampling instead. However, it is still useful to use PPS sampling at the first stage as a way of approximating an overall “self-weighting” design across the stages of sampling. This is because “self-weighting” designs provide sampling weights that are close to constant within strata; this approach protects against highly variable weights that can increase the variability of estimates and can negatively affect precision and power. Finally, in PPS sampling, it is best to use “the number of households in the EA” rather than “the number of individuals in the EA” as a size measure, as this will facilitate self-weighting designs. This is described in more detail in Section 10.1.5.

The type of PPS sampling that Feed the Future recommends for use at the first stage of sampling is called **systematic PPS sampling**. Systematic PPS sampling is simpler to implement than other types of PPS sampling and can be shown to lead to a self-weighting scheme, if combined with subsequent stages

⁵⁸ Even in countries where EA sizes vary considerably, it is typical that other geographic units, such as communities, will vary in size to an even greater extent. Therefore, it is always prudent to use EAs—rather than communities, say—as the sampling unit at the first stage of sampling.

of sampling where the same number of households per EA are sampled within each sampled EA and where all eligible individuals within a sampled household are interviewed.

5.3 Systematic PPS Sampling

The steps to select a sample of EAs separately within each stratum using systematic PPS sampling are given below. The steps can be carried out using any appropriate software. Using software has many advantages, including automating and documenting the selection, and automating the provision of first stage sampling weights. An example of a manual selection is provided below to illustrate the concept. The syntax provided is what would be used in Microsoft Excel.

STEP 1. Create a list of all EAs in a given stratum. This is essentially the first stage EA frame described in Section 3.1. It includes a count of the number of households in each EA.

STEP 2. Calculate a cumulative total number of households within the stratum. Create a new column on the first stage EA frame that contains a cumulative total number of households⁵⁹ per EA. This column of cumulative totals is used for selecting the sample of EAs. The first row of the cumulative total equals the number of households in the first EA on the list. The second row of the cumulative total equals the number of households in the second EA **plus** the number from the first row. This pattern of accumulation continues in the same way through to the end of the list. The following is an example.

List of all EAs in Ciril (stratum h)				
EA number	EA name	Number of households per EA		Cumulative total of households
1	Kvothe	53		53
2	Gumbo	60	+ 53 =	113
3	Pancho	48	+ 113 =	161
4	Glokta	42	+ 161 =	203
5	Rainbow's End	51	+ 203 =	254
6	Furculita	39	+ 254 =	293
7	Stanka	65	+ 293 =	358
8	Stormlight	52	+ 358 =	410
9	Deepness	55	+ 410 =	465
10	Black Dow	50	+ 465 =	515
...
41	Logan	54	+ 2028 =	2082
42	Tul Duru	61	+ 2082 =	2143
43	Bast	49	+ 2143 =	2192
44	Kaladin	47	+ 2192 =	2239
45	Arya	25	+ 2239 =	2264
46	Cashin	25	+ 2264 =	2289

⁵⁹ In the absence of information on the number of households per EA, one can also use the number of dwellings per EA instead, although this could compromise self-weighting.

Here, the example from Table 8 in Section 4.3 and Example I in Section 5.1 using the stratum of Ciril is being continued. An overall sample of $n_{final-overall} = 2,760$ households is required for the entire survey, but within the stratum of Ciril, a sample of 333 households is initially allocated. Within Ciril, there are $N_h = 2,289$ households and, at the first stage, $m_h = 14$ EAs are to be sampled. The decision is made to sample 25 households per EA in Ciril, for a total of $n_h = 14 * 25 = 350$ households to be sampled within Ciril.

STEP 3. Calculate a sampling interval for the stratum. The sampling interval (denoted by k_h) is calculated by dividing the total number of households in stratum h (denoted by N_h) by the number of EAs to select in stratum h (denoted by m_h). The formula is given by:

$$sampling\ interval = k_h = \frac{number\ of\ households\ in\ stratum\ h\ (N_h)}{number\ of\ EAs\ to\ select\ in\ stratum\ h\ (m_h)}$$

For instance, in Ciril, $N_h = 2,289$ households and $m_h = 14$ EAs, so the sampling interval (using unrounded value) is $k_h = 163.5$.

STEP 4. Calculate a random start for the stratum. The random start (denoted by RS) determines the first EA to select. It is calculated by generating a random number greater than or equal to 0 and less than the sampling interval (k). The Microsoft Excel function $rand()$ generates a random (fractional) number greater than or equal to 0 and less than 1. To compute the random start, this random number ($rand()$) is multiplied by the sampling interval (k_h). The following is the formula to use to calculate the random start using the Microsoft Excel function $rand()$:

$$random\ start = RS = rand() * k_h$$

For instance, if $rand() = 0.7146$ and the sampling interval within Ciril is $k_h = 163.5$ (from above), then $RS = 0.7146 * 163.5 = 116.84$.

STEP 5. Select the first EA within the stratum. The first EA to select within the stratum Ciril according to this scheme will be the one that corresponds to the value of the random start. To do this, identify the pair of consecutive EAs in the list for which the cumulative total corresponding to the first EA is less than the random start and for which the cumulative total corresponding to the second EA is greater than or equal to the random start. Choose the second EA in the pair. The following chart provides an example.

Total number of households across district of Ciril (stratum h)	N_h	2289		
Number of EAs to select in Ciril	m_h	14		
Random number	rand()	0.7146		
Sampling interval	$k_h = N_h/m_h$	163.50		
Random start	$RS = rand() * k_h$	116.84		

List of all EAs in stratum			
EA number	EA name	Number of households per EA	Cumulative total of households
1	Kvothe	53	53
2	Gumbo	60	113
3	Pancho	48	161

1ST EA TO SELECT	$a_1 = RS$	$= 116.84 =$	116.84	→	←	116.84 is greater than 113 but less than 161
------------------	------------	--------------	--------	---	---	--

In the example above, since the RS (116.84) is greater than 113 (corresponding to EA 2) and less than 161 (corresponding to EA 3), EA 3 (Pancho) is selected as the first EA in the sample.

Note that if, by chance, $rand()$ generates the number 0, then RS is also 0. In this case, simply choose the first EA on the list to be the first EA in the sample.

STEP 6. Select the second EA within the stratum. Determine the second EA to select for the sample within the stratum. Compute a number a_2 that corresponds to the number obtained by adding the sampling interval (k_h) to RS . Identify the pair of consecutive EAs in the list for which the cumulative total corresponding to the first EA is less than a_2 and for which the cumulative total corresponding to the second EA is greater than or equal to a_2 . Choose the second EA in the pair. The following chart provides an example.

Total number of households across district of Ciril (stratum h)	N_h	2289		
Number of EAs to select in Ciril	m_h	14		
Random number	rand()	0.7146		
Sampling interval	$k_h = N_h/m_h$	163.50		
Random start	$RS = rand() * k_h$	116.84		

List of all EAs in stratum			
EA number	EA name	Number of households per EA	Cumulative total of households
1	Kvothe	53	53
2	Gumbo	60	113
3	Pancho	48	161
4	Glokta	42	203
5	Rainbow's End	51	254
6	Furculita	39	293

1ST EA TO SELECT	$a_1 = RS$	$= 116.84 =$	116.84	→	←	116.84 is greater than 113 but less than 161
2ND EA TO SELECT	$a_2 = RS + k_h$	$= 116.84 + 163.5 =$	280.34	→	←	280.34 is greater than 254 but less than 293

In this example, the sampling interval ($k_h = 163.5$) is added to the random start ($RS = 116.84$) to obtain $a_2 = 280.34$. Since $a_2 = 280.34$ is greater than 254 (corresponding to EA 5) and less than 293 (corresponding to EA 6), EA 6 (Furculita) is selected as the second EA to be selected in the stratum Ciril.

STEP 7. Select the third EA within the stratum. Create a number a_3 by adding twice the sampling interval (k_h) to the random start to determine the third EA to select for the sample within the stratum. Use the resultant number in exactly the same way as in Step 6 above. The following chart provides an example.

Total number of households across district of Ciril (stratum h)	N_h	2289				
Number of EAs to select in Ciril	m_h	14				
Random number	rand()	0.7146				
Sampling interval	$k_h = N_h/m_h$	163.50				
Random start	$RS = rand() * k_h$	116.84				
			List of all EAs in stratum			
			EA number	EA name	Number of households per EA	Cumulative total of households
			1	Kvothe	53	53
			2	Gumbo	60	113
1ST EA TO SELECT	$a_1 = RS$	$= 116.84 =$	3	Pancho	48	161
			4	Glokta	42	203
			5	Rainbow's End	51	254
2ND EA TO SELECT	$a_2 = RS + k_h$	$= 116.84 + 163.5 =$	6	Furculita	39	293
			7	Stanka	65	358
			8	Stormlight	52	410
3RD EA TO SELECT	$a_3 = RS + 2 * k_h$	$116.84 + (2 * 163.5) =$	9	Deepness	55	465

In this example, twice the sampling interval ($2 * k_h = 2 * 163.5 = 327$) is added to the random start ($RS = 116.84$) to obtain $a_3 = 443.84$. Since $a_3 = 443.84$ is greater than 410 (corresponding to EA 8) and less than 465 (corresponding to EA 9), EA 9 (Deepness) is selected as the third EA to be selected in the stratum Ciril.

STEP 8. Continue in a similar fashion until the number of EAs (m_h) is reached. The following chart provides the final results of the selection.

Total number of households across district of Ciril (stratum h)	N_h	2289				
Number of EAs to select in Ciril	m_h	14				
Random number	rand()	0.7146				
Sampling interval	$k_h = N_h/m_h$	163.50				
Random start	$RS = rand() * k_h$	116.84				
			List of all EAs in stratum			
			EA number	EA name	Number of households per EA	Cumulative total of households
			1	Kvothe	53	53
			2	Gumbo	60	113
1ST EA TO SELECT	$a_1 = RS$	$= 116.84 =$	3	Pancho	48	161
			4	Glokta	42	203
			5	Rainbow's End	51	254
2ND EA TO SELECT	$a_2 = RS + k_h$	$= 116.84 + 163.5 =$	6	Furculita	39	293
			7	Stanka	65	358
			8	Stormlight	52	410
3RD EA TO SELECT	$a_3 = RS + 2 * k_h$	$116.84 + (2 * 163.5) =$	9	Deepness	55	465
			10	Black Dow	50	515
...
			41	Logan	54	2082
			42	Tul Duru	61	2143
			43	Bast	49	2192
			44	Kaladin	47	2239
14TH EA TO SELECT	$a_{14} = RS + 13 * k_h$	$116.84 + (13 * 163.5) =$	45	Arya	25	2264
			46	Cashin	25	2289

In this example, there are $m_h = 14$ EAs to select in the Ciril stratum. Therefore, 13 times the sampling interval ($13 * k_h = 13 * 163.5 = 2,125.5$) is added to the random start ($RS = 116.84$), resulting in $a_{14} = 2,242.34$. Since 2,242.34 is greater than 2,239 (corresponding to EA 44) and less than 2,264 (corresponding to EA 45), EA 45 (Arya) is selected as the 14th and last EA in stratum Ciril. In this example, the sampled EAs between the 3rd and 14th sampled EAs are not displayed.

Finally, it is also important to note that it is possible and acceptable to select the same EA more than once using systematic PPS sampling. This can happen if the number of households in a specific EA is very large and the sampling interval is relatively small (e.g., the sampling interval is less than half the number of households in a particular EA). The treatment of this situation will be dealt with at the third stage of sampling of households, which is discussed in Section 8.2.

5.3.1 Additional Issues to Consider for the First Stage Selection of EAs

The Treatment of Inaccessible EAs

In the context of many developing countries, it is often the case that after drawing a first stage systematic PPS sample of EAs, survey implementers discover while in the field or when about to undertake fieldwork that a few of the EAs are inaccessible. This can happen for a number of reasons, including physical limitations (e.g., a rainy season that washes out the access roads to the EA) and security issues (e.g., political instability) that make it unsafe for interviewers. Ultimately, it may be determined that interviewing cannot take place in the affected EAs.

Consider that in a survey of $n_{final-overall} = 2,760$ households, where there is a plan to survey in 111 EAs (i.e., 2,760 households to be sampled overall divided by 25 households to be sampled per EA), several EAs might fall into this category. Suppose we have a case where 6 of the 111 EAs are inaccessible. If sampling in these six EAs is simply abandoned, there will be a significant shortfall of sample: 150 households in the case of interviewing in 25 households per EA. To compensate for this, one should adopt a protective strategy by drawing a “random-generated reserve sample.” The recommendation is to apply a “two-phase” approach for the first stage of sampling. That is, 117 EAs are sampled using systematic PPS sampling at the first phase of the first stage. Here the intention is to interview in only 111 EAs, but 6 additional EAs are sampled and are kept in “reserve.” At a second phase of the first stage, 6 out of 117 EAs are randomly subsampled using fractional interval systematic sampling (described in detail in Section 8.4). Interviewing takes place in the 111 EAs that are not selected at the second phase. The six EAs in reserve are assigned numbers 1 through 6 (in accordance with the order in which they were randomly sampled) to define the order of release, one by one. That is to say, if only one reserve EA is needed, the reserve EA labeled with the number 1 replaces the first EA from among the original 111 in which there are access issues. EAs having access issues continue to be replaced with reserve EAs in this fashion as needed, with the ultimate aim of ensuring that the overall number of required EAs (111) is achieved. A detailed description of the necessary sample weight adjustments needed is given in Section 10.1.2.⁶⁰

⁶⁰ For more details on reserve samples, see: Lin, Li; Krenzke, Tom; and Mohadjer, Leyla. 2014. “Considerations for Selection and Release of Reserve Samples for In-Person Surveys.” Available at: http://ww2.amstat.org/sections/srms/Proceedings/y2014/files/311099_86830.pdf.

The above example is simplified to illustrate how to treat inaccessible EAs in general. However, for PBSs, stratification creates additional complexity. For instance, we revisit the PBS illustrated in Table 8 with $n_{final-overall} = 2,760$ households, where proportional allocation is applied to six strata. There, the individual sample sizes are 333, 327, 701, 550, 439, and 410 households, respectively, in the six strata, which translates to 14, 13, 28, 22, 18, and 16 EAs selected within each of the six strata, respectively (assuming 25 households sampled per EA). In most cases, local survey implementers will have an advance idea of which strata are likely to have EAs that might be a concern in terms of accessibility. In this case, drawing a reserve sample should be limited to those strata only. So, in the above example, reserve samples may need to be drawn in only one or two strata, depending on the situation. Finally, the number of additional EAs to include in the reserve sample is subjective, and is based on knowledge of the local survey implementer with regard to which and how many EAs are likely to be problematic.

Sampling EAs at the Second Point in Time of a Comparative Analytical PBS

A final issue to consider surfaces in relation to comparative analytical PBSs, when the main aim is applying a statistical test of differences over two time points, namely, whether the survey implementer should employ the same set of sampled EAs for the end-line PBS that was used for the baseline PBS or, instead, a fresh set of EAs should be drawn for the end-line. One of the major advantages of keeping the same set of sampled EAs over the two time points, which is akin to having a longitudinal panel of EAs, is the ability to measure change over time within the same EAs. Even though the same sampled individuals are not tracked over time, sampling the same EAs over time increases the ability to capture change because the EA-level effects are “fixed.” However, there are a number of disadvantages of and considerations in relation to adopting such a strategy, as well. These are highlighted below.

- If there is a substantial change or movement in the population over the two time points, there will likely be an erosion of representativity in the target population covered by the sampled EAs.
- If there is substantial growth or reduction in the number of households in the EAs on the first stage sampling frame over the two time points, then the size measures that were used to sample the EAs for the baseline PBS will be inaccurate at the time of the end-line PBS.⁶¹ There is no practical way of adjusting for these inaccuracies to obtain a more accurate reflection of the population at the time of the end-line. Thus, these inaccuracies will be reflected on the first stage sampling weights, which will also be used for the end-line PBS if the same EAs are used.
- Additionally, if a listing of households (discussed in the next chapter) in the sampled EAs was undertaken at the time of the baseline PBS, the listing will be inaccurate for the purposes of sampling households within those sampled EAs for the end-line PBS. In such cases, it would be prudent to relist the sampled EAs prior to the end-line PBS. The relisting would then be used as a sampling frame from which a subsequent (second or third) stage sample of households could be drawn for the end-line PBS.
- If projects are aware that sampled EAs from the baseline survey will be used again for the end-line survey, this fact might provoke an undue project focus on the sampled EAs by IPs.

⁶¹ Typically, size measures are obtained from the most recent census. It may well be that if the baseline PBS takes place well after the census, the size measures may be inaccurate for the baseline PBS as well. However, size measures become increasingly inaccurate as one moves forward in time from the original census year.

Therefore, if such a strategy is to be adopted, it is best not to make public the fact that the same EAs will be revisited at a later time.

- If a different third-party firm conducts the end-line PBS than the one that conducted the baseline PBS, it is essential that all the required information and relevant files relating to the sampled EAs be made available. Among other things, the exact location of the EAs sampled for the baseline PBS will need to be known, in order to relocate the same set of EAs for the end-line PBS. In developing countries, locating EAs can sometimes be a challenge without high-quality information, including maps, EA names, and GPS information. Obtaining such information could present a challenge if considerable time has passed between the two survey occasions, and highlights the need for comprehensive and complete documentation and record keeping.

In consideration of the above issues, Feed the Future recommends that, in most cases, survey implementers draw a fresh first stage set of EAs for the end-line PBS to ensure proper representativity at both time points. However, survey practitioners can, in some cases, decide to use same set of EAs for the end-line PBS if there are compelling reasons to do so and if it can be demonstrated that there have not been considerable changes in the population and EA composition over time.

6. Preparation for Second and Third Stage Sampling: Listing Exercise for Sampled EAs

Prior to conducting a second stage sampling of segments (if necessary) or a third stage sampling of households, a field-based listing operation within each sampled EA must take place. The listing of households for each sampled EA obtained through this exercise is used as the second and third stage frames in the selection of segments and households to be included in the PBS. In addition to listing households, one of the main benefits of this exercise is that it serves to record and update measures of size (e.g., number of households in the sampled EA) used for sampling at the second and third stages. There is no substitute for a proper listing operation prior to the main fieldwork; it is strongly recommended that survey implementers not avoid this step by using “random walk,”⁶² “spin the pen” (also known as “spin the bottle”), or similar non-random methods of household selection within selected EAs. Additionally, it is common practice among some implementers to replace the listing process, which is perceived as lengthy and resource-intensive, by obtaining an updated number of households in each selected EA from local sources (e.g., community headmen). While systematic sampling based on such counts is indeed probability-based, this practice is discouraged because such counts are likely inaccurate and the sampling weights based on them will lead to either a sample shortfall or sample overage in many EAs.

The listing operation of the selected EAs consists of two steps to be undertaken by the listing team.⁶³ In the first step, listers locate the boundaries of the sampled EAs using an EA base map or location map⁶⁴ in relation to the larger surrounding geography, to facilitate interviewers in locating the EA at a later time. In the second step, listers typically create or update existing EA sketch maps,⁶⁵ including indications of all dwelling units located in the EA and other major landmarks to help interviewers locate the dwelling units at a later time. During this second step, listers start at a well-defined location in the EA and follow a systematic pattern through EA, making sure to exclusively and exhaustively cover the entire EA. Listers stop at each dwelling unit to record the following information: a description of the exterior of the dwelling unit, whether one or more households are resident within the dwelling unit, and the name of a responsible adult household member for each household affiliated with the dwelling unit. It is also recommended that PBSs take GPS readings of the location of each dwelling unit in the sampled EA (to help interviewers locate each dwelling unit in the sampled EA at a later time). If resources are limited or the timeline is restricted, PBSs may alternatively choose to undertake only one GPS reading at the center of the EA (to help interviewers locate the EA at a later time).

⁶² The random-walk method entails: i) randomly choosing a starting point and a direction of travel within a sampled EA, ii) conducting an interview in the nearest household, and iii) continuously choosing the next nearest household for an interview until the target number of interviews has been obtained within the EA.

⁶³ Note that the listing exercise generally constitutes the beginning of fieldwork. As such, it is important to develop field manuals, tracking logs, and quality control templates to ensure that high-quality household listings and data collections are produced from the fieldwork.

⁶⁴ An *EA base map* is a map that shows the geographical location and boundaries of an EA, whereas an *EA location map* is a map that provides a more detailed view of an EA. The latter map shows roads and landmarks located within the EA. It may also include important roads and landmarks in neighboring areas. Both maps are typically provided by the local census authority.

⁶⁵ An *EA sketch map* is a map that is created by the census or survey cartographers using the location map. Each structure in the cluster is marked on the sketch map. The cartographers also indicate physical features and landmarks that are not on the location map, including mountains, rivers, roads, and electrical poles.

Finally, it is important to consider that, because there may be more than one household per dwelling unit and because the second and third stage sampling frames for a PBS (as defined in Section 3.2 and Section 3.3) are articulated in relation to the household rather than the dwelling unit, separate listings should be created for all households within the sampled EA. This means that a household listing form (rather than a dwelling unit listing form) should be used, and there should be a separate entry on the form for each household in the EA. This may necessitate repeating information on some of the dwelling units in the EA, in the case of multi-household dwelling units.⁶⁶ All relevant information in relation to each listed household, as indicated in Section 3.3, should be entered on the form.

In some developing countries, special consideration must be given to polygamous arrangements, or any other modalities of living arrangements known to exist that don't align precisely with the definition of "household" used by Feed the Future. In the case of polygamy, the delineation of dwelling units can become complicated if, for instance, different wives and their common husband live in fenced compounds with multiple physical structures within, as these separate structures may ultimately serve as rooms within one household rather than as separate dwelling units. The concepts of eating from the same pot and the recognition of a lead decision maker tend to be most critical in defining a household in these cases, although at times these two concepts may be at odds with each other. For instance, in some countries where polygamy is prevalent, different wives with a common husband cook for their own children and direct relatives from separate pots, while in other countries wives may rotate cooking for entire extended polygamous households. In both cases, the husband may be considered the lead decision maker. However, in the first case, the different wives (along with their children and relatives) may be considered as different households (each having the same lead decision maker), while in the second case, all the wives and their children and relatives may be considered as one household (again having the same lead decision maker). In all cases, it is important to consult with in-country partners, such as the NSO, to determine appropriate country-specific guidelines on how to handle these more complex types of living arrangements.

⁶⁶ For more information on listing, see: "Feed the Future Household Listing Manual." 2018. Available at: <https://agrilinks.org/post/feed-future-zoi-survey-methods>.

7. Potential Second Stage Sampling of Segmented Enumeration Areas

Census EAs, as defined and delineated by the NSO or the census authority of the country in which the PBS is taking place, are typically roughly uniform in size to facilitate equal census enumerator workloads. As mentioned earlier, this is one of the great advantages of adopting EAs as first stage sampling units. Despite this, sometimes there is rapid change (growth or reduction) in the size of some EAs' populations between the time of the last census and the time of the PBS, as ascertained by the listing operation. If a sampled EA has grown too large by the time of the listing operation, field teams need to divide the EA and subsample one part of it, a process called "segmentation."

A rapid count of the number of households in a sampled EA on the listing pass should give a sense of how large the EA is. Alternatively, if satellite imagery (with distinct structures visible) and EA shapefiles are both available, it may be possible (and certainly preferable) to avoid segmentation as part of the fieldwork listing exercise, and instead undertake segmentation at the survey office, before the listing exercise commences.

Segmentation need not take place in all sampled EAs, but rather only for those EAs that are deemed so large that fieldwork is deemed unwieldy. There is no exact rule to determine when an EA should be segmented. In cases where the average EA size is small (e.g., fewer than 150 households per EA), Feed the Future recommends segmenting any EA that is at least twice the size of the average EA in the survey area.⁶⁷ In cases where the average EA size is large (e.g., at least 150 households per EA), the recommendation (following DHS protocol) is to segment EAs that exceed 300 households in size.

Once it has been determined that an EA should be segmented, it should be divided according to the above rule as many times as needed so that each segment is roughly equal to the average EA size and, at the same time, so that the resultant segments are roughly equal in size. Typically, an EA will not need to be divided more than once or twice to achieve this. For instance, if the average EA size for a survey area is 150 and a selected EA has size 312, it could be divided into two segments of 156. If the average EA size is 160, for example, and a selected EA in the same survey area has size 500, it could be divided into two segments of size 167 and one segment of size 166. Each segment should be formed in such a way that it is as compact as possible, so that it contains households that are contiguous to one another, and so that the segments within the EA are of roughly equal size. Then, one segment from among all the segments within the EA should be randomly selected using PPS sampling.⁶⁸

To illustrate how PPS sampling is undertaken, consider an EA consisting of 610 households. Suppose the decision has been made to divide the EA into three segments of roughly equal size. Table 10 illustrates the segmentation. In this example, the three segments have been formed so that the number of

⁶⁷ Calculating the average EA size can be done from the field office before fieldwork commences. The average EA size can be computed directly using the size measures (number of households for each EA) from the first stage sampling frame.

⁶⁸ Technically speaking, it is preferable to randomly select two segments from among all the segments within the EA, because to compute the SEs for each stage of sampling, a minimum of two units must be drawn from each unit at the previous stage of sampling. However, due to the complexity of randomly selecting two segments per EA without replacement using PPS, the number of segments to sample is limited to one. Furthermore, it is assumed that segmentation will take place only in a small number of sampled EAs and, therefore, that the contribution to the overall SE from this stage of sampling will be small and can effectively be ignored.

households within the segments is 190, 200, and 220, respectively. Note that the division of this EA into three segments is not exactly equal—and this is usually necessary. This is because the division of the EA should take into account where households are located within the EA, and each segment should have households grouped together in a geographically logical way. It does not make sense to add households that are distant from the main grouping, just to ensure the segments are equal in size.

Table 10. Illustration of Segmentation

Segment Number	Number of Households	Percent of Total Households	Cumulative Proportion of Total Households
1	190	31.1	0.31
2	200	32.8	0.64
3	220	36.1	1.00
Total	610	100	

In the table, the proportion of total households represented by each segment is computed. The last column is the cumulative total of households expressed as a decimal number between 0 and 1. To randomly select one of the segments for interviewing using PPS, a random number is generated between 0 and 1. For instance, the Microsoft Excel function *rand()* generates a random (fractional) number greater than or equal to 0 and less than 1. This random number is compared with the last column in the table. The first segment for which the cumulative size in the last column is greater than or equal to the random number is selected as part of the survey. For instance, if the random number generated is 0.42, then segment number 2 would be selected because 0.31 is less than 0.42, which is less than 0.64. In this case, household selection and interviewing takes place in segment 2 only, rather than the entire sampled EA.

8. Third Stage Sampling of Households within Sampled EAs or Segments

The third stage of sampling consists of randomly selecting the households in which to conduct interviews within each sampled EA (or segment if applicable). For simplicity of presentation, in this chapter, it is assumed that no segmentation has taken place and households are selected within sampled EAs at the next stage. The listing exercise provides an ordered “list” of households for each sampled EA (with locations or GPS coordinates) to serve as a household sampling frame. The process of randomly selecting households is undertaken from the survey office after the listing exercise and before data collection commences, using an equal probability variant of systematic sampling called **fractional interval systematic sampling**. It is preferable to use systematic sampling rather than SRS because systematic sampling spreads the sample of selected households throughout the EA and, in doing so, captures more within-EA variation than SRS. Note that the so-called “random-walk” or “spin the pen” (also known as “spin the bottle”) method should never be used in this context as these are non-probability-based methods. Sampling weights cannot be computed, and biased results will be obtained if these methods are used.

8.1 Fractional Interval Systematic Sampling

The steps to apply fractional interval systematic sampling can be carried out using any appropriate software. As mentioned before, using such software has many advantages, including automating the selection, documenting the selection, and automating the provision of third stage sampling weights. An example of a manual selection is provided below to illustrate the concept. The syntax provided is what would be used in Microsoft Excel.

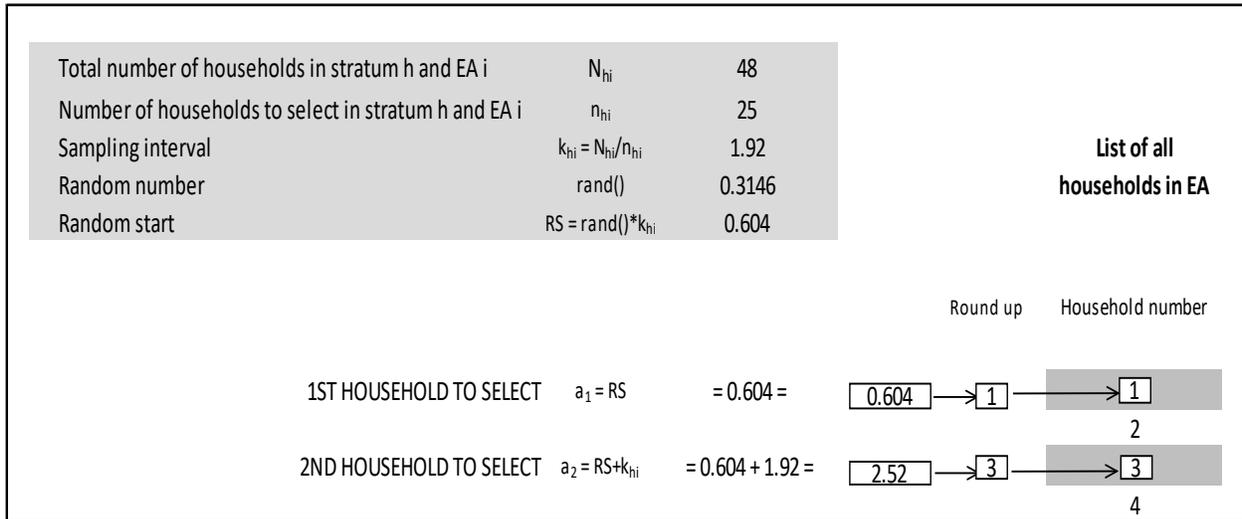
Continuing with the example from Section 5.2.1 to illustrate sampling within the selected EAs in the stratum of Ciril, suppose that one wishes to randomly select 25 households using fractional interval systematic sampling in the first EA selected at the first stage of sampling, EA #3 (Pancho), which has 48 households overall. Assume that there was no segmentation undertaken in EA #3 given that Pancho is not large, so there was no need for a second stage of sampling. Households can therefore be sampled directly within this EA.

STEP 1. Create a list of all households in the sampled EA. This is generated from the listing exercise described in Chapter 6. Make sure that the households appear in the order in which they were listed during the listing exercise, that is to say, households that are adjacent to each other should appear next to each other in the list.⁶⁹

STEP 2. Calculate a sampling interval. The sampling interval (k_{hi}) is calculated by dividing the total number of households in the sampled EA i within stratum h (N_{hi}) by the number of households to select in sampled EA i within stratum h (n_{hi}). The formula is given by:

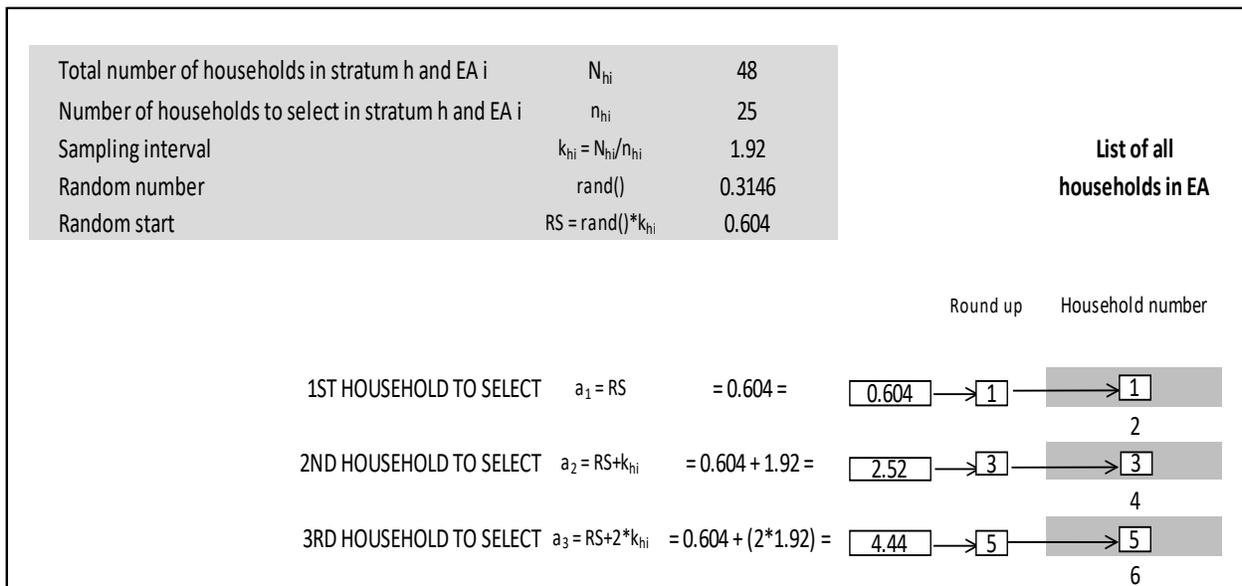
⁶⁹ In practice, the extent to which the home office can ascertain that households that are adjacent to each other appear next to each other in the listing is limited, unless this is made part of the quality control check of the listing exercise.

integer part and all decimals), rounded up to the nearest integer. The following chart provides an example.



In this example, the sampling interval ($k_{hi} = 1.92$) is added to the random start ($RS = 0.604$) to obtain 2.52. This is rounded up to 3, and therefore household 3 is selected as the second household in the sample within the EA of Pancho.

STEP 6. Select the third household in the EA. Add twice the sampling interval (k_{hi}) to the random start (RS) to determine the third household to select for the sample. Use the resultant number in exactly the same way as in Step 5 above. The following chart provides an example.



was not part of the selected sample, as a substitute. If a household is not available for interviewing on the first visit, the data collector should revisit the household up to three times to secure an interview. If, after three attempts, an interview still cannot be secured, then the household should be labeled as a “non-respondent household” and sample weight adjustments must be made after fieldwork to compensate for the data relating to the missing household.⁷⁰ While every effort must be made to obtain an interview with the selected households, recall that when the sample size was calculated, the initial sample size was inflated to compensate for an anticipated level of household non-response, i.e., to compensate for the fact that not all household interviews in the field would be secured as planned.

2. It was noted at the end of Section 5.2.1 that when using systematic PPS sampling at the first stage of sampling, it is technically possible to select the same EA more than once. Although this is rare, when this happens, the two (or more) selections of the same sampled EA should be treated separately. That is to say, when the same EA is selected twice at the first stage of sampling using systematic PPS sampling, and fractional interval systematic sampling is used at the third stage of sampling (assuming that there was no second stage sampling of segments), the list of households in the selected EA should be divided in two equal parts along geographic lines, and separate sampling of households using fractional interval systematic sampling should take place in each half of the EA. This is to ensure that there will not be any overlap in the two samples of households within the same sampled EA. In this case, if 25 households are to be sampled for each selected EA, then 25 households should be sampled in each of the two parts of the EA. This means that the sampling interval in each of the two parts will be half the size of what it would have been in the EA as a whole.
3. As mentioned earlier, selecting the same EA twice can happen if the number of households in a particular EA is very large and the sampling interval is relatively small (e.g., the sampling interval is less than half the number of households in a particular EA). Given that EAs that are selected twice tend to be large, they are also more prone to requiring segmentation. In the rare case where the same EA is selected twice at the first stage of sampling *and* it is deemed that the (doubly occurring) EA should be segmented at the second stage of sampling, the guidance is as follows: segment the EA and randomly select one segment at the second stage of sampling, and then divide the sampled segment into two equal parts and undertake independent third stage sampling of households using fractional interval systematic sampling in each half of the segment (as suggested in point 2 above.)

⁷⁰ Household non-response adjustments are discussed in more detail in Section 10.2.1. Even if a household is responsive, it may be the case that not all eligible respondents within the household agree or are available to be interviewed. Therefore, it still might be necessary to make three visits to a sampled household to collect data from all eligible respondents to be interviewed within the household. In this case, if after three attempts, interviews still cannot be secured with all the necessary individuals, then the non-responding sampled individuals should be labeled as such, and sample weight adjustments should be made after fieldwork to compensate for the data relating to the missing individual(s). Individual non-response adjustments are discussed in more detail in Section 10.2.2.

8.3 A Special Application of Fractional Interval Systematic Sampling: Joint Baseline and End-Line PBS

Section 2.1 (Box 1), Section 4.4, and Section 5.1 (Example 2) describe a scenario where both a baseline PBS and an end-line PBS are to be conducted jointly using one survey vehicle, and each survey requires $n_{final-overall} = 2,760$ households.⁷¹ The original and the new ZOIs are stratified using the districts in Table 9 in Section 4.4, and two separate proportional allocations require that 1,644 and 1,806 households for baseline and end-line, respectively, be allocated to the set of strata with common districts for both surveys. Given that the plan is to sample 25 households in each sampled EA, for the end-line PBS, $1,806 / 25$ or approximately 72 EAs first need to be sampled in the set of strata with common districts. Similarly, for the baseline PBS, $1,644 / 25$ or approximately 66 EAs need to be sampled in the set of strata with common districts. Therefore, as described earlier, in the set of strata with common districts, a two-phase sample design is employed at the first stage of sampling whereby 72 EAs are sampled at the first phase of the first stage of sampling, and then 66 of the 72 EAs are subsampled at the second phase of the first stage of sampling.

To implement this, systematic PPS sampling is used, as described in Section 5.2.1, to randomly select the 72 EAs. Then, the method of fractional interval systematic sampling is used, as described in Section 8.1, to randomly select 66 of the 72 EAs, but applying the method to EAs instead of to households. That is to say, the 72 EAs that are sampled at the first phase serve as the “frame” from which 66 EAs are sampled using fractional interval systematic sampling at the second phase. The 72 sampled EAs are used to determine results for the end-line PBS, while the 66 subsampled EAs are used to determine results for the baseline PBS. Finally, at the next stage of sampling, 25 households are sampled from each of the 72 sampled EAs using fractional interval systematic sampling, as described in Section 8.1. The 25 sampled households in each of the sampled EAs used for the end-line PBS are also used for the baseline PBS in the 66 subsampled EAs only. Sample weighting for this two-phase approach is described in Section 10.1.2.

8.4 A Special Application of Fractional Interval Systematic Sampling: The Treatment of Inaccessible EAs

In Section 5.2.2, a scenario is described where, after drawing a first stage systematic PPS sample of EAs, survey implementers discover while in the field or about to undertake fieldwork that a few of the EAs are inaccessible. The example states that 111 EAs are to be sampled in a PBS and, before fieldwork commences, it is suspected that up to 6 EAs may have access issues. The recommendation is to adopt a “two-phase” approach at the first stage of sampling whereby 117 EAs are selected using systematic PPS sampling at the first phase of the first stage, as described in Section 5.2.1. At a second phase of the first stage, 6 out of 117 EAs are subsampled using fractional interval systematic sampling as described in Section 8.1, applying the method to EAs instead of households. That is to say, the 117 EAs that are sampled at the first phase serve as the “frame” from which 6 EAs are sampled using fractional interval systematic sampling at the second phase. Special weighting considerations are described in Section 10.1.2.

⁷¹ Although the survey questionnaire content may not be identical for the baseline and end-line PBSs, it is worth undertaking joint administration of the two surveys only if there is significant overlap between the two surveys in the content of the questionnaires and indicators to be reported on.

9. Fourth Stage Sampling of Individuals within Sampled Households

As noted in Section 3.4, during the fieldwork and as part of the household interview, a roster of household members is established for each sampled household by obtaining information from a responsible adult member of the household; this serves as a fourth stage sampling frame of individuals for all sampled households.

Demographic information, including name, relationship to responsible adult who provided roster information on the household, age, sex, and any other information needed to establish eligibility to respond to one or more subsequent questionnaire modules, is collected on all individuals included in the roster.

As can be seen in Table I, data that are collected through a PBS in support of Feed the Future ZOI PBS indicators often span several sampling groups. For instance, children under 6 months (for the EBF indicator), children 6–23 months (for the MAD indicator), children under 5 years (for the stunting, healthy weight, and wasting indicators), non-pregnant women age 15–49 years (for the body mass index [BMI] indicator), women age 15–49 years (for the minimum dietary diversity for women [MDD-W] indicator), primary female and male decision makers in the household (for the Abbreviated Women’s Empowerment in Agriculture Index [A-WEAI] indicator), and producers (for the producers applying improved management practices or technologies and yield of agricultural commodities indicators). It is essential to ensure that only eligible household members are selected and interviewed for the appropriate questionnaire modules.

The issue remains how to select members of the household to be administered the appropriate questionnaire modules from among all eligible household members. The guidance is that for all sampling groups, **all** eligible individuals residing within the sampled household should be selected. For instance, if there are three children under the age of 5 in a sampled household, all of them should be selected for the purposes of taking anthropometric measures in support of the stunting, healthy weight, and wasting indicators. Although we select all eligible household members for interviewing with certainty, we still technically refer to this as a stage of sampling.

If a survey were to randomly select one eligible individual per household within each of the above sampling groups, there would need to be a separate individual-level weight associated with each of the associated sampling groups. Such a sampling strategy would be more complex to manage. The strategy of selecting all eligible individuals within a sampled household helps avoid this situation.

Furthermore, selecting all eligible household members simplifies fieldwork because a procedure for selecting individuals within sampled households (such as a Kish Grid in the case of a paper-based survey) does not need to be administered in the field.⁷²

⁷² Although a paper-based Kish Grid can be complex to administer in the field, the computer-assisted personal interviewing (CAPI)-based administration of a Kish Grid using a tablet or some other device would be much simpler due to automation of the procedures.

Additionally, for rarer sampling groups, such as children age 0–5 months, children age 6–23 months, and children age 0–59 months, this approach has the advantage of requiring the sampling of fewer households to achieve the desired number of sampled children.

Finally, the approach reduces the “unequal weighting” DEFF because all eligible members of the household are sampled with an equal probability of 1. If, alternatively, one member is selected for interviewing from among all eligible members within a household, then unequal weighting is introduced because different households have a different number of eligible members and this is reflected in the sampling weights. An unequal weighting design lowers the precision (i.e., raises the SEs) of survey estimates. In the case of less “rare” sampling groups (such as women age 15–49 years), selecting all eligible woman in a household does increase the interviewing burden on both the household and the interviewer. However, the modules for which data are collected on the members of this sampling group (relating to the MDD-W indicator) are relatively short to administer.

10. Sample Weighting

After data collection is completed, there are a number of activities that typically take place during and after fieldwork but prior to data analysis. If paper questionnaires were used to capture data, they are physically transmitted from the field to a central office, and data are entered into a database using double data entry⁷³ to help minimize errors in data entry. If computer-assisted personal interviewing (CAPI) was used and data were collected using a tablet or some other electronic device, then these data are electronically transmitted from the field to a central office for processing.

Once at the central office, the data are uploaded into a secure database and the data are then “cleaned.” Although field supervisors and quality control editors routinely check for different kinds of errors during fieldwork so that they can be resolved on the spot, some residual errors frequently remain after fieldwork is complete. This is true even with CAPI-based data collections, where a substantial portion of data cleaning is built into the software. The cleaning of the data involves a number of steps, including but not limited to: checking for valid data ranges (e.g., women of reproductive age must be between 15 and 49 years old), checking that questionnaire logic was adhered to (e.g., skips and filters respected), and creating flags for the resolution of any logical inconsistencies in the data (e.g., a 4-year-old married woman).

After data cleaning, a check is typically performed to make sure that there are no “outlier” values (e.g., a child under the age of 5 with weight of 35 kilograms). Sampling weights are then constructed to reflect the various stages of sampling as well as to reflect any residual household or individual non-response. A sampling weight is attached to each of the households and individuals on the cleaned data file.

Finally, data analysis takes place. This includes the production of estimates of the Feed the Future ZOI PBS indicators and their associated CIs and SEs in the case of a descriptive survey, and includes a statistical test of differences at the second time point for a comparative analytical survey. This chapter deals with the topic of creating sample weights.

The first step before data analysis is to calculate the sample weights associated with each of the households and individuals who have been randomly selected in the PBS and who have responded to the survey interview questions.⁷⁴ Sample weights for each selected individual (or household in the case of some indicators where the household is the sampling group) are calculated and applied to corresponding individual (or household) survey data record(s) to inflate the individual (or household) data values up to the level of the population of individuals (or households). In essence, sample weights are a means of

⁷³ Double data entry is a data entry quality control method where, in the first pass through a set of records, an operator enters data from all records into a database. Then, on a second pass through the batch, a verifier enters the same data. The contents entered by the verifier are compared with those of the original operator. If there are differences, the data fields or records where there are differences are flagged for follow-up and reconciliation.

⁷⁴ Although there are four stages of sampling suggested by this guide, the first stage sampling of EAs is implemented prior to the commencement of fieldwork, while the listing process and the last three stages of sampling (of segments, households, and individuals) are implemented during fieldwork. As such, it is important that at each stage of sampling in the field, information is captured on specialized templates (either on paper or electronically) that are then transmitted to a central office to facilitate the production of sampling weights. The necessary information depends on the stage of sampling. For instance, for the third stage sampling of households within EAs or segments, a record of all the selected households should be kept, along with all the associated information on the selected households outlined in Section 3.3. The information on these templates should be converted to electronic files (if they are paper-based) so that they can be easily merged with the files containing the corresponding data from interviews conducted in the field.

compensating for having collected data on a sampled subset of the population, instead of having conducted a full “census” of the entire population.

For the multi-stage clustered survey design discussed in earlier chapters, sample weights should be calculated and used in the construction of estimates of each indicator to account and compensate for the following:

- Probabilities of selection at each stage of sampling
- Non-response at the individual and household levels

10.1 Calculating Probabilities of Selection and Sampling Weights

All individuals and households included on their respective sampling frames have an underlying chance or probability of being included in the sample. To illustrate, suppose the indicator of interest is “Prevalence of Moderate and Severe Food Insecurity (based on the Food Insecurity Experience Scale [FIES]),” where the target sampling group is the household. If one household is randomly selected from among 100 possible households in an EA with equal probability, the probability of the household being selected is 1 in 100 and the associated sampling weight is 100, i.e., the inverse of the probability of selection. One interpretation of a sample weight is that the selected household represents all 100 households, that is to say, the sampled household, along with the 99 other households that were not selected in the survey. When the survey data on households are used to make inferences about the entire population of households, the survey weighted data from the sampled household used in the example above will have the effect of “being replicated” 100 times.

10.1.1 Overview of How to Calculate Probabilities of Selection

For survey designs where there are four stages of sampling (such as those suggested in this guide), the sampling weight associated with the probabilities of selection for each sampled household or individual is calculated, in general terms, using the steps outlined below.

STEP 1. Calculate the probability of selection at the first stage of sampling, which corresponds to the selection of EAs. This is done for each EA.

$$\text{probability of selection of EA } i \text{ in stratum } h \text{ at the first stage} = f_{1hi}$$

STEP 2. Calculate the probability of selection at the second stage of sampling; this corresponds to the “conditional” selection of a segment, assuming that the EA in which the segment is situated has been selected at the first stage of sampling. This step is only undertaken if segmentation is necessary.

$$\text{probability of selection of segment } j \text{ at the second stage, assuming EA } i \text{ in stratum } h \text{ selected at first stage} = f_{2hij}$$

STEP 3. Calculate the probability of selection at the third stage of sampling; this corresponds to the “conditional” selection of a household, assuming that the segment in which the household is situated has been selected at the second stage of sampling, and the EA in which the segment is situated has been selected at the first stage of sampling. If segmentation was not undertaken at the second stage, then assume the household was selected from the corresponding sampled EA.

probability of selection of household k at the third stage, assuming segment j selected at second stage,

$$\text{and EA } i \text{ in stratum } h \text{ selected at the first stage} = f_{3hijk}$$

STEP 4. Calculate the probability of selection at the fourth stage of sampling; this corresponds to the “conditional” selection of individuals within households, assuming that the household in which the individual resides has been selected at the third stage of sampling, the segment in which the household is situated has been selected at the second stage of sampling, and the EA in which the segment is situated has been selected at the first stage of sampling.

probability of selection of individual l at the fourth stage, assuming household k selected at third stage,

$$\text{segment } j \text{ selected at second stage, and EA } i \text{ in stratum } h \text{ selected at first stage} = f_{4hijkl}$$

STEP 5. To calculate the overall probabilities of selection for **individuals** selected for inclusion in the sample, f_{hijkl} , multiply the probability of selection at each of the four stages together.

$$f_{hijkl} = f_{1hi} * f_{2hij} * f_{3hijk} * f_{4hijkl}$$

To calculate the overall probabilities of selection for **households** selected for inclusion in the sample, f_{hijk} , multiply the probability of selection at each of the first three stages together.

$$f_{hijk} = f_{1hi} * f_{2hij} * f_{3hijk}$$

STEP 6. To calculate the overall sampling weights for **individuals** that reflects the probabilities of selection at each stage (to be used in the computation of individual-level indicators in Tables I and 2), take the inverse of the associated quantity calculated in step 5:

$$\text{overall sample weight (IND)} = w_{hijkl} = \frac{1}{f_{hijkl}} = \frac{1}{f_{1hi} * f_{2hij} * f_{3hijk} * f_{4hijkl}}$$

To calculate the overall sampling weights for **households** that reflects the probabilities of selection at each stage (to be used in the computation of household-level indicators in Tables I and 2), take the inverse of the associated quantity calculated in step 5:

$$\text{overall sample weight (HH)} = w_{hijk} = \frac{1}{f_{hijk}} = \frac{1}{f_{1hi} * f_{2hij} * f_{3hijk}}$$

Last, two non-response adjustments are made to the overall sampling weights to compensate for: i) the selected households that did not respond to the survey and ii) the eligible individuals who did not respond to the survey. More details on this will be given in the sections that follow. Only adjustment i) is made to w_{hijk} , whereas both adjustments i) and ii) are made to w_{hijkl} .

The previous section provides general formulas for computing sample weights, taking into account each stage of sampling. The purpose of sampling weights is to take into account the distortion imposed by the unequal sampling probabilities for different units in the population. In the following sections, specific formulas are provided for probabilities of selection and sampling weights corresponding to each of the four stages that have been discussed earlier. Table I I provides a summary of the methods of sampling recommended for each of the four stages of sampling described in this guide. See Section 5.2.1, Chapter 7, Section 8.1, and Chapter 9 for more details on each of the methods.

Table 11. Summary of Methods for Each Stage of Sampling

	Stage 1 Selection of EAs	Stage 2 Selection of Segments	Stage 3 Selection of Households	Stage 4 Selection of Individuals
Method of Sampling	Systematic PPS	PPS	Fractional Interval Systematic	Take all

10.1.2 Calculating the Probability of Selection at the First Stage Sampling of EAs

For the first stage of sampling, when **systematic PPS sampling** is used to sample EAs (as discussed in Section 5.2.1), the probability of selecting the *i*th EA in stratum *h* is calculated as follows:

$$f_{1hi} = \frac{\text{(number of EAs selected in stratum } h * \text{total number of households in selected EA } i \text{ in stratum } h)}{\text{total number of households in all EAs in stratum } h} = \frac{m_h * N_{hi}}{N_h}$$

In the above formula, m_h is the number of EAs selected in stratum *h* (as computed in Section 5.1), N_{hi} is the total population number of households in selected EA *i* in stratum *h*, and N_h is the total population number of households across all EAs in stratum *h*. Because EAs are selected prior to fieldwork, both N_{hi} and N_h are determined before the listing exercise takes place and are typically based on counts from the latest national census or some other administrative source. Therefore, these counts may differ from analogous counts obtained through the listing exercise during fieldwork. This is accepted practice in survey sampling, and no adjustment needs to be made to compensate for this discrepancy after the fact. The following illustrates the calculation of the probabilities of selection for systematic PPS sampling, continuing the example from Section 5.2.1.

Total number of households across district of Cyril (stratum h)	N_h	2289							
Number of EAs to select in Cyril	m_h	14							
Random number	rand()	0.7146							
Sampling interval	$k_h = N_h/m_h$	163.50							
Random start	$RS = rand() * k_h$	116.84							
			List of all Eas in stratum						
				$f_{1hi} = (m_h * N_{hi}) / N_h$					
			EA number	EA name					
			Number of households per EA (N_{hi})	Cumulative total of households					
			Probability of Selection (first stage)						
			1	Kvothe	53	53			
			2	Gumbo	60	113			
1ST EA TO SELECT	$a_1 = RS$	$= 116.84 =$	116.84	3	Pancho	48	161	0.293	$= (14 * 48) / 2289$
				4	Glokta	42	203		
				5	Rainbow's End	51	254		
2ND EA TO SELECT	$a_2 = RS + k_h$	$= 116.84 + 163.5 =$	280.34	6	Furculita	39	293	0.238	$= (14 * 39) / 2289$
				7	Stanka	65	358		
				8	Stormlight	52	410		
3RD EA TO SELECT	$a_3 = RS + 2 * k_h$	$116.84 + (2 * 163.5) =$	443.84	9	Deepness	55	465	0.336	$= (14 * 55) / 2289$
				10	Black Dow	50	515		
...		
				41	Logan	54	2082		
				42	Tul Duru	61	2143		
				43	Bast	49	2192		
				44	Kaladin	47	2239		
14TH EA TO SELECT	$a_{14} = RS + 13 * k_h$	$116.84 + (13 * 163.5) =$	2242.34	45	Arya	25	2264	0.152	$= (14 * 25) / 2289$
				46	Cashin	25	2289		

Special Scenario of a Joint Baseline PBS and End-Line PBS

Recall the scenario described in Section 2.1 (Box 1), Section 4.4, Section 5.1 (Example 2), and Section 8.3 where both a baseline PBS and an end-line PBS are conducted using one survey vehicle. In this case, a two-phase sample is employed at the first stage of sampling in the set of strata with common districts. First, 72 EAs are sampled at the first phase of the first stage of sampling using systematic PPS sampling, and then 66 of the 72 EAs are subsampled at the second phase of the first stage of sampling using fractional interval systematic sampling. Although computing the probabilities of selection in the set of strata with original districts and new districts is straightforward, computing the probabilities of selection in the strata with common districts requires special consideration. The probability of selection at the first phase of the first stage is the same as above, namely, $(\frac{m_h * N_{hi}}{N_h})$. In the above example, using the values from Table 9 in Section 4.4, $m_h = 72$ and $N_h = 4,804 + 3,769 = 8,573$, but the value for N_{hi} is not provided in the table. Since the 72 sampled EAs are used to determine the results of the end-line PBS, this is the first stage selection probability for this sub-survey. The probabilities of selection for the second phase of the first stage of sampling (described next) are not used for the end-line PBS.

Since the EAs at the second phase of the first stage are selected using fractional interval systematic sampling, the probability of selection at that phase is given by:

$$\frac{\text{number of EAs subsampled at the second phase (from the EAs sampled at the first phase) in stratum } h}{\text{number of EAs sampled at the first phase in stratum } h} = \frac{m_h^*}{m_h}$$

In the above example, $m_h = 72$ and $m_h^* = 66$.

Combining the phase one and phase two selection probabilities using multiplication, a combined first stage selection probability is given by:

$$f_{1hi} = \frac{m_h * N_{hi}}{N_h} * \frac{m_h^*}{m_h} = \frac{m_h^* * N_{hi}}{N_h}$$

Since, in the example above, the 66 sampled EAs are used to determine the results of the baseline PBS, the above formula is the final first stage selection probability for this sub-survey. In contrast to the end-line sub-survey, the probabilities of selection from both phases of sampling are used for the baseline sub-survey.

Special Scenario of Inaccessible EAs

In Section 5.2.2 and Section 8.4, a scenario is described where, after drawing a first stage systematic PPS sample of EAs, survey implementers discover while in the field or about to undertake fieldwork that some of the EAs are inaccessible. An example is given where 111 EAs are to be sampled in a PBS, and, before fieldwork commences, it is suspected that up to 6 EAs may have access issues. The recommendation is to adopt a “two-phase” approach at the first stage of sampling whereby 117 EAs are selected using systematic PPS sampling at the first phase of the first stage. At a second phase of the first stage, 6 of 117 EAs are subsampled using fractional interval systematic sampling to create a reserve sample. Interviewing is conducted in the original 111 EAs that are not among the 6 subsampled EAs. If no reserve EAs are released by the end of the survey (because there wind up being no inaccessible EAs in the original sample of 111 EAs), then the selection probabilities from the first phase of the first stage are adjusted by multiplying them by 111 / 117, because only 111 of the 117 sampled EAs have been

released. Similarly, if two reserve EAs are released by the end of the survey (because there are two inaccessible EAs), then the EA-level selection probabilities for the 113 EAs where interviewing took place are adjusted by 113 / 117, because 113 of the 117 sampled EAs have been released overall. It is important to understand that in this second case, it is a different set of 113 EAs from those originally sampled that will have their weights adjusted, because two EAs from the original 113 EAs have been dropped and replaced by two others.

To obtain a more general representation of the selection probabilities at both phases of the first stage sampling of EAs under this scenario, it can be assumed more generally that EAs are selected within strata. Then, as given above, the selection probability at the first phase of the first stage is given by $\frac{m_h * N_{hi}}{N_h}$. Assume the “reserve EAs” at the second phase of the first stage are selected using fractional interval systematic sampling within each stratum. If it is assumed that, overall, m_h^{**} EAs are released for sampling, the selection probability at the second phase of the first stage is given by:

$$\frac{\text{number of EAs released for interviewing in stratum } h}{\text{total number of EAs sampled at the first phase in stratum } h} = \frac{m_h^{**}}{m_h}$$

In the above example, it is assumed that prior to fieldwork there may be up to six EAs with access issues, but during fieldwork it is discovered that only two EAs have access issues, and therefore two of the six reserve EAs are released to replace two EAs that are inaccessible. Then $m_h = 117$ and $m_h^{**} = 113$. This selection probability should be computed only in those strata where reserve EAs were released.

Multiplying the phase one and phase two selection probabilities together, a combined first stage selection probability is given by:

$$f_{1hi} = \frac{m_h * N_{hi}}{N_h} * \frac{m_h^{**}}{m_h} = \frac{m_h^{**} * N_{hi}}{N_h}$$

Finally, in the example above, although 113 EAs have been released overall, 2 were dropped because they were inaccessible, and interviewing took place in only 111 EAs (as originally planned). Therefore, it is possible to make an EA-level non-response adjustment to account for the two dropped inaccessible EAs. An alternative approach is to undertake no further adjustments and to simply assume that any inference that is made reflects the total population excluding the two inaccessible EAs. Feed the Future recommends adopting the latter approach.

10.1.3 Calculating the Probability of Selection at the Second Stage Sampling of Segments

For the second stage of sampling, when **PPS** is used to sample one segment per segmented EA (as discussed in Chapter 7), the probability of selecting the j th segment in sampled EA i in stratum h is calculated as follows:

$$f_{2hij} = \frac{\text{total number of households in selected segment } j \text{ selected in EA } i \text{ in stratum } h}{\text{total number of households in selected EA } i \text{ in stratum } h} = \frac{N_{hij}}{N_{hi}}$$

where N_{hij} is the total number of households listed in selected segment j in EA i and stratum h .

To illustrate this, we continue the example at the end of Chapter 7 where segment 2 has been selected from an EA that has been divided into three segments. In this case, since $N_{hij} = 200$ households and $N_{hi} = 610$ households, then the selection probability is $f_{2hij} = \frac{200}{610} = 0.327$. If no segmentation takes place in EA i , then $f_{2hij} = 1$.

10.1.4 Calculating the Probability of Selection at the Third Stage Sampling of Households

For the third stage of sampling, when **fractional interval systematic sampling** is used to sample households (as discussed in Section 8.1), the probability of selecting the k th household in the j th sampled segment in the i th sampled EA in stratum h is calculated as follows:

$$f_{3hijk} = \frac{\text{number of households selected for sampling in segment } j \text{ in EA } i \text{ in stratum } h}{\text{total number of households listed in segment } j \text{ in EA } i \text{ in stratum } h} = \frac{n_{hij}}{N_{hij}}$$

where n_{hij} is the number of households selected for sampling in segment j in EA i in stratum h .

It is important to understand that the number of households selected for sampling in segment j in EA i and stratum h , denoted by n_{hij} , is the same as the number of households selected for sampling in EA i , and stratum h , denoted by n_{hi} (which, in previous examples in this guide, has been 25 households). This is because exactly one segment is randomly selected and all of the households to be sampled are within the selected segment. Therefore, $n_{hij} = n_{hi}$.

Finally, if there is no second stage of sampling (i.e., there is no segmentation within sampled EA i), then the subscript j is dropped and $n_{hij} = n_{hi}$ and $N_{hij} = N_{hi}$. In this case, the above formula reduces to:

$$f_{3hijk} = \frac{n_{hi}}{N_{hi}}$$

The calculation of the probabilities of selection at the third stage is illustrated below, continuing the example from Section 8.1, assuming no segmentation at the second stage of sampling.

Total number of households in stratum h and EA i	N_{hi}	48					
Number of households to select in stratum h and EA i	n_{hi}	25					
Sampling interval	$k_{hi} = N_{hi}/n_{hi}$	1.92					
Random number	rand()	0.3146					
Random start	$RS = \text{rand()} * k_{hi}$	0.604					

				Round up	Household number	Probability of Selection (Third Stage)	
1ST HOUSEHOLD TO SELECT	$a_1 = RS$	$= 0.604 =$	0.604	→ 1	1	0.521	$= 25/48$
					2		
2ND HOUSEHOLD TO SELECT	$a_2 = RS + k_{hi}$	$= 0.604 + 1.92 =$	2.52	→ 3	3	0.521	$= 25/48$
					4		
3RD HOUSEHOLD TO SELECT	$a_3 = RS + 2 * k_{hi}$	$= 0.604 + (2 * 1.92) =$	4.44	→ 5	5	0.521	$= 25/48$
					6		
...		
					46		
25TH HOUSEHOLD TO SELECT	$a_{25} = RS + 24 * k_{hi}$	$= 0.604 + (24 * 1.92) =$	46.68	→ 47	47	0.521	$= 25/48$
					48		

10.1.5 Calculating the Probability of Selection at the Fourth Stage Sampling of Individuals

For the fourth stage of sampling, when **take all sampling** is used to sample individuals within sampled households (as discussed in Chapter 9), the probability of selecting the *l*th individual in the *k*th sampled household in the *j*th sampled segment in the *i*th sampled EA in stratum *h* is calculated as:

$$f_{4hijkl} = 1$$

For take all sampling, because the number of individuals selected for sampling in any household is always the same as the number of eligible individuals in that household, f_{4hijkl} always equals 1.

10.1.6 Calculating the Overall Probability of Selection

As noted in step 5 of Section 10.1.1, once the probability of selection at all four stages of sampling is calculated, the overall probability of selection for an **individual** in the sample can be calculated by multiplying the probability of selection at all four stages together.

$$f_{hijkl} = f_{1hi} * f_{2hij} * f_{3hijk} * f_{4hijkl}$$

Similarly, by applying the first three stages of sampling only, the overall probability of selecting a **household** can be expressed as:

$$f_{hijk} = f_{1hi} * f_{2hij} * f_{3hijk}$$

This overall probability of selecting a household is the same as the overall probability of selecting an individual because $f_{4hijkl} = 1$.

The following illustrates the calculation of the overall probabilities of selection, continuing the example above where systematic PPS sampling is used at the first stage of sampling, it is assumed that there is no segmentation at the second stage of sampling, fractional interval systematic sampling is used at the third stage of sampling, and take all sampling is used at the fourth stage of sampling. The calculation is performed for one of the sampled first stage EAs (Pancho) in the stratum of Ciril only. In the illustration, probabilities of selection are given for all selected households in the EA (Pancho), but since all eligible individuals within sampled households are selected for interviewing (i.e., $f_{4hijkl} = 1$), the probabilities of selection for individuals are the same as the probabilities of selection for households.

		Number of EAs to select in stratum Ciril							
		m_h	14						
		N_h	2,289						
				$f_{1hi} = (m_h * N_{hi}) / N_h$	$f_{2hij} = 1$	$f_{3hijk} = n_{hi} / N_{hi}$	$f_{4hijkl} = 1$	$f_{hijkl} = f_{1hi} * f_{2hij} * f_{3hijk} * f_{4hijkl}$	
EA number	Cluster name	Number of Households (N_{hi})	Number of Households to Select (n_{hi})	Probability of selection (first stage)	No segmentation (second stage)	Household Number	Probability of selection (third stage)	Probability of selection (fourth stage)	Probability of selection (overall)
3	Pancho	48	25	0.2930	1	1	0.521	1	0.1527
						3	0.521	1	0.1527
						5	0.521	1	0.1527
					
						47	0.521	1	0.1527

This section finishes with a discussion of the concept of a “self-weighting” sample and why it is important. But first, it is important to understand what self-weighting is. When **systematic PPS sampling** is used in the first stage, **PPS** is used at the second stage, **fractional interval systematic sampling** is used at the third stage, and **take all sampling** is used at the fourth stage, then, using the formula above for the overall probability of selection and plugging in the values at each stage, the following is obtained:

$$f_{hijkl} = \frac{m_h * N_{hi}}{N_h} * \frac{N_{hij}}{N_{hi}} * \frac{n_{hij}}{N_{hij}} * 1 = \frac{m_h * n_{hij}}{N_h}$$

As can be seen, the terms N_{hi} and N_{hij} both cancel out from the numerator and the denominator, simplifying the above expression. Because the number of households to be sampled within each segment (or EA), denoted by n_{hij} , is typically a constant number, such as 25, and is the same for every segment (or EA) within a stratum, then $n_{hij} = n_h$, and the above expression for the overall probability of

selection of an individual (or a household) depends only on the subscript h and is constant within each stratum h . Therefore, in the case where the same number of households is selected in each EA (e.g., 25) across all selected EAs in the stratum, the above expression reduces to:

$$f_{hijkl} = f_h = \frac{m_h * n_h}{N_h}$$

This is an illustration of what is commonly referred to as a “self-weighting design,” where the overall probability of selection is a constant within each stratum h . Self-weighting is preserved whether or not there is segmentation at the second stage of sampling; furthermore, even if an EA is divided into segments of somewhat differing sizes, self-weighting will be preserved, provided PPS sampling (rather than SRS sampling) is used to select one segment. Such a design is useful because one of the critical features of self-weighting—insisting that the number of households to be sampled within each segment (or EA) is a constant number (such as 25), rather than a proportional fraction of the EA—ensures roughly equal interviewer workloads within selected EAs or segments. Another advantage to a self-weighting design is that it diminishes the overall DEFF (which is a component of the SE) associated with the estimates of indicators—in particular, the contribution to the DEFF related to unequal weighting. Constant weights within a given stratum minimizes what is called “the unequal weighting design effect.” A larger unequal weighting design effect reduces the precision (i.e., increases the SE) of an estimate.

It is important to acknowledge that, although self-weighting implies that selection probabilities are constant in principle, in practice they may vary somewhat. For instance, it may be the case that the cancellation of N_{hi} from the numerator and denominator of the above expression does not truly happen in practice because N_{hi} used at the first stage sampling of EAs is based on figures from the census or other administrative sources, while N_{hi} from the second stage sampling of segments is based on figures from the listing exercise. These two figures may and usually do differ, sometimes significantly, particularly when there has been substantial growth or reduction in the size of the EA between the time of the previous census and the listing exercise. Regardless, a self-weighting design helps keep the weights as close to constant as possible within a stratum.

A second way in which self-weighting can potentially be compromised is through non-response adjustments (see Section 10.2). Although this guide recommends applying both household and individual-level non-response adjustments at the stratum level to preserve self-weighting, some implementers may choose to do so at lower levels of geography if they believe that non-response differs across those lower levels of geography—and this can compromise self-weighting.⁷⁵

In summary, contrary to what is often popular practice, self-weighting is **not** a license to dispense with sample weighting at the time of analysis, because ultimately a design that is self-weighted in theory may be compromised to some degree by realities in the field that need to be accounted for. Indeed, the sampling weights should always be integrated into the data analyses and estimation procedures. This is discussed further in Chapters 11 and 12.

⁷⁵ Self-weighting can also be compromised if a design is adopted whereby one (rather than all) eligible members of a household is randomly selected using a Kish Grid or some other selection mechanism (not recommended in this guide). This is true because the additional step of selection of individuals within households is done at a non-constant rate with unequal probabilities.

10.1.7 Calculating the Sampling Weights to Account for Probabilities of Selection

At the final step, the sampling weights to account for the probabilities of selection at all stages of sampling are calculated by taking the inverse of the total probability of selection. The formula is given by:

$$w_{hijkl} = w_h = \frac{1}{f_h} = \frac{1}{f_{1hi} * f_{2hij} * f_{3hijk} * f_{4hijkl}} = \frac{N_h}{m_h * n_h}$$

The following illustration demonstrates the computation of the sample weights, continuing the example above using **systematic PPS sampling** at the first stage of sampling, assuming no segmentation at the second stage of sampling, using **fractional interval systematic sampling** at the third stage of sampling, and using **take all sampling** at the fourth stage of sampling. Once again, the illustration focuses on the sampled EA of Pancho within the stratum of Ciril, and, once again, the sampling weights for all individuals selected within households are the same as those for the selected households themselves.

		Number of EAs to select in stratum Ciril									
		m_h	14								
		Total number of households across stratum Ciril									
		N_h	2,289								
EA number	Cluster name	Number of Households (N_h)	Number of Households to Select (n_h)	Probability of selection (first stage)	No segmentation (second stage)	Household Number	Probability of selection (third stage)	Probability of selection (fourth stage)	Probability of selection (overall)	Sampling Weight	
3	Pancho	48	25	0.2930	1	1	0.521	1	0.1527	6.55	
						3	0.521	1	0.1527	6.55	
						5	0.521	1	0.1527	6.55	
						
						47	0.521	1	0.1527	6.55	

Alternatively, for this example, the sampling weights for individuals can be computed directly from the formula above as:

$$w_{hijkl} = w_h = \frac{N_h}{m_h * n_h} = \frac{2,289}{14 * 25} = 6.55$$

10.2 Adjusting Survey Weights for Household and Individual Non-Response

It is to be expected that some proportion of households and/or individuals randomly selected for the survey will be unreachable, unavailable, or unwilling to respond to the survey questions. When entire households do not respond, this is called “household non-response”; when one or more eligible individuals within a sampled household do not respond, this is called “individual non-response.” The recommended survey protocol is that interviewers return to households up to three times to complete interviews with the selected individuals who reside within the sampled households. Despite the best efforts of interviewers, however, there is frequently some non-response that remains even after three attempts to complete interviews. When non-response happens, adjustments to the sample weights need to be applied to compensate for the non-response at both levels.⁷⁶

10.2.1 Adjusting for Household Non-Response

To calculate the weight adjustments for household non-response, the survey must track both the selected households that do not respond and the selected households that do respond. Both responding and non-responding households have probabilities of selection. However, since no interview has taken place for the non-responding selected households, the sample weights of the responding selected households are adjusted to compensate for those that did not respond.

The weight adjustment for household non-response is calculated as:

$$w_{HH_NR_h} = \frac{\text{sum of } w_h \text{ over all households selected to be interviewed in stratum } h}{\text{sum of } w_h \text{ over all households responding in stratum } h}$$

where w_h is the sampling weight across all stages of sampling given in Section 10.1.7. The above formula implies that a weight adjustment for household non-response should be calculated separately for each stratum. The adjustment is computed at the stratum level to preserve self-weighting.⁷⁷ (See Section 10.1.6 for more details.)

For instance, Table 8 in Section 4.3 provides an example of a survey with six strata. The weight adjustments for household non-response will vary among strata given that strata will likely experience different household non-response rates. However, for all sampled households in a particular stratum, the same weight adjustment for household non-response can be used. After the weight adjustment is

⁷⁶ Sometimes a sampled household or individual may respond to most but not all questions on the questionnaire. In this case, the household or individual is deemed a “partial respondent” and the missing data points are called “item non-responses.” One approach to solving this issue is to “impute” (using special statistical methods) the missing data points related to key indicators for a household or individual (such as “age” as a component part of the underweight indicator, for children under 5 years of age). However, another common practice is to leave the missing data points blank and to compute the indicators without the inputs from the missing respondent(s). Since a discussion on methods of imputation is beyond the scope of this guide, the latter strategy of leaving missing data points blank should be adopted for Feed the Future PBSs, and no weight adjustment should be made to compensate for item non-response.

⁷⁷ In principle, non-response adjustments can be made at levels that differ from the stratum; in fact, they need not even be geographic in nature. Such levels are called “weighting classes” and are determined on the basis that units respond differently across these classes. For instance, it may be believed that household non-response differs by gendered household type. In this case, the four gendered household types would form the weighting classes and household non-response adjustments could be made separately within each of the four gendered household types. However, this choice of weighting class would compromise self-weighting.

made, the empty records for the non-responding sampled households can then be dropped for the purposes of analysis, although they should be retained in the database to keep track of all households that were initially sampled.

To provide an empirical illustration, we continue the example used throughout the guide based on the stratum of Ciril from Table 8 in Section 4.3 and Example 1 in Section 5.1, where 350 households are to be sampled (14 EAs each with 25 households), and assuming that 20 households do not respond across the entire stratum. Finally, we assume that there is a constant sampling weight of 6.55 across the entire stratum of Ciril, as in the example in Section 10.1.7.

Number of households in stratum h	N_h	2,289
Number of households selected for interviewing in stratum h	n_h	350
Number of households that do not respond in stratum h (not found, not present, or who refuse)	NR	20
Number of households responding in stratum h	$n_h - NR$	330
Sampling weight across all stages in stratum h	w_h	6.55
Weight adjustment to compensate for household non-response for stratum h	$w_{HH_NR_h} = \frac{(n_h * w_h)}{((n_h - NR) * w_h)}$	1.061

The household-level non-response adjustment should be taken into account in the computation of all the indicators given in Tables 1 and 2, even the individual-level indicators, since household-level weighting comprises part of the overall weights for individuals and households alike.

10.2.2 Adjusting for Individual Non-Response

Individual non-response occurs when an eligible individual within a sampled household is not available for, refuses, or does not respond to a request for an interview. For individual non-response, a separate adjustment must be made for each sampling group to which the individual belongs, corresponding to the relevant associated indicators. For instance, in Table 1, the collection of Feed the Future ZOI PBS indicators span eight sampling groups: children age 0–5 months, children age 6–23 months, children age 0–59 months, non-pregnant women age 15–49 years, women age 15–49 years, primary female and male decision makers in the household, producers, and the household itself. Suppose that the caregiver of a specific child age 7 months does not respond to any modules of the questionnaire for which the child is eligible. In this case, two sampling groups are affected by the missing data for that child (children age 6–23 months and children age 0–59 months). The indicators in Table 1 implicated by these sampling groups are: MAD, stunting, healthy weight, and wasting. That means that the computation of these indicators, which involves sampling weights for these sampling groups, will need to take into account the individual non-response adjustment relating to this missing child, as well as all other children whose data are missing for the associated sampling groups. More generally, in a PBS where data are collected in support of all indicators in Table 1, there are potentially seven separate individual non-response adjustments

(excluding the sampling group consisting of the household itself), assuming one or more individuals belonging to these sampling groups do not respond to a request for an interview.

To calculate the weight adjustments for individual non-response for any of the sampling groups in question, the survey must track both the selected individuals who do not respond and the selected individuals who do respond. Both respondent and non-respondent individuals have probabilities of selection. But because no interview takes place for the non-responding selected individuals, the sample weights of the responding selected individuals are adjusted to compensate for those that did not respond.

The weight adjustment for individual non-response for a specific sampling group is calculated as:

$$w_{IN_NR_h} = \frac{\text{sum of } w_h \text{ over all individuals from sampling group selected to be interviewed in stratum } h}{\text{sum of } w_h \text{ over all individuals from sampling group responding in stratum } h}$$

where w_h is the sampling weight across all stages of sampling given in Section 10.1.7. The above formula implies that, similar to the case of household non-response adjustments, a weight adjustment for individual non-response should be calculated separately for each stratum. The adjustment is computed at the stratum level to preserve self-weighting. (See Section 10.1.6 for more details.) For all sampled individuals belonging to a sampling group in a specific stratum, the same weight adjustment for individual non-response can be used. After the weight adjustment is made, the empty records for the non-responding sampled individuals can then be dropped for the purposes of analysis, although they should be retained in the database to keep track of all individuals that were initially sampled. This process should be repeated for each sampling group where there has been individual-level non-response.

The individual-level non-response adjustment should be taken into account in the computation of all individual-level indicators given in Tables 1 and 2 only, since household-level indicators do not require this adjustment and require only the household-level non-response adjustment.

Given that the computation for household and individual non-response adjustments are very similar and an empirical example of household non-response was provided earlier, a related example of the individual non-response adjustment is not provided.

10.3 Calculating the Final Sampling Weights

The final sampling weights to be used in all data analysis are calculated differently, depending on whether the indicators to which they will be applied have underlying sampling groups at the individual level or the household level. For instance, in Table 1, there are 20 indicators and the last 10 are defined at the individual level. For such **individual**-level indicators, the final weights (computed within each stratum separately) are formed by multiplying the sampling weights for all four stages by the weight adjustment for non-response at both the household and individual levels:

$$w_{final,INDIV,h} = w_h * w_{HH_NR_h} * w_{IN_NR_h}$$

In Table 1, the first 10 indicators are defined at the household level. For such **household**-level indicators, the final weights (computed within each stratum separately) are formed by multiplying the sampling weights for the first three stages (which is the same as the sampling weight for all four stages)

by the weight adjustment for non-response at the household level only, as the adjustment at the individual level is not required:

$$W_{final,HH,h} = w_h * W_{HH_NR_h}$$

As mentioned earlier, because (in theory) the design is self-weighting, the sampling weight for all four stages, w_h , is a constant at the stratum level. Since the non-response adjustments at both the household and individual levels are also made at the stratum level (i.e., the same adjustment is applied to all sampled individuals and households in a given stratum), then self-weighting is preserved at the stratum level, even after one or both non-response adjustments are made.

II. Data Analysis for Descriptive Surveys: Producing Single-Point-in-Time Estimates of Indicators along with Their Standard Errors and Confidence Intervals

The main purpose of descriptive surveys is to produce high-precision single-point-in-time estimates of indicators of proportions and means. In the previous chapter, procedures for producing final sampling weights to be used in data analysis were described. For a descriptive survey, the next step is to produce estimates, along with their SEs and CIs, for the indicators in Tables 1 and 2, integrating the final sampling weights.

II.1 Producing Estimates of the Indicators

The aim of descriptive PBSs is to facilitate the production of estimates that represent the entire population, not just the individuals and households interviewed in the survey sample. To do so, the sampling weights attached to data points are used to “inflate” the data from each of the sampled units (individuals or households) that respond, so that a sample-weighted representation of the data from the surveyed units provides an estimate of the proportion or mean (depending on the indicator in question) for the entire population of units. Estimates of all indicators given in Tables 1 and 2 can be produced this way. In Table 1, all indicators except four (i.e., “Ability to Recover from Shocks and Stresses Index,” “Index of Social Capital at the Household Level,” A-WEAI, and “Yield of Targeted Agricultural Commodities within Target Areas”) are proportions. Of those that are proportions, eight are at the individual level and eight are at the household level.

The production of SEs and CIs of the estimates must take into account the complex survey design. Therefore, survey implementers must use a statistical software package, such as SAS, SPSS, or STATA, to generate the estimates and their associated SEs and CIs that appropriately take into account the complex survey design. More detail on this is provided in Section II.2. However, to provide the reader with a sense of how indicator estimates are computed, the formulas for estimates of indicators of proportions and means are provided here. Formulas for estimates of indicators that are indexes are not provided here, as each such indicator of indexes has its own specific formula.

The formula for an estimate of an indicator at the **individual** level that is a **proportion** (such as “Prevalence of Stunted Children under Five”) is:

$$\begin{aligned} \text{estimate of indicator of proportion} &= P_{\text{actual}} \\ &= \sum (w_{\text{final,INDIV},h} * y_{ijkl}) / \sum (w_{\text{final,INDIV},h} * z_{ijkl}) \end{aligned}$$

where:

$y_{ijkl} = 1$ if the sampled individual l is in the sampling group corresponding to the indicator and has the attribute of the indicator (e.g., a sampled child 0–59 months who is stunted); 0 otherwise

$z_{ijkl} = 1$ if the sampled individual l is in the sampling group corresponding to the indicator (e.g., a sampled child 0–59 months); 0 otherwise

Σ refers to the sum overall the entire sample (i.e., across all strata h , all EAs i , all segments j , all households k , and all individuals l)

Note that for a sample weighted estimate of a proportion, weights contribute to both the numerator and the denominator of the estimate, as can be seen in the formula above.

The formula for an estimate of an indicator at the **household** level that is a **proportion** (such as PP) is:

$$\begin{aligned} \text{estimate of indicator of proportion} &= P_{\text{actual}} \\ &= \sum (w_{\text{final,HH},h} * y_{hijk}) / \sum (w_{\text{final,HH},h}) \end{aligned}$$

where

$y_{hijk} = 1$ if the sampled household k has the attribute of the indicator (e.g., a sampled household that is in poverty); 0 otherwise

Σ refers to the sum overall the entire sample (i.e., across all strata h , all EAs i , all segments j , and all households k)

Unlike indicators at the individual level, z_{hijk} does not appear in the denominator of the formula because every household is involved in the computation of all household-level indicators and so $z_{hijk} = 1$.

Feed the Future survey implementers are also responsible for producing estimates of disaggregates for all Feed the Future PBS indicators (see the list of disaggregates required in Tables 1 and 2). For example, if an indicator requires disaggregate estimates by sex (male versus female), then the computation is slightly different from what is given above. For instance, to obtain an estimate for stunting for females, the above formula should be used, but the definition of “sampling group” changes somewhat for the numerator and denominator.⁷⁸ That is:

$y_{hijkl} = 1$ if the sampled individual l is in the sampling group corresponding to the **disaggregated** indicator and has the attribute of the indicator (e.g., a sampled **female** child 0–59 months who is stunted); 0 otherwise

$z_{hijkl} = 1$ if the sampled individual l is in the sampling group corresponding to the **disaggregated** indicator (e.g., a sampled **female** child 0–59 months); 0 otherwise

Similarly, the formula for an estimate of an indicator at the **individual** level that is a **mean** (such as “Yield of Targeted Agricultural Commodities within Target Areas”) is:

$$\text{estimate of indicator of mean} = \bar{X}_{\text{actual}}$$

⁷⁸ Technically speaking, to compute disaggregates, the denominators should not change from the definition at the overall level. However, the “Feed the Future Indicator Handbook: Definition Sheets,” which is located at https://feedthefuture.gov/sites/default/files/resource/files/Feed_the_Future_Indicator_Handbook_Sept2016.pdf, has defined disaggregation in a way that corresponds to the above computation. The net result is that the estimates of the proportions for the disaggregated categories (e.g., male and female) will not add up to the estimate of the overall proportion.

$$= \sum (w_{final,INDIV,h} * x_{hijkl}) / \sum (w_{final,INDIV,h} * z_{hijkl})$$

where:

x_{hijkl} = the individual value of the indicator for sampled individual l in the sampling group corresponding to the indicator (e.g., yield of targeted agricultural commodities within target areas for a sampled producer)

z_{hijkl} = 1 if the sampled individual l is in the sampling group corresponding to the indicator (e.g., a sampled producer); 0 otherwise

Note the subtle difference in the estimator of a proportion and the estimator of a mean. For the latter, the numerator includes quantitative values corresponding to each sampled individual (e.g., yield of a targeted commodity in metric tons per hectare), whereas for the former, the numerator includes simply a value of 1, indicating membership to a sampling group with a specific attribute (e.g., a 1 if the individual is deemed stunted).

Finally, the formula for an estimate of an indicator at the **household** level that is a **mean** (such as PCE from Table 2) is:

$$\begin{aligned} \text{estimate of indicator of mean} &= \bar{X}_{actual} \\ &= \sum (w_{final,HH,h} * x_{hijk}) / \sum (w_{final,HH,h}) \end{aligned}$$

where:

x_{hijk} = the individual value of the indicator for sampled household k (e.g., daily expenditure for a sampled household)

11.2 Producing Confidence Intervals and Standard Errors Associated with the Estimates of the Indicators

An important step in data analysis is to calculate CIs and SEs for all estimates from the data collected through a PBS. A CI is a measure of the precision of an estimate of an indicator and is expressed as a range of numbers that have a specific interpretation. The formula for CIs involves the SE of the estimate. A SE quantifies how precisely the true population value of the proportion or mean is known, and is computed as the standard deviation divided by the square root of the sample size.⁷⁹ In the context of descriptive surveys, both CIs and SEs should be produced and reported, to provide measures of precision of the indicator estimates. More detailed interpretations of SEs and CIs are provided in Section 11.2.4.

⁷⁹ A distinction should be made between the interpretation and use of a standard deviation of a sampling distribution versus the SE of a sample estimate. The standard deviation of a sampling distribution describes how individual or household data points vary from one another across the distribution of individual or household values in the sample. In effect, the standard deviation is a measure of scatter of the sampling distribution of individual or household data points. It is a descriptive statistic in relation to the sampling distribution and is not used to describe the precision of the estimate. On the other hand, the SE of the sample estimate provides a measure of precision for the estimate (of an indicator) and is a companion measure to the CI. The SE describes the amount of fluctuation from the true population value that we can expect in sample estimates.

Survey implementers must use a specialized statistical software package that can take into account the complex design features of PBSs, such as clustering and unequal probabilities of selection, to generate the estimates of CIs and SEs. The most widely used statistical software packages are SAS, SPSS, and STATA.⁸⁰ Each of these packages has its own specialized syntax for entering information on complex survey design features (such as clustering and sampling weights) that permits the production of survey-based estimates of proportions or means, along with their associated CIs and SEs. It is critical that the correct syntax for complex survey designs be used, and therefore users should familiarize themselves with such software before undertaking any data analysis. See Table 12 for details on some statistical software packages that take into account complex survey designs.⁸¹ Further details on the syntax for these packages can be found in Appendix B.

Table 12. Statistical Software Packages for the Analysis of Complex Survey Data

Statistical Software Package	For Analyses of Complex Survey Data, Use...
SAS	Specialized survey procedures
SPSS	SPSS Complex Samples module
STATA	svy commands

11.2.1 Calculating Confidence Intervals and Standard Errors Associated with Estimates of Indicators of Proportions

Although CIs and SEs associated with the estimates of indicators of proportions must be computed using appropriate specialized statistical software, the formulas are provided here to convey a sense of how they are computed.

The formula to calculate the SE associated with the estimate of a proportion, P_{actual} , is:

$$SE(P_{actual}) = \left(\frac{\sqrt{D_{actual}} * \sqrt{P_{actual} * (1 - P_{actual})}}{\sqrt{n_{actual}}} \right)$$

The formula to calculate a CI with a confidence level of 95% for the estimate of a proportion (denoted by P_{actual}) involves the SE of the estimate, and is given by the following:

$$CI_{proportion} = P_{actual} \pm (z_{1-\alpha/2} * SE(P_{actual})) = P_{actual} \pm \left(z_{1-\alpha/2} * \left(\frac{\sqrt{D_{actual}} * \sqrt{P_{actual} * (1 - P_{actual})}}{\sqrt{n_{actual}}} \right) \right)$$

where:

P_{actual} = the sample-weighted estimate of the proportion computed from the survey data (discussed in Section 11.1)

⁸⁰ Estimates of indicators and their associated CIs and SEs should never be computed using formulas found in spreadsheet software, such as Microsoft Excel, as these do not appropriately account for complex survey designs.

⁸¹ For a reference with detailed information on how to conduct data analysis on data from complex survey designs, see: Heeringa, Steven G.; West, Brady T.; and Berglund, Patricia A. 2010. *Applied Survey Data Analysis*. Boca Raton, FL: CRC Press.

$z_{1-\alpha/2}$ = the critical value from the Normal Probability Distribution

D_{actual} = the DEFF corresponding to the indicator, computed by statistical software using the survey data⁸²

n_{actual} = the actual sample size realized after fieldwork

For a confidence level of 95% (with $\alpha = 0.05$), the corresponding critical value, $z_{1-\alpha/2}$, is equal to 1.96 (see Table 4a). Survey implementers should use a confidence level of 95% (and a critical value of 1.96) for calculating CIs, although values for critical value based on other confidence levels can be found from tables, statistical software, and spreadsheet software (such as Microsoft Excel).

In terms of the DEFF corresponding to the indicator, recall that an estimate of the design effect, D_{est} , is used in the calculation of the target sample size ($n_{initial}$) prior to survey implementation, as discussed in Section 2.3.1. In contrast, the DEFF that should be used in the computation of the CI in the formula above is one that is computed by the statistical software using data from the fieldwork corresponding to the indicator in question and is denoted by D_{actual} .

In the above formula, n_{actual} represents the actual sample size realized after fieldwork. The value of n_{actual} may represent the actual number of households interviewed (in the case of a CI corresponding to an indicator at the household level, such as PP) or the actual number of individuals interviewed (in the case of a CI corresponding to an indicator at the individual level, such as “Prevalence of Stunted Children under Five”). In the case where the indicator is at the individual level, n_{actual} is the number of individuals in the appropriate sampling group (e.g., children age 0–59 months) for whom data were collected.

11.2.2 Calculating Confidence Intervals and Standard Errors Associated with Estimates of Indicators of Means

Although CIs and SEs associated with the estimates of indicators of means must be computed using appropriate specialized statistical software, the formulas are provided here to provide the reader with a sense of how they are computed.

The formula to calculate the SE associated with the estimate of a mean, \bar{X}_{actual} , is:

$$SE(\bar{X}_{actual}) = \left(\frac{\sqrt{D_{actual}} * \sigma_{X_{actual}}}{\sqrt{n_{actual}}} \right).$$

The formula to calculate a CI with a confidence level of 95% for the estimate of a mean (denoted by \bar{X}_{actual}) is:

⁸² The DEFF represents the ratio of the statistical variance under the current multi-stage cluster sampling design to the statistical variance under SRS. To properly compute D_{actual} from the survey data, the numerator of the ratio should take into account all sources of variation under the complex design including: stratification, allocation, clustering and any unequal probability sample weighting.

$$CI_{mean} = \bar{X}_{actual} \pm (z_{1-\alpha/2} * SE(\bar{X}_{actual})) = \bar{X}_{actual} \pm \left(z_{1-\alpha/2} * \left(\frac{\sqrt{D_{actual}} * \sigma_{X_{actual}}}{\sqrt{n_{actual}}} \right) \right)$$

where:

\bar{X}_{actual} = the sample-weighted estimate of the mean computed from the survey data (discussed in Section 11.1)

$z_{1-\alpha/2}$ = the critical value from the Normal Probability Distribution

D_{actual} = the DEFF corresponding to the indicator, computed by statistical software using the survey data

$\sigma_{X_{actual}}$ = the standard deviation of X_{actual} computed from the survey data

n_{actual} = the actual sample size realized after fieldwork

In terms of the standard deviation of the distribution, an estimate, $\sigma_{X_{est}}$, used in the calculation of the sample size ($n_{initial}$) prior to survey implementation, is discussed in Section 2.3.2. In contrast, the standard deviation that should be used in the computation of the CI in the formula above is one that is computed by the statistical software using data from the fieldwork corresponding to the indicator in question and is denoted by $\sigma_{X_{actual}}$. All other input parameters (D_{actual} , n_{actual} , and $z_{1-\alpha/2}$) are defined similarly to those described in Section 11.2.1.

11.2.3 An Example of Calculating a Confidence Interval and a Standard Error for an Estimate of an Indicator of Proportions

Although CIs and SEs associated with the estimates of indicators of means must be computed using appropriate specialized statistical software, we provide an example of a manual computation here to convey to the reader a sense of what actual values might look like.

The example below illustrates the computation of a CI and a SE for an estimate of the “Prevalence of Stunted Children under Five” indicator. Suppose that data from a descriptive PBS have been collected using the minimum recommended sample size of 1,000 responding households (as per the guidance in Section 2.3.3), which has resulted in an actual sample size of $n_{actual} = 800$ children age 0–59 months. The collected data are used to compute a sample-weighted estimate of the proportion ($P_{actual} = 0.375$). The DEFF computed from the data is $D_{actual} = 2.5$. A 95% CI around $P_{actual} = 0.375$ is then given by (0.322, 0.428) and the SE of the estimate is computed as 0.027.

Survey-weighted estimate of Prevalence of Stunted Children	P_{actual}	0.375 = <i>derived from survey data</i>
Actual sample size	n_{actual}	800 = <i>derived from survey data</i>
Confidence level	$1-\alpha$	0.95
Z statistic	$Z_{1-\alpha/2}$	1.96
Actual design effect	D_{actual}	2.5 = <i>derived from survey data</i>
CONFIDENCE INTERVAL(CI) AND STANDARD ERROR		
CI: Lower limit = LL	$0.322 = P_{actual} - \left(z_{1-\alpha/2} * \left(\frac{\sqrt{D_{actual}} * \sqrt{P_{actual} * (1 - P_{actual})}}{\sqrt{n_{actual}}} \right) \right)$	$= 0.375 - \left(1.96 * \left(\frac{\sqrt{2.5} * \sqrt{0.375 * (1 - 0.375)}}{\sqrt{800}} \right) \right)$
CI: Upper limit = UL	$0.428 = P_{actual} + \left(z_{1-\alpha/2} * \left(\frac{\sqrt{D_{actual}} * \sqrt{P_{actual} * (1 - P_{actual})}}{\sqrt{n_{actual}}} \right) \right)$	$= 0.375 + \left(1.96 * \left(\frac{\sqrt{2.5} * \sqrt{0.375 * (1 - 0.375)}}{\sqrt{800}} \right) \right)$
CI: (Lower limit, Upper limit) = (LL, UL)	(0.322; 0.428)	
Standard Error of the Estimate	$0.027 = \left(\frac{\sqrt{D_{actual}} * \sqrt{P_{actual} * (1 - P_{actual})}}{\sqrt{n_{actual}}} \right)$	$= \left(\frac{\sqrt{2.5} * \sqrt{0.375 * (1 - 0.375)}}{\sqrt{800}} \right)$

11.2.4 Interpreting Standard Errors and Confidence Intervals

The SE of the estimate is used as a measure of precision of the estimate and describes the amount of fluctuation from the population parameter that we can expect in sample estimates. Under assumptions of normality in the sampling distribution, we can expect sample estimates to vary within ± 1 SE 68% of the time, within ± 2 SE 95% of the time, and within ± 3 SE 98% of the time.

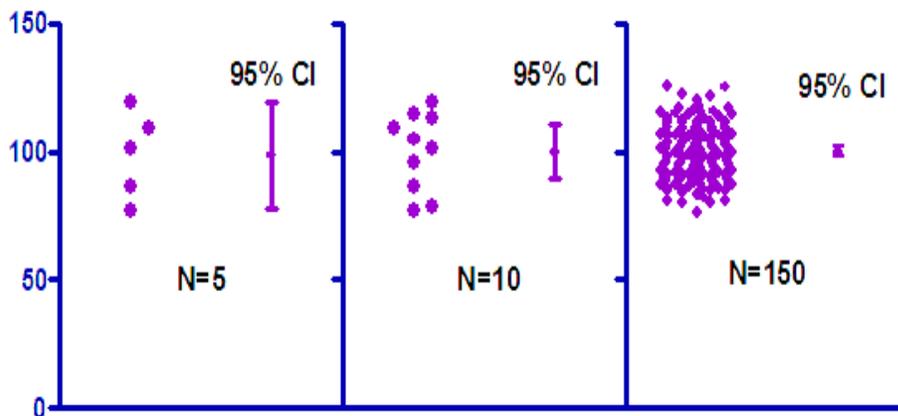
The interpretation of a CI is nuanced. The correct way to interpret the CI is as follows: If a large number of surveys were repeatedly conducted on the same population and if CIs were calculated for each survey conducted, 95% of the CIs would contain the true (target population) value of the indicator; the CI from the given PBS is one such interval.

A specific CI represents the amount of uncertainty, around the point estimate of the indicator (proportion or mean). A CI tells us that it would not be unusual at all for other random samples drawn from the same target population to obtain different sample estimates of the indicator. These other sample estimates of the indicator all suggest different values for the true indicator value, but if there is a high degree of confidence associated with the CI (e.g., 95%), then there should be a great degree of overlap in such CIs coming from different samples. Hence, the CI represents the inherent uncertainty that comes with using sample data.

It is instructive to note that, for the same sample size, 99% CIs are wider than 95% intervals, and 90% intervals are narrower than 95% intervals. If more confidence that an interval contains the true value of the indicator is desired, then the intervals must be wider. In the extreme, if 100% confidence that an interval contains the true value of the indicator is desired, the interval must contain every possible value, so it must be very wide.

It is important to understand that a CI does **not** mean that the probability is 95% that the true value of the indicator is contained in the CI. This is often incorrectly used as the interpretation for such a CI. This is because a specific CI either does or doesn't contain the true value of the indicator (with probability 1 or 0). The true value of the indicator has one value, which is unknowable. What can be said is that if a number of different samples are drawn from different PBSs, the CIs from each of these would not all be the same (but should have a large degree of overlap). One would expect 95% of them to contain the true value of the indicator, but one would never know whether the interval from a specific survey drawn contained the true value of the indicator or not.

Finally, a CI should not be interpreted as a range that contains 95% of the values from the sample drawn. In other words, a CI does not quantify variability of the sample. The graph below emphasizes this distinction.



12. Data Analysis for Comparative Analytical Surveys: Statistical Tests of Differences

For comparative analytical surveys, the main aim is to conduct statistical tests of differences between estimates—typically where the underlying data are collected at two points in time (e.g., through baseline and end-line PBSs) for indicators of proportions or means. Procedures for producing final sample weights to be used in data analysis were described in Chapter 10. For a comparative analytical survey, the next step is to conduct the statistical test of differences. For the Feed the Future ZOI PBS indicators listed in Tables 1 and 2, this occurs after data collection is complete for the survey at the second point in time. For such statistical tests, the final sampling weights are integrated into the indicator estimates and the corresponding SE estimates, as in the previous chapter, and are used in the corresponding statistics that underpin the statistical testing procedures.

Survey implementers must use a statistical software package, such as SAS, SPSS, or STATA, to undertake statistical tests of differences, so that elements of the complex survey design are appropriately taken into account. Further details on the syntax for these packages can be found in Appendix B. The two sections that follow describe the statistical tests of differences for indicators of means and proportions, respectively. The case of indicators of means is presented first since it is the simpler of the two cases.

12.1 Statistical Tests of Differences for Indicators of Means

Suppose there is a desire to establish a test of differences for the prevalence of an indicator that is a mean (such as “Yield of Targeted Agricultural Commodities within Target Areas”). Assume that \bar{X}_1 represents the true mean value of the indicator in the population at time point 1 and that \bar{X}_2 represents the true mean value of the indicator in the population at time point 2. If the project is attempting to influence an improvement in an indicator such as “Yield of Targeted Agricultural Commodities within Target Areas,” then one would expect to see an increase in the mean over time. Therefore, the null hypothesis would be stated as:

$$H_0: \bar{X}_2 - \bar{X}_1 < \delta$$

versus the alternative hypothesis:

$$H_A: \bar{X}_2 - \bar{X}_1 \geq \delta$$

which is equivalent to the formulation given in Section 2.2.2. To implement the statistical test, a “test statistic” is computed and compared to a predefined cutoff value to see if the test statistic is greater than or equal to the cutoff value. If the test statistic is greater than or equal to the cutoff value, the null hypothesis, H_0 , is rejected in favor of H_A ; otherwise, the null hypothesis is not rejected and the results are deemed inconclusive.

Although statistical tests of differences must be undertaken using appropriate specialized statistical software, the formulas are provided here to provide the reader with a sense of how the mechanics of such tests work. For an indicator of means, the test statistic (using a pooled SE over the two PBSs) is given by:

$$T_{means} = \frac{\bar{X}_{2,actual} - \bar{X}_{1,actual}}{\sqrt{\sigma_{X_{pooled}}^2 * \left(\frac{1}{n_{1,actual}} + \frac{1}{n_{2,actual}}\right)}}$$

where:

$$\sigma_{X_{pooled}} = \sqrt{\frac{((n_{1,actual} - 1) * D_{1,actual} * \sigma_{X_{1,actual}}^2) + ((n_{2,actual} - 1) * D_{2,actual} * \sigma_{X_{2,actual}}^2)}{(n_{1,actual} + n_{2,actual} - 2)}}$$

where:

$\bar{X}_{1,actual}$ = the sample-weighted estimate of the mean from the survey at baseline

$\bar{X}_{2,actual}$ = the sample-weighted estimate of the mean from the survey at end-line

$n_{1,actual}$ = the actual sample size (either individuals or households depending on the indicator) realized after fieldwork at baseline

$n_{2,actual}$ = the actual sample size (either individuals or households depending on the indicator) realized after fieldwork at end-line

$\sigma_{X_{1,actual}}$ = the standard deviation of X_1 computed from the survey data at baseline

$\sigma_{X_{2,actual}}$ = the standard deviation of X_2 computed from the survey data at end-line

$D_{1,actual}$ = the DEFF corresponding to the indicator using survey data from the baseline, and computed by statistical software

$D_{2,actual}$ = the DEFF corresponding to the indicator using survey data from the end-line, and computed by statistical software

The cutoff value against which the test statistic, T_{means} , should be compared is $t_{1-\alpha}(DF)$ where $t_{1-\alpha}(DF)$ = the value of the t-distribution with $\alpha = 0.05$ and where:

DF = degrees of freedom = (# EAs sampled across the baseline and end-line PBSs - # strata across the baseline and end-line PBSs)

For the above test of hypothesis, if $T_{means} \geq t_{1-\alpha}(DF)$, then the null hypothesis should be rejected in favor of the alternative hypothesis. On the other hand, if $T_{means} < t_{1-\alpha}(DF)$, then the null hypothesis should not be rejected, and the results should be deemed inconclusive.

The following example illustrates the test, assuming that two comparative analytical surveys (e.g., a baseline PBS and an end-line PBS) have been implemented. A test of differences is to be conducted on the indicator for “Yield of Targeted Agricultural Commodities within Target Areas.” Continuing the example from Section 2.2.2, both surveys aim to collect data on $n_{initial} = 149$ producers of maize at both baseline and end-line. Because the overall sample size for the survey is based on the stunting indicator, the number of households sampled at each time point is 2,760, which results in many more than the required 149 producers of maize sampled for at each time point. The results from the baseline

PBS gives $\bar{X}_{1,actual} = 1.52$ hectares of maize based on collecting data on $n_{1,actual} = 2,500$ producers. The results from the end-line PBS gives $\bar{X}_{2,actual} = 1.65$ based on collecting data on $n_{2,actual} = 2,550$ producers of maize. For both surveys, $D_{1,actual} = D_{2,actual} = 3.3$. The computed standard deviations from each survey are $\sigma_{X_{1,actual}} = 0.45$ and $\sigma_{X_{2,actual}} = 0.50$, respectively. The number of EAs in the survey at each time point is 111 and the number of strata at each time point is 6. Plugging these values into the above formula gives the following:

$$\sigma_{X_{pooled}} = \sqrt{\frac{((2,500 - 1) * 3.3 * 0.45^2) + ((2,550 - 1) * 3.3 * 0.50^2)}{(2,500 + 2,550 - 2)}} = 0.8645$$

and:

$$T_{means} = \frac{1.65 - 1.52}{\sqrt{0.8645^2 * (\frac{1}{2,500} + \frac{1}{2,550})}} = 5.35$$

This is then compared to $t_{1-\alpha}(DF) = t_{0.95}(2 * (111 - 6)) = t_{0.95}(210) = 1.65$.

Since $T_{means} = 5.35 > t_{0.95}(210) = 1.65$, the null hypothesis is rejected. In other words, at 80% power, there is evidence to suggest that there has been a statistically significant increase in the yield of maize between the two time periods.

It is very important to exhibit caution when interpreting the results of a statistical test of differences. In the above example, given that the design of the two surveys was limited to a simple pre-post “adequacy” design, the conclusion should not be that this improvement is due to the project intervention. It can be definitively stated that change has occurred, but statements of attribution solely to the project should not be made. There may have been external factors that contributed to the improvement in yield (e.g., cessation or diminution of a number of previous negative factors, such as climatic conditions [drought, floods, earthquakes], inhibiting government policies, civil strife/instability, and/or price or other economic fluctuations; or the addition of positive factors, such as related project interventions by other organizations). To be able to make statements regarding attribution to project interventions, more complex designs, involving control groups, randomization of project interventions to clusters or individuals, and/or systematic control of confounding factors, would be needed.

12.2 Statistical Tests of Differences for Indicators of Proportions

Suppose there is a desire to establish a test of differences for the “Prevalence of Stunted Children under Five” indicator, an indicator of proportions. Assume that P_1 represents the true prevalence of stunted children under 5 in the population at baseline and that P_2 represents the true prevalence of stunted children under 5 in the population at end-line. If the project is attempting to influence an improvement in the “Prevalence of Stunted Children under Five” indicator, then one would expect to see a decrease in the prevalence over time. Therefore, the null hypothesis:

$$H_0: P_1 - P_2 \leq \delta$$

needs to be tested against the alternative hypothesis:

$$H_A: P_1 - P_2 > \delta$$

To implement the statistical test, a “test statistic” is computed and compared to a predefined cutoff value to see if the test statistic exceeds the cutoff value. If the test statistic exceeds the cutoff value, the null hypothesis, H_o , is rejected in favor of H_A ; otherwise, the null hypothesis is not rejected. As mentioned earlier in the guide, the above test of hypothesis uses what is called a “one-sided test.”

For an indicator of proportions where the underlying values are dichotomous (i.e., where the underlying values are zeros and ones), the appropriate statistical test to use is based on the chi-square distribution, the details of which can be found in any basic statistics textbook. However, since the data for an indicator of proportions has been collected through a PBS where the underlying design is complex with stratification, multiple stages of sampling, and clustering, the chi-square statistic must be “corrected” to take into account the DEFF of the indicator of proportions. The resulting adjusted test is called the Rao-Scott chi-square test.⁸³ The functional form of this test is very complex and beyond the scope of this guide, and so will not be presented here. However, the results of empirical examples based on the Rao-Scott chi-square test using SAS, SPSS, and STATA software packages are presented in Appendix B.

⁸³ For more details on the Rao-Scott chi-square test, see: Rao, J.N.K. and Scott, A.J. 1981. “The analysis of categorical data from complex sample surveys: chi-squared tests for goodness-of-fit and independence in two-way tables.” *Journal of the American Statistical Association*. 76: 221–230.

Appendix A. Derivation of Inflation Factor for Number of Households to Contact Given an Initial Sample Size of Individuals

Section 2.2.1 of this guide provides guidance on computing the sample size for a PBS whose main aim is to undertake statistical tests of differences over time for indicators of proportions; Section 2.2.2 does the same for indicators of means. In many cases, the overall sample size for the PBS is driven by an indicator at the individual level, such as the “Prevalence of Stunted Children under Five,” which requires collecting data on children under 5 years of age. In such cases, the initial sample size calculated reflects the number of children in this age category for which data are required. However, most surveys use households rather than children as the basis for sampling, making it necessary to express the required sample size in terms of the number of households to contact. Although some households will have exactly one eligible child, other households will have more than one eligible child and some households will have no eligible children at all. In all cases, one cannot know the current composition (number and age) of children in a sampled household until the household is contacted and such information is obtained through the creation of a household roster. This creates a challenge for sample size calculation, because the correspondence between households and children is not always one-to-one. When basing the overall sample size of the PBS on an indicator such as stunting, it is essential therefore to have not only an estimate of the number of eligible children that must be sampled, but also an estimate of the number of households that need to be visited to obtain the required sample of eligible children.

The approach described in Appendix A is summarized in Section 2.2.4 and can be extended to any individual-level indicator, not just indicators involving children under 5 years of age. For instance, it can be extended to indicators involving women of reproductive age (e.g., women age 15–49). The indicator “Prevalence of Stunted Children under Five” is used for illustrative purposes only for the remainder of this appendix.

A.1 Inflating the Initial Sample Size to Account for Households with No Eligible Children to Arrive at an Initial Adjusted Sample Size

To increase the likelihood that the required sample size for children will be met in advance of fieldwork, previous approaches have suggested inflating the required sample size by an amount equal to the inverse of the estimated average number of eligible children per household. For example, assume the required sample size of children is $n_{initial} = 1,403$, as in the example provided in Section 2.2.3 for the “Prevalence of Stunted Children under Five” indicator, and assume the average household size is 4.27 and the proportion of children that are 0–59 months is no more than 0.16 (equivalent to 16%).⁸⁴ Then, the estimated average number of children 0–59 months of age per household is $4.27 * 0.16 = 0.6832$. We refer to this factor, 0.6832, the average number of children age 0–59 months per household, as λ . To obtain the appropriate number of households that need to be sampled to ensure that the required sample size of 1,403 children is achieved, past approaches have suggested dividing $n_{initial} = 1,403$ by λ

⁸⁴ Figures for both the average household size and the proportion of children in the target age group are typically obtained from the most recent national census or from some other national or internationally sponsored survey.

= 0.6832, to obtain a sample size of $n_{initial}/\lambda$ households. In this example, we calculate this as $1,403 / 0.6832 = 2,054$ households.

However, past field experience for some baseline and end-line PBSs conducted by FFP DFSAs has shown that this approach can underestimate the number of households that should be visited to obtain the required sample size of children. This in turn has resulted in surveys falling short of achieving the required sample size of children while conducting fieldwork. In such instances, some FFP DFSAs have opted to augment the number of households sampled using “on the fly” non-probability-based sampling techniques.⁸⁵ Such strategies should be avoided, and it is preferable to appropriately approximate the sample size of households truly needed in advance of conducting fieldwork.

Given the shortfall in sample size that has been experienced, it is recommended to use an alternative approach (introduced in Section 2.2.4) that will more closely approximate the number of households that need to be visited to ensure that the required child sample size is achieved. The approach involves inflating the required sample size by the inverse of the proportion of households that have at least one eligible child (rather than the average number of eligible children per household). This approach results in a household sample size that is greater than that suggested in the aforementioned approach that has been used in the past, but is more likely to result in the required sample size for children being achieved (and sometimes exceeded).

The alternative approach involves a sample size inflation factor that is approximated using the Poisson distribution.⁸⁶ Using this distribution, it can be estimated that the proportion of households having at least one eligible child (i.e., at least 1 or 2 or 3 or ... eligible children) is given by $adj_0 = 1 - e^{-\lambda}$, when, on average, there are λ eligible children per household. Here, e refers to the exponential function, found on any scientific hand calculator under the symbol “exp” or “ex.” In the above example with $\lambda = 0.6832$, we have that $1 - e^{-\lambda} = 1 - e^{-0.6832} = 0.4950$. Thus, if $n_{initial}$ is the original required sample size as calculated using one of the formulas given in Sections 2.2.1 and 2.2.2, then an adjusted sample size that takes into account this inflation factor is given by:

$$n_{adj_0} = \frac{n_{initial}}{adj_0} = \frac{n_{initial}}{(1 - e^{-\lambda})} \quad (A.1)$$

The technical details of the derivation of n_{adj_0} are given in Section A.3 of this appendix. Continuing the above example as an illustration of this adjustment, under the new approach, the original required sample size of $n_{initial} = 1,403$ children is adjusted to

$$n_{adj_0} = \frac{n_{initial}}{(1 - e^{-\lambda})} = \frac{1,403}{(1 - e^{-0.6832})} = \frac{1,403}{0.4950} = 2,834$$

⁸⁵ Such “on the fly” techniques have included repeatedly visiting additional adjacent households until the required sample size is achieved. Using such techniques, households are not drawn using a random mechanism and therefore the results are not probability-based.

⁸⁶ The Poisson distribution is a discrete statistical distribution defined for integers 0, 1, 2, 3 ... that gives the probability (or proportion) of the number of times (0, 1, 2, ...) a random variable occurs, when it is known to occur an average of λ times.

households. Using previous approaches, it is assumed that 651 (or 2,054 – 1,403) of the households sampled will have no eligible children. On the other hand, under the new approach recommended here, it is assumed that 1,431 (or 2,834 – 1,403) of the households sampled will have no eligible children.

A.2 Deflating the Initial Adjusted Sample Size to Account for Households with Two or More Eligible Children to Arrive at a Final Adjusted Sample Size

Although the above approach more correctly adjusts for the number of households with no eligible children, the adjusted sample size, n_{adj_0} , does not account for the fact that some households may have two or more eligible children age 0–59 months. Furthermore, surveys can opt to collect information on either all or a subsample of the eligible children within a sampled household. However, as described in Chapter 9, it is strongly recommended that the strategy of selecting **all** eligible children within a household be adopted, rather than subsampling one or more such children. In light of this, n_{adj_0} should be deflated slightly to account for households that contribute two or more children toward the overall required sample size of children.⁸⁷ Once again, the Poisson distribution is used to approximate the required deflation factor, and the sample size inflation from the previous section is used as a starting point. The formula for the deflation adjustment is shown below, where n_{adj_0} is the result of the earlier sample size inflation and n_{adj_1} is the result of the following deflation adjustment:

$$\begin{aligned} n_{adj_1} &= [A * n_{adj_0}] + [0.5 * (1 - A) * n_{adj_0}] \\ &= n_{initial} * \left(\left[\frac{A}{(1 - e^{-\lambda})} \right] + 0.5 * \left[\frac{1-A}{(1 - e^{-\lambda})} \right] \right) \\ &= n_{initial} * adj_1 \end{aligned} \tag{A.2}$$

where:

$$A = (1 + \lambda) * e^{-\lambda}$$

Here, adj_1 is the same as formula (4) given in Section 2.2.4. The details of the derivation of n_{adj_1} can be found in Section A.4 of this appendix. Continuing the example from above where $n_{adj_0} = 2,834$ and $\lambda = 0.6832$, we obtain:

$$e^{-\lambda} = 0.5050 \text{ and } A = (1 + 0.6832) * 0.5050 = 0.8500$$

and finally:

$$n_{adj_1} = [0.8500 * 2,834] + [0.5 * (1 - 0.8500) * 2,834] = 2,622 \text{ households}$$

⁸⁷ It is important to note that the deflation factor described in this section relies on strict adherence to the strategy of sampling **all** eligible children in a sampled household. If, instead, a strategy of subsampling one or more eligible children in a sampled household is adopted, then this deflation factor is not required and the inflation factor in Section A.1 is all that is required.

The deflation adjustment results in the sample size decreasing from 2,834 households to 2,622 households. This means that approximately $2,834 - 2,622 = 212$ households are expected to contribute two or more children to the sample of children.

A.3 Derivation of n_{adj_0}

To derive the first inflation factor, we use the Poisson distribution, which is a discrete distribution defined for integers 0, 1, 2, 3, ... and which gives the probability, denoted by Pr, of the number of occurrences (x) of a particular event (X), given that it is known that the **average** number of times the event occurs is λ . The distribution looks as follows:

$$\Pr(X = x) = \frac{(e^{-\lambda}\lambda^x)}{x!} \quad x = 0, 1, 2, 3 \dots \quad (\text{A.3})$$

where “ $x!$ ” is called “ x factorial” and is defined as $x! = x * (x - 1) * (x - 2) * \dots * 1$. Note that $0! = 1$ and $\lambda^0 = 1$.

If, for example, we define the event, X , as “number of children age 0–59 months in a household” and we define λ as “the average number of children age 0–59 months per household,” then the Poisson distribution gives the probability (or proportion) that the number of children age 0–59 months in a given household is x .

For example, if we want the probability that there are 0 (or no) children age 0–59 months in a household, using equation (A.3) with $x = 0$, we compute:

$$\Pr(X = 0) = \frac{(e^{-\lambda}\lambda^0)}{0!} = e^{-\lambda} \quad (\text{A.4})$$

Assuming that $n_{initial}$ completed interviews on children age 0–59 months are required, we need to know how many households to visit, including those where there are no children of eligible age. Therefore, we wish to know the probability of households that will have **at least one** (i.e., one or more) child age 0–59 months. Using equation (A.4), we can see that this is given by:

$$\Pr(X > 0) = 1 - \Pr(X = 0) = 1 - e^{-\lambda} \quad (\text{A.5})$$

To obtain the number of households to visit, we should inflate n (the sample size calculated for children under 5 years of age) by the inverse of the proportion given in equation (A.5). Therefore, we have:

$$n_{adj_0} = \frac{n_{initial}}{(1 - e^{-\lambda})} \quad (\text{A.6})$$

Equation (A.6) is the same as the inflation factor given by equation (A.1) at the end of Section A.1.

Note: The Poisson distribution spreads probabilities across whole numbers that range in value from 0 to infinity. However, there are not an infinite number of children age 0–59 months within a household. Therefore, to be most technically correct, this derivation should be based on a “truncated Poisson” distribution that does not permit values greater than some reasonable number of children of eligible age per household (e.g., 5) and that defines the distribution for discrete values 0, 1, 2, 3, 4, and 5 only. However, it can be shown that for small values of λ (e.g., $\lambda < 1.5$), $\Pr(X > 5)$ is close to 0 and therefore is negligible. So, it was deemed that the added accuracy in using the truncated Poisson distribution does

not warrant the additional complexity in the formula, and the usual Poisson distribution was used instead of the truncated Poisson distribution in the above derivation.

A.4 Derivation of n_{adj_1}

The adjusted sample size, n_{adj_0} , from the last section gives the number of households to sample to achieve the required sample size of children, $n_{initial}$, taking into account households with no eligible children. Therefore, n_{adj_0} accounts for households that have exactly one child of eligible age and households with two or more children of eligible age. If one were to sample only one child of eligible age per household (not recommended), then sampling in households having two or more children of eligible age would not be a concern. However, the recommended guidance is to sample **all** children of eligible age within a selected household. As a result, it is possible to achieve the overall desired sample size of children, $n_{initial}$, by visiting fewer than n_{adj_0} households because some households accounted for by n_{adj_0} will contain two or more children of eligible age, all of whom will be sampled. To appropriately discount for households with two or more children of eligible age, we must deflate n_{adj_0} accordingly.

To derive the deflator for this, we use equation (A.3) with $x = 1$ and note that:

$$\Pr(X = 1) = \frac{(e^{-\lambda}\lambda^1)}{1!} = \lambda e^{-\lambda}. \quad (\text{A.7})$$

Additionally:

$$\Pr(X \geq 2) = 1 - \Pr(X = 0) - \Pr(X = 1) = 1 - e^{-\lambda} - (\lambda e^{-\lambda}) \quad (\text{A.8})$$

using equations (A.4) and (A.7).

Combining terms, we have:

$$\Pr(X \geq 2) = 1 - [(1 + \lambda) * e^{-\lambda}] = 1 - A \quad (\text{A.9})$$

where:

$$A = (1 + \lambda) * e^{-\lambda} \quad (\text{A.10})$$

Using equations (A.8) and (A.9), it is useful to note that:

$$A = \Pr(X = 0) + \Pr(X = 1) \quad (\text{A.11})$$

Next, we use the tautology:

$$1 = \Pr(X = 0) + \Pr(X = 1) + \Pr(X \geq 2) \quad (\text{A.12})$$

Equation (A.12) is true because the sum of the Poisson (or any other discrete) distribution across all possible values is equal to 1. Using equations (A.4), (A.7), and (A.9), we can see that equation (A.12) can be rewritten as:

$$1 = e^{-\lambda} + (\lambda e^{-\lambda}) + [1 - ((1 + \lambda) * e^{-\lambda})] = [(1 + \lambda) * e^{-\lambda}] + [1 - ((1 + \lambda) * e^{-\lambda})] \quad (\text{A.13})$$

We multiply each term in equation (A.13) by n_{adj_0} and obtain:

$$\begin{aligned} n_{adj_0} &= [(1 + \lambda) * e^{-\lambda}] * n_{adj_0} + [1 - ((1 + \lambda) * e^{-\lambda})] * n_{adj_0} \\ &= [A * n_{adj_0}] + [(1 - A) * n_{adj_0}] \end{aligned} \quad (A.14)$$

We obtain this last expression by applying the definition of A given in equation (A.10).

Equation (A.14) essentially breaks n_{adj_0} into a composite sum with two component parts given by:

$$A * n_{adj_0} \text{ and } (1 - A) * n_{adj_0}$$

From equation (A.11) above, we see that $A = \Pr(X = 0) + \Pr(X = 1)$, and so the first component part of equation (A.14) takes into account the households to be visited that contain either no children or one child of eligible age.

From equation (A.9) above, we see that $1 - A = \Pr(X \geq 2)$, and so the second component part of equation (A.14) takes into account the households to be visited that contain two or more children of eligible age.

The aim of the deflator is to reduce the number of households that comes from the second component. We therefore wish to diminish to a half the number of households to sample containing two children of eligible age, and to diminish to a third the number of households to sample containing three children of eligible age, and so on. However, we can assume that the number of households having three or more children of eligible age is negligibly small, relatively speaking. Therefore, for simplicity, we “bundle them” with households having two children of eligible age. What is meant by “bundling” is that we do not, for instance, diminish to a third the number of households to sample having three children of eligible age, because the added complexity of the computation is not worth the negligible difference this would make. Instead, we diminish the number of such households to sample to a half. Similarly, we diminish to a half the number of households to sample having four children of eligible age.

Thus, we create a new adjustment called n_{adj_1} , where we halve the second component of equation (A.14):

$$n_{adj_1} = [A * n_{adj_0}] + [0.5 * (1 - A) * n_{adj_0}] \quad (A.15)$$

Equation (A.15) is the deflation factor given by equation (A.2) at the end of Section A.2.

Appendix B. Syntax for Statistical Software Packages SAS, SPSS, and STATA

B.1 Syntax for SAS Software Users

The syntax is provided separately for analyzing indicators of proportions and indicators of means for both descriptive surveys (described in Chapter 11) and comparative analytical surveys (described in Chapter 12).

B.1.1 Analyzing Indicators of Proportions for Either Descriptive or Comparative Analytical PBSs

If one is undertaking an analysis of an **indicator of proportions** over time using a **comparative analytical PBS**, the first step is to combine the two PBS datasets corresponding to the two time points by appending one onto the end of the other. The two datasets should have the common variables: **STRATUMVAR**, **CLUSTERVAR**, **WEIGHTVAR**, and **OUTCOME**. The combined dataset must include a variable **ROUND** that identifies whether a record comes from the first or the second PBS occasion. For the purposes of the example below, we use **ROUND = 0** and **ROUND = 1**.

PROC SURVEYFREQ in SAS can be used to statistically test the difference in proportions between PBSs over two time points under complex survey designs by using the first-order Rao-Scott chi-square test (by specifying “CHISQ” in the syntax). The procedure will also produce single-point-in-time estimates of the proportions at each time point, along with the corresponding SEs and CIs (by specifying “CL” in the syntax) at each time point (i.e., for each value of **ROUND**). Note that if the sample design is stratified with multiple stages of sampling, one need identify only the first-stage units in the **CLUSTER** statement.⁸⁸ To simplify the output, the “NOFREQ NOCELLPERCENT NOWT” statements are included to suppress some nonessential output. In the syntax below, anything in bold is a variable name or dataset name that the user needs to specify.

```
PROC SURVEYFREQ DATA=DATASETNAME;  
  STRATA STRATUMVAR;  
  CLUSTER CLUSTERVAR;  
  WEIGHT WEIGHTVAR;  
  TABLES ROUND*OUTCOME / ROW CHISQ CL NOFREQ NOCELLPERCENT NOWT;  
  TITLE 'TITLE FOR THE OUTPUT';  
RUN;
```

To provide an example of typical output, a synthetic dataset was created with 4,294 observations (2,147 for each of the two PBS occasions), 2 strata per PBS occasion, and 128 clusters or EAs per PBS occasion, and where **OUTCOME** is a dichotomous variable for which estimates of proportions are desired. The output from the above SAS code when applied to the dataset is given below. When reading

⁸⁸ Some simplifying assumptions are used here: Clusters are considered to be selected with replacement from the first stage strata. Multi-stage sampling within selected clusters is ignored for the purposes of variance estimation, and clusters of observations are assumed instead. This greatly simplifies variance estimation because clusters within strata are assumed to be the dominant source of variance in sample estimates. Any finite population correction for the first stage sample is ignored. The resulting estimates of sampling variance will be slight overestimates in this case, however.

the output, the user should report the proportions (and associated SEs and confidence limits) in the column labelled ROW PERCENT for the rows in which the OUTCOME has value 1 only (highlighted). For ROUND = 0, the estimate of the proportion is 60.56%, the SE is 2.83, and the CI is (54.95%, 66.16%). For ROUND = 1, the estimate of the proportion is 88.09%, the SE is 2.26, and the CI is (83.61%, 92.58%). The (first order) Rao-Scott chi-square statistic (highlighted) is 55.15, which is found to be highly significant (<0.0001). This implies that there is a statistically significant difference in the proportions between ROUND = 0 and ROUND = 1.

<p align="center">'TITLE FOR THE OUTPUT' The SURVEYFREQ Procedure</p>								
Data Summary								
Number of Strata		2						
Number of Clusters		128						
Number of Observations		4294						
Sum of Weights		10418587.7						
Table of ROUND by OUTCOME								
ROUND	OUTCOME	95% Confidence Limits for Percent*		Row Percent	Standard Error of Row Percent	95% Confidence Limits for Row Percent		
0	0	16.9179	22.5217	39.4397	2.8317	33.8359	45.0435	
	1	27.4783	33.0821	60.5603	2.8317	54.9565	66.1641	
	Total	50.0000	50.0000	100.000				
1	0	3.7081	8.1934	11.9014	2.2665	7.4161	16.3867	
	1	41.8066	46.2919	88.0986	2.2665	83.6133	92.5839	
	Total	50.0000	50.0000	100.000				
Total	0	21.5013	29.8398					
	1	70.1602	78.4987					
	Total							
Rao-Scott Chi-Square Test								
Pearson Chi-Square		426.6567						
Design Correction		7.7362						
Rao-Scott Chi-Square		55.1509						
Degrees Freedom		1						
Pr > ChiSq		<.0001						
F Value		55.1509						
Numerator DF		1						
Denominator DF		126						
Pr > F		<.0001						
Sample Size = 4294								

* Note that this first set of reported confidence limits does not take into account the complex survey design, and therefore should not be used. Rather, the second set of confidence limits at the far right of the table is the correct ones to use.

If one is undertaking an analysis of an **indicator of proportions** using a **descriptive PBS**, datasets at each time point are analyzed individually and therefore there is no need to combine the datasets or to create the variable ROUND. No statistical test is performed because the test of differences over two time points is not relevant. In this case, the same syntax as above can be used but reference to ROUND

can be eliminated. Additionally, reference to ROW and CHISQ can be eliminated. Using the following SAS syntax, one can produce a single-point-in-time estimate of the indicator of proportions, along with the corresponding SE and CI (by specifying “CL” in the syntax):

```
PROC SURVEYFREQ DATA=DATASETNAME;
  STRATA STRATUMVAR;
  CLUSTER CLUSTERVAR;
  WEIGHT WEIGHTVAR;
  TABLES OUTCOME / CL NOFREQ NOCELLPERCENT NOWT;
  TITLE 'TITLE FOR THE OUTPUT';
  RUN;
```

To provide an example of typical output, the same synthetic dataset is used as before, but for one time point only (ROUND = 0); that is, with 2,147 observations, 2 strata, and 128 clusters or EAs. The output from the above SAS code when applied to the dataset is given below. When reading the output, the user should report the percent, SE, and confidence limits for the row in which OUTCOME = 1 (highlighted). For this dataset, the estimate of proportion is 60.56%, the SE is 2.83, and the CI is (54.95%, 66.16%).

'TITLE FOR THE OUTPUT' The SURVEYFREQ Procedure				
Data Summary				
Number of Strata				2
Number of Clusters				128
Number of Observations				2147
Sum of Weights				5209293.87
Table of OUTCOME				
OUTCOME	Percent	Standard Error of Percent	95% Confidence Limits for Percent	
0	39.4397	2.8317	33.8359	45.0435
1	60.5603	2.8317	54.9565	66.1641
Total	100.000			

B.1.2 Analyzing Indicators of Means for Either Descriptive or Comparative Analytical PBSs

If one is undertaking an analysis of an **indicator of means** over time using a **comparative analytical PBS**, the first step is to combine the two PBS datasets corresponding to the two time points by appending one onto the end of the other. The two datasets should have common variables **STRATUMVAR**, **CLUSTERVAR**, **WEIGHTVAR**, and **OUTCOME**. The combined dataset must include a variable **ROUND** that identifies whether a record comes from the first or the second PBS occasion. For the purposes of the example below, we use ROUND = 0 and ROUND = 1.

PROC SURVEYMEANS in SAS produces single-point-in-time estimates of an indicator of means, along with the corresponding SEs and CIs at each time point (i.e., for each value of ROUND). The procedure SURVEYMEANS is run prior to SURVEYREG (syntax follows) because the former procedure

provides estimates of confidence limits while the latter does not. However, the latter provides the results of a test of differences.

```
PROC SURVEYMEANS DATA=DATASETNAME;
STRATA STRATUMVAR;
CLUSTER CLUSTERVER;
WEIGHT WEIGHTVAR;
DOMAIN ROUND;
VAR OUTCOME;
TITLE 'TITLE FOR THE OUTPUT';
RUN;
```

To provide an example of typical output, the same synthetic dataset is used as before, with 4,294 observations (2,147 for each of the two PBS occasions), 2 strata per PBS occasion, and 128 clusters or EAs per PBS occasion, and where OUTCOME is a continuous variable for which estimates of means are desired. The output from the above SAS code when applied to the dataset is given below. For ROUND = 0, the estimate of the mean is 13.70, the SE is 0.07, and the CI is (13.56, 13.85). For ROUND = 1, the estimate of the mean is 13.76, the SE is 0.07, and the CI is (13.62, 13.90). These values are highlighted in the output.

The SURVEYMEANS Procedure						
Data Summary						
Number of Strata	2					
Number of Clusters	128					
Number of Observations	4294					
Sum of Weights	10418587.7					
Statistics						
Variable	N	Mean	Std Error of Mean	95% CL for Mean		
OUTCOME	4294	13.735713	0.070032	13.5971217	13.8743033	
The SURVEYMEANS Procedure						
Domain Statistics in Time						
ROUND	Variable	N	Mean	Std Error of Mean	95% CL for Mean	
0	OUTCOME	2147	13.706885	0.074014	13.5604135	13.8533563
1	OUTCOME	2147	13.764540	0.072610	13.6208466	13.9082337

PROC SURVEYREG in SAS uses the t-test to statistically test the difference in means between PBSs over two time points under complex survey designs.

```

PROC SURVEYREG DATA=DATASETNAME;
STRATA STRATUMVAR;
CLUSTER CLUSTERVER;
WEIGHT WEIGHTVAR;
CLASS ROUND;
MODEL OUTCOME = ROUND / NOINT SOLUTION
      VADJUST=NONE;
LSMEANS ROUND / DIFF;
TITLE 'TITLE FOR THE OUTPUT';
RUN;

```

The SAS syntax is applied to the same synthetic dataset as before. To correctly interpret whether the mean OUTCOME variable significantly differs between ROUND = 0 and ROUND = 1, consider the final output table, “Differences of ROUND least squares means.” The value for ROUND = 0 is compared to that for ROUND = 1. The negative value of the estimate (−0.05) means that the value of OUTCOME is lower for ROUND = 0 than for ROUND = 1; however, the difference is not statistically significant (Pr = 0.1863).

'TITLE FOR OUTPUT' The SURVEYREG Procedure Regression Analysis for Dependent Variable OUTCOME			
Data Summary			
Number of Observations	4294		
Sum of Weights	10418588		
Weighted Mean of OUTCOME	13.73571		
Weighted Sum of OUTCOME	143106726		
Design Summary			
Number of Strata	2		
Number of Clusters	128		
Fit Statistics			
R-Square	0.9911		
Root MSE	1.3005		
Denominator DF	126		
Class Level Information			
CLASS Variable	Levels	Values	
ROUND	2	0 1	
Tests of Model Effects			
Effect	Num DF	F Value	Pr > F
Model	2	19266.4	<.0001
ROUND	2	19266.4	<.0001
The denominator degrees of freedom for the F tests is 126.			

Estimated Regression Coefficients					
Parameter	Estimate	Standard Error	t Value	Pr > t	
ROUND 0	13.7068849	0.07401393	185.19	<.0001	
ROUND 1	13.7645402	0.07261024	189.57	<.0001	

The degrees of freedom for the t tests is 126.

ROUND Least Squares Means					
ROUND	Estimate	Standard Error	Degrees Freedom	t Value	Pr > t
0	13.7069	0.07401	126	185.19	<.0001
1	13.7645	0.07261	126	189.57	<.0001

Differences of ROUND Least Squares Means						
ROUND	_ROUND	Estimate	Standard Error	Degrees Freedom	t Value	Pr > t
0	1	-0.05766	0.04339	126	-1.33	0.1863

If one is undertaking an analysis of an **indicator of means** using a **descriptive PBS**, datasets at each time point are analyzed individually and therefore there is no need to combine the datasets or to create the variable ROUND. No statistical test is performed because the test of differences over two time points is not relevant. In this case, the same syntax as above for PROC SURVEYMEANS can be used but reference to DOMAIN ROUND can be eliminated. Using the following SAS syntax, one can produce a single-point-in-time estimate of the indicator of means, along with the corresponding SE and CI:

```
PROC SURVEYMEANS DATA=DATASETNAME;
STRATA STRATUMVAR;
CLUSTER CLUSTERVAR;
WEIGHT WEIGHTVAR;
VAR OUTCOME;
TITLE 'TITLE FOR THE OUTPUT';
RUN;
```

To provide an example of typical output, the same synthetic dataset is used as before, but for one time point only (ROUND = 0), that is, with 2,147 observations, 2 strata, and 128 clusters or EAs. The output from the above SAS code when applied to the dataset is given below. The estimate of mean is 13.70, the SE is 0.07, and the CI is (13.56, 13.85); these values are highlighted.

The SURVEYMEANS Procedure					
Data Summary					
Number of Strata	2				
Number of Clusters	128				
Number of Observations	2147				
Sum of Weights	5209293.87				
Statistics					
Variable	N	Mean	Standard Error of Mean	95% CL for Mean	
OUTCOME	2147	13.706885	0.074014	13.5604135	13.8533563

B.2 Syntax for SPSS Software Users

To analyze complex survey data, one must use the Complex Samples add-on module, which is available only for SPSS version 21.0.0 and above.⁸⁹ For more information, see: IBM. IBM SPSS Complex Samples 24. Available at: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/en/client/Manuals/IBM_SPSS_Complex_Samples.pdf.

Before any analysis can take place, an “analysis plan” that provides details of the survey design must be defined using a CSPLAN ANALYSIS statement. A Wizard function walks the user through the steps of defining an analysis plan.

To illustrate, assume a stratified three-stage sample where clusters are randomly drawn at the first stage of sampling using PPS, households are randomly drawn at the second stage of sampling using systematic sampling, and all individuals are selected at the third stage of sampling. In this case, to analyze data from such a design, one would define an analysis plan using the syntax below, where anything in bold is a variable name or dataset name that the user needs to specify.

```
CSPLAN ANALYSIS  
/PLAN FILE='DATASETNAME.csaplan'  
/PLANVARS ANALYSISWEIGHT=WEIGHTVAR  
/PRINT PLAN  
/DESIGN STRATA= STRATUMVAR CLUSTER= CLUSTERVAR  
/ESTIMATOR TYPE=WR90
```

Once the analysis plan has been defined, analysis can take place. In what follows, the syntax is provided separately for analyzing indicators of proportions and indicators of means for both descriptive surveys (described in Chapter 11) and comparative analytical surveys (described in Chapter 12). For more information, see: http://www.spss.ch/upload/1071150823_SPSS%2012%20Complex%20Samples.pdf.

B.2.1 Analyzing Indicators of Proportions for Either Descriptive or Comparative Analytical PBSs

If one is undertaking an analysis of an **indicator of proportions** over time using a **comparative analytical PBS**, the first step is to combine the two PBS datasets corresponding to the two time points by appending one onto the end of the other. The two datasets should have common variables **STRATUMVAR**, **CLUSTERVAR**, **WEIGHTVAR**, and **OUTCOME**. The combined dataset must include a variable **ROUND** that identifies whether a record comes from the first or the second PBS occasion. For the purposes of the example below, we use **ROUND = 0** and **ROUND = 1**.

CSTABULATE in SPSS uses the Rao-Scott chi-square test (by specifying **/TEST INDEPENDENCE**) to statistically test the difference between indicators of proportions between PBSs over two time points

⁸⁹ Empirical examples with outputs are not provided for the case of SPSS because FHI 360 does not have a license for the SPSS Complex Samples add-on module.

⁹⁰ Some simplifying assumptions are used here: Clusters are considered to be selected with replacement from the first stage strata. Multi-stage sampling within selected clusters is ignored for the purposes of variance estimation, and clusters of observations are assumed instead. This greatly simplifies variance estimation because clusters within strata are assumed to be the dominant source of variance in sample estimates. Any finite population correction for the first stage sample is ignored. The resulting estimates of sampling variance will be slight overestimates in this case, however.

under complex survey designs. The procedure also produces single-point-in-time estimates of indicators at each time point (i.e., for each value of ROUND), along with the corresponding SEs, 95% CIs, and DEFFs (by specifying /STATISTICS SE CIN (95) DEFF). In the syntax below, anything in bold is a variable name or dataset name that the user needs to specify.

```
CSTABULATE  
/PLAN FILE='file path for DATASETNAME.csaplan'  
/TABLES VARIABLES = OUTCOME BY ROUND  
/CELLS ROWPCT  
/STATISTICS SE CIN (95) DEFF  
/TEST INDEPENDENCE
```

If one is undertaking an analysis of an **indicator of proportions** using a **descriptive PBS**, datasets at each time point are analyzed individually and therefore there is no need to combine the datasets or create the variable ROUND. No statistical test is performed because the test of differences over two time points is not relevant. In this case, the same syntax as above can be used but reference to BY ROUND can be eliminated in the /TABLES line. Additionally, the /CELLS ROWPCT and the /TEST INDEPENDENCE lines can be eliminated. Using the following SPSS syntax, one can produce a single-point-in-time estimate of the proportion indicator, along with the corresponding SE, 95% CI, and DEFF.

```
CSTABULATE  
/PLAN FILE='file path for DATASETNAME.csaplan'  
/TABLES VARIABLES = OUTCOME  
/STATISTICS SE CIN (95) DEFF
```

B.2.2 Analyzing Indicators of Means for Either Descriptive or Comparative Analytical PBSs

If one is undertaking an analysis of an **indicator of means** over time using a **comparative analytical PBS**, the first step is to combine the two PBS datasets corresponding to the two time points by appending one onto the end of the other. The two datasets should have common variables **STRATUMVAR**, **CLUSTERVAR**, **WEIGHTVAR**, and **OUTCOME**. The combined dataset must include a variable **ROUND** that identifies whether a record comes from the first or the second PBS occasion. For the purposes of the example below, we use ROUND = 0 and ROUND = 1.

CSDESCRIPTIVES in SPSS produces single-point-in-time estimates of an indicator of means at each time point, along with the corresponding SEs, 95% CIs, and DEFFs at each time point (i.e., for each value of ROUND). The procedure CSDESCRIPTIVES is run prior to CSGLM (syntax follows) because the former procedure provides estimates of confidence limits while the latter does not. The latter provides the results of a test of differences only. In the syntax below, anything in bold is a variable name or dataset name that the user needs to specify.

```
CSDESCRIPTIVES  
/PLAN FILE='file path for DATASETNAME.csaplan'  
/SUMMARY VARIABLES = OUTCOME  
/SUBPOP TABLE=ROUND DISPLAY = LAYERED  
/MEAN  
/STATISTICS SE CIN (95) DEFF
```

CSGLM in SPSS uses the t-test to statistically test the difference between an indicator of means over two time points under a complex design. For more information, see:

https://www.ibm.com/support/knowledgecenter/en/SSLVMB_23.0.0/spss/complex_samples/syn_csglm.html and <http://www-01.ibm.com/support/docview.wss?uid=swg21477793>.

CSGLM OUTCOME BY ROUND

```
/PLAN FILE='file path for DATASETNAME.csaplan'  
/STATISTICS SE TTEST CINTERVAL DEFF
```

If one is undertaking an analysis of an **indicator of means** using a **descriptive PBS**, datasets at each time point are analyzed individually and therefore there is no need to combine the datasets or to create the variable ROUND. No statistical test is performed because the test of differences over two time points is not relevant. In this case, the same syntax as above for CSDESCRIPTIVES can be used, but the /SUBPOP line can be eliminated. Using the following SPSS syntax, one can produce a single-point-in-time estimate of the mean indicator, along with the corresponding SE, 95% CI, and DEFF.

CSDESCRIPTIVES

```
/PLAN FILE='file path for DATASETNAME.csaplan'  
/SUMMARY VARIABLES = OUTCOME  
/MEAN  
/STATISTICS SE CIN (95) DEFF
```

B.3 Syntax for STATA Software Users

SVYSET in STATA is used for survey datasets and specifies variable(s) for stratification, sampling weighting, and/or cluster variables at the various stages of sampling. For example, if a three-stage sampling design has been used, one specifies the design as follows⁹¹:

```
SVYSET CLUSTERVAR [PW=WEIGHTVAR], STRATA(STRATUMVAR)92
```

After declaring the survey design, the command **SVYDESCRIBE** gives summary details on the survey design.

Once the survey design has been defined, analysis can take place and any of the survey estimation commands can be used by including “SVY:” before the estimation command. In what follows, the syntax is provided separately for analyzing indicators of proportions and indicators of means for both descriptive surveys (described in Chapter 11) and comparative analytical surveys (described in Chapter 12). For more information, see: <https://www.stata.com/manuals13/svy.pdf>.

⁹¹ Note that commands in STATA should be entered in lowercase, not uppercase. They are displayed in uppercase here to distinguish them from the remainder of the text only.

⁹² Some simplifying assumptions are used here: Clusters are considered to be selected with replacement from the first stage strata. Multi-stage sampling within selected clusters is ignored for the purposes of variance estimation, and clusters of observations are assumed instead. This greatly simplifies variance estimation because clusters within strata are assumed to be the dominant source of variance in sample estimates. Any finite population correction for the first stage sample is ignored. The resulting estimates of sampling variance will be slight overestimates in this case, however.

B.3.1 Analyzing Indicators of Proportions for Either Descriptive or Comparative Analytical PBSs

If one is undertaking an analysis of an **indicator of proportions** over time using a **comparative analytical PBS**, the first step is to combine the two PBS datasets corresponding to the two time points by appending one onto the end of the other. The two datasets should have common variables **STRATUMVAR**, **CLUSTERVAR**, **WEIGHTVAR**, and **OUTCOME**. The combined dataset must include a variable **ROUND** that identifies whether a record comes from the first or the second time point. For the purposes of the example below, we use $ROUND = 0$ and $ROUND = 1$.

In STATA, the APPEND command is used to add new observations to an existing dataset, assuming both datasets are saved as STATA data files. Load the first dataset (DATASET1) in STATA then use the following command:

```
APPEND USING DATASET2.DTA
```

TABULATE in STATA uses the Pearson chi-square with the second-order Rao-Scott correction (by specifying PEARSON in the syntax) to statistically test the difference in proportions between PBSs over two time points under a complex design. This also produces single-point-in-time estimates of indicators at each time point (i.e., for each value of ROUND), along with the corresponding 95% CIs and SEs (by specifying CI SE in the syntax). In the syntax below, anything in bold is a variable name or dataset name that the user needs to specify.

```
SVY: TABULATE OUTCOME ROUND, PERCENT ROW CI SE PEARSON
```

To provide an example of typical output, the same synthetic dataset is used as before, with 4,294 observations (2,147 for each of the two PBS occasions), 2 strata per PBS occasion, and 128 clusters or EAs per PBS occasion, and where OUTCOME is a dichotomous variable for which estimates of proportions are desired. The output from the above STATA code when applied to the dataset is given below. When reading the output, the user should report the proportions (and associated SEs and CIs) for which the OUTCOME has value 1 (highlighted). For $ROUND = 0$, the estimate of the proportion is 60.56%, the SE is 2.83, and the CI is (54.84%, 66.00%). For $ROUND = 1$, the estimate of the proportion is 88.1%, the SE is 2.26, and the CI is (82.84%, 91.91%). The (second order) Rao-Scott chi-square statistic (highlighted and labeled as “design-based F”) is 64.29, which is found to be highly statistically significant ($P = 0.0000$). This implies that there is a statistically significant difference in the proportions between $ROUND = 0$ and $ROUND = 1$.

```

Number of strata = 2
Number of PSUs = 128
Number of obs = 4294
Population size = 10418588
Design df = 126

```

ROUND	OUTCOME		Total
	0	1	
0	39.44 (2.832) [34, 45.16]	60.56 (2.832) [54.84, 66]	100
1	11.9 (2.266) [8.094, 17.16]	88.1 (2.266) [82.84, 91.91]	100
Total	25.67 (2.107) [21.73, 30.06]	74.33 (2.107) [69.94, 78.27]	100

Key: row percentages
(linearized standard errors of row percentages)
[95% confidence intervals for row percentages]

Pearson:

```

Uncorrected chi2(1) = 426.6567
Design-based F(1, 126) = 64.2943 P = 0.0000

```

If one is undertaking an analysis of an **indicator of proportions** using a **descriptive PBS**, datasets at each time point are analyzed individually and therefore there is no need to combine the datasets or to create the variable ROUND. No statistical test is performed because the test of differences over two time points is not relevant. In this case, the same syntax as above can be used but reference to “ROUND” can be eliminated. Additionally, reference to ROW and PEARSON should be eliminated given that this is a one-way tabulation and no statistical testing will be performed. Using the following STATA syntax, one can produce a single-point-in-time estimate of the indicator of means, along with the corresponding CI and SE:

```
SVY: TABULATE OUTCOME, PERCENT CI SE
```

To provide an example of typical output, the same synthetic dataset is used, but for one time point only (ROUND = 0); that is, with 2,147 observations, 2 strata, and 128 clusters or EAs. The output from the above STATA code when applied to the dataset is given below. When reading the output, the user should report the percentage, SE (se), and CI (lb and ub) for the row in which the OUTCOME = 1 (highlighted). The estimate of proportion is 60.56%, the SE is 2.83, and the CI is (54.84%, 66.00%).

```

Number of strata = 2
Number of PSUs = 128
Number of obs = 2147
Population size = 5209293.9
Design df = 126

```

OUTCOME	percentages	se	lb	ub
0	39.44	2.832	34	45.16
1	60.56	2.832	54.84	66
Total	100			

```

Key: percentages = cell percentages
     se          = linearized standard errors of cell percentages
     lb          = lower 95% confidence bounds for cell percentages
     ub          = upper 95% confidence bounds for cell percentages

```

B.3.2 Analyzing Indicators of Means for Either Descriptive or Comparative Analytical PBSs

If one is undertaking an analysis of an **indicator of means** over time using a **comparative analytical PBS**, the first step is to combine the two PBS datasets corresponding to the two time points by appending one onto the end of the other. The two datasets should have common variables **STRATUMVAR**, **CLUSTERVAR**, **WEIGHTVAR**, and **OUTCOME**. The combined dataset must include a variable **ROUND** that identifies whether a record comes from the first or the second time point. For the purposes of the example below, we use **ROUND = 0** and **ROUND = 1**.

In STATA, the **APPEND** command is used to add new observations to an existing dataset, assuming both datasets are saved as STATA data files. Load the first dataset (**DATASET1**) in STATA then use the following command:

```
APPEND USING DATASET2.DTA
```

MEAN in STATA produces single-point-in-time estimates of indicators, along with the corresponding SEs and 95% CIs at each time point (i.e., for each value of **ROUND**). The procedure **MEAN** is run prior to **LINCOM** (syntax follows) because the former procedure provides estimates of confidence limits at each time point while the latter does not. The latter provides the results of a test of differences only. In the syntax below, anything in bold is a variable name or dataset name that the user needs to specify.

```
SVY: MEAN OUTCOME, OVER (ROUND)
```

To provide an example of typical output, the same synthetic dataset is used as before, with 4,294 observations (2,147 for each of the two PBS occasions), 2 strata per PBS occasion, and 128 clusters or EAs per PBS occasion, and where **OUTCOME** is a continuous variable for which estimates of means are desired. The output from the above STATA code when applied to the dataset is given below. For **ROUND = 0**, the estimate of the mean is 13.70, the SE is 0.07, and the CI is (13.56, 13.85). For

ROUND = 1, the estimate of the mean is 13.76, the SE is 0.07, and the CI is (13.62, 13.90). These values are highlighted in the output.

```
Number of strata =      2      Number of obs   =      4294
Number of PSUs  =     128     Population size = 10418588
                                   Design df      =      126
```

```
0: time = 0
1: time = 1
```

Over	Mean	Linearized Std. Err.	[95% Conf. Interval]	
OUTCOME				
0	13.70688	.0740139	13.56041	13.85336
1	13.76454	.0726102	13.62085	13.90823

LINCOM in STATA uses the t-test to statistically test the difference in means between PBSs over two time points under a complex survey design.

```
LINCOM [OUTCOME]0- [OUTCOME]1
```

The STATA syntax is applied to the same synthetic dataset as before. To correctly interpret whether the mean OUTCOME variable significantly differs between ROUND = 0 and ROUND = 1, one notes that the negative value of the estimate (Coef. = -0.05) means that the value of OUTCOME was lower for ROUND = 0 than for ROUND = 1; however, the difference is not statistically significant (Pr = 0.186). Furthermore, the 95% CI includes the value 0, confirming that the difference between the two means is not statistically significant.

Mean	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
(1)	-.0576553	.0433916	-1.33	0.186	-.143526	.0282154

If one is undertaking an analysis of an **indicator of means** using a **descriptive PBS**, datasets at each time point are analyzed individually and therefore there is no need to combine the datasets or to create the variable ROUND. No statistical test is performed because the test of differences over two time points is not relevant. In this case, the LINCOM statement can be dropped. The same syntax as above for MEAN can be used, but OVER (ROUND) should not be included given that data from only one time point are being analyzed. Using the following STATA syntax, one can produce a single-point-in-time estimate of the indicator of means, along with the corresponding SE and CI:

```
SVY: MEAN OUTCOME
```

To provide an example of typical output, the same synthetic dataset is used, but for one time point only (ROUND = 0); that is, with 2,147 observations, 2 strata, and 128 clusters or EAs. The output from the

above STATA code when applied to the dataset is given below. The estimate of mean is 13.70, the SE is 0.07, and the CI is (13.56, 13.85).

```
Number of strata =      2      Number of obs   =    2147
Number of PSUs   =    128     Population size = 5209294
                                   Design df       =    126
```

	Mean	Linearized Std. Err.	[95% Conf. Interval]	
OUTCOME	13.70688	.0740139	13.56041	13.85336