



**FEED<sup>THE</sup>FUTURE**

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

# Survey Implementation

## Document

### Guide to

### Feed the Future Statistics for Phase Two Zone of Influence Midline Surveys

### Zone of Influence Survey

September 2023



**USAID**  
FROM THE AMERICAN PEOPLE

This guide is made possible by the support of the American People through the U.S. Agency for International Development, under the terms of OASIS IDIQ No. GS00Q14OADU209, Task Order No. 7200AA21M00009. The opinions expressed herein are those of ICF and do not necessarily reflect the views of the U.S. Agency for International Development or the United States Government.

**Recommended Citation**

ICF. 2023. *Guide to Feed the Future Statistics for Phase Two Zone of Influence Midline Surveys*. Reston, VA: ICF

**Contact Information**

Feed the Future  
1300 Pennsylvania Ave, NW  
Washington, DC 20004  
[www.feedthefuture.gov](http://www.feedthefuture.gov)  
[rfs.ald@usaid.gov](mailto:rfs.ald@usaid.gov)

## ACKNOWLEDGMENTS

Feed the Future is grateful to all who have made essential contributions to updating and expanding the *Guide to Feed the Future Statistics* for phase two Zone of Influence midline analyses and creating this guide: Johanna Andrews-Trevino, Lindsey Anna, Mahmoud Elkasabi, Patricia Freitag, Madeleine Gauthier, Tesfayi Gebreselassie, Nathan Ives, Azam Khan, Nizam Khan, Mandy McCleary, Rachel Orlowski, James Parr, Mousumi Sakar, Michelle Saunders, Anne Swindale, and Kirsten Zalisk.

## TABLE OF CONTENTS

Acknowledgments .....	i
Abbreviations.....	ix
Part I. Introduction to Phase Two Zone of Influence Midline survey data and variables.....	1
1. Introduction.....	2
1.1. Feed the Future population-based P2-ZOI indicators at midline .....	3
1.2. Feed the Future P2-ZOI Midline Survey core questionnaires.....	5
1.3. Updates to the Guide to Statistics for P2-ZOI Midline Surveys.....	6
1.3.1. Deletions .....	6
1.3.2. Revisions.....	7
1.3.3. Additions .....	7
1.3.4. Forthcoming .....	8
1.4. Results reporting.....	8
2. Organization of ZOI Survey data .....	9
2.1 Structure of the Feed the Future ZOI Survey data files.....	9
2.1.1 Hierarchical files .....	10
2.1.2 Flat files.....	10
2.1.3 Formats and naming convention of datasets exported from CSPro .....	10
2.1.4 Matching relationships.....	11
2.2 Variable name conventions .....	12
2.2.1 General variable naming .....	12
2.2.2 Select-all questions.....	15
2.2.3 Special coding conventions .....	16
2.2.4 Other variables exported from CSPro .....	17
3. Data analysis information .....	18
3.1 Sample design, sampling weights, and response rates .....	18
3.1.1 Sample design .....	18
3.1.2 Sample weighting .....	20
3.1.3. Response rates.....	22
3.2 Descriptive statistics and statistical tests of difference.....	26
3.2.1 Analyzing indicators of proportions .....	28
3.2.2 Analyzing indicators of means .....	38
3.2.3 Analyzing indicators of proportions or means if the data are not available.....	47
3.3 Data analysis folder structure and file naming conventions.....	49
3.3.1 Folder structure .....	49
3.3.2 File naming conventions.....	50
4. Key individual-level analytic variables .....	52

4.1	Household member status (de jure, de facto) and sex variables.....	52
4.1.1	De jure household member .....	52
4.1.2	De facto household member.....	53
4.1.3	Sex .....	53
4.2	Children under 5 years age-in-month variables.....	54
4.2.1	Age in days, children under 5 years of age.....	54
4.2.2	Age in months, children under 5 years of age.....	56
4.2.3	Children 0-5 months of age, overall and by sex .....	57
4.2.4	Children 6-23 months of age, overall and by sex.....	57
4.2.5	Children 0-23 months of age, overall and by sex.....	58
4.2.6	Children 0-59 months of age, overall and by sex.....	59
4.2.7	Children 6-23 months age category—6 month intervals .....	60
4.2.8	Children 0-59 months age categories—0-23 and 24-59 months .....	61
4.3	Household member age-in-year variables .....	62
4.3.1	Age in years .....	62
4.3.2	Adults (18 years of age or older), overall and by sex .....	63
4.3.3	Youth (15-29 years of age), overall and by sex.....	64
4.3.4	Children 5-17 years of age in household roster, overall and by sex .....	64
4.4	Women of reproductive age variables.....	65
4.4.1	Women of reproductive age (15-49 years of age).....	65
4.4.2	Women of reproductive age's age category—5-year age categories.....	66
4.4.3	Women of reproductive age's age categories—15-19 years and 20-49 years.....	67
4.5	Primary adult decision-maker variables .....	67
4.5.1	Primary adult decision-makers, by sex.....	67
4.5.2	De jure household member, primary adult decision-makers, by sex.....	68
4.5.3	Age category, primary adult female decision-makers (de jure only) .....	69
4.5.4	Youth, primary adult female decision-makers (de jure only) .....	70
4.5.5	Marital status, primary adult female decision-makers (de jure only).....	70
4.5.6	Participate in agricultural production, non-farm work, and wage or salary work, primary adult female decision-makers (de jure only) .....	71
4.6	Producer variables .....	73
4.6.1	Targeted VCC producer, any VCC .....	73
4.6.2	Targeted VCC producer, specific VCCs .....	74
4.6.3	Targeted VCC producer who completed relevant agriculture module .....	75
5.	Key household-level variables .....	77
5.1	Household composition .....	77
5.1.1	Number of usual (de jure) household members .....	77
5.1.2	Number of adults who are de jure household members in household, overall and by sex .....	78

5.1.3	Number of women of reproductive age who are de jure household members in household.....	78
5.1.4	Number of children under 2 years of age who are de jure household members in household.....	79
5.1.5	Number of children under 5 years of age who are de jure household members in household.....	79
5.1.6	Number of children 5-17 years of age who are de jure household members in household.....	80
5.1.7	Number of youth 15-29 years of age who are de jure household members in household.....	80
5.1.8	Number of producers of targeted value chain commodities who are de jure household members in household .....	81
5.1.9	Household cultivates or raises targeted VCCs.....	82
5.1.10	Household size category—de jure household members.....	83
5.2	Household-level disaggregates .....	84
5.2.1.	Gendered household type—de jure household members.....	84
5.2.2.	Wealth quintile .....	85
5.2.3.	Poverty status.....	85
5.2.4.	Shock exposure severity .....	85
6.	ZOI population estimates.....	87
6.1	Background .....	87
6.2	Guidelines to estimating the ZOI population, sub-populations, and households .....	88
6.3	Step-by-step instructions for populating the population estimate tables .....	90
Part II.	Indicators.....	94
7.	Demographic indicators.....	95
7.1	Household demographic characteristics.....	95
7.1.1	Mean household size .....	95
7.1.2	Mean number of children under 2 years .....	96
7.1.3	Mean number of children under 5 years .....	96
7.1.4	Mean number of children 5-17 years of age .....	97
7.1.5	Mean number of youth 15-29 years of age .....	97
7.1.6	Mean number of women of reproductive age.....	98
7.1.7	Mean number of adult male household members.....	99
7.1.8	Mean number of adult female household members.....	99
7.1.9	Mean number of producers of any targeted value chain commodity and each targeted value chain commodity.....	100
7.1.10	Percent distribution of households by household size.....	101
7.2	Characteristics of primary adult female decision-makers .....	101
7.2.1	Percent distribution of primary adult female decision-makers by age group .....	102

7.2.2	Percent distribution of primary adult female decision-makers by their current marital status .....	102
7.2.3	Percentage of primary adult female decision-makers who participate in economic activities .....	103
7.2.4	Percentage of primary adult female decision-makers who participate in farm work, non-farm work, and wage or salaried employment .....	104
8.	Water, sanitation, and hygiene indicators .....	106
8.1	Percent of households using an improved water source .....	106
8.2	Percent distribution of households by the number of trips household members made to obtain drinking water during the past 7 days .....	108
8.3	Percent of households using a correct water treatment practice or technology .....	109
8.4	Percent of households with soap and water at a handwashing station on premises (Feed the Future phase two indicator) .....	110
8.5	Percent of households using an improved but shared sanitation facility .....	112
8.6	Percent of households using an improved facility that is not shared (basic sanitation service) (Feed the Future phase two indicator) .....	113
8.7	Percent of households using an unimproved sanitation facility .....	115
8.8	Percent of households practicing open defecation .....	116
8.9	Mean number of household members who regularly defecate in the open .....	117
8.10	HWISE-4 Scale indicator .....	118
9.	Dwelling and household characteristic indicators .....	122
9.1	Percent of households using solid fuels for cooking .....	122
9.2	Mean number of de jure household members per sleeping room .....	123
9.3	Percent distribution of households by dwelling roof materials .....	124
9.4	Percent distribution of households by dwelling exterior wall materials .....	125
9.5	Percent distribution of households by dwelling floor materials .....	126
9.6	Percent of households that have electricity .....	127
9.7	References .....	127
10.	The consumption aggregate and the poverty indicators .....	128
10.1	Guidelines to construct the consumption aggregate and the poverty indicators .....	128
10.1.1	Overview of the consumption aggregate .....	128
10.1.2	Handling missing values and outliers .....	129
10.1.3	Analyzing data comprising the consumption aggregate .....	131
10.1.4	Computing the consumption aggregate .....	136
10.1.5	Poverty indicators .....	137
10.2	Step-by-step procedure to calculate the poverty indicators .....	141
10.2.1	Protocol a: to identify outlier values in the dataset .....	142
10.2.2	Protocol b: to replace identified outliers with an acceptable imputed median value .....	143
10.2.3	Protocol c: to change the dataset from wide to long format .....	144

10.2.4	Food consumption expenditures of the past 1 week (7 days) .....	145
10.2.5	Regular non-food, non-durable goods expenditures of the past 1 week (7 days) ..	157
10.2.6	Occasional non-food, non-durable goods expenditures of the past 1 month.....	161
10.2.7	Occasional non-food, non-durable goods of the past 3 months .....	164
10.2.8	Occasional non-food, non-durable goods of the past 1 year (12 months) .....	167
10.2.9	Consumption expenditures on durable goods owned by the household .....	172
10.2.10	Consumption expenditures on housing.....	176
10.2.11	Calculating the consumption aggregate .....	182
10.2.12	Prevalence of poverty (\$1.90 per day 2011 PPP) indicator .....	184
10.2.13	Depth of poverty of the poor (\$1.90 per day 2011 PPP) indicator .....	186
10.2.14	Percent of people who are 'near-poor' (\$1.90 per day 2011 PPP) indicator .....	187
10.2.15	Prevalence of poverty (\$1.25 per day 2005 PPP) indicator .....	189
10.2.16	Depth of poverty of the poor (\$1.25 per day 2005 PPP) indicator .....	190
10.2.17	Percent of people who are 'near-poor' (\$1.25 per day 2005 PPP) indicator .....	192
10.2.18	Prevalence of poverty (\$2.15 per day 2017 PPP) indicator .....	193
10.2.19	Depth of poverty of the poor (\$2.15 per day 2017 PPP) indicator .....	194
10.2.20	Percent of people who are 'near-poor' (\$2.15 per day 2017 PPP) indicator .....	196
10.3	References .....	197
11.	The comparative wealth index.....	198
11.1	Guidelines to construct Feed the Future's CWI indicator .....	198
11.2	Step-by-step procedures to calculate the CWI indicator .....	203
11.3	References .....	227
12.	Resilience indicators .....	228
12.1	Guidelines to construct the resilience indicators.....	228
12.1.1	Ability to recover from shocks and stresses index.....	228
12.1.2	Index of social capital at the household level .....	230
12.1.3	Percent of households that believe local government will respond effectively to future shocks and stresses .....	232
12.1.4	Percent of households participating in group-based savings, micro-finance, or lending programs .....	232
12.2	Step-by-step procedures to calculate resilience indicators.....	233
12.2.1	Ability to recover from shocks and stresses index.....	233
12.2.2	Index of social capital at the household level .....	239
12.2.3	Percent of households that believe local government will respond effectively to future shocks and stresses .....	242
12.2.4	Percent of households participating in group-based savings, micro-finance, or lending programs .....	243
12.3	References .....	245
13.	The five domains of empowerment.....	247



13.1	Background .....	247
13.2	Guidelines to construct the 5DE.....	248
13.2.1	Part 1: Preparing the data and calculating the A-WEAI indicators.....	248
13.2.2	Part 2: Computing the 5DE .....	253
13.2.3	Part 3: Calculating the 5DE standard error (jackknife approach) and performing a test of difference .....	255
13.2.4	Part 4: Calculating adequate achievement in each A-WEAI indicator.....	257
13.3	Step-by-step instructions for computing the 5DE.....	257
13.3	References .....	293
14.	Agricultural productivity indicator .....	295
14.1	Guidelines to construct the improved management practices and technologies indicator .	295
14.2	Step-by-step procedure to calculate the targeted improved management practices and technologies indicator .....	298
14.2.1	Maize calculations.....	299
14.2.2	Fishpond calculations.....	310
14.2.3	Dairy cow calculations .....	317
14.2.4	Sheep calculations .....	327
14.2.5	Overall indicator calculations.....	335
14.3	References .....	344
15.	Household food insecurity indicator .....	345
15.1	Guidelines to construct the FIES indicator .....	345
15.2	Step-by-step procedures to calculate the FIES indicator .....	348
15.3	References .....	373
16.	Children’s and women’s dietary intake indicators .....	375
16.1	Guidelines to construct the indicators .....	375
16.1.1	Prevalence of exclusive breastfeeding of children under 6 months of age.....	375
16.1.2	Percent of children 6-23 months of age receiving a minimum acceptable diet.....	375
16.1.3	Percent of women of reproductive age consuming a diet of minimum diversity ....	377
16.2	Step-by-step procedures to calculate nutrition indicators.....	377
16.2.1	Prevalence of exclusive breastfeeding of children under 6 months of age.....	377
16.2.2	Percentage of children 6-23 months of age receiving a minimum acceptable diet..	381
16.2.3	Percent of women of reproductive age consuming a diet of minimum dietary diversity .....	389
17.	Children’s and women’s nutritional status indicators .....	396
17.1	Guidelines to construct the indicators using secondary data.....	396
17.1.1	Point-to-polygon geospatial overlay method to identify DHS clusters in the Feed the Future ZOI .....	397
17.1.2	Children’s nutritional status indicators.....	397
17.1.3	Prevalence of underweight women of reproductive age.....	399

17.2	Step-by-step procedures to calculate nutritional status indicators.....	399
17.2.1	Geospatial overlay to identify DHS clusters in the ZOI.....	399
17.2.2	Prevalence of stunted children under 5 years of age.....	403
17.2.3	Prevalence of wasted children under 5 years of age .....	405
17.2.4	Prevalence of healthy weight children under 5 years of age.....	406
17.2.5	Prevalence of underweight women of reproductive age.....	408
17.2.6	Additional children’s anthropometric indicators.....	410
17.2.7	Additional women’s anthropometric indicators.....	412
17.3	References .....	413
18.	Climate adaptation .....	415
18.1	Background .....	415
18.2	Step-by-step instructions for populating the climate tables.....	415
18.2.1	Table 5.5.1: Knowledge, Perception, and Concern about Impacts of Climate Change on Daily Life and Income among Households in the ZOI.....	416
18.2.2	Table 5.5.2: Percent of Households in the ZOI that Received Information to Help Prepare for or Manage Weather-related Challenges During the 12 months Preceding the Survey.....	417
18.2.3	Table 5.5.3: Percent of Households in the ZOI that Received Training or Support or Adopted Practices to Help Manage Weather-related Challenges Related to Crop Production During the 12 Months Preceding the Survey .....	418
18.2.4	Table 5.5.4: Percent of Households in the ZOI that Received Training or Support or Adopted Practices to Manage Weather-related Challenges Related to Raising Livestock or Other Farm Animals During the 12 Months Preceding the Survey....	420
18.2.5	Table 5.5.5: Percent of Households in the ZOI that Faced Barriers to Adopting Crop or Farm Animal Practices to Help Manage the Impacts of Weather-related Problems During the 12 Months Preceding the Survey.....	421
18.2.6	Table 5.5.6: Percent of Households in the ZOI that Participated in a Group or Committee Working to Strengthen the Community’s Ability to Manage Weather-related Problems During the 12 Months Preceding the Survey .....	422
18.2.7	Table 5.5.7: Percent of Households in the ZOI that Obtained Insurance, Type of Insurance and Whether Support was Received, and Whether Insurance for Crops or Farm Animals Was Obtained to Protect Against Weather-related Problems During the 12 Months Preceding the Survey .....	424
18.2.8	Table 5.5.8: Percent of Households in the ZOI that Obtained Payments for Weather-related Problems that Affected Their Crop or Farm Animal Production During the 12 Months Preceding the Survey .....	426
18.2.9	Table 5.5.9: Percent of Households in the ZOI that Participated in a Group or Committee Working to Monitor or Influence Land Use During the 12 Months Preceding the Survey.....	428
18.2.10	Table 5.5.10: Percent of Households in the ZOI that Participated in a Training or Received Support to Help Diversify Their Sources of Income During the 12 Months Preceding the Survey, Feed the Future Phase Two ZOI Midline Survey ...	430

18.3	References .....	432
	Appendix A. Abbreviated Women’s Empowerment in Agriculture Index summary information .....	433
	Appendix B. 5DE confidence interval and design effect calculations .....	436

## LIST OF FIGURES

Figure 1: Setting up a Working Directory in R Studio .....	350
Figure 2: Example Rasch Module Output .....	355

## LIST OF TABLES

Table 1: List of Population-based P2-ZOI Midline Indicators .....	3
Table 2: Modules included in the Core Phase 1 Endline/Phase 2 Baseline and Phase 2 Midline Main Survey Questionnaires .....	5
Table 3: Data File Names by Units of Analysis and File Format.....	10
Table 4: Special Response Categories and Their Codes, by Question Type .....	16
Table 5: HWISE-4 Indicator Questions and Variables.....	119
Table 6: PPP Conversion Factors (Local Currency Unit per International \$) for Private Consumption for Feed the Future Target Countries .....	138
Table 7A: Example of Flat Format Data: Consumption Amount.....	144
Table 7B: Example of Rectangular Format Data: Consumption Amount .....	145
Table 8: Example Weight Unit Conversion Table .....	153
Table 9: Example Volume Unit Conversion Table.....	153
Table 10: Example Count Unit Conversion Table.....	154
Table 11: Unsatisfied Basic Needs Categories and Criteria.....	201
Table 12: Reference Survey Anchoring Point Values .....	201
Table 13: Example Cases that Show Variables with 0 or 1 ‘Yes’ Responses .....	214
Table 14: Frequencies and Percentages by UBN Score .....	221
Table 15: A-WEAI Domains, Indicators, and Indicator Weights .....	247
Table 16: Activities Included in the A-WEAI Time Use Module.....	253
Table 17: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Maize Cultivation, by Category .....	300
Table 18: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Fishpond Aquaculture, by Category.....	312
Table 19: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Raising Dairy Cows, by Category.....	319
Table 20: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Raising Sheep, by Category.....	328
Table 21: Eight Food Groups Used to Calculate Minimum Dietary Diversity for Children .....	382
Table 22: Ten Food Groups Used to Generate the Women’s Food Score .....	390
Table 23: Z-score Cutoffs for Children’s Nutritional Status Indicators.....	398
Table 24: Women’s Nutritional Status Category by BMI .....	399
Table A1: Summary of Abbreviated Women’s Empowerment in Agriculture Index Domains, Indicators, Survey Questions, Variables, Definitions, and Weights.....	433

## ABBREVIATIONS

5DE	five domains of empowerment
A-WEAI	Abbreviated Women's Empowerment in Agriculture Index
ARSSI	ability to recover from shocks and stresses index
ATR	ability to recover
AWI	asset-based wealth index
BMI	body mass index
CI	confidence interval
CPI	Consumer Price Index
CSPro	Census and Survey Processing System
CWI	comparative wealth index
DEFF	design effect
DHS	Demographic and Health Surveys
EA	enumeration area
FAO	Food and Agriculture Organization of the United Nations
FIES	Food Insecurity Experience Scale
GNI/p	gross national income per capita
HAZ	height-for-age z-score
HWISE-4	Brief Household Water Insecurity Experiences
JRR	jackknife repeated replication
LCU	local currency unit
MOE	margin of error
P2-ZOI	phase two Zone of Influence
PCA	principal component analysis
PCD	per capita daily
PPP	purchasing power parity

PSU	primary sampling unit
RFS	Bureau for Resilience and Food Security
SD	standard deviation
SEI	shock exposure index
UBN	Unsatisfied Basic Needs
USAID	United States Agency for International Development
VCC	value chain commodity
WHO	World Health Organization
WHZ	weight-for-height z-score
ZOI	Zone of Influence

**PART I.**  
**INTRODUCTION TO PHASE TWO ZONE OF INFLUENCE**  
**MIDLINE SURVEY DATA AND VARIABLES**

## I. INTRODUCTION

Feed the Future seeks to sustainably reduce poverty, hunger, and malnutrition among women and children; strengthen agriculture and food systems; increase income, resilience, women's empowerment, dietary diversity, and appropriate feeding practices; and improve hygienic environments. Program efforts are designed to impact the population in Zones of Influence (ZOIs) in Feed the Future target countries. Progress in achieving Feed the Future's objectives is tracked using population-based ZOI-level indicators. Feed the Future ZOI Surveys provide U.S. Government interagency partners, the United States Agency for International Development (USAID) Bureau for Resilience and Food Security, USAID Missions, host country governments, and development partners with information on the status of the population-based ZOI-level indicators.

Phase two ZOI (P2-ZOI) Midline Surveys (hereafter referred to as "Midline Surveys"), which are being implemented between 2022 and 2024, are designed to collect data to monitor the short-term progress of population-based ZOI-level indicators. Midline Surveys are not designed to detect statistically significant changes in indicator values since baseline. However, it will be possible to detect statistically significant differences if the difference between baseline and midline values is large enough.

The priority of ZOI Surveys is to collect high-quality data that are reliable and comparable across all Feed the Future countries. To ensure comparable information across Midline Surveys, Feed the Future developed the *Feed the Future ZOI Survey Methods Toolkit—Midline* (hereafter referred to as "Midline Toolkit"),<sup>1</sup> which provides detailed guidance and templates covering all aspects of the planning and implementation of these surveys and analysis of the data that are collected, ensuring that robust methods are implemented across Midline Surveys.

This *Guide to Feed the Future Statistics for Phase Two Zone of Influence Midline Surveys* is part of the Midline Toolkit. It contains a detailed description of the process and steps needed to analyze Midline Survey data and calculate the population-based ZOI-level indicators and related statistics. The Guide is divided into two parts:

Part I provides an overview of the organization and use of the Midline Survey data, describes the process for weighting and performing statistical tests on the data, and defines both key variables used to calculate indicators and variables by which indicator results are disaggregated in the midline report.

---

<sup>1</sup> Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods-toolkit-midline-2021>



Part II is devoted to the population-based ZOI-level indicators and related statistics. Each chapter in Part II focuses on a specific indicator or set of related indicators and provides the precise definition of the indicators and then the step-by-step process to calculate these indicators and statistics using data collected with the Midline Survey core questionnaires included in the Midline Toolkit.<sup>2</sup> The final chapter in Part II focuses on the analysis of Demographic and Health Survey data to calculate the population-based anthropometry indicators for the ZOI. Findings related to demographic and dwelling characteristics, program participation, and climate adaptation, as well as the brief household water insecurity scale indicator, which are not Feed the Future indicators but are important descriptors of ZOI populations and reported in midline indicator assessment reports, are also described in Part II of this Guide.

This Guide should be used in conjunction with the *Feed the Future Indicator Handbook*. In addition, Stata syntax files to compute ZOI-level indicators and other key statistics to populate the results table shells, which are available as a complement to this Guide. Both the *Indicator Handbook* and syntax files are available in the Midline Toolkit.

### 1.1. Feed the Future population-based P2-ZOI indicators at midline

This Guide pertains to a subset of all Feed the Future indicators and is concerned only with the population-based ZOI-level indicators relevant to P2-ZOI Midline Surveys.<sup>3</sup> The midline data analysis entails generating midline indicator estimates and comparing these estimates to baseline estimates when data are available for both points in time.<sup>4</sup> **Table I** presents the ZOI-level indicators that are calculated at midline.

**Table I: List of Population-based P2-ZOI Midline Indicators**

Indicators calculated using Midline Survey data	
1	Prevalence of moderate and severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES)
2	Percent of households below the comparative threshold for the poorest quintile of the Asset-Based Comparative Wealth Index
3	Percent of producers who have applied targeted improved management practices or technologies
4	Percent of households with access to a basic sanitation service
5	Percent of households with soap and water at a handwashing station on premises
6	Percent of children 6-23 months receiving a minimum acceptable diet
7	Prevalence of exclusive breastfeeding of children under 6 months of age
8	Prevalence of women of reproductive age consuming a diet of minimum diversity

<sup>2</sup> There are two Midline Survey questionnaires—one for the main survey and one for the parallel survey. All indicator estimates and other results reported are calculated using data collected using the main survey questionnaire. The parallel survey data are used only to validate the survey-to-survey imputation model used to calculate the poverty indicators. The questionnaires in the Midline Toolkit will be customized for each survey; therefore, the analysis syntax will also have to be customized, as needed, to reflect the customization.

<sup>3</sup> Information on the universe of Feed the Future indicators is available at:  
[https://www.agrilinks.org/sites/default/files/revised\\_ftf\\_indicator\\_handbook\\_clean\\_version\\_20190926.pdf](https://www.agrilinks.org/sites/default/files/revised_ftf_indicator_handbook_clean_version_20190926.pdf)

<sup>4</sup> To the extent possible, baseline estimates should be recalculated as part of the midline analysis using the most updated syntax used to analyze midline data to ensure maximum comparability.

Indicators calculated using Midline Survey data	
9	Ability to recover from shocks and stresses index
10	Index of social capital at the household level
11	Percent of households that believe local government will respond effectively to future shocks and stresses
12	Percent of households participating in group-based savings, micro-finance, or lending programs
13	Percent of households that are water secure, based on the Brief Household Water Insecurity Experiences (HWISE-4) scale
Indicator components calculated using Midline Survey data	
14	Abbreviated Women's Empowerment in Agriculture Index (A-WEAI) Five Domains of Empowerment
	Five Domains of Empowerment score
	Percent of women who are empowered
	Average adequacy score of disempowered women
	Percent of disempowered women who are adequate in each A-WEAI indicator
Indicators calculated using secondary data	
15	Prevalence of stunted (HAZ < -2) children under 5 (0-59 months)
16	Prevalence of wasted (WHZ < -2) children under 5 (0-59 months)
17	Prevalence of underweight (BMI < 18.5) women of reproductive age
18	Prevalence of healthy weight (WHZ ≤ 2 and ≥ -2) among children under 5 (0-59 months)
Indicators calculated using survey-to-survey imputation methods	
19	Prevalence of poverty: Percent of people living on less than \$1.90/day 2011 PPP
20	Depth of poverty of the poor: Mean percent shortfall of the poor relative to the \$1.90/day 2011 PPP poverty line
21	Percent of people who are 'near-poor', living on 100 percent to less than 125 percent of the \$1.90 2011 PPP poverty line

Missions may add questions to the Midline Survey questionnaires to calculate a limited number of additional population-based indicators. For example, a ZOI-level market-related outcome indicator could be added to assist a USAID Mission in linking beneficiary-level outcomes to population-level impacts and tracking progress toward higher-level impacts. Generally, questions included in the Midline Survey should have been included in the Baseline Survey to enable comparison of indicator estimates over time.

Midline Surveys, however, include a number of questions that were not included in Baseline Surveys—to calculate additional descriptive statistics related to water and sanitation, climate change knowledge and adaptation services, knowledge and effects of COVID-19, and participation in programs of the type that Feed the Future promotes in the survey country.

Guidance to calculate results related to the new climate and water and sanitation questions included in Midline Surveys is included in this Guide, but guidance to calculate Mission-specific indicators and descriptive statistics related to COVID-19 and program participation questions is not included in this Guide.<sup>5</sup>

<sup>5</sup> Information relevant to calculating descriptive statistics for the program participation questions in the core Midline Survey is included in the core results table shells available in the Midline Toolkit. The information includes relevant core questionnaire question numbers, sampling weights to apply, numerators, and denominators.

## 1.2. Feed the Future P2-ZOI Midline Survey core questionnaires

The Feed the Future ZOI Survey Methods Toolkits include core phase one Endline/phase two Baseline and phase two Midline Survey questionnaires and guidance on how to adapt the core questionnaires to the country context by incorporating country-specific education levels, assets, foods, shocks and stresses, agricultural practices, and program participation questions.

There are two Midline Survey questionnaires: a main survey questionnaire and a parallel survey questionnaire. The main survey questionnaire is used to collect data to calculate ZOI-level indicator estimates at midline. The parallel survey questionnaire is used to collect consumption expenditure and crop yield data,<sup>6</sup> but these data are not used to directly calculate indicator estimates. The consumption expenditure data are used to validate a baseline prediction model used for survey-to-survey imputation and provide a backup method to compute midline estimates for poverty indicators if needed. The crop yield data are shared with USAID for separate analyses.

**Table 2** presents the modules included in the core questionnaires in addition to household identification and informed consent information.

**Table 2. Modules included in the Core Phase I Endline/Phase 2 Baseline and Phase 2 Midline Main Survey Questionnaires**

Module	Endline/Baseline	Midline Main	Midline Parallel <sup>a</sup>
1	Household roster and demographics	Household roster and demographics	Household roster and demographics
2	Dwelling characteristics  Includes household assets and WASH	Dwelling characteristics and household assets  Includes expenditures on utilities, WASH, and HWISE	Dwelling characteristics and household assets  Streamlined version of main survey version that does not include utility expenditures or all dwelling characteristic questions
2A	n/a	Climate adaptation  Includes climate change knowledge and services accessed	n/a
3	Food security and resilience	Food security and resilience  Includes COVID-19	n/a
4	Women's nutrition and anthropometry	Women's nutrition  Includes women's health and nutrition program participation	n/a
5	Children's nutrition and anthropometry	Children's nutrition  Includes children's health and nutrition program participation	n/a

<sup>6</sup> The parallel survey is always used to collect consumption expenditure data, but crop yield data may not be collected; the data collected depend on the midline indicator assessment scope of work agreed to by USAID and the contractor implementing the assessment.

Module	Endline/Baseline	Midline Main	Midline Parallel <sup>a</sup>
6	Women's Empowerment in Agriculture (women and men)	Women's Empowerment in Agriculture (women only)	n/a
7	Agricultural technologies  Includes questions about targeted value chain practices and production, plot measurement, and soil assessment	Agricultural technologies  Includes questions about only targeted value chain practices	Agricultural technologies  Includes question about only targeted crop value chain yield (plot information and measurement)
8	Household consumption expenditures	n/a	Household consumption expenditures
9	n/a	Program participation <sup>b</sup>	n/a

HWISE=Household Water Insecurity Experiences, n/a=not applicable, WASH=water, sanitation, and hygiene

<sup>a</sup> The parallel survey is used to collect data to validate the survey-to-survey imputation model used to compute the consumption aggregate, which is used to estimate the three Feed the Future P2-ZOI poverty indicators.

<sup>b</sup> Although Module 9 in the core P2-ZOI Midline Survey questionnaire includes water, sanitation, and hygiene; agriculture; and resilience program participation questions, during the questionnaire customization process, the questions are moved to other survey modules—most to Modules 2 and 3. Health and nutrition program participation questions do not have to be moved; women's health and nutrition program participation questions are included in Module 4 of the core questionnaire, and children's health and nutrition program participation questions are included in Module 5 of the core questionnaire.

### 1.3. Updates to the Guide to Statistics for P2-ZOI Midline Surveys

This Guide has been developed from the *Guide to Feed the Future Statistics* available in the *Feed the Future ZOI Survey Methods Toolkit—Baseline*<sup>7,8</sup> and incorporates a number of updates, which are summarized in this section.

#### 1.3.1. Deletions

The following phase one indicators are not phase two indicators and have been removed:

- Depth of poverty: Mean percentage shortfall relative to the \$1.25/day (2005 PPP) poverty line
- Daily per capita expenditures (as a proxy for incomes) (2010 USD)
- Women dietary diversity: Mean number of food groups consumed by women of reproductive age (WDDS)
- Prevalence of underweight (WAZ<-2) children under age 5 years
- Prevalence of moderate and severe hunger in the population, based on the Household Hunger Scale (HHS)

The following phase two indicator is not calculated at midline and has been removed:

- Yield of targeted agricultural commodities within target areas

Key analytic variables and disaggregate variables that are not required for the midline analysis have been removed from Chapter 4 and Chapter 5.

<sup>7</sup> <https://agrilinks.org/post/feed-future-zoi-survey-methods>

<sup>8</sup> First made publicly available in August 2019 and last updated in December 2021

### **I.3.2. Revisions**

The information on sampling design, sampling weights, and response rates has been expanded (see Section 3.1.)

Updates to the methods used to calculate the following indicators have been reflected:

- Percent of households with soap and water at a handwashing station on premises (see Section 8.4)
- Percent of households below the comparative threshold for the poorest quintile of the Asset-Based Comparative Wealth Index (see Chapter 11)
- Ability to recover from shocks and stresses index (see Section 12.1.1 and Section 12.2.1)
- Index of social capital at the household level (see Section 12.1.2. and Section 12.2.2)
- Percent of households participating in group-based savings, micro-finance, or lending programs (see Section 12.1.4 and Section 12.2.4)
- Percent of children ages 6-23 months receiving a minimum acceptable diet (see Section 16.1.2 and Section 16.2.2)
- Prevalence of moderate and severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES) (see Chapter 15)

Chapter 13 on the Abbreviated Women's Empowerment in Agriculture Index (A-WEAI) has been revised to include the calculation of only the five domains of empowerment and its components. A-WEAI data are not collected from primary adult male decision-makers at midline, so the gender parity index and A-WEAI score at midline cannot be calculated. The chapter presents how to calculate the five domains of empowerment, the percentage of women who are empowered, the average adequacy score of the disempowered, and the percentage of women with adequate in each A-WEAI indicator among three populations: (1) only women with data for all six A-WEAI indicators, (2) only women who are disempowered, and (3) all women. The chapter also presents how to create additional variables required for the tables in Section 6.4 of the midline indicator assessment report.

Chapter 17 on the anthropometry indicators has been revised to describe the calculation of the Feed the Future ZOI-level anthropometry indicators using Demographic and Health Survey data rather than ZOI Survey data.

### **I.3.3. Additions**

The following new indicator, although not a Feed the Future indicator, is calculated at midline and has been added:

- Percent of households that are water secure, based on the Brief Household Water Insecurity Experiences (HWISE-4) scale (see Section 8.10)

Information describing the calculation of the following additional water and sanitation statistics that are new at midline has been added:

- Percent distribution of households by the number of trips household members made to obtain drinking water during the past seven days (see Section 8.2)

- Mean number of household members who regularly defecate in the open (see Section 8.9)

A chapter to describe the calculation of the descriptive climate knowledge and adaptation findings has been added (see Chapter 18).

Information on calculating the three poverty indicators at the previous international threshold (\$1.25 per day at 2005 purchasing power parity) and the new international poverty threshold (\$2.15 per day at 2017 purchasing power parity) has been added (see Chapter 10).

#### **1.3.4. Forthcoming**

The following updates are forthcoming in a future edition of this Guide:

- Guidance on how to estimate the consumption aggregate and calculate the three poverty indicators at midline using survey-to-survey imputation

### **1.4. Results reporting**

For the ZOI Surveys, design weights and sampling weights are calculated for individuals and households. Point estimates for all indicators and disaggregates are calculated and tabulated, and each point estimate is presented with the unweighted N, standard deviation (if applicable), 95 percent confidence interval, design effect, and indicator incompleteness rate. In P2-ZOI baseline reports, statistical tests of differences are conducted to determine if there is a bivariate association between the indicator and selected background characteristics (e.g., gendered household type, wealth quintile, shock exposure severity, and household education). In P2-ZOI midline reports, statistical tests of difference are conducted between P2-ZOI baseline and midline values to determine if there was a statistically significant change in indicator estimates over time. However, it is important to note that P2-ZOI Midline Surveys are not designed to detect statistically significant changes in indicator values since the P2-ZOI Baseline Survey, but it will be possible to detect statistically significant differences if the difference between baseline and midline values is large enough.

ZOI Survey findings are presented in the survey reports. Standardized templates for the phase two baseline report and phase two midline report can be found in the *Feed the Future ZOI Survey Methods Toolkits*.<sup>9</sup>

After USAID Mission and the Bureau for Resilience and Food Security approval of the survey reports, the survey contractor prepares and submits public, restricted, and non-public access datasets and supporting materials according to Feed the Future's *Protocol for Preparing Non-Public, Restricted, and Public Access Datasets* (available as Item 4.1 in the Midline Toolkit).

---

<sup>9</sup> Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>

## 2. ORGANIZATION OF ZOI SURVEY DATA

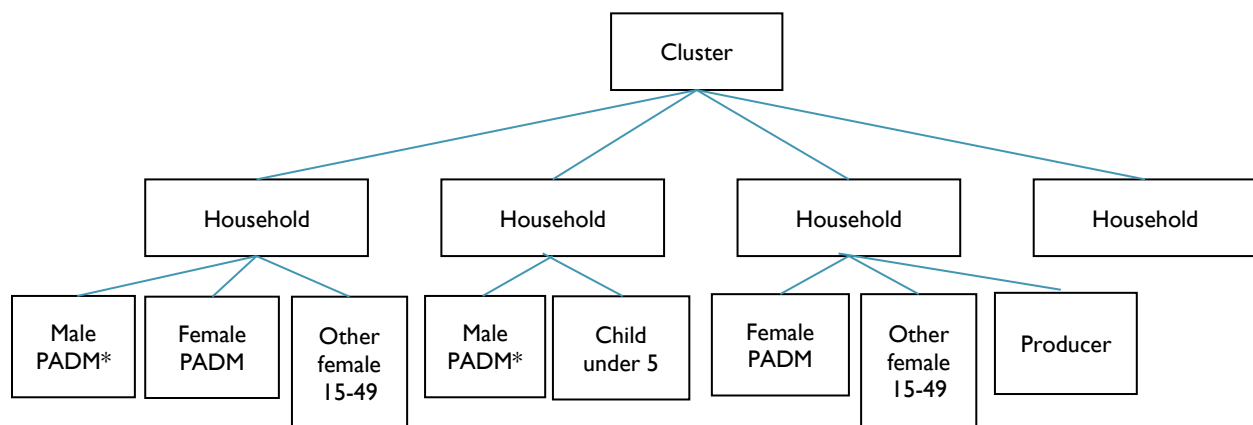
### 2.1 Structure of the Feed the Future ZOI Survey data files

Feed the Future ZOI Surveys use a computer-assisted personal interviewing system developed using a public domain software package called Census and Survey Processing System (CSPro) to collect, edit, and tabulate data during fieldwork. This system is developed and supported by the U.S. Census Bureau and ICF to support data collection and transfer on Android and Windows devices. CSPro allows for exporting data for use in various statistical software programs, such as R, SPSS, SAS, and Stata. It can also transfer the data as a plain text file, such as in a comma-separated values file.

Data files can be hierarchical or flat. Although CSPro creates hierarchical files, common statistical software programs, such as R, SPSS, SAS, and Stata, cannot use this file structure, and CSPro hierarchical files must be converted into flat files for statistical analysis.

The base unit for data collection in the ZOI Surveys is the household. A cluster number and household number uniquely identify a household. Some data are also associated with household members. Within a household, information is collected from primary adult female decision-makers, women of reproductive age (15-49 years of age), and individuals 15 years of age or older who are responsible for making decisions about the production of select value chains (crop, livestock, or fish) included in the survey. In addition, information about children under 5 years of age is collected from their primary caregiver. For individual-level data, such as in the nutrition modules, an additional identifying variable exists—the line number of the individual in the household roster—which, in combination with the cluster number and household number, allows all information about the individual to be uniquely linked to the appropriate household.

The following diagram illustrates the data structure used for ZOI Surveys. Note that households may exist in the data file without individual-level data if, for example, no household members were eligible for any of the individual-level modules, or if the dwelling was destroyed, vacant, or not found.



PADM=primary adult decision-maker

\* Baseline Survey only

### 2.1.1 Hierarchical files

Hierarchical data files maintain the structure of data collection; each record contains a section of the questionnaire, and identifying variables appear at the beginning of each record. Records can be single-occurring (e.g., dwelling characteristics and household assets, which are unique to the household) or multiple-occurring (e.g., household roster and nutrition information, which may have multiple entries per household—one per household member [roster] or eligible household member [nutrition]). Records for different sections of the questionnaire can each have a different number of variables, so the data are not rectangular. All data are ordered in the data file according to the order of the questionnaire. The advantage of hierarchical data files, such as those produced by CSPro, is that all data are in one file. Users can therefore analyze the data using different units of analysis, such as households, women of reproductive age, children under 5 years of age, and primary adult decision-makers. As mentioned previously, however, hierarchical files cannot be read by statistical software such as Stata and SPSS; thus, they must be transformed into flat files so that they can be loaded into statistical software programs for analysis.

### 2.1.2 Flat files

Flat data files contain one record for each case. In ZOI Surveys, two flat rectangular files are exported from CSPro for analysis: one for household-level data and another for individual-level data. There is no structural relationship among the records, or rows of data, within the data files, so each record must contain a separate variable for each piece of data collected. Variables for any sections of the questionnaire that repeat are placed sequentially in the record, and variable names are indexed by case—from 1 to the maximum number of cases allowed. Each record, therefore, has the same variables in it and is the same length, which is why the files are rectangular.

### 2.1.3 Formats and naming convention of datasets exported from CSPro

ZOI Survey datasets are exported from CSPro as R, SAS, SPSS, and Stata file types. Datasets include survey documentation and syntax files that describe the data and allow users to easily import the data into the statistical software programs. The naming conventions for ZOI Survey data files exported from CSPro are presented in **Table 3**.

**Table 3: Data File Names by Units of Analysis and File Format**

Analysis unit	File format	
	Hierarchical (CSPro)	Flat (R, SAS, SPSS, Stata)
Household	CFTF_HOUSEHOLD_YYYY	CFTF_HOUSEHOLD_YYYY
Individual <sup>a</sup>	CFTF_PERSONS_YYYY	CFTF_PERSONS_YYYY

C=placeholder for the first letter of the country where the ZOI Survey is conducted, FTF=Feed the Future, YYYY=year in which the ZOI Survey fieldwork is started

<sup>a</sup> The individual-level file includes records for all individuals in the household roster, women 15-49 years of age, children 0-59 months of age, primary adult female decision-makers, and producers.

All data files exported from CSPro begin with “CFTF,” in which “C” is a placeholder for the first letter of the country where the ZOI Survey was implemented. The file name also contains the four-digit year of data collection (“YYYY”) (or the start of data collection if data collection spans two calendar years)



and indicates whether it is the household-level or individual-level data file. All associated documentation and syntax for labeling variables and value sets are included in the exported data files. More information on exporting data from CSPro can be found in the *Feed the Future ZOI Survey Data Processing Manual*.<sup>10</sup>

### 2.1.4 Matching relationships

When combining data files, it is necessary to determine the relationship between the data in the files being combined and the type of output dataset needed. With ZOI Survey data files, two relationships are possible: a one-to-one relationship and a one-to-many relationship.

In a one-to-one relationship, one record in one file relates to exactly one record in another file, if available. An example is the relationship between the primary adult male decision-maker and the primary adult female decision-maker in each household. Because there can be a maximum of one of each in a household, the relationship is one-to-one, although it is possible that a household does not have one or either of these individuals. If combining a file of primary adult male decision-makers and another file of primary adult female decision-makers, if one does not exist in a household, no match is made; but when both exist, the match is one-to-one, and each primary adult male decision-maker is matched to exactly one primary adult female decision-maker, and vice versa. Because the relationship is one-to-one, either data file may be used as the base and the other unit merged into it. It is important to understand what type of analysis is needed to determine which dataset makes sense to use as the base.

In a one-to-many relationship, one data file may have more than one record that matches to a single record in the second data file. For example, the relationship between households and women 15-49 years of age is one-to-many. For each household, there may be zero, one, or multiple women 15-49 years of age. The one-to-many relationship requires special attention to the output dataset desired. In this example, the difference between using the household data file as the base versus using the women 15-49 years of age data file determines the size and format of the output dataset. If analysis is being done on women 15-49 years of age, the file with the women's information in it should be the base, and the information from the household file should be merged into it. If, however, the analysis is being done on households, and household-level information about women is required, the women's information must first be aggregated by household and then merged into the household file.

All statistical packages can merge data files in any direction—meaning that users must pay special attention to the order of merging and the number of records in the combined dataset. To merge, users must take the following actions:

- Determine the identifier variables necessary to match the data files.  
In ZOI Survey data files, the identifier variables are usually cluster and household number for household-level data and cluster, household number, and roster line number for individual-level data.
- Sort each data file by the matching identifier variables.
- Determine the base data file (the data file that has the unit of analysis required in the output dataset).
  - The other data file will be merged into the base data file.

---

<sup>10</sup> Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>

- In a one-to-one merge, typically, the base data file is the one with the fewest number of records.
- In a one-to-many merge, typically, the base data file is the one with the larger number of records.
- In both a one-to-one merge and a one-to-many merge, the output dataset will contain, at most, a number of matches equal to the number of records in the base data file.
- Specify the inputs (e.g., type of match, file names, and matching identifier variables) required by the analysis software package syntax to merge the files according to the specified inputs and matching identifier variables.
- Review the resulting data file to ensure it has the expected content.

## 2.2 Variable name conventions

ZOI Survey variables created in CSPro use standard naming conventions that are described in this section. Please see the country-customized codebook for the survey for detailed information about all variables, including variable names and labels, variable types, response codes, question text, and a description of the population that answers each question.

### 2.2.1 General variable naming

ZOI Midline Survey variables generated for the main survey are named according to the following conventions:<sup>11</sup>

- Household identification (cover page)<sup>12</sup> c+question identifier (c04, c05, c06, ...)
- Household roster (Module 1) v+question number (v102, v103, v104, ...)
- Dwelling characteristics and household assets (Module 2) v+question number (v201, v202, v203, ...)
- Climate adaptation (Module 2A) v+question number (v2a01, v2a02, v2a03, ...)
- Food security and resilience (Module 3) v+question number (v301, v302, v303, ...)
- Women's nutrition (Module 4)<sup>13</sup> v+question number (v401, v402, v403, ...)
- Children's nutrition (Module 5)<sup>14</sup> v+question number (v501, v502, v503, ...)
- Women's Empowerment in Agriculture (Module 6)<sup>15</sup> v+question number (v6101, v6102, v6103, ...)
- Part I (Respondent information)

<sup>11</sup> Although some module names and content slightly differ between Baseline and Midline Surveys, the naming conventions are the same for the surveys. As Midline Surveys were being implemented, the core questionnaires were updated so that all question numbers included leading consonants, so "C," "PP," or "V" became part of the question number. Therefore, in several Midline Surveys, the only difference between the question number and the survey variable name is that the leading consonants in the question numbers are capitalized and the leading constants in the survey variable names exported from CSPro are lowercase.

<sup>12</sup> The questions on the cover page of the ZOI Survey questionnaire are not numbered, so the household identification variables are named using a question identifier, rather than a question number as is done for the rest of the ZOI Survey variables.

<sup>13</sup> In the Baseline Survey questionnaire, Module 4 is *Women's nutrition and anthropometry*.

<sup>14</sup> In the Baseline Survey questionnaire, Module 5 is *Children's nutrition and anthropometry*.

<sup>15</sup> In the Baseline Survey, Module 6 was also administered to primary adult male decision-makers; variables created for data collected from men all start with "m" instead of "v."

Part 2 (Role in decision-making)	v+question number (v6201, v6202, v6203, ...)
Part 3 (Access to capital and credit)	v+question number (v6301, v6302, v6303, ...)
Part 4 (Group membership)	v+question number (v6404, v6405, ...)
Part 6 (Time allocation)	v+question number (v6601)

For questions that are asked about multiple items (e.g., economic activities, sources of credit, assets) in Module 6, the corresponding variables for each item are indexed sequentially starting with 1, regardless of how the items are numbered in the ZOI Survey questionnaire. If there are 10 or more items, there is a leading zero for items 1-9 (e.g., \_01, \_02, \_03). If there are less than 10 items, there is no leading zero (e.g., \_1, \_2, \_3). In both cases, the indexing number is preceded by an underscore. Therefore, the variables for question v6201 are named v6201\_01, v6201\_02, etc., and v6201\_01 corresponds to the first economic activity (6.2\_1—food crop farming) and v6202\_02 corresponds to the second economic activity (6.2\_2—cash crop farming).

Part 6 (Time allocation) is an activity tracker for a 24-hour period preceding the interview. Respondents provide information on their primary activities during the day preceding the survey, beginning at 4 a.m. Separate variables are created for the primary activities performed during 15-minute intervals over a 24-hour period, which results in 96 unique variables. These variables have the format v6601p\_[minutes]\_[hour], where [minutes] can have a value of 15, 30, 45, or 60, and [hour] can have a value between 1 and 24.<sup>16</sup> Following is an example using hour 1 (4:00 a.m.-4:59 a.m.):

v6601p_15_01	primary activity, minutes 0-14 of hour 1
v6601p_30_01	primary activity, minutes 15-29 of hour 1
v6601p_45_01	primary activity, minutes 30-44 of hour 1
v6601p_60_01	primary activity, minutes 45-59 of hour 1

- Agricultural technologies (Module 7)

Module 7.1 (maize)	v+question number (v7101, v7102, v7103, ...)
Module 7.2 (beans)	v+question number (v7201, v7202, v7203, ...)
Module 7.3 (coffee)	v+question number (v7301, v7302, v7303, ...)
Module 7.4 (groundnuts)	v+question number (v7401, v7402, v7403, ...)
Module 7.5 (wheat)	v+question number (v7501, v7502, v7503, ...)
Module 7.6 (soybeans)	v+question number (v7601, v7602, v7603, ...)
Module 7.7 (paddy rice)	v+question number (v7701, v7702, v7703, ...)
Module 7.8 (cow pea)	v+question number (v7801, v7802, v7803, ...)
Module 7.9 (chick pea)	v+question number (v7901, v7902, v7903, ...)
Module 7.10 (millet)	v+question number (v71001, v71002, ...)
Module 7.12 (mango)	v+question number (v71201, v71202, ...)
Module 7.50 (dairy cows)	v+question number (v75001, v75002, ...)
Module 7.51 (goats)	v+question number (v75101, v75102, ...)

<sup>16</sup> At baseline, some surveys captured data about secondary activities in Module 6.6.A. Variables for secondary activities include an “s,” for secondary, instead of a “p,” for primary, in their name but follow the same naming conventions.

Module 7.52 (sheep)	v+question number (v75201, v75202, ...)
Module 7.80 (fish)	v+question number (v78001, v78002, ...)
• Program participation questions <sup>17</sup>	
Agriculture (Module 2)	pp+question number (pp101, pp102, pp103, ...)
Resilience (Module 3)	pp+question number (pp201, pp202, pp203, ...)
Water, sanitation, and hygiene (Module 2)	pp+question number (pp301, pp302, pp303, ...)
Women's health and nutrition (Module 4)	pp+question number (pp401, pp402, pp403, ...)
Children's health and nutrition (Module 5)	pp+question number (pp501, pp502, pp503, ...)

ZOI Midline Survey variables generated for the parallel survey are named according to the following conventions:<sup>18</sup>

- Household identification (cover page) c+question identifier (c03, c04, c05, ...)
- Household roster (Module 1) v+question number (v101, v102, v103, ...)
- Dwelling characteristics and household assets (Module 2) v+question number (v201, v202, v203, ...)
- Module 7.90 (land map) v+question number (v79000a, v79000b, ...)
- Module 7.91 (plot information) v+question number (v79101, v79102, v79103, ...)
- Module 7.92 (plot area) v+question number (v79201, v79202, v79203, ...)
- Household consumption expenditure (Module 8)
  - Part 1 (food consumption) v+question number (v8101, v8102, v8103, ...)
  - Part 2 (non-food past 7 days) v+question number (v8201, v8202, v8203, ...)
  - Part 3 (non-food past 1 month) v+question number (v8301, v8302, v8303, ...)
  - Part 4 (non-food past 3 months) v+question number (v8401, v8402, v8403, ...)
  - Part 5 (non-food past 12 months) v+question number (v8501, v8502, v8503, ...)
  - Part 6 (housing expenditures) v+question number (v8601, v8602, v8603, ...)
  - Part 7 (durable goods) v+question number (v8701, v8702, v8703, ...)

For most Module 8 sections, the respondent is asked a set of questions for every item in a list—similar to sections of Module 6 in which the same questions are asked about multiple items. In such cases, the variable names include an underscore and item code at the end and are structured as: “v”+question number+“\_”+item code. For example, for question V8101 in Sub-module 8.1, *Food consumption over the past 7 days*, the variable for the first food item, 8001. Normal maize flour, is v8101\_001, and the variable for the second food item, 8002. Fine maize flour, is v8101\_002. This is true for all sections of Modules 6 and 8 in which a set of questions is asked for a list of items. The list of items in each section

<sup>17</sup> In the country-customized questionnaire, the agriculture and water, sanitation, and hygiene program participation questions are likely to be included in Module 2 and the resilience program participation questions are likely to be included in Module 3. However, their placement may vary; be sure to review the customized questionnaire carefully. Regardless of where in the questionnaire they are included, the same numbering convention is used.

<sup>18</sup> Although some module names and content slightly differ between Baseline and Midline Surveys, the naming conventions are the same for the surveys.

can be found in the final survey-customized ZOI Survey questionnaire used for data collection and in the associated survey-customized codebook.

## 2.2.2 Select-all questions

There are two types of questions in ZOI Surveys: select one and select all that apply. Response options for select one questions are coded as numeric values. Response options for select all that apply questions are coded as letters.<sup>19</sup> In the data files exported from CSPro, these select all that apply questions are coded in two ways: in alphanumeric form and in binary form. Analysts can use either form, depending on what works best for their purpose. In the alphanumeric form, only one string variable is created for the question, with a value set indicating the choice of responses. In the binary form, as many binary variables are created as there are response options, each assigned a value of 0 if not selected or 1 if selected. The alphanumeric variables are identified by the question number, per the naming conventions described in Section 2.2.1. The binary variables are identified by the question number followed by the letter corresponding to the response category. The binary variables are appended to the end of the data for that module.

For example, question V217 in Module 2, *Dwelling characteristics and household assets*, has multiple response options: the respondent can indicate more than one method used by the household to make water safer to drink.

217	What do you usually do to make the water safer to drink?	BOIL .....	A
		ADD BLEACH/CHLORINE .....	B
		STRAIN THROUGH A CLOTH .....	C
	Anything else?	USE WATER FILTER (CERAMIC/ SAND/COMPOSITE/ETC) .....	D
		SOLAR DISINFECTION .....	E
		LET IT STAND AND SETTLE .....	F
	SELECT ALL THAT APPLY	OTHER .....	X
		(SPECIFY)	
		DONT KNOW .....	Z

In its alphanumeric form, the variable appears as **v217**, with a value set equal to the letter or letters corresponding to the response given by the respondent. For example, if the respondent indicated that the household adds bleach to its water and also uses solar disinfection, the value for **v217** would be “BE.” In its binary form, eight variables are created with a value set of 0/1 and indicated as **v217a**, **v217b**, **v217c**, **v217d**, **v217e**, **v217f**, **v217x**, and **v217z**. For the example above, all the binary variables would have a value of 0, except for **v217b** and **v217e**, which would both have a value of 1.

If the response to a select all that apply question is missing (e.g., the question was not asked because of the skip patterns set up in the CSPro data collection application), the alphanumeric variable and all corresponding binary variables are also coded as missing—that is, the alphanumeric and binary variables for that question will not have values in the dataset for that case.

<sup>19</sup> Note that there are often one or more response options that prevent the interviewer from selecting another response option, for example if the respondent says that he or she does not know the answer.

### 2.2.3 Special coding conventions

This section describes the coding conventions for special response options, such as “don’t know,” “refused,” and “other,” as well as missing observations and questions that are not applicable. Any variables in the dataset that deviate from these conventions should be clearly noted in the survey codebook and variable value labels. **Values that indicate special coding conventions should be excluded when means and medians are calculated.** For example, when calculating the mean number of goats a household owns, any households in which the respondent replied, “Don’t know,” which is coded as “98,” should be excluded from the calculation; otherwise, the mean will be artificially inflated. In some cases, values that indicate special coding conventions should be recoded to another value before calculating an estimate. For example, “Don’t know” responses may be grouped with “No” responses—particularly when calculating percentages. In such cases, “Don’t know” and “No” responses should both be recoded into an analytic variable with value of “0” for these to responses, leaving “Yes” responses coded as “1.” Please see individual variable and indicator definitions in Part II of this Guide for more information on how special response options and missing observations are treated.

Questions that do not apply to a household or respondent based on skip patterns and filters in the questionnaire are considered “not applicable” (not “missing”). Data for questions that are not applicable appear as blank variables in the dataset (i.e., no special code is given to questions that are not applicable based on the flow of the questionnaire). Data that are truly missing, however, are designated as such with a code that designates missing. In ZOI Surveys, data are considered missing if information should have been collected during the interview but was not due to interviewer error or a programming error. The questionnaire does not allow for missing data for select key variables, such as cluster and household identifiers, geographical variables that are obtained from the sample design (region, district, urban/rural), and interview outcome codes. Missing data are indicated by a numeric value that ends in ‘9’ for single response questions (i.e., response options are listed as numbers in the questionnaire) and by ‘?’ for questions allowing for multiple responses (i.e., response options are listed as letters in the questionnaire).

**Table 4** shows how special response options and missing observations are typically coded, depending on whether the question has a single response option or allows multiple responses.

**Table 4: Special Response Categories and Their Codes, by Question Type**

Special response option	Response option code	
	Single response question	Multiple response question
Refusals <sup>a</sup>	7, 97, 997, 9997, ...	Y
Other (specify)	6, 96, 996, 9996, ...	X
Don’t know, unsure	8, 98, 998, 9998, ...	Z
Missing values <sup>b</sup>	9, 99, 999, 9999, ...	?

<sup>a</sup> Refusals to answer specific questions in the survey are assigned values that end in ‘7.’ However, the refused household interview and module outcome codes are assigned a value of ‘5.’

<sup>b</sup> A value is coded as missing during data processing if a variable should have a value but does not.

## 2.2.4 Other variables exported from CSPro

There are several variables that capture data that are not collected as part of the survey but are included in the CSPro export files and are important for analysis, such as those that geographically define the households or sample. They include:

- **ahtype** Urban or rural designation (referred to as “residence” throughout the Guide)
- **hhea** Cluster
- **hhnum** Household number
- **strata** Survey-specific categories used in the sampling process
- **vfarm** Flags whether an individual was eligible for any of the agricultural technologies modules
- **vtype** Flags whether a household has a primary adult female decision-maker or primary adult male decision-maker, both, or neither

In addition, some variables that are collected as part of the survey have names in the CSPro export file that do not follow the standard conventions. They include:

- **ahresult** Result of the household interview
- **m1\_line** Household roster line number

All applicable sample weight variables will also be included with the data exported from CSPro; these variables will all start with “wgt\_” as follows, although the weights for specific value chain commodities will vary across surveys:

- **wgt\_hh** Household sampling weight for the main survey
- **wgt\_hh\_p** Household sampling weight for the parallel survey
- **wgt\_fdm** Sampling weight for primary adult female decision-makers
- **wgt\_wra** Sampling weight for women of reproductive age (15-49 years)
- **wgt\_c0\_5m** Sampling weight for children 0-5 months
- **wgt\_c6\_23m** Sampling weight for children 6-23 months
- **wgt\_c0\_59m** Sampling weight for children 0-59 months
- **wgt\_vcc** Sampling weight for producers of one or more value chain commodities
- **wgt\_maize** Sampling weight for maize producers
- **wgt\_dairy** Sampling weight for dairy cow producers
- **wgt\_sheep** Sampling weight for sheep producers
- **wgt\_fish** Sampling weight for fishpond producers



## 3. DATA ANALYSIS INFORMATION

### 3.1 Sample design, sampling weights, and response rates

ZOI Surveys are population-based surveys using probability sampling methods with sample design features, such as stratification, clustering and unequal probabilities of selection, to make inference to the target population for key indicators and statistics. Therefore, it is important that analysts understand and be familiar with certain aspects of how the sample is designed and how sampling weights and response rates are derived for ZOI Surveys because these aspects play a role in generating the Feed the Future data on which estimates of indicators described in the following chapters are based. The aim of this section is not to provide a detailed description of the sampling methodology for ZOI Surveys, but rather to give an overview of the key elements that analysts should understand when analyzing Feed the Future data. For a more detailed description of the sampling methodology, please refer to the *Feed the Future Population-Based Survey Sampling Guide*.<sup>20</sup>

#### 3.1.1 Sample design

ZOI Survey samples are based on stratified multi-stage cluster sample designs with up to four stages of sampling. The first-stage sample is typically selected from an existing sampling frame, generally the most recent census frame of enumeration areas (EAs). A sampling frame is a complete list of all sampling units that entirely covers the target population.

Stratification is a sample design feature used to divide the sampling frame units into sub-groups or strata that are as homogeneous as possible using certain criteria. Within each stratum, the sample is designed and selected independently. The principal objective of stratification is to reduce sampling errors. In a stratified sample, the sampling errors depend on the population variance existing within the strata but not between the strata. In some cases, a single ZOI Survey is used to collect information on more than one target population (for example, a survey needing to collect Feed the Future indicators from the ZOI target population as well as resilience indicators from the Resilience Focus Zone target population). Stratification also enables the frame to be divided into these different target populations to enable reporting for each target population. Typically, Feed the Future population-based survey samples are stratified by the relevant administrative and geographic hierarchy (including urban or rural residence) used in the country and available on the sampling frame within each target population.

Sample allocation is another sample design feature used to allocate the sample across (i.e., divide among) the different strata after the sampling frame has been stratified. The principal objective of sample allocation is to allocate the sample appropriately across the strata such that optimal estimates (i.e., lowest possible variance or highest possible precision for a fixed sample size) can be produced based on the reporting objectives of the survey. Usually, ZOI Surveys are multipurpose, in which there are many indicators of interest that span a variety of sampling groups and in which the interest lies in producing optimal estimates (i.e., estimates with the lowest possible variance and highest possible precision) at the overall ZOI level only. In this case, a proportional allocation of the sample across strata is the most appropriate method to use to achieve optimal estimates at the overall ZOI level.

---

<sup>20</sup> Stukel, D.M. (2018). *Feed the Future population-based survey sampling guide*. Washington, DC: Food and Nutrition Technical Assistance Project, FHI 360. Available at: <https://agrilinks.org/post/feed-future-zoi-survey-methods>.



Proportional allocation allocates the sample to each stratum proportional to the stratum size, using a size measure, such as number of households in each stratum. However, some ZOI Surveys may require optimal estimates at stratum level rather than at the overall level (for example, when a single ZOI Survey is used to collect information on more than one target population [e.g., the overall ZOI population and the Resilience Focus Zone population within the ZOI]). In this case, an equal allocation of the sample across strata is the most appropriate method to use to achieve optimal estimates at the stratum level. Equal allocation allocates the sample equally across the strata. In some cases, a power allocation, which is a compromise between proportional and equal allocation methods, may also be used and is appropriate when optimality is desired at both the overall level as well as the stratum level. Power allocation allocates the sample to each stratum proportional to the size measure raised to a certain power. Therefore, the allocation method used will be survey specific, and analysts should refer to Appendix 2 of the ZOI midline indicator assessment report for the specific allocation method used for a particular survey.

After the sample allocation has been completed, the first stage of sample selection is undertaken. In the first stage of selection, the primary sampling units (PSUs) are selected with systematic probability proportional to size sampling within each stratum. The PSUs are typically census EAs. The *Feed the Future Population-Based Survey Sampling Guide* makes provision for the allocation of a reserve sample of EAs that could be used if some of the EAs from the main sample are found to be inaccessible using a two-phase sampling approach during this first stage of selection. During the first phase of the first stage of sampling, a total number of EAs is sampled from each stratum. The total number of EAs constitutes both the number of EAs required for the main sample and the reserve number of EAs expected to be required. At the second phase of the first stage of sampling, the subset of EAs for the reserve sample is selected from the first phase sample using fractional interval systematic sampling. The reserve EAs are released during fieldwork as and when required. If any EAs from the reserve sample are used as replacements in the main sample, the selection probabilities and sampling weights for the first stage are adjusted accordingly to account for the total number of EAs released for the survey, including the reserve sample EAs used.

In some cases, it may be necessary for PSUs to be segmented, and one segment will be selected for the survey through an additional stage of sampling using probability proportional to size sampling. This happens when an EA has grown too large to undertake a listing in the entire EA. Therefore, the final survey clusters could either be an EA or a segment of an EA.

In the second stage (or third stage if segmentation was undertaken), a complete household listing is conducted in each of the selected clusters. Following the listing of the households, a fixed number of households is selected by fractional interval systematic sampling in the selected clusters. Every effort must be made to obtain an interview from all selected households, following the data collection protocol for the survey. No replacement households are selected as part of the sample design for ZOI Surveys; the sample size determination process for the survey already incorporates an adjustment to inflate the sample size to compensate for any anticipated household non-response (i.e., to compensate for the fact that not all household interviews will be completed as planned). Any non-responding households will be accounted for through a non-response adjustment in the sampling weights.

At the final stage of sampling, individuals are selected within the sampled households into different sampling groups using the household roster. The selection into different sampling groups is based on the criteria defined by the analysis requirements for the indicators and statistics for the different study populations covered by the survey. For example, from an analysis perspective, ZOI Surveys need to produce indicators and statistics on dietary intake of women 15-49 years of age in the population. Therefore, the criteria for being included in the survey will be all women 15-49 years of age, who spent the previous night in the household (i.e., the de facto population of women 15-49 years of age).<sup>21</sup>

Typically, ZOI Surveys implement a take-all approach for the different individual sampling groups (i.e., selecting all individuals in the sampled household that meet the criteria as defined by the analysis requirements). However, in some cases, a sub-sampling of individuals for some sampling groups (e.g., children under 5 years of age, women of reproductive age, or producers) may occur, in which one individual is randomly selected from all who are eligible in the household. The random selection of an individual within a household is undertaken using a Kish selection grid.

The analytic selection probability for each household in the sample is the probability of selecting the cluster multiplied by the probability of selecting the household within the cluster.<sup>22</sup> The overall selection probability for individuals is the overall selection probability for each household multiplied by the probability of selecting the individual (based on the criteria defined by the analysis requirements) for the specified sampling group within the household. The overall probability of selection of households will differ from cluster to cluster, and the overall probability of selection of individuals will differ from household to household within each cluster. The Sampling Design Document for the ZOI Survey will provide a detailed description of the sample design specific to that particular survey, and analysts are therefore encouraged to review it to gain a better understanding of the underlying sample design that generated the survey data. If the Sampling Design Document is not available, similar information can be found in Appendix A of the ZOI Survey report.

ZOI Survey dataset users should be aware that to make valid inferences to the population, the survey data must be weighted during analysis. The following section describes how sampling weights are constructed for ZOI Surveys.

### 3.1.2 Sample weighting

The first step before analysis of the survey data can begin is to calculate the sampling weights associated with each of the households and individuals who have been randomly selected into the survey and have responded to the interview questions. Sampling weights for each selected individual (or household in the case in which the household is the sampling group) are calculated and applied to corresponding individual (or household) survey data records 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

---

<sup>21</sup> To avoid the potential of double-counting individuals in a survey, it is important to ensure from a data analysis perspective that either a de facto population or a de jure population is selected for analysis of individual-level data. The choice of whether to use a de facto or de jure population depends on the analysis requirements and objectives for that indicator or disaggregate.

<sup>22</sup> If segmentation was undertaken, the probability of selecting the cluster will be the probability of selecting the PSU multiplied by the probability of selecting the segment.

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

Sampling weights are adjustment factors applied to each record during analysis to account for the selection probabilities at each stage of sampling and any non-response at the household and individual levels during survey implementation. Due to the sample design features, the final sample of completed interviews will not necessarily be representative of the entire population by itself. Sampling weights must be applied to the data to create representative results. For example, in some ZOI Surveys covering more than one target population (e.g., the overall ZOI population and the Resilience Focus Zone population within the ZOI), to obtain a sufficient number of cases to produce reliable estimates for each population, the sample size may be increased for some portions of the overall population to accommodate these needs (e.g., an oversample is taken). In that case, the weights are applied so that the final results are representative of the overall population and not exceedingly influenced by the data from the oversampled portion of the population.

In general, ZOI Surveys collect data for indicators that span at least eight sampling groups: children 0-5 months of age, children 6-23 months of age, children 0-59 months of age, women 15-49 years of age, primary adult female decision-makers, producers, and the household itself. ZOI Surveys generally produce separate sampling weights for each sampling group. The number of weights for producers depends on whether the ZOI Survey includes specific producer-targeted value chain commodities (VCCs). If it does, there will be a weight for producers of any targeted VCC and separate weights for producers of each targeted VCC covered in the survey. If the survey does not include targeted VCCs but does include a module for crop or livestock/farm animal producers, a weight will be created for each producer sample (i.e., crop producers or livestock/farm animal producers). There may be additional sampling weights created for indicators covering other sampling groups based on the survey analysis requirements for a particular survey.

The household weight for a particular household is the household design weight (i.e., the inverse of the overall household selection probability) multiplied by the inverse of the weighted household response rate (i.e., weighted by the household design weight) in the stratum. The sample weight for a particular individual of a given sampling group is the individual design weight (i.e., the inverse of the individual's overall selection probability) multiplied by the inverse of the weighted household response rate and weighted individual response rate for the given sampling group (i.e., weighted by the individual design weight) in the stratum. A weighted response rate is used for non-response adjustment in deriving the sample weights, because a non-responding sample unit with a larger design weight will have a larger impact on the survey estimates than a non-responding sample unit with a smaller design weight. This difference in design weights between sample units needs to be accounted for when determining the non-response adjustment required for a particular sampling group.

The Sampling Statistician for the ZOI Survey will calculate the relevant weights based on data analysis requirements, taking into consideration the sample design, sample implementation results, and any overlap between target populations if the survey is covering more than one target population. The relevant sampling weights will be added to the data files during the data processing phase for applicable respondent groups. Refer to the *Feed the Future Population-Based Survey Sampling Guide* for detailed

information on calculating weights and the *Feed the Future Data Processing Manual* for detailed information on adding the sampling weights to the data files that are exported from CSPro.

Appendix A of the ZOI Survey reports will provide a detailed description of how the sampling weights were derived for that particular survey, and analysts are encouraged to review it to gain a better understanding of the sampling weights that will be used during data analysis.

### 3.1.3. Response rates

The response rate is an important survey parameter. It provides information on the survey coverage across the different sampling groups. This information is used in survey monitoring and is used as one of the quality indicators during survey implementation.

Non-response has the potential to introduce bias into survey data, and therefore response rate information is also used to adjust sample weights to account for differential non-response for the sampling groups in the survey, as mentioned in Section 3.1.2. The response rate for a particular sampling group can be defined as the proportion of eligible sampling units (households or individuals) that have been interviewed for that sampling group (at the household or individual level), expressed as a percentage. In practice, it is to be expected that some proportion of households or individuals randomly selected for the survey will be unreachable, unavailable, or unwilling to respond to the survey. It is therefore important to systematically categorize in detail the resulting final survey outcome status for each household and individual sampled for the survey. For ZOI Surveys, standardized final result codes are used on questionnaires and field records, and the response rates are calculated based on these codes.

In ZOI Surveys, the following final household interview result codes (**AHRESULT**) are used:

- 01 Completed<sup>23</sup>
- 02 No household member at home
- 03 Entire household absent for extended period of time
- 05 Refused
- 06 Dwelling vacant
- 07 Address not a dwelling
- 08 Dwelling destroyed
- 09 Dwelling not found
- 10 Household member too ill to respond/cognitively impaired
- 12 No household member at least age 15 years

Note that the household above refers to the household found in the dwelling at the time of the interview during the survey fieldwork. This may not necessarily be the same household named at the time of the listing operation, as the household occupying the dwelling unit may have changed during the period between household listing and time of interview during survey fieldwork.

---

<sup>23</sup> These are households for which at least Module 1, *Household roster and demographics*, and Module 2, *Dwelling characteristics and household assets*, have been completed.

For ZOI Survey reports, the household response rate is calculated by:

$$RR_H = \frac{01}{01 + 02 + 05 + 09 + 10 + 12}$$

The reason to include 09 in the denominator is that if a dwelling is not found at the time of the fieldwork, it does not necessarily mean that the dwelling and its household do not exist; it may be that the dwelling was not found because of an error that occurred during the survey implementation. Therefore, including these cases will provide a more conservative estimate of the household response rate instead of assuming that all these cases are invalid and excluding them from the denominator.

At the individual level, the final result codes of the specific modules covering the individual sampling group in question are used. The following final result codes are used for each module:

- 01 Completed
- 03 Unavailable after all callbacks made
- 05 Refused
- 10 Household member too ill to respond/Cognitively impaired
- 12 Age ineligible<sup>24</sup>

Therefore, the individual response rate for a particular sampling group and presented in ZOI Survey reports is calculated by:

$$RR_I = \frac{01}{01 + 03 + 05 + 10}$$

It should also be noted that these response rates (i.e.,  $RR_H$  and  $RR_I$ ) are different from the weighted response rates calculated in Section 3.1.2. The response rates presented in ZOI Survey reports are calculated as ratios of the number of interviewed units over the number of eligible units because the objective is just to show the results of survey implementation (i.e.,  $RR_H$ ). However, in Section 3.1.2, the weighted ratios are used, because the objective is to adjust the sample weights to compensate for non-response, taking into account the sample design.

To calculate response rates correctly, it is also important to understand the underlying eligibility criteria for the survey across the different sampling groups, because the eligibility criteria will define which sampled units (households or individuals) are valid and in scope for the survey and which are not. This information is then used with the outcome from the survey implementation to calculate response rates for each sampling group after the survey is conducted using the previous formulas. In addition, it should be noted that response rates are calculated for sampled units in a survey, such as households or individuals, because they provide a measure of the percentage or proportion of sampled units that have responded out of all eligible sampled units identified. The eligibility criteria and how response rates are calculated for the eight sampling groups covered in ZOI Surveys are described as follows.

---

<sup>24</sup> This final result code is relevant only for modules based on age eligibility in which the household member's age is confirmed in the module for which they are eligible (i.e., Module 4, *Women's nutrition*, Module 5, *Children's nutrition*, and Module 6, *Women's empowerment in agriculture*).

## Household response rate

**Eligibility:** All households except for vacant dwellings, dwellings in which the household is absent for an extended period of time, structures at an address that are not dwellings, and dwellings that have been destroyed.

**Numerator:** Number of households with a completed household interview. These are households for which Module 1, *Household roster*, and Module 2, *Dwelling characteristics and household assets*, were completed (i.e., **ahresult** is 01).

**Denominator:** Number of households with a completed household interview, households that live in the dwelling but no competent respondent was at home, households with permanently postponed or refused interviews, and households for which the dwelling was not found (**ahresult** is 01, 02, 05, 09, or 10).

## Women of reproductive age (15-49 years) response rate

**Eligibility:** Women whose age is between 15 and 49 years in the household roster who are de facto household members, excluding those determined to be age ineligible in Module 4, *Women's nutrition*.

**Numerator:** Number of eligible women with a completed the Module 4 interview (i.e., Module 5 result [**v400x**]=01=completed).

**Denominator:** Number of eligible women with a completed the Module 4 interview, eligible women with permanently postponed or refused interviews, eligible women for whom an interview could not be completed due to incapacitation or being too cognitively impaired to respond (Module 4 result [**v400x**] is 01, 03, 05, or 10).

## Primary adult female decision-makers (18+ years) response rate

**Eligibility:** Adult females (18+ years) identified as primary decision-makers in the household roster (line 02) who are de jure household members, excluding those determined to be age ineligible in Module 6, *Women's empowerment in agriculture*.

**Numerator:** Number of eligible primary adult female decision-makers with a completed Module 6 interview (i.e., Module 6 outcome [**v600x**]=01=completed).

**Denominator:** Number of eligible primary adult female decision-makers with a completed Module 6 interview, eligible primary adult female decision-makers with permanently postponed or refused interviews, eligible primary adult female decision-makers for whom an interview could not be completed due to incapacitation or being too cognitively impaired to respond (Module 6 result [**v400x**] is 01, 03, 05, or 10).

## Children under 5 years response rate

**Eligibility:** Children whose age is between 0 and 4 years in the household roster who are de facto household members, excluding those determined to be age ineligible in Module 5, *Children's nutrition*.

**Numerator:** Number of eligible children whose caregiver completed Module 5 (i.e., Module 5 outcome [v500r]=01=completed).

**Denominator:** Number of eligible children whose caregiver completed a Module 5 interview, eligible children whose caregiver was not interviewed because of permanently postponed or refused interviews, or interviews that could not be completed with the caregiver of eligible children due to incapacitation or the caregiver being too cognitively impaired to respond (Module 5 result [v500r] is 01, 03, 05, or 10).

### **Children under 2 years response rate**

**Eligibility:** Children whose age is 0 or 1 years in the household roster who are de facto household members, excluding those determined to be age ineligible in Module 5, *Children's nutrition*.

**Numerator:** Number of eligible children whose caregiver completed Module 5 (i.e., Module 5 outcome [v500r]=01=completed).

**Denominator:** Number of eligible children whose caregiver completed a Module 5 interview, eligible children whose caregiver was not interviewed because of permanently postponed or refused interviews, or interviews that could not be completed with the caregiver of eligible children due to incapacitation or the caregiver being too cognitively impaired to respond (Module 5 result [v500r] is 01, 03, 05, or 10).

### **Children 6-23 months response rate**

**Eligibility:** Children whose age is between 6 and 23 months in the household roster who are de facto household members, excluding those determined to be age ineligible in Module 5, *Children's nutrition*.

**Numerator:** Number of eligible children whose caregiver completed Module 5 (i.e., Module 5 outcome [v500r]=01=completed).

**Denominator:** Number of eligible children whose caregiver completed a Module 5 interview, eligible children whose caregiver was not interviewed because of permanently postponed or refused interviews, or interviews that could not be completed with the caregiver of eligible children due to incapacitation or the caregiver being too cognitively impaired to respond (Module 5 result [v500r] is 01, 03, 05, or 10).

### **Children under 6 months response rate**

**Eligibility:** Children whose age is between 0 and 5 months in the household roster who are de facto household members, excluding those determined to be age ineligible in Module 5, *Children's nutrition*.

**Numerator:** Number of eligible children whose caregiver completed Module 5 (i.e., Module 5 outcome [v500r]=01=completed).

**Denominator:** Number of eligible children whose caregiver completed a Module 5 interview, eligible children whose caregiver was not interviewed because of permanently postponed or refused interviews, or interviews that could not be completed with the caregiver of eligible children due to incapacitation or the caregiver being too cognitively impaired to respond (Module 5 result [v500r] is 01, 03, 05, or 10).



## Farmers of any targeted VCC response rate

**Eligibility:** De jure household members 15 years of age or older, listed in Module 2, *Dwelling characteristics and household assets*, as being responsible for cultivating at least one plot of land or raising at least one type of livestock or other farm animal for any of the targeted VCCs during the 12 months preceding the survey.

**Numerator:** Number of eligible household members who completed all of the targeted VCC modules applicable to them (i.e., VCC module outcome=01=completed, for all applicable VCC modules).

**Denominator:** Number of eligible household members with all applicable VCC module interviews completed, eligible household members with permanently postponed or refused interviews for at least one of the applicable VCC modules, or interviews that could not be completed due to incapacitation or being too cognitively impaired to respond for at least one of the applicable VCC modules (any VCC module result is 01, 03, 05, or 10).

## Farmers of specific targeted VCC response rate

The response rate for a specific targeted VCC is calculated in the same way but restricting the numerator and denominator to household members involved in producing that particular targeted VCC.

### 3.2 Descriptive statistics and statistical tests of difference

This section provides an overview of descriptive analysis and statistical tests of difference used when analyzing the ZOI Survey data, calculating the Feed the Future indicators, and tabulating the results.

Statistics presented in the ZOI Survey reports for each ZOI Survey indicator and its standard disaggregates (i.e., estimates presented in Table A1.1)<sup>25</sup> include the unweighted number of observations, the sample-weighted point estimate (for example, a mean or percentage) and its 95 percent confidence interval (CI), and the design effect. Standard deviations are also presented for means. Result tables throughout midline indicator assessment reports generally include the sample-weighted point estimates and the unweighted number of observations for the reported indicators and their disaggregates. Result tables that compare estimates over time also include CIs and differences in estimates over time, including the associated p-values and level of significance designation,<sup>26</sup> for the reported indicators and their disaggregates. Results tables that present results for one point in time include level of significance designations for statistical tests of difference conducted to assess the relationship between disaggregates and indicators. Note that statistics should not be reported for any indicator or disaggregate categories that have an unweighted number of observations less than 30.

The indicator estimates produced are either descriptive or comparative in nature. Indicator estimates for one point in time are descriptive. Descriptive results can be presented by selected disaggregates

---

<sup>25</sup> Standard disaggregates are those that are entered into the United States Agency for International Development's Development Information System. Other disaggregates are those that are not entered into the Development Information System but are included in the ZOI Survey reports.

<sup>26</sup> Differences significant at the  $p<0.05$ ,  $0.01$ , and  $0.001$  levels are indicated in the result tables as follows: one asterisk (\*)= $p<0.05$ , two asterisks (\*\*)= $p<0.01$ , three asterisks (\*\*\*)= $p<0.001$ . Differences that are not significant at the 0.05 level are indicated with "n/s."



(e.g., gendered household type, sex, or age group), with a p-value that indicates whether the association between the outcome and the selected characteristic is statistically significant. Indicator estimates comparing two points in time (e.g., baseline and midline) are comparative. Comparative results include the result of a statistical test of difference—a p-value that indicates whether the difference in indicator estimates at the two points in time is statistically significant. Comparative results can be presented for selected disaggregates such that each disaggregate category is compared individually between the two points in time. For example, the percentage of households with access to basic sanitation at midline can be compared to baseline, and the indicator at midline can also be compared to baseline separately for households in rural areas and households in urban areas, as long as the estimates being compared are calculated based on 30 or more observations.

Stukel's *Feed the Future Population-Based Survey Sampling Guide* (2018) provides additional information and background about conducting descriptive and comparative analysis using population-based survey data in Chapter 11 (Data Analysis for Descriptive Surveys: Producing Single Point-in-Time Estimates of Indicators along with Their Standard Errors and Confidence Intervals) and Chapter 12 (Data Analysis for Comparative Analytical Surveys: Statistical Tests of Differences). Stukel also provides guidance and syntax in Appendix B for conducting descriptive and comparative analysis using SAS, SPSS, and Stata statistical software packages. Stukel, however, does not include guidance or syntax for analyzing indicators by selected disaggregates or for obtaining statistics beyond CIs and standard errors—information that is presented in this section for Stata.

In Stata, complex survey designs should be specified using the `'svyset'` command, which allows analysts to take into account specific components of the survey design, such as clustering and unequal probabilities of selection, by specifying the PSU, the sampling weight, and the strata (if applicable). Analysis results should be subsequently sample-weighted, taking into account the complex survey design by specifying `'svy'` before the analysis command (e.g., `'svy: mean,'` `'svy: prop,'` or `'svy: tab'`).

Throughout this section, examples of Stata output are included for an indicator of proportions and an indicator of means. The variables used in the examples include the following:

- **i\_scap\_index**: total rescaled social capital index, continuous variable (scale: 0-100)
- **locgov\_resp**: household believes that the local government will help the community cope with future difficult times, dichotomous variable (0=No, 1=Yes)
- **survey**: survey round, dichotomous variable (1=baseline, 2=midline)
- **genhhtype\_dj**: gendered household type disaggregate, based on de jure household members, categorical variable (1=male and female adult household, 2=female adult-only household, 3=male adult-only household, 4=children-only household).

### 3.2.1 Analyzing indicators of proportions

#### Descriptive, overall

Descriptive estimates for an indicator of proportions can be calculated using the following syntax:<sup>27,28</sup>

```
svy: tab outcome, percent
```

where **outcome** is the variable for the indicator of proportions

Using the syntax above, sample-weighted estimates are presented as percentages for the two indicator variable values (usually 0 and 1), along with the number of observations included in the calculation (i.e., the unweighted sample size) and the corresponding “population size” (i.e., the weighted sample size) (see **Example 1**). Note that if the ‘**percent**’ option is not specified, the estimates will be presented as proportions, which must then be multiplied by 100 to obtain percentages for inclusion the results tables.

---

<sup>27</sup> If the outcome is dichotomous and takes a value of 0 or 1, **svy: mean outcome** can also be used to calculate the estimate; however, **svy: mean** generates confidence intervals that can extend outside of the 0-1 range because the command does not use logit transformation to generate confidence intervals. To ensure that confidence intervals stay within the expected 0-1 range, **svy: tab** (or **svy: prop**) should be used. Another option is to use the option ‘**cotype(logit)**’ with **svy: mean**, but the option does not work with all versions of Stata.

<sup>28</sup> Another option is to use **svy: prop outcome, percent**. The estimates, standard errors, and confidence intervals should be the same as those produced using **svy: tab outcome, percent**; however, the p-values calculated using **svy: tab** and **svy: prop** (or **svy: mean**) when an independent variable is added to the command differ slightly because they are generated from different tests **svy: prop** and **svy: mean** followed by the **lincom** command generate a p-value using a survey design-adjusted t-test, and **svy: tab** generates a p-value using a design-adjusted Person’s chi-squared F-test, or Rao-Scott test. These approaches may result in slightly different p-values, so it is advisable to use **svy: tab** for all indicators of proportions for consistency.

### Example 1: Indicators of proportions: Descriptive, overall output

```
. svy: tab loggov_resp, percent col stubwidth (16) format(%6.1f)
(running tabulate on estimation sample)
```

Number of strata = 22  
Number of PSUs = 88

Number of obs = 2,079  
Population size = 1,793,879  
Design df = 66

HH believes local govt will help the community cope with future difficult times	column
No	45.1
Yes	54.9
Total	100.0

Key: column = Column percentage

Note: The “`stubwidth()`” option is not otherwise discussed in this document, but it can be used to increase the length of the variable and values displayed so that they can be more easily understood.

### Descriptive, by selected disaggregates

Descriptive estimates for an indicator of proportions by a selected disaggregate, in addition to the overall indicator estimate, can be calculated using the following syntax:

```
svy: tab outcome disaggregate, percent col
```

where: **outcome** is the variable for the indicator of proportions

**disaggregate** is the variable for the selected characteristic

Using the syntax above, sample-weighted estimates are presented as percentages for the two indicator of proportions variable values for each disaggregate category, along with the total number of observations included in the calculation (i.e., the unweighted sample size across all disaggregate categories) and the corresponding total population size (i.e., the weighted sample size across all disaggregate categories) (see **Example 2**). A survey design-adjusted p-value is also presented. This p-value indicates whether there is a statistically significant association between the selected disaggregate and the indicator of proportions (e.g., between gendered household type and households that believe the government will help the community cope with difficult times). The p-value indicates whether a difference exists between categories if the disaggregate has only two categories (e.g., whether the government will help or not). However, the p-value does not indicate whether differences exist between specific categories of the disaggregate if the disaggregate has more than two categories (e.g., whether there is a difference in beliefs that the government will help cope with difficult times).

between male and female adult households and female-adult only households); to determine whether any differences exist between two disaggregate categories, survey design-adjusted t-tests would have to be performed.<sup>29</sup>

Note that by specifying 'col' as a **svy: tab** option when the disaggregate is specified as the second variable in the tab command, following the indicator variable, the percentages in each column sum to 100, and the output includes the desired indicator estimates by disaggregate category. In **Example 2**, for instance, the output table shows that 44.9 percent of male and female adult households do not believe that the local government will help the community cope with future difficult times. If 'row' is specified instead of 'col,' the percentages in the output table are presented for the indicator categories in each row (i.e., the percentages would add to 100 for each yes/no category instead of for each gendered household type category), which is not desired. If neither 'col' nor 'row' is specified, the percentages in the output table are presented across all indicator and disaggregate categories (i.e., the percentages sum to 100 only across all indicator and disaggregate categories—neither the row totals nor the column totals would equal 100), which is also not desired.

**Example 2: Indicators of proportions: Descriptive, by selected disaggregates output**

```
. label values genhhtype_dj genhhtype_dj1

. svy: tab locgov_resp genhhtype_dj, percent col stubwidth (16) format (%6.1f)
(running tabulate on estimation sample)
```

Number of strata = 22	Number of obs = 2,079
Number of PSUs = 88	Population size = 1,793,879
	Design df = 66

HH believes local govt will help the community cope with future difficult times	Gendered HH type, de jure HH members				
	M&F	FNM	MNF	I	Total
0	44.9	47.8	30.6	27.0	45.1
1	55.1	52.2	69.4	73.0	54.9
Total	100.0	100.0	100.0	100.0	100.0

Key: Column percentage

Pearson:

Uncorrected	chi2(3)	=	7.1154	
Design-based	F(2.43, 160.49)	=	3.1000	P = 0.0381

<sup>29</sup> Survey design-adjusted t-tests are not required for Feed the Future descriptive reports.

M&F=male and female adult households, FNM=female adult-only households (no adult males), MNF=male adult-only households (no adult females), I=children-only households (no adults)

## Comparative, overall

Comparative estimates for an indicator of proportions can be calculated using the following syntax, which is the same as for indicators of proportions by a selected disaggregate:

```
svy: tab outcome round, percent col
```

where: **outcome** is the variable for the indicator of proportions

**round** is the variable for the survey round (e.g., baseline or midline)

Using the syntax above, sample-weighted estimates are presented as a percentages for the two indicator variable values for each survey round, the total number of observations included in the calculation (i.e., the unweighted sample size across the two survey rounds), and the corresponding total population size (i.e., the weighted sample size across the two survey rounds) (see Example 3). A survey design-adjusted p-value is also presented, which is obtained from a Rao-Scott chi-squared test that accounts for complex survey designs. This p-value indicates whether the difference between the estimates being compared at the two points in time is statistically significant.

Note that by specifying '**col**' as a **svy: tab** option when the survey round is specified as the second variable in the **tab** command, following the indicator variable, the percentages in each column sum to 100, and the output includes the desired indicator estimates by survey round. In Example 3, for instance, the output table shows that 94.8 percent of households believe that the local government will help the community cope with future difficult times at baseline, and 54.9 percent of households believe that the local government will help the community cope with future difficult times at midline. If '**row**' is specified instead of '**col**,' the percentages in the output table are presented for the disaggregate categories in each row (i.e., the percentages would add to 100 for each Household Hunger Scale disaggregate category instead of for each survey round), which is not desired. If neither '**col**' nor '**row**' is specified, the percentages in the output table are presented across both indicator categories and survey rounds (i.e., the percentages sum to 100 only across all indicator categories and survey rounds—neither the row totals nor the column totals would equal 100), which is also not desired.

### Example 3: Indicators of proportions: Comparative, overall output

```
. svy: tab locgov_resp survey, percent col stubwidth (16) format(%6.1f)  
(running tabulate on estimation sample)
```

Number of strata = 36  
Number of PSUs = 253

Number of obs = 4,129  
Population size = 3,118,123  
Design df = 217

HH believes local govt will help the community cope with future difficult times	survey		
	Baseline	Midline	Total
No	5.2	45.1	28.2
Yes	94.8	54.9	71.8
Total	100.0	100.0	100.0

Key: Column percentage

Pearson:

Uncorrected chi2(1) = 791.8917  
Design-based F(1, 217) = 300.3515 P = 0.0000

### Comparative, by selected disaggregates

Comparative estimates for an indicator of proportions by a selected disaggregate can be calculated using the following syntax:

```
svy, subpop(sub_1): tab outcome round, percent col  
svy, subpop(sub_x): tab outcome round, percent col
```

where: **outcome** is the variable for the indicator of proportions

**round** is the variable for the survey round (e.g., baseline or midline)

**sub\_1** is the variable that defines the sampled observations included in the first category of the disaggregate being analyzed (e.g., male and female adult households for the gendered household type disaggregate)

**sub\_x** is the variable that defines the sampled observations included in category 'x' of the disaggregate being analyzed (e.g., for the gendered household type disaggregate, there would be three additional sub-population variables: **sub\_2** for female adult-only

households, **sub\_3** for male adult-only households, and **sub\_4** for children-only households)<sup>30</sup>

Subpopulations are created for each disaggregate category, and **svy, subpop(sub\_x) : tab outcome round, percent col** is run separately for each disaggregate category.

Using the syntax above, for each disaggregate category, sample-weighted estimates are presented as percentages for the indicator variables for each survey round, along with the total number of observations included in the calculation (i.e., the total unweighted sample size across the two survey rounds) and the corresponding total population size (i.e., the weighted sample size across the two survey rounds) (see **Example 4**). A survey design-adjusted p-value is also presented. This p-value, the result of a Rao-Scott chi-squared test that accounts for complex survey designs, indicates whether the difference between the round 1 and round 2 estimates is statistically significant.

When defining sub-populations (e.g., **sub\_1** and **sub\_2**), the sub-population variable is set to 1 for the observations included in the disaggregate category being compared over time, and it is set to 0 for the non-missing observations not included in the disaggregate category. It is important to set the sub-population variable to missing if the observation is missing a value for the selected disaggregate (e.g., if the child's sex is missing). Sub-populations can be set up by creating a dichotomous variable or by using an IF statement; see the sub-population estimation section of the Stata manual (pages 67-72) for more information.<sup>31</sup>

**Example 4** presents the output for a comparative analysis of the mean belief in local government to help the community cope with future difficulties variable over time for the male and female adult household category of the gendered household type disaggregate. The male and female adult household category is specified as the sub-population for analysis using an IF statement (i.e., **if genhhstype\_dj==1**).

---

<sup>30</sup> Note, however, that if a disaggregate category has fewer than 30 observations in it, that category should not be analyzed.

<sup>31</sup> StataCorp. (2021). Stata Survey Data Reference Manual Release 17. College Station, TX: StataCorp LLC, available at: <https://www.stata.com/manuals/svy.pdf>.

#### Example 4: Indicators of proportions: Comparative, by selected disaggregate output

```
. svy, subpop(if genhhtype_dj==1): tab locgov_resp survey, percent col stubwidth (16) format(%6.1f)
(running tabulate on estimation sample)
```

Number of strata = 36  
Number of PSUs = 253

Number of obs = 4,216  
Population size = 3,172,769  
Subpop. no. obs = 3,167  
Subpop. size = 2,378,170  
Design df = 217

HH believes local govt will help the community cope with future difficult times	survey		
	Baseline	Midline	Total
No	5.3	44.9	27.9
Yes	94.7	55.1	72.1
Total	100.0	100.0	100.0

Key: Column percentage

Pearson:  
Uncorrected chi2(1) = 808.2721  
Design-based F(1, 217) = 290.2601 P = 0.0000

#### Supplementary statistics for both descriptive and comparative proportions

Options can be specified with the **svy: tab** command to display additional information related to the indicator estimates (see **Examples 5-9**). Note that the output for each option is shown separately in the examples that follow, but multiple options can be specified in the same command.

**Confidence intervals:** Confidence intervals are included in the output when the 'ci' option is specified. For example:

```
svy: tab outcome, percent ci or
svy: tab outcome disaggregate, percent ci or
svy: tab outcome round, percent ci
```



**Example 5: Indicators of proportions: Supplementary statistics output, CI**

```
. svy: tab locgov_resp survey, percent col ci stubwidth (16) format(%6.1f)
(running tabulate on estimation sample)
```

Number of strata = 36	Number of obs = 4,129
Number of PSUs = 253	Population size = 3,118,123
	Design df = 217

HH believes local govt will help the community cope with future difficult times	survey		
	Baseline	Midline	Total
No	5.2 [3.8,7.2]	45.1 [41.4,48.8]	28.2 [26.0,30.5]
Yes	94.8 [92.8,96.2]	54.9 [51.2,58.6]	71.8 [69.5,74.0]
Total	100.0	100.0	100.0

Key: Column percentage  
[95% confidence interval for column percentage]

Pearson:

Uncorrected	chi2(1)	=	791.8917	
Design-based	F(1, 217)	=	300.3515	P = 0.0000

**Standard error:** Standard errors are included in the output when the 'se' option is specified. For example:

svy: tab outcome, percent se	or
svy: tab outcome disaggregate, percent se	or
svy: tab outcome round, percent se	

**Example 6: Indicators of proportions: Supplementary statistics output, standard errors**

```
. svy: tab locgov_resp survey, percent col se stubwidth (16) format (%6.1f)
(running tabulate on estimation sample)
```

Number of strata = 36  
Number of PSUs = 253

Number of obs = 4,129  
Population size = 3,118,123  
Design df = 217

HH believes local govt will help the community cope with future difficult times	survey		
	Baseline	Midline	Total
No	5.2 (0.8)	45.1 (1.9)	28.2 (1.1)
Yes	94.8 (0.8)	54.9 (1.9)	71.8 (1.1)
Total	100.0	100.0	100.0

Key: Column percentage  
(Linearized standard error of column percentage)

Pearson:

Uncorrected chi2(1) = 791.8917  
Design-based F(1, 217) = 300.3515 P = 0.0000

**Number of observations (unweighted *n*) for disaggregate categories or survey rounds:** The number of observations, or the unweighted sample size, for each disaggregate category or survey round is included in the output when the 'obs' option is specified. For example:

```
svy: tab outcome, percent obs or
svy: tab outcome disaggregate, percent obs or
svy: tab outcome round, percent obs
```

Note that this is different from the number of observations that is included in the output of all **svy: tab** commands, which presents the total number of observations included in the calculation across all disaggregate categories or rounds; the 'obs' option presents the number of observations included in the calculation for each outcome value and disaggregate category or survey round.

**Example 7: Indicators of proportions: Supplementary statistics output, number of observations by survey round**

```
. svy: tab locgov_resp survey, percent col obs stubwidth (16) format (%6.1f)
(running tabulate on estimation sample)
```

Number of strata = 36  
Number of PSUs = 253

Number of obs = 4,129  
Population size = 3,118,123  
Design df = 217

HH believes local govt will help the community cope with future difficult times	survey		
	Baseline	Midline	Total
No	5.2 107.0	45.1 934.0	28.2 1041.0
Yes	94.8 1943.0	54.9 1145.0	71.8 3088.0
Total	100.0 2050.0	100.0 2079.0	100.0 4129.0

Key: Column percentage  
Number of observations

Pearson:

Uncorrected chi2(1) = 791.8917  
Design-based F(1, 217) = 300.3515 P = 0.0000

**Design effect:** The design effect for the overall indicator, or for each characteristic disaggregate category or survey round, is included in the output when the 'deff' option is specified. For example:

```
svy: tab outcome, percent deff or
svy: tab outcome disaggregate, percent deff or
svy: tab outcome round, percent deff
```

Note that multiple options can be specified in the same command, but depending on the options selected, there are limitations on the other options that can be specified at the same time. For example, only 'count' or 'col' can be specified when 'se' is specified. Therefore, the same **svy: tab** command may have to be run multiple times with different options to obtain all desired results.

```
svy: tab outcome disaggregate, col percent ci se deff obs
plus
```

`svy: tab outcome disaggregate, percent count`

**Example 8: Indicators of proportions: Supplementary statistics output, design effects**

`. svy: tab locgov_resp survey, percent col deff stubwidth (16) format(%6.1f)`  
(running tabulate on estimation sample)

Number of strata = 36  
Number of PSUs = 253

Number of obs = 4,129  
Population size = 3,118,123  
Design df = 217

HH believes local govt will help the community cope with future difficult times	survey		
	Baseline	Midline	Total
No	5.2 2.5	45.1 3.4	28.2 2.7
Yes	94.8 2.5	54.9 3.4	71.8 2.7
Total	100.0	100.0	100.0

Key: Column percentage  
DEFF for variance of column percentage

Pearson:

Uncorrected chi2(1) = 791.8917  
Design-based F(1, 217) = 300.3515 P = 0.0000

Mean generalized deff = 2.1968  
CV of generalized deffs = 0.0000

### 3.2.2 Analyzing indicators of means

#### Descriptive, overall

Descriptive estimates for an indicator of means can be calculated using the following syntax:

`svy: mean outcome`

where `outcome` is the variable for the indicator of means

Using the syntax above, the sample-weighted mean for the indicator is presented, along with the mean's standard error and 95 percent CI, the number of observations included in the mean calculation (i.e., the unweighted sample size), and the corresponding population size (i.e., the weighted sample size) (see **Example 9**).

**Example 9: Indicators of means: Descriptive, overall output**

```
. svy: mean i_scap_index, cformat(%6.1f)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata = 36      Number of obs = 4,544
Number of PSUs   = 253    Population size = 3,367,202
Design df        = 217
```

	Linearized			
	Mean	std. err.	[95% conf. interval]	
i_scap_index	65.2	0.9	63.4	67.0

**Descriptive, by selected disaggregates**

Descriptive estimates for an indicator of means by a selected disaggregate can be calculated using the following syntax:

```
svy: mean outcome, over(disaggregate)
```

where: **outcome** is the variable for the indicator of means

**disaggregate** is the variable for the selected disaggregate

Using the syntax above, the sample-weighted means are presented for each disaggregate category, along with the means' standard errors and 95 percent CIs, the total number of observations included in the calculation (i.e., the unweighted sample size across all disaggregate categories), and the total population size (i.e., the weighted sample size across all disaggregate categories) (see **Example 10**).

### Example 10: Indicators of means: Descriptive, by selected disaggregates output #1

```
. svy: mean i_scap_index, over(genhhtype_dj) cformat(%6.1f)  
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata = 36  
Number of PSUs = 253

Number of obs = 4,544  
Population size = 3,367,202  
Design df = 217

	Linearized			
	Mean	std. err.	[95% conf. interval]	
c.i_scap_index@genhhtype_dj				
1. Male and Female adults	65.9	0.9	64.0	67.7
2. Female adults only	62.3	1.3	59.7	64.9
3. Male adults only	68.8	3.5	61.9	75.7
4. Children only	65.5	8.1	49.6	81.5

To obtain a survey design-adjusted p-value (or in this case, F-statistic), a second step must be taken; a regression is run using the following syntax:

```
svy: regress outcome i.disaggregate
```

where: **outcome** is the variable for the indicator of means

**disaggregate** is the variable for the selected disaggregate

In general, a t-test is used to determine whether there is a significant difference between the means of two groups, and analysis of variance (ANOVA) is used to determine whether there is a significant difference among the means of three or more groups. In Stata, however, t-test and ANOVA commands that can be run with **svy** do not exist, but there are other ways to obtain p-values for continuous outcomes and changes over time, including linear regression (**svy: regress**), as shown in **Example 11**, and the post-estimation command **lincom**, which is used in **Example 15** to compare means over time.<sup>32</sup>

Note that **i.** is required in the syntax if the selected characteristic is a categorical variable with more than two categories. If the selected characteristic is a continuous variable or a categorical variable with only two categories, **i.** is not required.

The resulting F-statistic indicates whether there is a statistically significant association between the selected disaggregate and the indicator of means (see **Example 11**). The F-statistic, however, does not indicate whether differences exist between the selected disaggregate categories or whether associations exist between specific disaggregate categories and the outcome indicator.

<sup>32</sup> UC REGENTS. (2019). UCLA Institute for Digital Research and Education. *How can I do a t-test with survey data?* | Stata FAQ. Available at: <https://stats.idre.ucla.edu/stata/faq/how-can-i-do-a-t-test-with-survey-data/>.

Note that a sub-population of households is specified in **Example 11**; households in the children-only household category of the gendered household type disaggregate are excluded from the regression because the numbers of observations in this category are too small both at baseline (n=8) and midline (n=3) to produce valid estimates. The male and female adult household category does not appear in the output because it is the reference category.

**Example 11: Indicators of means: Descriptive, by selected disaggregates output #2**

```
. svy, subpop(if genhhtype_dj<4) : regress i_scap_index i.genhhtype_dj, cformat(%6.1f)
(running regress on estimation sample)
```

Survey: Linear regression

Number of strata =	36	Number of obs =	4,544
Number of PSUs =	253	Population size =	3,367,202
		Subpop. no. obs =	4,526
		Subpop. size =	3,354,372
		Design df =	217
		F(2, 216) =	5.29
		Prob > F =	0.0057
		R-squared =	0.0027

i_scap_index	Linearized		t	P> t	[95% conf. interval]	
	Coefficient	std. err.				
genhhtype_dj						
2. Female adults only	-3.6	1.2	-2.89	0.004	-6.1	-1.1
3. Male adults only	2.9	3.4	0.87	0.384	-3.7	9.5
_cons	65.9	0.9	70.38	0.000	64.0	67.7

## Comparative, overall

Comparative estimates for an indicator of means can be calculated using the following syntax:

```
svy: mean outcome, over(round)
where: outcome is the variable for the indicator of means
round is the variable for the survey round (e.g., baseline or midline)
```

Using the syntax above, the sample-weighted means are presented for each survey round, along with the means' standard errors and 95 percent CIs, the total number of observations included in the calculation (i.e., the unweighted sample size across rounds), and the total population size (i.e., the weighted sample size across rounds) (see **Example 12**).

### Example 12: Indicators of means: Comparative, overall output #1

```
. svy: mean i_scap_index, over(survey) cformat(%6.1f)  
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata =	36	Number of obs =	4,544
Number of PSUs =	253	Population size =	3,367,202
		Design df =	217

	Linearized			
	Mean	std. err.	[95% conf. interval]	
c.i_scap_index@survey				
Baseline	69.9	1.2	67.6	72.2
Midline	61.2	1.3	58.6	63.8

To determine whether there is a statistically significant difference in the indicator of mean estimates between survey rounds, a second step must be taken. The `lincom` command, which uses linear combinations of coefficients generated by a `svy: mean` command, can be used to obtain a survey design-adjusted p-value:

```
lincom _b[outcome1] - _b[outcome0]
```

where: `_b[outcome0]` is the coefficient for the indicator of means in the first survey round  
`_b[outcome1]` is the coefficient for the indicator of means in the second survey round

The resulting p-value indicates whether the difference in the indicator estimates between the two time points is statistically significant (see **Example 13**).

Note that to obtain the names of the coefficients to be used in the `lincom` command, the '`coeflegend`' option can be added to the `svy: mean` estimation command.



### Example 13: Indicators of means: Comparative, overall output #2

```
. svy: mean i_scap_index, over(survey) coeflegend cformat(%6.1f)  
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata =	36	Number of obs =	4,544
Number of PSUs =	253	Population size =	3,367,202
		Design df =	217

	Mean	Legend
c.i_scap_index@survey		
Baseline	69.9	_b[c.i_scap_index@0bn.survey]
Midline	61.2	_b[c.i_scap_index@1.survey]

### Comparative, by selected disaggregates

Comparative estimates for an indicator of means by a selected disaggregate can be calculated using the following syntax:

```
svy: mean outcome, over(disaggregate round)
```

where: **outcome** is the variable for the indicator of means

**disaggregate** is the variable for the selected disaggregate

**round** is the variable for the survey round (e.g., baseline or midline)

Using the syntax above, the sample-weighted means are presented for each disaggregate category and survey round combination, along with the means' standard errors and 95 percent CIs, and the total number of observations included in the calculation (i.e., the unweighted sample size across all disaggregate categories and both survey rounds) (see **Example 14**).

#### Example 14: Indicators of means: Comparative, by selected disaggregates, output #1

```
. svy: mean i_scap_index, over(genhhtype_dj survey) cformat(%6.1f)
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata = 36  
Number of PSUs = 253

Number of obs = 4,544  
Population size = 3,367,202  
Design df = 217

	Linearized			
	Mean	std. err.	[95% conf. interval]	
c.i_scap_index@genhhtype_dj#survey				
1. Male and Female adults#Baseline	70.1	1.3	67.6	72.5
1. Male and Female adults#Midline	62.2	1.3	59.5	64.8
2. Female adults only#Baseline	68.9	1.8	65.2	72.5
2. Female adults only#Midline	57.1	1.9	53.4	60.8
3. Male adults only#Baseline	73.3	3.8	65.9	80.8
3. Male adults only#Midline	65.5	5.2	55.3	75.8
4. Children only#Baseline	66.9	9.8	47.7	86.2
4. Children only#Midline	63.5	13.7	36.5	90.4

To obtain a survey design-adjusted p-value for each characteristic category, additional steps must be taken using the following syntax for each characteristic category:

```
lincom _b[outcomeX.1] - _b[outcomeX.0]
```

where: **\_b[outcome]1** is the coefficient for the indicator of means for disaggregate category X in the second survey round

**\_b[outcome]0** is the coefficient for the indicator of means for disaggregate category X in the first survey round

The resulting p-value indicates whether the difference the indicator estimates for each disaggregate between the two time points is statistically significant (see **Example 15**).

Note that to obtain the names of the coefficients to be used in the **lincom** command, the **'coeflegend'** option can be added to the **svy: mean** estimation command.

### Example 15: Indicators of means: Comparative, by selected characteristics, output #2

```
. svy: mean i_scap_index, over (genhhstype_dj survey) coeflegend cformat(%6.1f)
(running mean on estimation sample)
```

Survey: Mean estimation

```
Number of strata = 36      Number of obs   =    4,544
Number of PSUs   = 253      Population size = 3,367,202
                        Design df    =    217
```

	Mean	Legend
c.i_scap_index@genhhstype_dj#survey		
1. Male and Female adults#Baseline	70.1	_b[c.i_scap_index@1bn.genhhstype_dj#0bn.survey]
1. Male and Female adults#Midline	62.2	_b[c.i_scap_index@1bn.genhhstype_dj#1.survey]
2. Female adults only#Baseline	68.9	_b[c.i_scap_index@2.genhhstype_dj#0bn.survey]
2. Female adults only#Midline	57.1	_b[c.i_scap_index@2.genhhstype_dj#1.survey]
3. Male adults only#Baseline	73.3	_b[c.i_scap_index@3.genhhstype_dj#0bn.survey]
3. Male adults only#Midline	65.5	_b[c.i_scap_index@3.genhhstype_dj#1.survey]
4. Children only#Baseline	66.9	_b[c.i_scap_index@4.genhhstype_dj#0bn.survey]
4. Children only#Midline	63.5	_b[c.i_scap_index@4.genhhstype_dj#1.survey]

```
. lincom _b[c.i_scap_index@1bn.genhhstype_dj#1.survey] - _b[c.i_scap_index@1bn.genhhstype_dj#0bn.survey], cformat(%6.1f)
```

```
( 1) - c.i_scap_index@1bn.genhhstype_dj#0bn.survey + c.i_scap_index@1bn.genhhstype_dj#1.survey = 0
```

Mean	Coefficient	Std. err.	t	P> t	[95% conf. interval]
(1)	-7.9	1.8	-4.28	0.000	-11.5 -4.3

```
. lincom _b[c.i_scap_index@2.genhhstype_dj#1.survey] - _b[c.i_scap_index@2.genhhstype_dj#0bn.survey], cformat(%6.1f)
```

```
( 1) - c.i_scap_index@2.genhhstype_dj#0bn.survey + c.i_scap_index@2.genhhstype_dj#1.survey = 0
```

Mean	Coefficient	Std. err.	t	P> t	[95% conf. interval]
(1)	-11.7	2.6	-4.47	0.000	-16.9 -6.6

```
. lincom _b[c.i_scap_index@3.genhhstype_dj#1.survey] - _b[c.i_scap_index@3.genhhstype_dj#0bn.survey], cformat(%6.1f)
```

```
( 1) - c.i_scap_index@3.genhhstype_dj#0bn.survey + c.i_scap_index@3.genhhstype_dj#1.survey = 0
```

Mean	Coefficient	Std. err.	t	P> t	[95% conf. interval]
(1)	-7.8	6.4	-1.21	0.228	-20.5 4.9

```
. lincom _b[c.i_scap_index@4.genhhstype_dj#1.survey] - _b[c.i_scap_index@4.genhhstype_dj#0bn.survey], cformat(%6.1f)
```

```
( 1) - c.i_scap_index@4.genhhstype_dj#0bn.survey + c.i_scap_index@4.genhhstype_dj#1.survey = 0
```

Mean	Coefficient	Std. err.	t	P> t	[95% conf. interval]
(1)	-3.5	16.8	-0.21	0.835	-36.6 29.6

## Supplementary statistics for both descriptive and comparative means

Design effects, standard deviations, and numbers of observations and population sizes by disaggregate categories and survey rounds can be obtained using the '**estat**' command after running a '**svy: mean**' command (see **Example 16**).

**Design effect:** The design effect for indicators of means can be obtained by running the following command after running a **svy: mean** estimation command:

```
estat effects
```

**Standard deviation:** The standard deviation for indicators of means can be obtained by running the following command after running a **svy: mean** estimation command:

```
estat sd
```

**Number of observations (unweighted *n*) and population size (weighted *n*):** The number of observations included in the mean calculations by disaggregate category and survey round, and the corresponding population sizes, can be obtained by running the following command after running a **svy: mean** estimation command:

```
estat size
```

### Example 16: Indicators of means, supplementary statistics output

```
. svy: mean i_scap_index, over (survey) cformat(%6.1f)  
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata =	36	Number of obs =	4,544
Number of PSUs =	253	Population size =	3,367,202
		Design df =	217

	Linearized			
	Mean	std. err.	[95% conf. interval]	
c.i_scap_index@survey				
Baseline	69.9	1.2	67.6	72.2
Midline	61.2	1.3	58.6	63.8

```
. estat sd
```

Over	Mean	Std. dev.
c. i_scap_index@ survey		
Baseline	69.88792	30.99113
Midline	61.18866	28.51229

```
. estat size
```

Over	Mean	Linearized std. err.	Obs	Size
c. i_scap_index@ survey				
Baseline	69.88792	1.160536	2,442	1,560,486.143
Midline	61.18866	1.306612	2,102	1,806,715.527

### 3.2.3 Analyzing indicators of proportions or means if the data are not available

If the original data used to calculate the original estimates are not available, the z-score and the corresponding p-value can be manually calculated using the two indicator point estimates and either their CIs or standard errors, for instance, in a spreadsheet program, such as Excel. Note, however, that the preferred approach is to use statistical software to assess changes or associations whenever possible. Although the approach in this section accounts for complex survey designs by using the standard errors or CIs associated with the estimates, the approach is limited by the precision of the point estimates and statistics available. The approach also assumes that Stata's `svy` command or a

similar command in another statistical software program that accounts for the survey design was used to compute the standard errors. Results in reports or results tables are usually presented to only one or two decimal places; this rounding will result in p-values that differ from those obtained from statistical software packages, which use many more decimals in their p-value calculations.<sup>33</sup> The p-values should not differ substantially, however. If a result is extremely significant or extremely non-significant, this will be true regardless of the approach used. If a result is near a level-of-significance cutoff, the level **may** differ by approach, but it provides a good indication of level of significance in the absence of the original data. In all cases, the approach used should be documented.

Assuming that the two indicator estimates being compared (e.g., baseline and midline or group 1 and group 2) follow the standard normal distribution,<sup>34</sup> let  $\hat{\theta}_1$  denote the estimate computed for baseline (or group 1), and  $\hat{\theta}_2$  denote the estimate computed for midline (or group 2). To test the null hypothesis,  $H_0: \hat{\theta}_1 = \hat{\theta}_2$ , versus the alternative hypothesis,  $H_a: \hat{\theta}_1 \neq \hat{\theta}_2$ , the test statistic (z-score) can be calculated as follows:<sup>35</sup>

$$Z = \frac{|\hat{\theta}_1 - \hat{\theta}_2|}{\sqrt{SE_1^2 + SE_2^2}}$$

where  $SE_1$  is the standard error for  $\hat{\theta}_1$   
 $SE_2$  is the standard error for  $\hat{\theta}_2$

If the standard errors are not available, the upper bound of the CIs can be used to calculate the standard errors:

$$SE_1 = \frac{(UCI_1 - \hat{\theta}_1)}{1.96}$$

$$SE_2 = \frac{(UCI_2 - \hat{\theta}_2)}{1.96}$$

where  $\hat{\theta}_1$  is the estimate computed for baseline (or group 1)  
 $\hat{\theta}_2$  is the estimate computed for midline (or group 2)  
 $UCI_1$  is the upper CI bound for baseline (or group 1)  
 $UCI_2$  is the upper CI bound for midline (or group 2)

Using the standard normal tables, the associated two-tailed p-value of the z-score can be found as  $2 * P(Z > Z_\alpha)$ , where  $Z_\alpha$  is the critical z value for the chosen  $\alpha$  level significance (e.g.,  $Z_\alpha = 1.96$  for the 95 percent confidence level and  $Z_\alpha = 1.645$  for the 90 percent confidence level). The p-value is then assessed in comparison with the significance level (0.05 for 95 percent confidence level, or 0.10 for

<sup>33</sup> An additional benefit of using the original round 1 data for comparative analyses is that the round 1 estimates can be validated and recomputed using the same methods used to generate the round 2 estimates to ensure comparability.

<sup>34</sup> This is also what Stata and other statistical packages assume when conducting the statistical tests described in the previous sections.

<sup>35</sup> Heeringa, S., West, B., & Berglund, P. (2010). *Applied survey data analysis*. New York: Chapman and Hall/CRC. <https://doi.org/10.1201/9781420080674>

90 percent confidence level), where the null hypothesis is accepted if the p-value is larger than the significance level, and rejected otherwise. If the null hypothesis is accepted, any observed difference is due to sampling or experimental error, but if the null hypothesis is rejected, it is likely that the indicator estimates being compared are truly different.

The information from **Example 12** in Section 3.2.2 that would be available in a comparative report shows that the mean rescaled social capital index score is 69.9 percent at baseline, with a standard error of 1.2, a 95 percent CI lower bound of 67.6 percent, and a 95 percent CI upper bound of 72.2 percent, and is 61.2 percent at midline, with a standard error of 1.3, a 95 percent CI lower bound of 58.6 percent, and a 95 percent CI upper bound of 63.8 percent. Using the estimates and standard errors, the z-score can be calculated as follows:

$$Z = \frac{|\hat{\theta}_1 - \hat{\theta}_2|}{\sqrt{SE_1^2 + SE_2^2}} = \frac{|69.9 - 61.2|}{\sqrt{1.2^2 + 1.3^2}} = \frac{8.7}{\sqrt{1.44 + 1.69}} = 4.9175$$

If the standard errors are not available, the upper bounds of the CI can be used to calculate the standard error. Note that due to rounding, reported standard errors will not match exactly the standard errors obtained using the CIs. Likewise, due to rounding, standard errors obtained using reported CIs will not match exactly the standard errors available in statistical software programs.

$$SE_1 = \frac{UCI_1 - \hat{\theta}_1}{1.96} = \frac{(72.2 - 69.9)}{1.96} = 1.1734 \quad SE_2 = \frac{UCI_2 - \hat{\theta}_2}{1.96} = \frac{(63.8 - 61.2)}{1.96} = 1.3265$$

By plugging the z-score into an online p-value from z-score calculator<sup>36</sup> and specifying a two-tailed hypothesis and significance level of 0.05, a p-value of 0.000 is obtained, which indicates that the result is statistically significant at  $p < 0.001$ . Alternately, a z table could be used.

### 3.3 Data analysis folder structure and file naming conventions

Data analysts can set up their directories and name their files according to what works best for them. In the core Stata template syntax files, however, certain conventions, which are described in this section, were used.

#### 3.3.1 Folder structure

The folder structure assumes that data analysts set up a main folder for the ZOI Survey on their C-drive and store the syntax files and analytic data files in separate sub-folders.

Syntax	"C:\FTF ZOI Survey [COUNTRY] [YEAR]\Syntax"
Data	"C:\FTF ZOI Survey [COUNTRY] [YEAR]\Data"

<sup>36</sup> An example of one such calculator is available at:  
<https://www.socscistatistics.com/pvalues/normaldistribution.aspx>

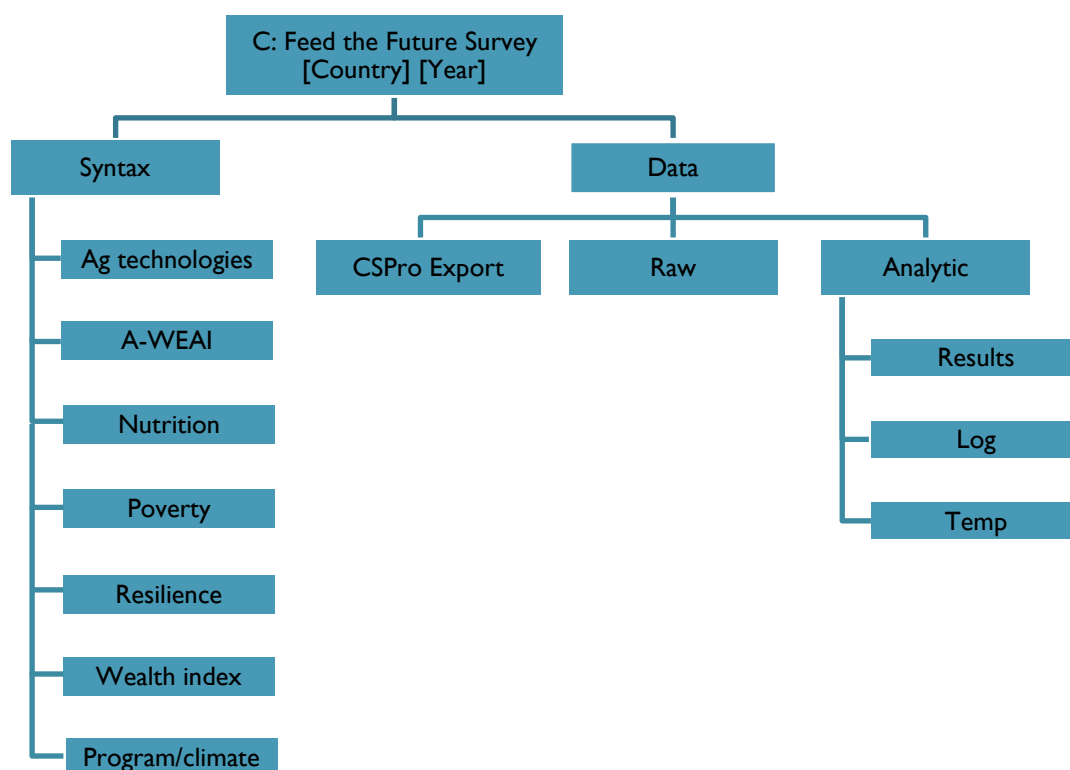
The Syntax folder has several topical sub-folders that contain the relevant core Stata template syntax files; they can be adapted as applicable for the survey. The individual-level and household-level analytic syntax files, which include code to create analytic variables used throughout the data analysis, are stored in the main Syntax folder.

The Data folder is further divided into sub-folders for the CPro export files, the raw data files (i.e., the data files with variable and value set labels created from the CPro export files), and all the analytic files.

CPro export	"C:\FTF ZOI Survey [COUNTRY] [YEAR]\Data\CPro Export"
Raw	"C:\FTF ZOI Survey [COUNTRY] [YEAR]\Data\Raw"
Analytic	"C:\FTF ZOI Survey [COUNTRY] [YEAR]\Data\Analytic"

In the Analytic sub-folder, there are three additional sub-folders:

Results	for storing all the calculation results
Log	for storing all the Stata log files
Temp	for storing any temporary data files created during data analysis



### 3.3.2 File naming conventions

The following naming conventions are used for the raw, analytic, log, and syntax files throughout the core Stata template syntax files. [COUNTRY] should be replaced by the name of the ZOI Survey country. [YEAR] should be replaced by the year of the ZOI Survey fieldwork.<sup>37</sup> [X] should be replaced

<sup>37</sup> If fieldwork spans two calendar years, use year in which fieldwork started.



by a term that identifies the indicator, group of indicators, or component of indicator being calculated. These terms are specified in the syntax file names and in the syntax files themselves. The results files should include only variables that will be added to the final dataset and identifier variables to enable merging the data.

Raw data files:	\Raw\FTF ZOI Survey [COUNTRY] [YEAR] persons data raw.dta \Raw\FTF ZOI Survey [COUNTRY] [YEAR] household data raw.dta
Analytic data files:	\Analytic\FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic.dta \Analytic\FTF ZOI Survey [COUNTRY] [YEAR] household data analytic.dta
Log files:	\Analytic\Log\FTF ZOI Survey [COUNTRY] [YEAR] [X].log
Syntax files:	\Syntax\FTF ZOI Survey [COUNTRY] [YEAR] syntax [X].do
Temp data files:	\Analytic\Temp\[X].dta
Results files:	\Analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] [X].dta

## 4. KEY INDIVIDUAL-LEVEL ANALYTIC VARIABLES

In this chapter, key individual-level analytic variables used to calculate or disaggregate indicators are defined and their calculations described. The chapter is divided into six sections: Section 4.1 describes three analytic variables related to all household members—de jure status, de facto status, and sex; Section 4.2 describes analytic age-in-month variables for children under 5 years of age; Section 4.3 describes analytic age-in-year variables for all household members; Section 4.4 describes analytic variables for women of reproductive age; Section 4.5 describes analytic variables for primary adult decision-makers; and Section 4.6 describes analytic variables for producers.

All variables created in this chapter require variables in the individual-level data file created from the Census and Survey Processing System exported data (FTF ZOI Survey [COUNTRY] [YEAR] persons data raw). The analytic variables created are saved to an analytic version of the individual data file (FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic), which can be used across data analysts so that they do not have to re-create these analytic variables as part of their analysis. The step-by-step procedures to create the key individual-level analytic variables included in this chapter follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax persons analytic* do file.

### 4.1 Household member status (de jure, de facto) and sex variables

This section describes analytic variables that identify de jure household members, de facto household members, and household members' sex.

#### 4.1.1 De jure household member

This variable identifies all de jure household members—that is, usual household members—according to information collected in the household roster. The household roster includes individuals who are usual members of the household, as well as those who are not usual members but stayed in the household the night preceding the survey. This variable identifies only the subset who are usual household members.

#### Definitions

Treatment of missing data	Not applicable. Missing values are not allowed in the household roster for the question that asks if individuals are usual household members (v105a).
Survey variables used	v105a
Analytic variables used	n/a
Analytic variables created	hhmem_dj

#### Calculations

**Step 1.** Create a binary de jure household member analytic variable (hhmem\_dj).

```
Set          hhmem_dj=0
Replace      hhmem_dj=1 if v105a=1
Label values 0 "No"
              1 "Yes"
Label variable "De jure (usual) HH member"
```

### 4.1.2 De facto household member

This variable identifies all de facto household members—that is, all individuals who stayed in the household the night before the survey, regardless of whether they are usual household members according to information collected in the household roster. De facto household members include guests of the household, but they will not include usual household members who did not stay in the household the night before the survey.

#### Definitions

Treatment of missing data	Not applicable. Missing values are not allowed in the household roster for the question that asks if individuals stayed in the household the night before the survey ( <b>v105b</b> ).
Survey variables used	<b>v105b</b>
Analytic variables used	n/a
Analytic variables created	<b>hhmem_df</b>

#### Calculations

**Step 1.** Create a binary de facto household member analytic variable (**hhmem\_df**).

```
Set          hhmem_df=0
Replace      hhmem_df=1 if v105b=1
Label values 0 "No"
              1 "Yes"
Label variable "De facto HH member (stayed in HH night prior to survey) "
```

### 4.1.3 Sex

This variable identifies the sex of all individuals according to information collected in the household roster.

#### Definitions

Treatment of missing data	Missing values for sex ( <b>v102</b> ) are not allowed in the household roster. However, if <b>v102</b> and sex in Module 5, <i>Children's nutrition</i> , ( <b>v501</b> ) are found to be different for children under 5 years of age during data processing, the child's sex in both the household roster and Module 5 will be set to '7' to indicate inconsistent values.
Survey variables used	<b>v102</b>
Analytic variables used	n/a
Analytic variables created	<b>sex</b>

#### Calculations

**Step 1.** Create a binary sex analytic variable (**sex**).

```
Set          sex=missing
Replace      sex=1 if v102=1
Replace      sex=2 if v102=2
```

Label values      1 "Male"  
                          2 "Female"  
 Label variable    "Sex of HH member"

## 4.2 Children under 5 years age-in-month variables

This section describes two analytic age variables for children under 5 years of age: age in days and age in years, variables that identify children 0-5 months, 6-23 months, 0-23 months, and 0-59 months of age; and two categorical age variables that are used as disaggregates for children's results.

### 4.2.1 Age in days, children under 5 years of age

This variable identifies the age in days of all children under 5 years of age according to information collected in the Module 5, *Children's nutrition*.

#### Definitions

Treatment of missing data	Children whose day of birth (v502 and v506) is missing or unknown are assigned '15' as their birth day. The <b>cage_days</b> analytic variable is set to missing for children who are missing a valid month of birth (v502m and v506m) or year of birth (v502y and v506y).
Survey variables used	hhea, hhnum, v502d, v502m, v502y, v505, v506d, v506m, v506y, ahintd, ahintm, ahinty
Analytic variables used	n/a
Analytic variables created	bday, bmon, byear, bdate, intdate, cage_days

#### Calculations

**Step 1.** Create a birth day variable (**bday**) using the day reported by the caregiver in the nutrition module if the day is valid—that is, 31 or less.

Set                      **bday=missing**  
 Replace                **bday=v502d if v502d≤31**

**Step 2.** Set **bday** equal to the birth day on the child's vaccination card if the caregiver did not report a valid birth day, the interviewer saw the vaccination card, and the day on the vaccination card is valid—that is, 31 or less.

Replace                **bday=v506d if v505=1 and v506d≤31 and bday=missing**

**Step 3.** Set **bday** equal to 15 if the caregiver did not report a valid birth day, and there was not a valid birth day on the child's vaccination card. Note that **bday** should be set to 15 only if the child's caregiver consented to respond to the children's nutrition module (that is, if **v500c=1**) and the child had a valid birth month and birth year. Ensure that [YYYY] in the syntax reflects that the maximum values for **v502y** and **v506y** are the year of the ZOI Survey fieldwork, or the later year if fieldwork spanned 2 years.

Replace                **bday=15 if v502d≥98 and v506d≥98 and v500c=1 and  
                                  (v502m≤12 or v506m≤12) and (v502y≤[YYYY] or  
                                  v506y≤[YYYY]) and bday=missing**

```
Label variable    "Child's day of birth, derived"
```

**Step 4.** Create a birth month variable (**bmon**) using the month reported by the caregiver in the nutrition module if the month is valid—that is, 12 or less. If the month reported by the caregiver is not valid or is missing, set the birth month to be the birth month on the child's vaccination card if the month on the card is valid.

```
Set              bmon=missing
Replace          bmon=v502m if v502m≤12
Replace          bmon=v506m if v505=1 and v506m≤12 and
                 bmon=missing
Label variable    "Child's month of birth, derived"
```

**Step 5.** Create a birth year variable (**byear**) using the year reported by the caregiver in the nutrition module if the year is valid—that is, less than or equal to the year of the survey data collection. If the year reported by the caregiver is not valid or is missing, set the birth year to be the birth year on the child's vaccination card if the year on the card is valid. Ensure that the syntax reflects that the maximum values for **v502y** and **v506y** are the year of the ZOI Survey fieldwork, or the later year if fieldwork spanned 2 years.

```
Set              byear=missing
Replace          byear=v502y if v502y≤[YYYY]
Replace          byear=v506y if v505=1 and v506y≤[YYYY] and
                 byear=missing
Label variable    "Child's year of birth, derived"
```

**Step 6.** Combine the birth day, month, and year variables into a variable formatted as a date that can be used in date arithmetic (**bdate**).

```
Set              bdate=date(bmon,bday,byear)
Label variable    "Child's date of birth, derived"
```

**Step 7.** Add the variables that hold the final date of interview information (**ahintd**, **ahintm**, **ahinty**) in the household-level data file into the individual-level data file using cluster (**hhea**) and household number (**hhnum**) as the key matching variables.

```
Add variables    ahintd ahintm ahinty
from              "FTF ZOI Survey [COUNTRY] [YEAR] household data
                 raw"
matching on       hhea and hhnum
```

**Step 8.** Combine the final interview day, month, and year variables into a variable formatted as a date that can be used in date arithmetic (**intdate**).

```
Set              intdate=date(ahintm,ahintd,ahinty)
Label variable    "Date of interview"
```

**Step 9.** Create an age in days variable by subtracting the birthdate variable from the date of interview variable (`cage_days`).

```
Set                cage_days=intdate-bdate if bdate#missing
Label variable    "Child's age in days"
```

#### 4.2.2 Age in months, children under 5 years of age

This variable identifies the age in months of all children under 5 years of age according to information collected in Module 5, *Children's nutrition*, and the household roster.

#### Definitions

Treatment of missing data	Children's age in months reported by the caregiver interviewed ( <code>v508</code> ) is used for children who are missing a value for <code>cage_days</code> . Set <code>cage_months</code> to missing for children who are also missing a value for <code>v508</code> . If Module 5 is not completed for a child, they will not have a value for <code>cage_months</code> . If a child does not have a value for <code>cage_months</code> , and they are less than 1 year of age according to the household roster, <code>v104a</code> will be used for their age in months. If a child does not have a value for <code>cage_months</code> , and they are between 1 and 4 years of age according to the household roster, <b>they will not have an age for <code>cage_months_int</code>.</b>
Survey variables used	<code>v104a</code> , <code>v508</code>
Analytic variables used	<code>cage_days</code>
Analytic variables created	<code>cage_months</code> , <code>cage_months_int</code>

#### Calculations

**Step 1.** Create a variable that holds the age in months (`cage_months`) of children whose primary caregiver consented to respond to the children's nutrition module by dividing the child's age in days by the number of days in a year (365.25) and multiplying by 12 months.

```
Set                cage_months=cage_days÷365.25*12
Label variable    "Child's age in months, including decimal"
```

**Step 2.** Create an integer version of the child's age in months variable (`cage_months_int`) using the variable created in the previous step. If the child does not have a value for `cage_months` and completed Module 5, use the child's caregiver-reported age (`v508`). If child did not complete Module 5 but is less than 1 year of age in the household roster, use the child's age in months (`v104a`).

```
Set                cage_months_int=integer(cage_months)
Replace            cage_months_int=v508 if cage_months_int=missing
                  and v508<60
Replace            cage_months_int=v104a if cage_months_int=missing
                  and v104a#missing
Label variable    "Child's age in months, excluding decimal"
```

### 4.2.3 Children 0-5 months of age, overall and by sex

These variables identify all children 0-5 months of age—overall and by sex, according to the `cage_months_int` variable created in Section 4.2.2 and information collected in the household roster.

#### Definitions

Treatment of missing data	Not applicable for <code>c0_5m</code> . If the analytic variable <code>sex</code> is set to '7' (see Section 4.1.3), analytic variables <code>c0_5mm</code> and <code>c0_5mf</code> will be '0.'
Survey variables used	n/a
Analytic variables used	<code>sex</code> , <code>cage_months_int</code>
Analytic variables created	<code>c0_5m</code> , <code>c0_5mm</code> , <code>c0_5mf</code>

#### Calculations

**Step 1.** Create a binary variable that identifies all children 0-5 months of age (`c0_5m`) using the derived age-in-months variable (`cage_months_int`).

```
Set          c0_5m=0
Replace      c0_5m=1 if cage_months_int<6
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months"
```

**Step 2.** Create a binary variable that identifies all male children 0-5 months of age (`c0_5mm`) using the `c0_5m` and `sex` variables.

```
Set          c0_5mm=0
Replace      c0_5mm=1 if c0_5m=1 and sex=1
Label values 0 "No"
              1 "Yes"
Label variable "Male child 0-5 months"
```

**Step 3.** Create a binary variable that identifies all female children 0-5 months of age (`c0_5mf`) using the `c0_5m` and `sex` variables.

```
Set          c0_5mf=0
Replace      c0_5mf=1 if c0_5m=1 and sex=2
Label values 0 "No"
              1 "Yes"
Label variable "Female child 0-5 months"
```

### 4.2.4 Children 6-23 months of age, overall and by sex

These variables identify all children 6-23 months of age—overall and by sex, according to the `cage_months_int` variable created in Section 4.2.2 and information collected in the household roster.

## Definitions

Treatment of missing data	Not applicable for <code>c6_23m</code> . If the analytic variable <code>sex</code> is set to '7' (see Section 4.1.3), analytic variables <code>c6_23mm</code> and <code>c6_23mf</code> will be '0.'
Survey variables used	<code>v104</code>
Analytic variables used	<code>sex</code> , <code>cage_months_int</code>
Analytic variables created	<code>c6_23m</code> , <code>c6_23mm</code> , <code>c6_23mf</code>

## Calculations

**Step 1.** Create a binary variable that identifies all children 6-23 months of age (`c6_23m`) using the derived age-in-months variable (`cage_months_int`). If a child is missing a value for `cage_months_int` their age in years from the household roster (`v104`) is used.

```
Set          c6_23m=0
Replace      c6_23m=1 if cage_months_int≥6 and
              cage_months_int≤23
Replace      c6_23m=1 if v104=1 and c6_23m=0
Label values 0 "No"
              1 "Yes"
Label variable "Child 6-23 months"
```

**Step 2.** Create a binary variable that identifies all male children 6-23 months of age in the household roster (`c6_23mm`) using the `c6_23m` and `sex` variables.

```
Set          c6_23mm=0
Replace      c6_23mm=1 if c6_23m=1 and sex=1
Label values 0 "No"
              1 "Yes"
Label variable "Male child 6-23 months"
```

**Step 3.** Create a binary variable that identifies all female children 6-23 months of age in the household roster (`c6_23mf`) using the `c6_23m` and `sex` variables.

```
Set          c6_23mf=0
Replace      c6_23mf=1 if c6_23m=1 and sex=2
Label values 0 "No"
              1 "Yes"
Label variable "Female child 6-23 months"
```

### 4.2.5 Children 0-23 months of age, overall and by sex

These variables identify all children 0-23 months of age—overall and by sex, according to the `cage_months_int` variable created in Section 4.2.2 and information collected in the household roster.



## Definitions

Treatment of missing data	If the child does not have a value for the derived age-in-months variable ( <b>cage_months_int</b> ), the age ( <b>v104</b> ) variable from the household roster is used. Variable <b>v104</b> cannot contain missing data. If the analytic variable <b>sex</b> is set to '7' (see Section 4.1.3), analytic variables <b>c0_23mm</b> and <b>c0_23mf</b> will be '0.'
Survey variables used	<b>v104</b>
Analytic variables used	<b>sex, cage_months_int</b>
Analytic variables created	<b>c0_23m, c0_23mm, c0_23mf</b>

## Calculations

**Step 1.** Create a binary variable that identifies all children 0-23 months of age (**c0\_23m**) using the derived age-in-months variable (**cage\_months\_int**) and, if **cage\_months\_int** is missing, the age-in-years variable (**v104**) in the household roster.

```
Set          c0_23m=0
Replace      c0_23m=1 if cage_months_int≤23
Replace      c0_23m=1 if v104≤1 and c0_23m=0
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-23 months"
```

**Step 2.** Create a binary variable that identifies all male children 0-23 months of age in the household roster (**c0\_23mm**) using the **c0\_23m** and **sex** variables.

```
Set          c0_23mm=0
Replace      c0_23mm=1 if c0_23m=1 and sex=1
Label values 0 "No"
              1 "Yes"
Label variable "Male child 0-23 months"
```

**Step 3.** Create a binary variable that identifies all female children 0-23 months of age in the household roster (**c0\_23mf**) using the **c0\_23m** and **sex** variables.

```
Set          c0_23mf=0
Replace      c0_23mf=1 if c0_23m=1 and sex=2
Label values 0 "No"
              1 "Yes"
Label variable "Female child 0-23 months"
```

### 4.2.6 Children 0-59 months of age, overall and by sex

These variables identify all children 0-59 months of age—overall and by sex, according to the **cage\_months\_int** variable created in Section 4.2.2 and information collected in the household roster.

## Definitions

Treatment of missing data	Not applicable for <code>c0_59m</code> . If the analytic variable <code>sex</code> is set to '7' (see Section 4.1.3), analytic variables <code>c0_59mm</code> and <code>c0_59mf</code> will be '0.'
Survey variables used	<code>v104</code>
Analytic variables used	<code>sex</code> , <code>cage_months_int</code>
Analytic variables created	<code>c0_59m</code> , <code>c0_59mm</code> , <code>c0_59mf</code>

## Calculations

**Step 1.** Create a binary variable that identifies all children 0-59 months of age (`c0_59m`) using the derived age-in-months variable (`cage_months_int`) and, if `cage_months_int` is missing, the age-in-years variable (`v104`) in the household roster.

```
Set          c0_59m=0
Replace      c0_59m=1 if cage_months_int≤59
Replace      c0_59m=1 if v104≥1 and v104≤4 and c0_59m=0
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-59 months"
```

**Step 2.** Create a binary variable that identifies all male children 0-59 months of age in the household roster (`c0_59mm`) using the `c0_59m` and `sex` variables.

```
Set          c0_59mm=0
Replace      c0_59mm=1 if c0_59m=1 and sex=1
Label values 0 "No"
              1 "Yes"
Label variable "Male child 0-59 months"
```

**Step 3.** Create a binary variable that identifies all female children 0-23 months of age in the household roster (`c0_23mf`) using the `c0_59m` and `sex` variables.

```
Set          c6_59mf=0
Replace      c6_59mf=1 if c0_59m=1 and sex=2
Label values 0 "No"
              1 "Yes"
Label variable "Female child 0-59 months"
```

### 4.2.7 Children 6-23 months age category—6 month intervals

This variable categorizes all children 6-23 months of age by 6-month age categories according to the `cage_months_int` variable created in Section 4.2.2. This variable is used as a disaggregate for the percentage of children 0-23 months of age receiving a diverse diet indicator, so it is required for only children who completed Module 5; see Sections 16.1.2 and 16.2.2.

## Definitions

Treatment of missing data	Not applicable. All children who completed Module 5 will have a value for <code>cage_months_int</code> .
Survey variables used	<code>v500r</code>
Analytic variables used	<code>cage_months_int</code>
Analytic variables created	<code>cage_mad</code>

## Calculations

**Step 1.** Create a variable that identifies the age of all children 6-23 months of age in the children's nutrition module by 6-month age categories (`cage_mad`).

```
Set          cage_mad=missing
Replace      cage_mad=1 if cage_months_int≥6 and
              cage_months_int <12
Replace      cage_mad=2 if cage_months_int≥12 and
              cage_months_int <18
Replace      cage_mad=3 if cage_months_int≥18 and
              cage_months_int <24
Replace      cage_mad=missing if v500r≠1
Label values 1 "6-11 months"
              2 "12-17 months"
              3 "18-23 months"
Label variable "Child's age category (6-11,12-17,18-23 months)"
```

### 4.2.8 Children 0-59 months age categories—0-23 and 24-59 months

This variable categorizes children 0-59 months of age into two age categories: 0-23 months of age and 24-59 months of age according to the `cage_months_int` variable created in Section 4.2.2 and information collected in the household roster. This variable is used as a disaggregate for the children's health and nutrition program participation results, so it is required for only children who completed Module 5, *Children's nutrition*.

## Definitions

Treatment of missing data	Not applicable. All children who completed Module 5 will have a value for <code>cage_months_int</code> .
Survey variables used	<code>v104</code> , <code>v500r</code>
Analytic variables used	<code>cage_months_int</code>
Analytic variables created	<code>cage_pp</code>

## Calculations

**Step 1.** Create a variable that categorizes all children 0-59 months of age in the children's nutrition module into two age categories: 0-23 months of age and 24-59 months of age (`cage_pp`).

```
Set          cage_pp=missing
Replace      cage_pp=1 if cage_months_int≥0 and
              cage_months_int<24
```

```

Replace      cage_pp=2 if cage_months_int≥24 and
              cage_months_int<60
Replace      cage_pp=1 if v104<2 and cage_pp=missing and
              v500r=1
Replace      cage_pp=2 if v104≥2 and v104<5 and
              cage_pp=missing and v500r=1
Replace      cage_pp=missing if cage_months_int≥60 or v500r≠1
Label values 1 "0-23 months"
              2 "24-59 months"
Label variable "Child's age category (0-23,24-59 months)"

```

## 4.3 Household member age-in-year variables

### 4.3.1 Age in years

This variable identifies the age in years of all individuals included in the household roster. Individuals reported to be older than 95 years are listed as 95 years of age in the household roster. If there is a self-reported age for an individual in Module 4, *Women's nutrition*, or in Module 6, *Women's empowerment in agriculture*, the self-reported age will be used instead of the age in the household roster. If an individual responded to both Module 4 and Module 6, and the self-reported ages in the two modules do not match, check the household roster and use the age that matches the age in the household roster. If none of the ages match for an individual, investigate further to determine which age should be used. If a child has a value for `cage_months_int`, that age will be used instead of the household roster age.

#### Definitions

Treatment of missing data	Not applicable. The age survey variable ( <code>v104</code> ) in the household roster cannot contain missing values.
Survey variables used	<code>v104</code> , <code>v402</code> , <code>v6102</code>
Analytic variables used	<code>cage_months_int</code>
Analytic variables created	<code>age</code>

#### Calculations

**Step 1.** Create a continuous age-in-years analytic variable (`age`).

```

Set          age=v104
Replace      age=v402 if v402≠missing
Replace      age=v6102 if v6102≠missing
Replace      age=trunc(cage_months_int/12) if
              cage_months_int≠missing
Label variable "Age of HH member (years)"

```

**Step 2.** Explore whether there are any cases in which an individual responded to both Module 4 and Module 6, and their self-reported ages in the two modules (`v402` and `v6102`) do not agree. If their age in the household roster matches either `v402` or `v6102`, keep the age that matches their age in the household roster.

Count	if v402≠v6102 and v402≠missing and v6102≠missing
List	v104 v402 v6102 if v402≠v6102 and v402≠missing and v6102≠missing
Replace	age=v402 if v402≠v6102 and v402≠missing and v6102≠missing and v104=v402
Replace	age=v6102 if v402≠v6102 and v402≠missing and v6102≠missing and v104=v6102

### 4.3.2 Adults (18 years of age or older), overall and by sex

These variables identify all adults—that is, individuals who are 18 years of age or older according to the **age** variable created in Section 4.3.1—overall and by sex.

#### Definitions

Treatment of missing data	Not applicable. The variables <b>age</b> and <b>sex</b> cannot contain missing values for adults.
Survey variables used	n/a
Analytic variables used	<b>age</b> , <b>sex</b>
Analytic variables created	<b>adult</b> , <b>adult_m</b> , <b>adult_f</b>

#### Calculations

**Step 1.** Create a binary variable that identifies all adults in the household roster (**adult**).

Set	<b>adult</b> =0
Replace	<b>adult</b> =1 if <b>age</b> ≥18
Label values	0 "No" 1 "Yes"
Label variable	"Adult (18+ years), HH roster"

**Step 2.** Create a binary variable that identifies all adult males in the household roster (**adult\_m**).

Set	<b>adult_m</b> =0
Replace	<b>adult_m</b> =1 if <b>adult</b> =1 and <b>sex</b> =1
Label values	0 "No" 1 "Yes"
Label variable	"Male adult (18+ years), HH roster"

**Step 3.** Create a binary variable that identifies all adult females in the household roster (**adult\_f**).

Set	<b>adult_f</b> =0
Replace	<b>adult_f</b> =1 if <b>adult</b> =1 and <b>sex</b> =2
Label values	0 "No" 1 "Yes"
Label variable	"Female adult (18+ years), HH roster"

### 4.3.3 Youth (15-29 years of age), overall and by sex

These variables identify all youth—that is, individuals who are 15-29 years of age according to the **age** variable created in Section 4.3.1—overall and by sex.

#### Definitions

Treatment of missing data	Not applicable. The variables <b>age</b> and <b>sex</b> cannot contain missing values for youth.
Survey variables used	n/a
Analytic variables used	<b>age, sex</b>
Analytic variables created	<b>age15_29y, age15_29ym, age15_29yf</b>

#### Calculations

**Step 1.** Create a binary variable that identifies all youth in the household roster (**age15\_29y**).

```
Set          age15_29y=0
Replace      age15_29y=1 if age≥15 and age≤29
Label values 0 "No"
              1 "Yes"
Label variable "Youth (15-29 years), HH roster"
```

**Step 2.** Create a binary variable that identifies all male youth in the household roster (**age15\_29ym**).

```
Set          age15_29ym=0
Set          age15_29ym=1 if age15_29y=1 and sex=1
Label values 0 "No"
              1 "Yes"
Label variable "Male youth (15-29 years), HH roster"
```

**Step 3.** Create a binary variable that identifies all female youth in the household roster (**age15\_29yf**).

```
Set          age15_29yf=0
Set          age15_29yf=1 if age15_29y=1 and sex=2
Label values 0 "No"
              1 "Yes"
Label variable "Female youth (15-29 years), HH roster"
```

### 4.3.4 Children 5-17 years of age in household roster, overall and by sex

These variables identify all children 5 years of age or older—that is, between 5 and 17 years of age—according to the **age** variable created in Section 4.3.1—overall and by sex.

#### Definitions

Treatment of missing data	Not applicable. The variables <b>age</b> and <b>sex</b> cannot contain missing values for children 5-17 years of age.
Survey variables used	n/a

Analytic variables used	<b>sex, age</b>
Analytic variables created	<b>c5_17y, c5_17ym, c5_17yf</b>

## Calculations

**Step 1.** Create a binary variable that identifies all children between 5 and 17 years of age in the household roster (**c5\_17y**).

```
Set          c5_17y=0
Replace      Set c5_17y=1 if age≥5 and age≤17
Label values 0 "No"
              1 "Yes"
Label variable "Child 5-17 years, HH roster"
```

**Step 2.** Create a binary variable that identifies all male children between 5 and 17 years of age in the household roster (**c5\_17ym**).

```
Set          c5_17ym=0
Replace      c5_17ym=1 if c5_17y=1 and sex=1
Label values 0 "No"
              1 "Yes"
Label variable "Male child 5-17 years, HH roster"
```

**Step 3.** Create a binary variable that identifies all female children between 5 and 17 years of age in the household roster (**c5\_17yf**).

```
Set          c5_17yf=0
Replace      c5_17yf=1 if c5_17y=1 and sex=2
Label values 0 "No"
              1 "Yes"
Label variable "Female child 5-17 years, HH roster"
```

## 4.4 Women of reproductive age variables

This section describes an analytic variable that identifies women of reproductive age, as well as two categorical age variables that are used as disaggregates for the women's dietary intake indicator—the percent of women of reproductive age consuming a diet of minimum diversity.

### 4.4.1 Women of reproductive age (15-49 years of age)

This variable identifies all women of reproductive age—that is, women who are between 15 and 49 years of age—according to the **age** variable created in Section 4.3.1.

## Definitions

Treatment of missing data	Not applicable. The variables <b>age</b> and <b>sex</b> cannot contain missing values for women of reproductive age.
Survey variables used	n/a
Analytic variables used	<b>sex, age</b>
Analytic variables created	<b>wra</b>

## Calculations

**Step 1.** Create a binary variable that identifies all women of reproductive age in the household roster (**wra**).

```
Set          wra=0
Replace      wra=1 if age≥15 and age≤49 and sex=2
Label values 0 "No"
              1 "Yes"
Label variable "Woman of reproductive age (15-49 years)"
```

### 4.4.2 Women of reproductive age's age category—5-year age categories

This variable identifies the age of all women of reproductive age (15-49 years of age) interviewed for the women's nutrition module by 5-year age categories. This variable is used as a disaggregate for the percentage of women of reproductive age achieving dietary diversity indicator, so it is required for only women of reproductive age who completed Module 4, *Women's nutrition*; see Sections 16.1.3 and 16.2.3.

## Definitions

Treatment of missing data	Not applicable. The variables <b>age</b> and <b>wra</b> cannot contain missing values.
Survey variables used	<b>v400r</b>
Analytic variables used	<b>age, wra</b>
Analytic variables created	<b>agegrp_wra</b>

## Calculations

**Step 1.** Create a categorical variable that indicates the age of women of reproductive age by 5-year age categories (**agegrp\_wra**) using the **age**, **wra**, and **v400r** variables.

```
Set          agegrp_wra=1 if age≥15 and age≤19 and wra=1
Replace      agegrp_wra=2 if age≥20 and age≤24 and wra=1
Replace      agegrp_wra=3 if age≥25 and age≤29 and wra=1
Replace      agegrp_wra=4 if age≥30 and age≤34 and wra=1
Replace      agegrp_wra=5 if age≥35 and age≤39 and wra=1
Replace      agegrp_wra=6 if age≥40 and age≤44 and wra=1
Replace      agegrp_wra=7 if age≥45 and age≤49 and wra=1
Replace      agegrp_wra=missing if v400r≠1
Label values 1 "15-19"
              2 "20-24"
              3 "25-29"
              4 "30-34"
              5 "35-39"
              6 "40-44"
              7 "45-49"
Label variable "Woman of reproductive age's age (5-year categories)"
```



#### 4.4.3 Women of reproductive age's age categories—15-19 years and 20-49 years

This variable categorizes all women of reproductive age (15-49 years of age) interviewed for the women's nutrition module into two age categories: 15-19 years of age and 20-49 years of age. This variable is used as a disaggregate for the percentage of women of reproductive age achieving dietary diversity indicator, so it is required for only women of reproductive age who completed Module 4, *Women's nutrition*; see Sections 16.1.3 and 16.2.3.

##### Definitions

Treatment of missing data	Not applicable. The analytic variables <b>age</b> and <b>wra</b> cannot contain missing values.
Survey variables used	<b>400r</b>
Analytic variables used	<b>age, wra</b>
Analytic variables created	<b>wra_cage</b>

##### Calculations

**Step 1.** Create a categorical variable that categorizes women of reproductive age into two age categories: 15-19 years of age and 20-49 years of age (**wra\_cage**).

```
Set          wra_cage=1 if age≥15 and age≤19 and wra=1
Replace      wra_cage=2 if age≥20 and age≤49 and wra=1
Replace      wra_cage=missing if r400r≠1
Label values 1 "15-19 years"
              2 "20-49 years"
Label variable "Woman of reproductive age's age (15-19,20-49
              years)"
```

#### 4.5 Primary adult decision-maker variables

This section describes analytic variables that identify primary adult male and female decision-makers—overall and de jure household members only, as well as the following additional variables for primary adult female decision-makers: age; age category; whether the woman is 18-29 years of age (youth); marital status; and whether the woman participated in agricultural, non-agricultural, and wage or salary economic activities.

##### 4.5.1 Primary adult decision-makers, by sex

These variables identify primary adult decision-makers—that is, individuals who are 18 years of age or older according to the household roster who make the more important decisions in a household—in total and by sex.

##### Definitions

Treatment of missing data	Not applicable. Missing information on the presence of male and female primary adult decision-makers is not allowed in the household roster.
Survey variables used	<b>m1_line</b>

Analytic variables used	n/a
Analytic variables created	mdm, fdm, pdm

## Calculations

**Step 1.** Create a variable that indicates the primary adult male decision-maker in the household (**mdm**).

```
Set          mdm=0
Replace      mdm=1 if m1_line=1
Label values 0 "No"
              1 "Yes"
Label variable "Primary adult male decision-maker (Male PADM)"
```

**Step 2.** Create a variable that indicates the primary adult female decision-maker in the household (**fdm**).

```
Set          fdm=0
Replace      fdm=1 if m1_line=2
Label values 0 "No"
              1 "Yes"
Label variable "Primary adult female decision-maker (Female PADM)"
```

**Step 3.** Create a variable that indicates whether the household member is a primary adult male decision-maker or a primary adult female decision-maker (**pdm**).

```
Set          pdm=0
Replace      pdm=1 if mdm=1 or fdm=1
Label values 0 "No"
              1 "Yes"
Label variable "Primary adult decision-maker (PADM)"
```

### 4.5.2 De jure household member, primary adult decision-makers, by sex

These variables identify primary adult decision-makers who are de jure household members by sex.

## Definitions

Treatment of missing data	Not applicable. Missing information on usual household members and the presence of primary adult male and female decision-makers is not allowed in the household roster.
Survey variables used	n/a
Analytic variables used	mdm, fdm, hhmem_dj
Analytic variables created	fdm_dj, mdm_dj

## Calculations

**Step 1.** Create a variable that indicates whether the primary adult male decision-maker is a de jure household member (**mdm\_dj**).

```
Set          mdm_dj=missing
```

```

Replace      mdm_dj=0 if mdm=1
Replace      mdm_dj=1 if mdm=1 and hhmem_dj=1
Label values 0 "No"
              1 "Yes"
Label variable "Male PADM, de jure HH member"

```

**Step 2.** Create a variable that indicates whether the primary adult female decision-maker is a de jure household member (`fdm_dj`).

```

Set          fdm_dj=missing
Replace      fdm_dj=0 if fdm=1
Replace      fdm_dj=1 if fdm=1 and hhmem_dj=1
Label values 0 "No"
              1 "Yes"
Label variable "Female PADM, de jure HH member"

```

### 4.5.3 Age category, primary adult female decision-makers (de jure only)

This variable identifies primary adult female decision-makers who are de jure household members by age category, according to the `age` variable created in Section 4.3.1. Primary adult female decision-makers 25-59 years of age are assigned to 5-year age categories; those under 25 years of age are assigned to an 18-24 years of age category, and those 60 years of age or older are assigned to a 60+ years of age category. This variable is required for only women who completed Module 6, *Women's empowerment in agriculture*.

#### Definitions

Treatment of missing data	Not applicable. The variables <code>age</code> and <code>fdm_dj</code> cannot contain missing values for primary adult female decision-makers.
Survey variables used	<code>v600r</code>
Analytic variables used	<code>age</code> , <code>fdm_dj</code>
Analytic variables created	<code>agegrp_fdm_dj</code>

#### Calculations

**Step 1.** Create a variable that indicates primary adult female decision-makers' age by age category (`agegrp_fdm_dj`).

```

Set          agegrp_fdm_dj=1 if age≥18 and age≤24 and fdm_dj=1
Replace      agegrp_fdm_dj=2 if age≥25 and age≤29 and fdm_dj=1
Replace      agegrp_fdm_dj=3 if age≥30 and age≤34 and fdm_dj=1
Replace      agegrp_fdm_dj=4 if age≥35 and age≤39 and fdm_dj=1
Replace      agegrp_fdm_dj=5 if age≥40 and age≤44 and fdm_dj=1
Replace      agegrp_fdm_dj=6 if age≥45 and age≤49 and fdm_dj=1
Replace      agegrp_fdm_dj=7 if age≥50 and age≤54 and fdm_dj=1
Replace      agegrp_fdm_dj=8 if age≥55 and age≤59 and fdm_dj=1
Replace      agegrp_fdm_dj=9 if age≥60 and age≤95 and fdm_dj=1
Replace      agegrp_fdm_dj=missing if v600r≠1
Label values 1 "18-24"
              2 "25-29"

```

3 "30-34"  
 4 "35-39"  
 5 "40-44"  
 6 "45-49"  
 7 "50-54"  
 8 "55-59"  
 9 "60+"

Label variable "Age category of female PADM, de jure HH member"

#### 4.5.4 Youth, primary adult female decision-makers (de jure only)

This variable identifies primary adult female decision-makers who are de jure household members by whether they are youth (18-29 years of age) or non-youth (30 years of age or older) according to the **age** variable created in Section 4.3.1. This variable is required for only women who completed Module 6, *Women's empowerment in agriculture*.

##### Definitions

Treatment of missing data	Not applicable. The variables <b>age</b> and <b>fdm_dj</b> cannot contain missing values for primary adult female decision-makers.
Survey variables used	<b>v600r</b>
Analytic variables used	<b>fdm_dj, age</b>
Analytic variables created	<b>youth_fdm_dj</b>

##### Calculations

**Step 1.** Create a variable that indicates whether de jure primary adult female decision-makers are youth (**youth\_fdm\_dj**).

```
Set          youth_fdm_dj=missing
Replace     youth_fdm_dj=0 if fdm_dj=1
Replace     youth_fdm_dj=1 if age≥18 and age<30 and fdm_dj=1
Replace     youth_fdm_dj=missing if v600r≠1
Label values 0 "No"
              1 "Yes"
Label variable "Female PADM is 18-29 years (youth), de jure HH member"
```

#### 4.5.5 Marital status, primary adult female decision-makers (de jure only)

This variable identifies the marital status of primary adult female decision-makers who are de jure household members according to information in Module 6, *Women's empowerment in agriculture*. This variable is required for only women who completed Module 6.

##### Definitions

Treatment of missing data	This variable will not have a value for primary adult female decision-makers who completed Module 6 but did not answer the marital status questions.
Survey variables used	<b>v6105, v6106, v6107</b>
Analytic variables used	<b>fdm_dj</b>
Analytic variables created	<b>marstat_fdm_dj</b>

## Calculations

**Step 1.** Create a variable that indicates the marital status of primary adult female decision-makers (`marstat_fdm_dj`).

```
Set          marstat_fdm_dj=1 if fdm_dj=1 and v6105=1
Replace      marstat_fdm_dj=2 if fdm_dj=1 and v6105=2
Replace      marstat_fdm_dj=3 if fdm_dj=1 and v6105=3 and
              v6107=1
Replace      marstat_fdm_dj=4 if fdm_dj=1 and v6105=3 and
              (v6107=2 or v6107=3)
Replace      marstat_fdm_dj=5 if fdm_dj=1 and v6105=3 and
              v6106=3
Label values 1 "Married"
              2 "Living together"
              3 "Widowed"
              4 "Divorced or separated"
              5 "Never married or in union"
Label variable "Female PADM marital status, de jure HH member"
```

### 4.5.6 Participate in agricultural production, non-farm work, and wage or salary work, primary adult female decision-makers (de jure only)

These variables identify primary adult female decision-makers who are de jure household members by whether they participated in agricultural production (i.e., food crop farming, cash crop farming, livestock raising, or fishpond aquaculture), non-farm work (e.g., running a small business, self-employment), and wage or salaried work, which could be for cash or in-kind payment, according to information in Module 6, *Women's empowerment in agriculture*. These variables are required for only women who completed Module 6.

## Definitions

Treatment of missing data	These variables will not have a value for primary adult female decision-makers who completed Module 6 but did not answer the economic activity participation questions.
Survey variables used	<code>v6201_1</code> , <code>v6201_2</code> , <code>v6201_3</code> , <code>v6201_4</code> , <code>v6201_5</code> , <code>v6201_6</code> , <code>v6605</code>
Analytic variables used	<code>fdm_dj</code>
Temporary analytic variables created	<code>v6201_1x</code> , <code>v6201_2x</code> , <code>v6201_3x</code> , <code>v6201_4x</code> , <code>v6201_5x</code> , <code>v6201_6x</code> , <code>fdm_econ_miss</code>
Analytic variables created	<code>fdm_econ_farm</code> , <code>fdm_econ_nonfarm</code> , <code>fdm_econ_wage</code> , <code>fdm_econ_any</code>

## Calculations

**Step 1.** Create temporary new variables for the Module 6 economic activity participation variables (`v6201_1`–`v6201_6`) that recode no's (2) to 0 and inconsistent (7) and missing (9) values to missing (.) (`v6201_1x`–`v6201_6x`).

```
For variables v6201_1 v6201_2 v6201_3 v6201_4 v6201_5 v6201_6
Set          `var'x=1 if `var'=1
```

```

Replace    `var'x=0 if `var'=2
Replace    `var'x=missing if `var'=7 or `var'=9 or
           `var'=missing

```

**Step 2.** Create a variable that counts the number variables created in the previous step that are missing a value (0-6) and set the variable to missing if Module 6 was not completed (`fdm_econ_miss`).

```

Set          fdm_econ_miss=0
Replace      fdm_econ_miss=fdm_econ_miss+1 if v6201_1x=missing
Replace      fdm_econ_miss=fdm_econ_miss+1 if v6201_2x=missing
Replace      fdm_econ_miss=fdm_econ_miss+1 if v6201_3x=missing
Replace      fdm_econ_miss=fdm_econ_miss+1 if v6201_4x=missing
Replace      fdm_econ_miss=fdm_econ_miss+1 if v6201_5x=missing
Replace      fdm_econ_miss=fdm_econ_miss+1 if v6201_6x=missing
Replace      fdm_econ_miss=missing if v600r#1 or fdm_dj#1
Label variable "Number of activities female PADM is missing, de
               jure HH member"

```

**Step 3.** Create a variable to flag whether the primary adult female decision-maker participated in agricultural production (i.e., food crop farming, cash crop farming, livestock raising, or fishpond culture/fishing) (`fdm_econ_farm`).

```

Set          fdm_econ_farm=0 if v600r=1
Replace      fdm_econ_farm=1 if v6201_1=1 or v6201_2=1 or
               v6201_3=1 or v6201_4=1
Replace      fdm_econ_farm=missing if fdm_econ_miss =6 or
               fdm_dj#1
Label values  0 "No"
               1 "Yes"
Label variable "Female PADM partook in farm work, de jure HH
               member"

```

**Step 4.** Create a variable to flag whether the primary adult female decision-maker participated in non-farm work (`fdm_econ_nonfarm`)

```

Set          fdm_econ_nonfarm=0 if v600r=1
Replace      fdm_econ_nonfarm=1 if v6201_5=1
Replace      fdm_econ_nonfarm=missing if fdm_econ_miss =6 or
               fdm_dj#1
Label values  0 "No"
               1 "Yes"
Label variable "Female PADM partook in non-farm work, de jure HH
               member"

```

**Step 5.** Create a variable to flag whether the primary adult female decision-maker participated in wage or salary work (`fdm_econ_wage`).

```

Set          fdm_econ_wage=0 if v600r=1
Replace      fdm_econ_wage=1 if v6201_6=1

```

```

Replace      fdm_econ_wage=missing if fdm_econ_miss =6 or
              fdm_dj≠1
Label values  0 "No"
              1 "Yes"
Label variable "Female PADM partook in wage/salary work, de jure
              HH member"

```

**Step 6.** Create a variable to flag whether the primary adult female decision-maker participated in any economic work (**fdm\_econ\_any**) and delete all temporary analytical variables created.

```

Set          fdm_econ_any=0 if v600r=1
Replace      fdm_econ_any=1 if fdm_econ_farm=1 or
              fdm_econ_nonfarm=1 or fdm_econ_wage=1
Replace      fdm_econ_any=missing if fdm_econ_miss =6 or
fdm_dj≠1
Label values  0 "No"
              1 "Yes"
Label variable "Female PADM partook in any economic work, de
              jure HH member"

Delete       v6201_1x v6201_2x v6201_3x v6201_4x v6201_5x
              v6201_6x fdm_econ_miss

```

## 4.6 Producer variables

This section describes analytic variables that identify producers of any targeted value chain commodity (VCC) included in the survey, producers of each targeted VCC individually, and producers of each targeted VCC who completed the relevant VCC module.

### 4.6.1 Targeted VCC producer, any VCC

This variable identifies individuals who were primarily responsible for cultivating or raising at least one targeted VCC included in the ZOI Survey during the 12 months preceding the survey, according to information collected in the targeted VCC modules. The template syntax includes maize, dairy cows, sheep, and fishponds as the targeted VCCs. Be sure to adapt the syntax to reflect the VCCs included in the ZOI Survey being analyzed.

#### Definitions

Treatment of missing data	Not applicable. The outcome fields in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module.
Survey variables used*	v7100c, v7100r, v75000c, v75000r, v75200c, v75200r, v78000c, v78000r
Analytic variables used	n/a
Analytic variables created	vcc

\* Depends on the targeted VCCs included in the ZOI Survey

## Calculations

**Step 1.** Examine the consent and outcome variables for each VCC module to ensure that they align.

Tabulate	v7100d	v7100r	//Maize	(Module 7.1)
Tabulate	v75000d	v75000r	//Dairy	(Module 7.50)
Tabulate	v75200d	v75200r	//Sheep	(Module 7.52)
Tabulate	v78000d	v78000r	//Fishpond	(Module 7.80)

**Step 2.** Create a binary variable to flag household members who were responsible for cultivating or raising at least one targeted VCC in the 12 months preceding the survey (**vcc**).

Set	vcc=0
Replace	vcc=1 if v7100r <del>missing</del> or v75000r <del>missing</del> or v75200r <del>missing</del> or v78000r <del>missing</del>
Label values	0 "No" 1 "Yes"
Label variable	"Producer of 1+ targeted VCC, past 12 months"

### 4.6.2 Targeted VCC producer, specific VCCs

These variables identify individuals who were primarily responsible for cultivating or raising specific targeted VCCs included in the ZOI Survey during the 12 months preceding the survey, according to information collected in the targeted VCC modules. The template syntax includes maize, dairy cows, sheep, and fishponds as the targeted VCCs. Be sure to adapt the syntax to reflect the VCCs included in the ZOI Survey being analyzed.

## Definitions

Treatment of missing data	Not applicable. The outcome field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module.
Survey variables used*	v7100r, v75000r, v75200r, v78000r
Analytic variables used	n/a
Analytic variables created*	vcc_maize2, vcc_dairy2, vcc_sheep2, vcc_fish2

\* Depends on the targeted VCCs included in the ZOI Survey

## Calculations

**Step 1.** Create a variable to indicate if the household member was responsible for cultivating maize (**vcc\_maize2**).

Set	vcc_maize2=0
Replace	vcc_maize2=1 if v7100r <del>missing</del>
Label values	0 "No" 1 "Yes"
Label variable	"Maize producer, past 12 months"

**Step 2.** Create a variable to indicate if the household member was responsible for raising dairy cows (**vcc\_dairy2**).



```

Set          vcc_dairy2=0
Replace      vcc_dairy2=1 if v75000r#missing
Label values 0 "No"
              1 "Yes"
Label variable "Dairy cow producer, past 12 months"

```

**Step 3.** Create a variable to indicate if the household member was responsible for raising sheep (vcc\_sheep2).

```

Set          vcc_sheep2=0
Replace      vcc_sheep2=1 if v75200r#missing
Label values 0 "No"
              1 "Yes"
Label variable "Sheep producer, past 12 months"

```

**Step 4.** Create a variable to indicate if the household member was responsible for cultivating fish (vcc\_fish2).

```

Set          vcc_fish2=0
Replace      vcc_fish2=1 if v78000r#missing
Label values 0 "No"
              1 "Yes"
Label variable "Fishpond producer, past 12 months"

```

#### 4.6.3 Targeted VCC producer who completed relevant agriculture module

These variables identify individuals in the household roster who were primarily responsible for cultivating or raising each targeted VCC included in the ZOI Survey during the 12 months preceding the survey and who completed the relevant VCC module. The template syntax includes maize, dairy cows, sheep, and fishponds as the targeted VCCs. Be sure to adapt the syntax to include the VCCs included in the ZOI Survey being analyzed.

#### Definitions

Treatment of missing data	Not applicable. The outcome field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module.
Survey variables used*	v7100r, v75000r, v75200r, v78000r
Analytic variables used	n/a
Analytic variables created*	vcc_maize, vcc_dairy, vcc_sheep, vcc_fish

\* Depends on the targeted VCCs included in the ZOI Survey

#### Calculations

**Step 1.** Create a binary variable to indicate whether individuals were responsible for cultivating maize in the 12 months preceding the ZOI Survey and completed the maize module (vcc\_maize).

```

Set          vcc_maize=0
Replace      vcc_maize=1 if v7100r=1
Label values 0 "No"
              1 "Yes"
Label variable "Maize producer, completed maize module"

```

**Step 2.** Create a binary variable to indicate whether individuals were responsible for raising dairy cows in the 12 months preceding the ZOI Survey and completed the dairy cow module (**vcc\_dairy**).

```
Set          vcc_dairy=0
Replace      vcc_dairy=1 if v75000r=1
Label values 0 "No"
              1 "Yes"
Label variable "Dairy cow producer, completed dairy cow module"
```

**Step 3.** Create a binary variable to indicate whether individuals were responsible for raising sheep in the 12 months preceding the ZOI Survey and completed the sheep module (**vcc\_sheep**).

```
Set          vcc_sheep=0
Replace      vcc_sheep=1 if v75200r=1
Label values 0 "No"
              1 "Yes"
Label variable "Sheep producer, completed sheep module"
```

**Step 4.** Create a binary variable to indicate whether individuals were responsible for cultivating fish in fishponds in the 12 months preceding the ZOI Survey and completed the fishpond module (**vcc\_fish**).

```
Set          vcc_fish=0
Replace      vcc_fish=1 if v78000r=1
Label values 0 "No"
              1 "Yes"
Label variable "Fishpond producer, completed fishpond module"
```

## 5. KEY HOUSEHOLD-LEVEL VARIABLES

In this chapter, key household-level analytic variables used to calculate or disaggregate indicators are defined and their calculation described. The chapter is divided into two sections; Section 5.1 describes analytic variables related to household composition, and Section 5.2 describes analytic variables that are used as indicator disaggregates.

### 5.1 Household composition

This section describes household-level analytic variables that capture household composition. With the exception of the last variable in this section, the variables created in this section require variables created in Chapter 4 that have been added to the individual-level data file (FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic). After creating the analytic variables described in this section, they can be saved to an analytic version of the household data file (FTF ZOI Survey [COUNTRY] [YEAR] household data analytic), which can be used across data analysts so that they do not have to re-create these analytic variables as part of their analysis. The last variable in this section is created directly in the household-level data file using the variable created in Section 5.1.1. The step-by-step procedures to create the household composition variables included in this section follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax household analytic* do file.

#### 5.1.1 Number of usual (de jure) household members

This variable counts the number of de jure household members in each household. De jure household members include all individuals who usually live in the household. Household members can include servants, lodgers, and agricultural laborers, as well as family members—as long as they live under the same roof (or in the same compound), share cooking arrangements, and acknowledge the same primary decision-makers.<sup>38</sup> The variable counts all individuals listed in the household roster who are de jure household members.

#### Definitions

Treatment of missing data	Not applicable. The <code>hhmem_dj</code> analytic variable cannot have missing values.
Survey variables used	n/a
Analytic variables used	<code>hhmem_dj</code>
Analytic variables created	<code>hhsiz_dj</code>

#### Calculations

**Step 1.** Create a variable that counts the number of de jure household members in each household (`hhsiz_dj`) by summing the de jure household member variable created in Section 4.1.1.

```
For each      hhea hnum
Set           hhsiz_dj=sum(hhmem_dj)
Label variable "Number of de jure HH members"
```

<sup>38</sup> The definition of what constitutes a household may vary by country, especially if polygamous relationships exist.

### 5.1.2 Number of adults who are de jure household members in household, overall and by sex

These variables identify the number of adults who are de jure household members per household—in total and by sex, according to analytic variables created in Sections 4.1.1 and 4.3.2.

#### Definitions

Treatment of missing data	Not applicable. The analytic variables <code>adult</code> , <code>adult_m</code> , <code>adult_f</code> , and <code>hhmem_dj</code> cannot have missing values.
Survey variables used	n/a
Analytic variables used	<code>adult</code> , <code>adult_m</code> , <code>adult_f</code> , <code>hhmem_dj</code>
Analytic variables created	<code>nadult_dj</code> , <code>nadult_mdj</code> , <code>nadult_fdj</code>

#### Calculations

**Step 1.** Create a continuous variable that counts the total number of adults who are de jure household members in each household (`nadult_dj`) by summing the adult variable (`adult`) created in Section 4.3.2 for de jure household members by household.

```
For each      hhea hhnum
  Set         nadult_dj=sum(adult) if hhmem_dj=1
Label variable "Number of adults in HH (de jure only)"
```

**Step 2.** Create a continuous variable that counts the total number of adult males who are de jure household members in each household (`nadult_mdj`) by summing the adult variable (`adult_m`) created in Section 4.3.2 for de jure household members by household.

```
For each      hhea hhnum
  Set         nadult_mdj=sum(adult_m) if hhmem_dj=1
Label variable "Number of male adults in HH (de jure only)"
```

**Step 3.** Create a continuous variable that counts the total number of adult females who are de jure household members in each household (`nadult_f`) by summing the adult variable (`adult_f`) created in Section 4.3.2 for de jure household members by household.

```
For each      hhea hhnum
  Set         nadult_fdj=sum (adult_f) if hhmem_dj=1
Label variable "Number of female adults in HH (de jure only)"
```

### 5.1.3 Number of women of reproductive age who are de jure household members in household

This variable identifies the number of women of reproductive age who are de jure household members in each household according to analytic variables created in Sections 4.1.1 and 4.4.1.

#### Definitions

Treatment of missing data	Not applicable. The analytic variables <code>wra</code> and <code>hhmem_dj</code> cannot have missing values.
Survey variables used	n/a

Analytic variables used	wra, hhmem_dj
Analytic variables created	nwra_dj

## Calculations

**Step 1.** Create a continuous variable that counts the total number of women of reproductive age who are de jure household members in each household (nwra\_dj) by summing the woman of reproductive age variable (wra) created in Section 4.4.1 for de jure household members by household.

```
For each      hhea hhnum
Set          nwra_dj=sum(wra) if hhmem_dj=1
Label variable "Number of WRA 15-49 years in HH (de jure only)"
```

### 5.1.4 Number of children under 2 years of age who are de jure household members in household

This variable identifies the number of children under 2 years of age (0-23 months of age) who are de jure household members in each household according to analytic variables created in Sections 4.1.1 and 4.2.5.

## Definitions

Treatment of missing data	Not applicable. The analytic variables c0_23m and hhmem_dj cannot have missing values.
Survey variables used	n/a
Analytic variables used	c0_23m, hhmem_dj
Analytic variables created	ncu2_dj

## Calculations

**Step 1.** Create a continuous variable that counts the total number of children under 2 years of age who are de jure household members in each household (ncu2\_dj) by summing the children under 2 years of age variable (c0\_23m) created in Section 4.2.5 for de jure household members by household.

```
For each      hhea hhnum
Set          ncu2_dj=sum(c0_23m) if hhmem_dj=1
Label variable "Number of children <2 years in HH (de jure only)"
```

### 5.1.5 Number of children under 5 years of age who are de jure household members in household

This variable identifies the number of children under 5 years of age who are de jure household members in each household according to analytic variables created in Sections 4.1.1 and 4.2.8.

## Definitions

Treatment of missing data	Not applicable. The analytic variables c0_59m and hhmem_dj cannot have missing values.
Survey variables used	n/a

Analytic variables used	c0_59m, hhmem_dj
Analytic variables created	ncu5_dj

## Calculations

**Step 1.** Create a continuous variable that counts the total number of children under 5 years of age who are de jure household members in each household (ncu5\_dj) by summing the children under 5 years of age variable (c0\_59m) created in Section 4.2.8 for de jure household members by household.

```
For each      hhea hhnum
Set          ncu5_dj=sum(c0_59m) if hhmem_dj=1
Label variable "Number of children <5 years in HH (de jure
              only) "
```

### 5.1.6 Number of children 5-17 years of age who are de jure household members in household

This variable identifies the number of children who are between 5 and 17 years of age who are de jure household members in each household according to analytic variables created in Sections 4.1.1 and 4.3.4.

## Definitions

Treatment of missing data	Not applicable. The analytic variables c5_17y and hhmem_dj cannot have missing values.
Survey variables used	n/a
Analytic variables used	c5_17y, hhmem_dj
Analytic variables created	nc5_17y_dj

## Calculations

**Step 1.** Create a continuous variable that counts the total number of children 5-17 years of age, inclusive, who are de jure household members in each household (nc5\_17y\_dj) by summing the children 5-17 years of age variable (c5\_17y) created in Section 4.3.4 for de jure household members by household.

```
For each      hhea hhnum
Set          nc5_17y_dj=sum(c5_17y) if hhmem_dj=1
Label variable "Number of children 5-17 years in HH (de jure
              only) "
```

### 5.1.7 Number of youth 15-29 years of age who are de jure household members in household

This variable identifies the number of youth (15-29 years of age) who are de jure household members in each household according to analytic variables created in Sections 4.1.1 and 4.3.3.

## Definitions

Treatment of missing data	Not applicable. The analytic variables <b>age15_29y</b> and <b>hhmem_dj</b> cannot have missing values.
Survey variables used	n/a
Analytic variables used	<b>age15_29y</b> , <b>hhmem_dj</b>
Analytic variables created	<b>nage15_29y_dj</b>

## Calculations

**Step 1.** Create a continuous variable that counts the total number of youth who are de jure household members in each household (**nage15\_29y\_dj**) by summing the youth variable (**age15\_29y**) created in Section 4.3.3 for de jure household members by household.

```
For each      hhea hhnum
Set          nage15_29y_dj=sum(age15_29y) if hhmem_dj=1
Label variable "Number of youth 15-29 years in HH (de jure
              only)"
```

### 5.1.8 Number of producers of targeted value chain commodities who are de jure household members in household

This variable identifies the number of producers of targeted value chain commodities (VCCs) who are de jure household members in each household according to analytic variables created in Sections 4.1.1, 4.6.1, and 4.6.2. The template syntax includes maize, dairy cows, sheep, and fishponds as the targeted VCCs. Be sure to adapt the syntax to reflect the VCCs included in the ZOI Survey being analyzed.

## Definitions

Treatment of missing data	The analytic variables for VCC producers created in Section 4.5.1 and 4.5.2 and <b>hhmem_dj</b> cannot have missing values.
Survey variables used	n/a
Analytic variables used*	<b>vcc</b> , <b>vcc_maize2</b> , <b>vcc_dairy2</b> , <b>vcc_sheep2</b> , <b>vcc_fish2</b> , <b>hhmem_dj</b>
Analytic variables created*	<b>nvcc_dj</b> , <b>nvcc_maize_dj</b> , <b>nvcc_dairy_dj</b> , <b>nvcc_sheep_dj</b> , <b>nvcc_fish_dj</b>

\* Depends on the targeted VCCs included in the ZOI Survey

## Calculations

**Step 1.** Create a continuous variable that counts the total number of producers of any targeted VCC included in the ZOI Survey who are de jure household members (**nvcc\_dj**) by summing the producer variable (**vcc**) created in Section 4.6.1 for de jure household members by household.

```
For each      hhea hhnum
Set          nvcc_dj=sum(vcc) if hhmem_dj=1
Label variable "Number of producers of 1+ targeted VCC in HH (de
              jure only)"
```

**Step 2.** Create a continuous variable that counts the total number of maize producers who are de jure household members (**nvcc\_maize\_dj**) by summing the maize producer variable (**vcc\_maize2**) created in Section 4.6.2 for de jure household members by household.

```
For each      hhea hhnum
  Set         nvcc_maize_dj=sum(vcc_maize2) if hhmem_dj=1
Label variable "Number of maize producers in HH (de jure only)"
```

**Step 3.** Create a continuous variable that counts the total number of dairy cow producers who are de jure household members (**nvcc\_dairy\_dj**) by summing the dairy cow producer variable (**vcc\_dairy2**) created in Section 4.6.2 for de jure household members by household.

```
For each      hhea hhnum
  Set         nvcc_dairy_dj=sum(vcc_dairy2) if hhmem_dj=1
Label variable "Number of dairy cow producers in HH (de jure only)"
```

**Step 4.** Create a continuous variable that counts the total number of sheep producers who are de jure household members (**nvcc\_sheep\_dj**) by summing the sheep producer variable (**vcc\_sheep2**) created in Section 4.6.2 for de jure household members by household.

```
For each      hhea hhnum
  Set         nvcc_sheep_dj=sum(vcc_sheep2) if hhmem_dj=1
Label variable "Number of sheep producers in HH (de jure only)"
```

**Step 5.** Create a continuous variable that counts the total number of fishpond producers who are de jure household members (**nvcc\_fish\_dj**) by summing the fishpond producer variable (**vcc\_fish2**) created in Section 4.6.2 for de jure household members by household.

```
For each      hhea hhnum
  Set         nvcc_fish_dj=sum(vcc_fish2) if hhmem_dj=1
Label variable "Number of fishpond producers in HH (de jure only)"
```

### 5.1.9 Household cultivates or raises targeted VCCs

These variables identify households that cultivated or raised specific targeted VCCs included in the ZOI Survey during the 12 months preceding the survey, according to information collected in the targeted VCC modules. The template syntax includes maize, dairy cows, sheep, and fishponds as the targeted VCCs. Be sure to adapt the syntax to reflect the VCCs included in the ZOI Survey being analyzed.

#### Definitions

Treatment of missing data	Not applicable. The outcome field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module.
Survey variables used*	n/a
Analytic variables used	<b>vcc_maize2, vcc_dairy2, vcc_sheep2, vcc_fish2</b>
Analytic variables created*	<b>hh_maize, hh_dairy, hh_sheep, hh_fish</b>

\* Depends on the targeted VCCs included in the ZOI Survey



## Calculations

**Step 1.** Create a variable to indicate if the household member was responsible for cultivating maize (hh\_maize).

```
Set          hh_maize=0
Replace      hh_maize=1 if vcc_maize2>0 and vcc_maize2#missing
Label values 0 "No"
              1 "Yes"
Label variable "HH cultivated maize, past 12 months"
```

**Step 2.** Create a variable to indicate if the household member was responsible for raising dairy cows (hh\_dairy).

```
Set          hh_dairy=0
Replace      hh_dairy=1 if vcc_dairy2>0 and vcc_dairy2#missing
Label values 0 "No"
              1 "Yes"
Label variable "HH raised dairy cows, past 12 months"
```

**Step 3.** Create a variable to indicate if the household member was responsible for raising sheep (hh\_sheep).

```
Set          hh_sheep=0
Replace      hh_sheep=1 if vcc_sheep2>0 and vcc_sheep2#missing
Label values 0 "No"
              1 "Yes"
Label variable "HH raised sheep, past 12 months"
```

**Step 4.** Create a variable to indicate if the household member was responsible for cultivating fish (hh\_fish).

```
Set          hh_fish=0
Replace      hh_fish=1 if vcc_fish2>0 and vcc_fish2#missing
Label values 0 "No"
              1 "Yes"
Label variable "Fishpond producer, past 12 months"
```

### 5.1.10 Household size category—de jure household members

This variable identifies the household size category (small, medium, or large) into which the household falls, according to the analytic variable created in Section 5.1.1.

## Definitions

Treatment of missing data	Not applicable. The analytic variable <code>hhsize_dj</code> cannot have missing values.
Survey variables used	n/a
Analytic variables used	<code>hhsize_dj</code>
Analytic variables created	<code>hhsizegrp_dj</code>

## Calculations

**Step 1.** Create a categorical variable for household size (**hhsizegrp\_dj**) based on de jure household members.

```
Set          hhsizegrp_dj=missing
Replace      hhsizegrp_dj=1 if hhsiz_dj≥1 and hhsiz_dj≤5
Replace      hhsizegrp_dj=2 if hhsiz_dj≥6 and hhsiz_dj≤10
Replace      hhsizegrp_dj=3 if hhsiz_dj≥11 and
              hhsiz_dj≠missing
Label values 1 "1-5 de jure members"
              2 "6-10 de jure members"
              3 "11+ de jure members"
Label variable "Household size, categorical (de jure only)"
```

## 5.2 Household-level disaggregates

This section describes household-level disaggregates that are used to report Feed the Future indicators in the midline indicator assessment report. Each disaggregate variable is created in a different template syntax file (see each subsection), but after they are created, they are all added to both the household-level analytic data file (FTF ZOI Survey [COUNTRY] [YEAR] household data analytic) and the individual-level analytic data file (FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic), so they can be accessed by all data analysts.

### 5.2.1. Gendered household type—de jure household members

This variable identifies each household's gendered household type—that is, adult male and female, adult female-only (no adult males), adult male-only (no adult females), or child-only (no adults)—based on the sex and age of de jure household members. This variable is created in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax household analytic* do file.

## Definitions

Treatment of missing data	Not applicable. The <b>nadult_mdj</b> and <b>nadult_fdj</b> analytic variables cannot be missing values.
Survey variables used	n/a
Analytic variables used	<b>nadult_mdj</b> , <b>nadult_fdj</b>
Analytic variables created	<b>genhhtype_dj</b>

## Calculations

**Step 1.** Create a categorical variable that identifies each household by the sex of de jure adult household members living in the household, or as a household without de jure adult household members (**genhhtype\_dj**).

```
Set          genhhtype_dj=0
Replace      genhhtype_dj=1 if nadult_fdj≥1 and nadult_mdj≥1
Replace      genhhtype_dj=2 if nadult_fdj≥1 and nadult_mdj=0
Replace      genhhtype_dj=3 if nadult_mdj≥1 and nadult_fdj=0
```

```

Replace      genhhtype_dj=4 if nadult_fdj=0 and nadult_mdj=0
Replace      genhhtype_dj=missing if nadult_mdj=missing and
              nadult_fdj=missing
Label values  1 "Male and female adults"
              2 "Female adults only, no male adults"
              3 "Male adults only, no female adults"
              4 "Children only, no adults"
Label variable "Gendered household type (de jure HH members
              only)"

```

### 5.2.2. Wealth quintile

The wealth quintile variable (**awiquint**) is created as part of the comparative wealth index indicator. Please see Chapter 11 for step-by-step instructions on how to create the **awiquint** variable. This variable is created in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax wealthindex AWI* do file.

### 5.2.3. Poverty status

The poverty status variable (**poor190**) is created to calculate the prevalence of poverty indicator. Please see the survey-to-survey imputation step-by-step instructions on how to create the **poor190** variable for Midline Surveys.<sup>39</sup>

### 5.2.4. Shock exposure severity

The shock severity variable (**shock\_sev**) is a categorical variable that groups households into four categories based on the shock exposure index (SEI), or the self-perceived severity of shocks that the household faced during the year before the survey. The SEI is calculated as part of the Feed the Future ability to recover from shocks and stresses index indicator. Please see Section 12.2.1 for step-by-step instructions on how to create the SEI variable, which is the starting point to create the shock severity variable. This variable is created in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax resilience* do file.

## Definitions

Treatment of missing data	If a household has any missing or refused responses for the questions used to calculate the SEI, set the <b>shock_sev</b> variable to missing.
Survey variables used	n/a
Analytic variables used	<b>sei</b>
Analytic variables created	<b>shock_sev</b>

## Calculations

**Step 1.** Using the SEI variable (**sei**) created in Chapter 12, create a categorical variable with four categories to identify households by the severity of shocks experienced during the year prior to the survey. The first category captures households that did not experience any shocks, and the remaining

<sup>39</sup> Guidance, step-by-step instructions, and template syntax are forthcoming in an update to this edition of the *Guide to Feed the Future Statistics for P2-ZOI Midline Surveys*. The **poor190** variable created at baseline was created using data from Module 8, *Household consumption expenditure*, and the procedures are outlined in Section 10.2.

households are split into three categories with roughly equal numbers of households. Set the variable to missing for any households that are missing an SEI variable value.

```
Tabulate      sei if sei>0 and sei≠missing
Set           shock_sev=missing
Set           shock_sev=1 if sei≤X

              [where X is the SEI value at which the cumulative
              percentage of households is greater than or equal
              to 33.3%]

Set           shock_sev=2 if sei>X and sei≤Y

              [where Y is the SEI value at which the cumulative
              percentage of households is greater than or equal
              to 66.7%]

Set           shock_sev=3 if sei>Y and sei≠missing
Set           shock_sev=0 if sei=0
Label values  0 "No shocks"
              1 "Low"
              2 "Moderate"
              3 "Severe"
Label variable "Shock severity disaggregate"
```

## 6. ZOI POPULATION ESTIMATES

### 6.1 Background

In midline indicator assessment reports, population totals are presented for the ZOI overall as well as for selected sub-populations of interest—for example, women 15-49 years of age, primary adult female decision-makers 18 years of age or older, children 6-23 months of age, and individuals 15 years of age or older who are responsible for cultivating or raising specific value chains (see midline indicator assessment Table 1.2.1 and Table 1.2.2). In addition, the number of households in the ZOI—overall and by selected background characteristics (e.g., region, residence [urban/rural], and gendered household type)—are presented in midline indicator assessment Table 1.2.3.

There are two basic methods to estimating the overall ZOI population. One method is to use the survey sample with its sampling weights to produce an estimate of the overall ZOI population. This method, however, has a number of limitations. The sampling weights derived for a survey are a function of the sampling frame and can only scale to the population represented by the sampling frame in use. This method is, therefore, susceptible to the age of the sampling frame and any other deficiencies (e.g., under-coverage). The estimated population total determined using this first method will not account for any changes in population size not represented by the sampling frame.

In addition, the main purpose of the ZOI Survey is to produce estimates for key indicators within the ZOI with adequate reliability, either for comparative or descriptive analysis.<sup>40</sup> The surveys are not specifically designed to estimate population sizes in the ZOI. The underlying statistics composing ZOI indicators are means and proportions, which, in general, require smaller sample sizes compared to estimating population totals with the same level of precision. Even though estimates of the population size can be produced using the sampling weights in the ZOI Surveys, it should be noted that these estimates may not be reliable because the survey and sample size have not been specifically designed to estimate population totals.

The second method, which is not dependent on the survey sample, is to estimate the overall ZOI population using an external benchmark, such as a recent population census conducted in the country or the official demographic population projections from the country's national statistics agency.

Using the benchmark method to estimate the overall ZOI population means that the population estimates presented in Tables 1.2.1 to 1.2.3 will conform to an official source or external benchmark. For these reasons, the method of using an external benchmark, such as a recent population census or official demographic population projections, is favored as a more appropriate approach to estimate the ZOI population. However, when a reliable external benchmark to estimate the ZOI population is not available, the alternative method of using the survey sample with its sampling weights to estimate the overall ZOI population can be used, keeping in mind the limitations described previously.

After an estimate of the overall ZOI population is determined, the sample proportions from the survey for the selected sub-populations can be used to estimate the selected sub-population sizes in the ZOI,

---

<sup>40</sup> Stukel, D.M. (2018). *Feed the Future population-based survey sampling guide*. Washington, DC: Food and Nutrition Technical Assistance Project, FHI 360. Available at: <https://agrilinks.org/post/feed-future-zoi-survey-methods>.

and the number of households in the ZOI can be estimated using the average sample-weighted household size from the survey.

## 6.2 Guidelines to estimating the ZOI population, sub-populations, and households

### External benchmark method steps

**Step 1.** Determine whether estimates from a recent population census or official population projections are available from the country's national statistics agency for the year in which the ZOI Midline Survey fieldwork started.

**Step 2.** Ensure that the population census estimates or population projections are available at the administrative level at which the ZOI is defined. The estimates or projections can be population totals; they do not have to be disaggregated by sex or any other characteristic.

**Step 3.** If estimates from a recent population census or population projections are available but not for the year of survey fieldwork, determine whether a national (or appropriate subnational) annual population growth rate is available that could be applied to the population census estimates or projections to determine the population at the time of the survey. Possible sources include the country's national statistics agency. Note that the growth rate can be applied going forward or backward in time to extrapolate the population at the time of the survey from the benchmark population using an exponential growth assumption.<sup>41</sup>

**Step 4.** If neither appropriate population census estimates nor appropriate population projections can be found, consider using the sampling weight method. (See next section.)

**Step 5.** Determine whether the ZOI population estimates included in the ZOI Baseline Survey report need to be recalculated. Baseline ZOI population estimates may need to be recalculated when the method used at baseline is different from the method used at midline for ZOI population estimation for consistency during comparison between the two surveys or when the ZOI population estimation disaggregates required for Tables 1.2.1, 1.2.2, or 1.2.3 have changed since the Baseline Survey was conducted (e.g., the addition of the water programming area as a disaggregate at midline).

**Step 5.1.** If they need to be recalculated:

**5.1.1** Identify the source of the estimates used in the ZOI Baseline Survey report. (This information should be included in Chapter 1 of the report.)

**5.1.2** Determine whether the data source is available and can be accessed.

**5.1.3** If the data cannot be accessed, perform Step 1 to Step 3 for the year in which ZOI Baseline Survey fieldwork started.

---

<sup>41</sup> Exponential growth is a form of population growth in which the rate of change in the population is proportional to the population size at any given time (i.e., the population size follows an exponential distribution). Mathematically, the formula to determine  $P_t$ , the population at time  $t$  under an exponential population growth assumption, is defined as  $P_t = P_0 e^{rt}$  where  $P_0$  = population size at time  $t=0$  (beginning of the reference period),  $r$  = the expected population growth rate, and  $t$  = time period.

**Step 6.** Identify the administrative units that comprise the ZOI. Be sure to include all administrative units that comprise the ZOI, even if there are no sampled clusters in the administrative unit.

**Step 7.** Sum the population estimates for those administrative units identified in Step 6 to obtain the overall ZOI population estimate, and, if needed, apply the identified growth rate (see Step 3).

**Step 8.** Populate the overall ZOI population estimate calculated in Step 7 in the “Total number of individuals” row of Table 1.2.1 (cell B4 for baseline and cell E4 for midline).

**Step 9.** Calculate the sample-weighted proportion of each sub-population and background characteristic of interest (see Tables 1.2.1 and 1.2.2) for de jure household members using the persons-level analytic data file and the household sampling weight. For example:

```
use "FTF ZOI Survey [COUNTRY] [YEAR] analytic persons data"
svyset hhea [pw_wgt=wgt_hh], strata(strata)
svy, subpop(hhmem_dj): tab c0_5m
svy, subpop(hhmem_dj): tab genhhstype_dj
```

**Step 10.** Populate the sample-weighted proportions (multiplied by 100) to one decimal place for each sub-population or background characteristic calculated in Step 9 in Column C (baseline) and Column F (midline) of Tables 1.2.1 and 1.2.2.

**Step 11.** Multiply the “Total number of individuals” value in cell B4 by the percentages (divided by 100) in Column C and the “Total number of individuals” value in cell E4 by the percentages (divided by 100) in Column F to obtain the estimated number of individuals in the ZOI at baseline and midline by key sub-population and background characteristic in Tables 1.2.1 and 1.2.2.

**Step 12.** Calculate the sample-weighted average household size (de jure household members) using the household-level analytic data file and the household sampling weight.

```
use "FTF ZOI Survey [COUNTRY] [YEAR] analytic household data"
svyset hhea [pw_wgt=wgt_hh], strata(strata)
svy: mean hhsize_dj
```

**Step 13.** Divide the overall ZOI population estimate calculated in Step 7 by the average household size calculated in Step 12 and populate the result in the “Total number of households” row of Table 1.2.3 (cell B4 for baseline and cell E4 for midline).

**Step 14.** Calculate the sample-weighted proportion of each sub-population and background characteristic (see Table 1.2.3) using the household-level analytic data file and the household sampling weight. For example:

```
use "FTF ZOI Survey [COUNTRY] [YEAR] analytic household data"
svyset hhea [pw_wgt=wgt_hh], strata(strata)
svy: tab genhhstype_dj
svy: tab ahtype
```

**Step 15.** Populate the sample-weighted proportions (multiplied by 100) to one decimal place for each sub-population or background characteristic calculated in Step 14 in Column C (baseline) and Column F (midline) of Table 1.2.3.

### Sample weight method steps

1. Calculate the household sampling weight and add it to the household data file.
2. Multiply the household sampling weight by the number of household members for each household.
3. Sum the quantities calculated in Step 2 across all households in the survey sample to obtain the estimate of the overall ZOI population.
4. Go to Step 8 under the external benchmark method and implement Steps 8–15.

Because the external benchmark method is the preferred method that will be used in most cases to estimate the overall ZOI population size, estimating the individual sub-population sizes will be done the same way under the two methods for consistency. Another way to estimate sub-population sizes is to use the individual-level sample weights, but taking this approach would give different population sizes (and also different percentages of the overall ZOI population) for the sub-populations. In addition, for some sub-populations (e.g., children 0-5 months of age), sample sizes may be small so the population estimates using the sum of the sample weights for these sub-populations could be unreliable.

## 6.3 Step-by-step instructions for populating the population estimate tables

These step-by-step instructions pick up at Step 9 in Section 6.2 for both the external benchmark method and the sample weight method and include more detail on how to produce the population estimates for the ZOI Midline Survey population estimate tables in the midline indicator assessment report template:

- Table 1.2.1: Population of Individuals in the ZOI, by Category
- Table 1.2.2: Population of Farmers Responsible for Cultivating or Raising Targeted Value Chain Commodities in the ZOI, by Category
- Table 1.2.3: Number of Households in the ZOI, by Category

**Step 9.** Calculate the sample-weighted proportion of each sub-population and background characteristic (see Tables 1.2.1 and 1.2.2) for de jure household members using the persons-level analytic data file and the household sampling weight. (All template syntax in this step uses Stata syntax.)

**Step 9.1.** Load the persons-level analytic data file.

```
use "FTF ZOI Survey [COUNTRY] [YEAR] analytic persons data.dta",  
clear
```

**Step 9.2.** Apply the complex survey design with the household sample weight (`wgt_hh`), where `hhea` is the primary sampling unit or cluster and `strata` is the strata variable.

```
svyset hhea [pw=wgt_hh], strata(strata)
```



**Step 9.3.** Calculate the sample-weighted percentage of children 0-5 months of age overall and by sex.

```
svy, subpop(hhmem_dj): tab c0_5m, perc format(%6.1f)
svy, subpop(hhmem_dj): tab c0_5m sex, col perc format(%6.1f)
```

**Step 9.4.** Calculate the sample-weighted percentage of children 6-23 months of age overall and by sex.

```
svy, subpop(hhmem_dj): tab c6_23m, perc format(%6.1f)
svy, subpop(hhmem_dj): tab c6_23m sex, col perc format(%6.1f)
```

**Step 9.5.** Calculate the sample-weighted percentage of children 0-23 months of age overall and by sex.

```
svy, subpop(hhmem_dj): tab c0_23m, perc format(%6.1f)
svy, subpop(hhmem_dj): tab c0_23m sex, col perc format(%6.1f)
```

**Step 9.6.** Calculate the sample-weighted percentage of children 0-59 months of age overall and by sex.

```
svy, subpop(hhmem_dj): tab c0_59m, perc format(%6.1f)
svy, subpop(hhmem_dj): tab c0_59m sex, col perc format(%6.1f)
```

**Step 9.7.** Calculate the sample-weighted percentage of youth 15-29 years of age overall and by sex.

```
svy, subpop(hhmem_dj): tab age15_29y, perc format(%6.1f)
svy, subpop(hhmem_dj): tab age15_29y sex, col perc format(%6.1f)
```

**Step 9.8.** Calculate the sample-weighted percentage of women of reproductive age (15-49 years of age) overall and by age (15-19 and 20-49 years).

```
svy, subpop(hhmem_dj): tab wra, perc format(%6.1f)
svy, subpop(hhmem_dj): tab wra cage, col perc format
```

**Step 9.9.** Calculate the sample-weighted percentage of primary adult female decision-makers who are usual household members overall and by age (18-29 years and 30 years of age or older).

```
svy, subpop(hhmem_dj): tab fdm_dj, perc format(%6.1f)
svy, subpop(hhmem_dj): tab fdm_dj youth_fdm_dj, col perc
format(%6.1f)
```

**Step 9.10.** Calculate the sample-weighted percentage of primary adult female decision-makers who participate in household agricultural production overall.

```
svy, subpop(hhmem_dj): tab fdm_econ_farm, perc format(%6.1f)
svy, subpop(hhmem_dj): tab fdm_econ_farm youth_fdm_dj, col perc
format(%6.1f)
```

**Step 9.11.** Calculate the percent distribution of the overall ZOI population by residence (those living in urban areas and those living in rural areas).

```
svy, subpop(hhmem_dj): tab ahtype, perc format(%6.1f)
```

**Step 9.12.** Calculate the percent distribution of the overall ZOI population by administration level I in the country (e.g., province or region).

```
svy, subpop(hhmem_dj): tab c06, perc format(%6.1f)
```

**Step 9.13.** Calculate the percent distribution of the overall ZOI population by gendered household type.

```
svy, subpop(hhmem_dj): tab genhhtype_dj, perc format(%6.1f)
```

**Step 9.14.** Calculate the sample-weighted percentage of farmers responsible for cultivating or raising at least one targeted value chain commodity overall, by sex, and by age (15-29 years or 30 years of age or older).

```
svy, subpop(hhmem_dj): tab vcc_elig, perc format(%6.1f)
svy, subpop(hhmem_dj): tab vcc_elig sex, col perc format(%6.1f)
svy, subpop(hhmem_dj): tab vcc_elig age15_29y, perc format(%6.1f)
```

**Step 9.15.** Calculate the sample-weighted percentage of farmers responsible for cultivating each targeted value chain commodity overall, by sex and by age (15-29 years or 30 years of age or older).

```
svy, subpop(hhmem_dj): tab [VCC]_elig, perc format(%6.1f)
svy, subpop(hhmem_dj): tab [VCC]_elig sex, col perc format(%6.1f)
svy, subpop(hhmem_dj): tab [VCC]_elig age15_29y, perc
format(%6.1f)
```

**Step 10.** Populate the sample-weighted percentages to one decimal place for each sub-population or background characteristic calculated in Step 9 in Column C (baseline) and Column F (midline) of Tables I.2.1 and I.2.2, as applicable.

**Step 11.** Multiply the “Total number of individuals” value in cell B4 by the percentages (divided by 100) in Column C and the “Total number of individuals” value in cell E4 by the percentages (divided by 100) in Column F to obtain the estimated number of individuals in the ZOI by key sub-population and background characteristic in Tables I.2.1 and I.2.2.

**Step 12.** Calculate the sample-weighted average household size (de jure household members) using the household-level analytic data file and the household sampling weight. (All template syntax in this step uses Stata syntax.)

**Step 12.1.** Load the household-level analytic data file.

```
use "FTF ZOI Survey [COUNTRY] [YEAR] analytic household
data.dta", clear
```

**Step 12.2.** Apply the complex survey design with the household sample weight (**wgt\_hh**), where **hhea** is the primary sampling unit or cluster and **strata** is the strata variable.

```
svyset hhea [pw=wgt_hh], strata(strata)
```

**Step 12.3.** Calculate the sample-weighted average household size (de jure household members).

```
svy: mean hhsize_dj
```

**Step 13.** Divide the overall ZOI population estimate calculated in Step 7 by the average household size calculated in Step 12 and populate the result in the “Total number of households” row of Table 1.2.3 (cell B4 for baseline and cell E4 for midline).

**Step 14.** Calculate the sample-weighted percentage of each sub-population and background characteristic of interest (see Table 1.2.3) using the household-level analytic data file and the household sampling weight. For example:

```
svy: tab genhhtype_dj
svy: tab ahtype
svy: tab c06
svy: tab hh_[VCC]
```

**Step 15.** Populate the sample-weighted proportions (multiplied by 100) to one decimal place for each sub-population or background characteristic calculated in Step 14 in Column C (baseline) and Column F (midline) of Table 1.2.3.

## **PART II. INDICATORS**

## 7. DEMOGRAPHIC INDICATORS

In this chapter, indicators that provide important background information about the composition of households and the education of individuals in the ZOI are defined and their calculation described. The chapter is divided into two sections: household demographic characteristics and characteristics of de jure primary adult female decision-makers.

### 7.1 Household demographic characteristics

This section describes the following household-level demographic indicators:

- Mean household size
- Mean number of children under 2 years
- Mean number of children under 5 years
- Mean number of children 5 years or older (5-17 years)
- Mean number of youth (15-29 years)
- Mean number of women of reproductive age (15-49 years)
- Mean number of adult male household members
- Mean number of adult female household members
- Mean number of farmers of any targeted value chain
- Mean number of farmers of each value chain
- Percent distribution of households by household size category

The indicators in this section can be calculated using the household-level analytic data file (FTF ZOI Survey [COUNTRY] [YEAR] household data analytic). The step-by-step guidance follows the syntax in Household Demographics section of the *FTF ZOI Survey [COUNTRY] [YEAR] syntax household analytic* do file.

#### 7.1.1 Mean household size

This indicator is the sample-weighted mean number of de jure household members per household. It is calculated using the `hhsize_dj` variable created in Section 5.1.1.

#### Definitions

Numerator	Total number of de jure household members in surveyed households
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <code>hhsize_dj</code> analytic variable used cannot be missing values.
Survey variables used	<code>hhea</code> , <code>wgt_hh</code> , <code>strata</code>
Analytic variables used	<code>hhsize_dj</code>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the `hhsz_dj` variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: mean hhsz_dj
```

### 7.1.2 Mean number of children under 2 years

This indicator is the sample-weighted mean number of children under 2 years of age per household who are de jure household members. It is calculated using the variable, `ncu2_dj`, created in Section 5.1.4.

## Definitions

Numerator	Number of children under 2 years of age in surveyed households who are de jure household members
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <code>ncu2_dj</code> analytic variable used cannot be missing values.
Survey variables used	<code>hhea</code> , <code>wgt_hh</code> , <code>strata</code>
Analytic variables used	<code>ncu2_dj</code>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the `ncu2_dj` variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: mean ncu2_dj
```

### 7.1.3 Mean number of children under 5 years

This indicator is the sample-weighted mean number of children under 5 years of age per household who are de jure household members. It is calculated using the variable, `ncu5_dj`, created in Section 5.1.5.

## Definitions

Numerator	Number of children under 5 years of age in surveyed households who are de jure household members
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <code>ncu5_dj</code> analytic variable used cannot be missing values.

Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>ncu5_dj</b>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the **ncu5\_dj** variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: mean ncu5_dj
```

### 7.1.4 Mean number of children 5-17 years of age

This indicator is the sample-weighted mean number of children between 5 and 17 years of age, inclusive, per household who are de jure household members. It is calculated using the variable, **nc5\_17\_dj**, created in Section 5.1.6.

## Definitions

Numerator	Number of children between 5 and 17 years of age in surveyed households who are de jure household members
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <b>nc5_17_dj</b> analytic variable used cannot be missing values.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>nc5_17_dj</b>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the **nc5\_17y\_dj** variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: mean nc5_17y_dj
```

### 7.1.5 Mean number of youth 15-29 years of age

This indicator is the sample-weighted mean number of youth (15-29 years of age) per household who are de jure household members. It is calculated using the variable, **nage15\_29y\_dj**, created in Section 5.1.7.

## Definitions

Numerator	Number of youth between 15 and 29 years of age in surveyed households who are de jure household members
-----------	---

Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <b>nage15_29_dj</b> analytic variable used cannot be missing values.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>nage15_29y_dj</b>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the **nage15\_29\_dj** variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: mean nage15_29h_dj
```

### 7.1.6 Mean number of women of reproductive age

This indicator is the sample-weighted mean number of women of reproductive age (15-49 years) per household who are de jure household members. It is calculated using the variable, **nwra\_dj**, created in Section 5.1.3.

## Definitions

Numerator	Number of women of reproductive age in surveyed households who are de jure household members
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <b>nwra_dj</b> analytic variable used cannot be missing values.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>nwra_dj</b>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the **nwra\_dj** variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: mean nwra_dj
```



### 7.1.7 Mean number of adult male household members

This indicator is the sample-weighted mean number of de jure adult (18 years of age or older) male household members per household. It is calculated using the variable, **nadult\_mdj**, created in Section 5.1.2.

#### Definitions

Numerator	Number of de jure adult male household members in surveyed households
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <b>adult_mdj</b> analytic variable used cannot be missing values.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>nadult_mdj</b>
Analytic variables created	n/a

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the **nadult\_mdj** variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: mean nadult_mdj
```

### 7.1.8 Mean number of adult female household members

This indicator is the sample-weighted mean number of de jure adult (18 years of age or older) female household members per household. It is calculated using the variable, **nadult\_fdj**, created in Section 5.1.2.

#### Definitions

Numerator	Number of de jure adult female household members in surveyed households
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <b>nadult_fdj</b> analytic variable used cannot be missing values.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>nadult_fdj</b>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the `nadult_fdj` variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: mean nadult_fdj
```

### 7.1.9 Mean number of producers of any targeted value chain commodity and each targeted value chain commodity

These indicators are the sample-weighted mean number of producers (farmers) of any targeted value chain commodity (VCC) and each targeted VCC who are de jure household members per household. Targeted VCCs will vary by country. Producers are considered to be those who are responsible for making management decisions about cultivating one or more plots of targeted crops or about raising one or more types of targeted livestock. The template syntax includes maize, dairy cows, sheep, and fishponds as the targeted VCCs, and the indicators are calculated using variables created in Section 5.1.8. Be sure to adapt the syntax to reflect the VCCs included in the ZOI Survey.

## Definitions

Numerators	<ul style="list-style-type: none"><li>• Number of producers (farmers) of any targeted VCC in surveyed households who are de jure household members</li><li>• Number of maize producers (farmers) in surveyed households who are de jure household members</li><li>• Number of dairy cow producers (farmers) in surveyed households who are de jure household members</li><li>• Number of sheep producers (farmers) in surveyed households who are de jure household members</li><li>• Number of fishpond producers (farmers) in surveyed households who are de jure household members</li></ul>
Denominator	Total number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <code>nvcc</code> analytic variables used cannot be missing values.
Survey variables used	<code>hhea</code> , <code>wgt_hh</code> , <code>strata</code>
Analytic variables used*	<code>nvcc_dj</code> , <code>nvcc_maize_dj</code> , <code>nvcc_dairy_dj</code> , <code>nvcc_sheep_dj</code> , <code>nvcc_fish_dj</code>
Analytic variables created	n/a

\* Depends on the targeted VCCs included in the ZOI Survey

## Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the `nvcc_dj` variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: mean nvcc_dj
```

**Step 2.** Repeat Step 1 for each targeted VCC included in the ZOI Survey. For example, for maize, dairy cows, sheep, and fishponds, the Stata syntax would be as follows:

```
svy: mean nvcc_maize_dj
svy: mean nvcc_dairy_dj
svy: mean nvcc_sheep_dj
svy: mean nvcc_fish_dj
```

### 7.1.10 Percent distribution of households by household size

This indicator is the sample-weighted percentage distribution of households among three household size categories (small, medium, and large) and is calculated using the variable, **hhsizegrp\_dj**, created in Section 5.1.10. Small households have 1-4 de jure household members, medium households have 5-9 household members, and large households have 10 or more household members. It is based on information reported in the household roster.

#### Definitions

Numerator	Number of households in each household size category
Denominator	Total number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Not applicable. The <b>hhsizegrp_dj</b> analytic variable used cannot be missing values.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>hhsizegrp_dj</b>
Analytic variables created	n/a

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the percentage of households that are small, medium, and large using the **hhsizegrp\_dj** variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: tab hhsizegrp_dj
```

## 7.2 Characteristics of primary adult female decision-makers

This section describes indicators that characterize primary adult female decision-makers. Each of these characteristics is calculated for only primary adult female decision-makers who completed Module 6, *Women's empowerment in agriculture*. Primary adult female decision-makers are adult females (18 years of age or older) who make the most important decisions in the household and are identified in the household roster. The following indicators are described in this section:

- Percent distribution of primary adult female decision-makers by age group
- Percent distribution of primary adult female decision-makers by current marital status
- Percentage of primary adult female decision-makers who participate in any economic activity
- Percent distribution of primary adult female decision-makers by type of economic activity

The indicators in this section can be calculated using the individual-level analytic data file (FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic). The step-by-step guidance follows the syntax in the Primary Adult Female Decision-maker Characteristics section of the *FTF ZOI Survey [COUNTRY] [YEAR] syntax persons analytic* do file.

### 7.2.1 Percent distribution of primary adult female decision-makers by age group

This indicator is the percent distribution of primary adult female decision-makers (18 years of age or older) by age group in years: 18-24, 25-29, 30-39, 40-49, 50-59, and 60+. The indicator is calculated using the variable, `agegrp_fdm_dj`, created in Section 4.5.3.

#### Definitions

Numerator	Number of primary adult female decision-makers in each specified age group
Denominator	Number of primary adult female decision-makers in surveyed households
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Primary adult female decision-maker
Disaggregation levels	n/a
Treatment of missing data	The <code>agegrp_fdm_dj</code> analytic variable cannot have missing values for women who completed Module 6.
Survey variables used	<code>hhea</code> , <code>wgt_fpd</code> , <code>strata</code>
Analytic variables used	<code>agegrp_fdm_dj</code>
Analytic variables created	n/a

#### Calculations

**Step 1.** After applying the primary adult female decision-maker sampling weight, calculate the percent distribution of primary adult female decision-makers by age category using the `agegrp_fdm_dj` variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_fpd], strata(strata)
svy: tab agegrp_fdm_dj
```

### 7.2.2 Percent distribution of primary adult female decision-makers by their current marital status

This indicator is the percentage distribution of primary adult female decision-makers by their current marital status. The indicator is calculated using the variable, `marstat_fdm_dj`, created in Section 4.5.5.

## Definitions

Numerator	Number of primary adult female decision-makers with a marital status in each specified marital status category
Denominator	Number of primary adult female decision-makers in surveyed households
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Primary adult female decision-maker
Disaggregation levels	n/a
Treatment of missing data	If information on marital status is missing, the <code>marstat_fdm_dj</code> analytic variable is set to missing and will be excluded from the indicator numerator and denominator.
Survey variables used	<code>hhea</code> , <code>wgt_fpdm</code> , <code>strata</code>
Analytic variables used	<code>marstat_fdm_dj</code>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the primary adult female decision-maker sampling weight, calculate the percentage of primary adult female decision-makers by marital status category using the `marstat_fdm_dj` variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_fpdm], strata(strata)
svy: tab marstat_fdm_dj
```

### 7.2.3 Percentage of primary adult female decision-makers who participate in economic activities

This indicator is the percentage of primary adult female decision-makers who participate in any economic activity. Both paid and unpaid types of economic activity are included. Economic activity includes farm work (food crop farming, cash crop farming, livestock raising, or fishing/fishpond culture); non-farm work (running small businesses or self-employment); and wage/salaried employment (agriculture and non-agriculture-based work). Domestic work, such as caring for children or the elderly, cooking, and cleaning, is not included. This indicator is calculated using the variable, `fdm_econ_any`, created in Section 4.5.6.

## Definitions

Numerator	Number of primary adult female decision-makers who participate in any economic activity
Denominator	Number of primary adult female decision-makers in surveyed households
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Primary adult female decision-maker
Disaggregation levels	n/a

Treatment of missing data	If information on economic activity participation is missing, the <b>fdm_econ_any</b> analytic variable is set to missing and will be excluded from the indicator numerator and denominator.
Survey variables used	<b>hhea, wgt_fpdm, strata</b>
Analytic variables used	<b>fdm_econ_any</b>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the primary adult female decision-maker sampling weight, calculate the percentage of primary adult female decision-makers who participated in at least one economic activity using the **fdm\_econ\_any** variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_fpdm], strata(strata)
svy: tab fdm_econ_any
```

### 7.2.4 Percentage of primary adult female decision-makers who participate in farm work, non-farm work, and wage or salaried employment

These indicators reflect the percentage of primary adult female decision-makers who participate in three types of economic activities. Both paid and unpaid types of economic activities are included. Economic activities include farm work (food crop farming, cash crop farming, livestock raising, or fishing/fishpond culture); non-farm work (running small businesses or self-employment); and wage or salaried employment (agriculture and non-agriculture-based work). Domestic work, such as caring for children or the elderly, cooking, and cleaning, is not included. These indicators are calculated using variables created in Section 4.5.6.

## Definitions

Numerators	<ul style="list-style-type: none"> <li>Number of primary adult female decision-makers who participate in farm work</li> <li>Number of primary adult female decision-makers who participate in non-farm work</li> <li>Number of primary adult female decision-makers who participate in wage or salaried employment</li> </ul>
Denominator	Number of primary adult female decision-makers in surveyed households
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Primary adult female decision-maker
Disaggregation levels	n/a
Treatment of missing data	If information on economic activity participation is missing, the <b>fdm_econ_farm</b> , <b>fdm_econ_nonfarm</b> , and <b>fdm_econ_wage</b> analytic variables are set to missing and will be excluded from the indicator numerator and denominator.
Survey variables used	<b>hhea, wgt_fpdm, strata</b>
Analytic variables used	<b>fdm_econ_farm, fdm_econ_nonfarm, fdm_econ_wage</b>
Analytic variables created	n/a

## Calculations

**Step 1.** After applying the primary adult female decision-maker sampling weight, calculate the percentage of primary adult female decision-makers who participated in farm work using the `fdm_econ_farm` variable. Repeat for non-farm work using the `fdm_econ_nonfarm` variable and for wage or salaried work using the `fdm_econ_wage` variable. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_fpdm], strata(strata)
svy: tab fdm_econ_farm
svy: tab fdm_econ_nonfarm
svy: tab fdm_econ_wage
```

## 8. WATER, SANITATION, AND HYGIENE INDICATORS

In this chapter, indicators related to water, sanitation, and hygiene are defined and their calculation described. The following eight indicators are included in the chapter:

- Percent of households using an improved water source
- Percent of households using a correct water treatment practice or technology
- Percent of households with soap and water at a handwashing station commonly used by family members (Feed the Future phase two indicator)
- Percent of households with access to a basic sanitation service (Feed the Future phase two indicator)
- Percent of households using an improved but shared sanitation facility
- Percent of households using an unimproved sanitation facility
- Percent of households practicing open defecation
- Prevalence of water insecurity, based on the Brief Household Water Insecurity Experiences (HWISE-4) scale

All indicators in this chapter are based on information reported in Module 2, *Dwelling characteristics and household assets*. The step-by-step procedures for all but the last indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax WASH dwelling.do* file.

### 8.1 Percent of households using an improved water source

This indicator is the percentage of households using an improved water source that is regularly available. Improved water sources include piped water into the dwelling, piped water into the yard, piped water to the household's neighbor, a public tap or standpipe, a tubewell or borehole, a protected dug well, a protected spring, rainwater, tanker trucks, carts with a small tank, and bottled water. Regular availability of the water source means that it is available (1) all year round and (2) every day in the 2 weeks preceding the survey.

#### Definitions

Numerator	Number of households using an improved water source that is available year-round and that was available every day during the 2 weeks preceding the survey as their main source of drinking water
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Households missing information on water source or regular availability are considered to be not using an improved water source. They are included in the denominator but excluded from the numerator.
Survey variables used	<b>v211, v214, v215, hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	<b>n/a</b>
Analytic variables created	<b>h2o_improved, h2o_regular, h2o_imp_reg</b>



## Calculations

**Step 1.** Create a binary variable to flag households that reported use of an improved water source as their main source of drinking water (**h2o\_improved**).

```
Set          h2o_improved=0
Replace      h2o_improved=1 if
              v211=11,12,13,14,21,31,41,51,61,71,or 91
Label values 0 "No"
              1 "Yes"
Label variable "HH uses improved drinking water source"
```

**Step 2.** Create a binary variable to flag households that reported that their main source of drinking water was regularly available (**h2o\_regular**)—that is, it was available year-round and was available every day during the 2 weeks preceding the survey.

```
Set          h2o_regular=missing
Replace      h2o_regular=0 if v214#missing and v215#missing
Replace      h2o_regular=1 if v214=1 and v215=1
Label values 0 "No"
              1 "Yes"
Label variable "HH uses regularly available drinking water
              source"
```

**Step 3.** Create a final binary variable that indicates whether the households' main source of drinking water was improved and regularly available (**h2\_imp\_reg**).

```
Set          h2o_imp_reg=missing
Replace      h2o_regular=0 if h2o_improved#missing and
              h2o_regular#missing
Replace      h2o_imp_reg=1 if h2o_improved=1 and h2o_regular=1
Label values 0 "No"
              1 "Yes"
Label variable "HH uses regularly available improved drinking
              water source"
```

**Step 4.** After applying the household sampling weight, calculate the percentage of households that use a regularly available improved water source using the **h2o\_imp\_reg** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab h2_imp_reg
svy: tab h2_imp_reg ahtype, col
```

## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

UNICEF & World Health Organization. (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/publications-detail-redirect/9789241516235>

## 8.2 Percent distribution of households by the number of trips household members made to obtain drinking water during the past 7 days

This indicator is the percent distribution of households by the number of trips household members made to obtain drinking water from the household's main drinking water source during the 7 days preceding the survey. Households that have their main drinking water source in their dwelling, yard, or plot are included as having made no trips. The categories used can be adapted based on the survey data; they do not have to be 0, 1-4, 5-9, and 10 or more trips. The question required to calculate this indicator is not included in the ZOI Baseline Survey questionnaire; it was added to the ZOI Midline Survey main survey questionnaire, so the indicator estimates cannot be compared between baseline and midline.

### Definitions

Numerator	Number of surveyed households that made 0, 1-4, 5-9, or 10 or more trips* to collect water from the household's main drinking water source during the seven days preceding the survey. *Note: The categories used can be adapted based on the survey data; they do not have to be 0, 1-4, 5-9, and 10 or more trips.
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	n/a
Treatment of missing data	Households missing data for <b>v213a</b> that do not have their main drinking water source located in their dwelling, yard, or plot are excluded from both the indicator numerator and denominator.
Survey variables used	<b>v213a</b>
Analytic variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables created	<b>h2o_trips, h2o_trips_cat</b>

### Calculations

**Step 1.** Create a variable that captures the number of trips that households took during the 7 days preceding the survey to collect drinking water (**h2o\_trips**).

```
Set          h2o_trips=missing
Replace      h2o_trips=0 if v212<3 or if v211=11 or v211=12
Replace      h2o_trips=v213a if v213a≥0 and v213a<97
Label variable "Number of trips for drinking water, past 7 days
              (continuous) "
```

**Step 2.** Create a variable that assigns households to categories according to the number of trips that the household members took to collect drinking water during the 7 days preceding the survey

(h2o\_trips\_cat). INSTRUCTIONS: Adjust the categories used as needed, based on the data, so that households are reasonably spread across the categories.

```
Set          h2o_trips_cat=0 if h2o_trips=0
Replace      h2o_trips_cat=1 if h2o_trips=1,2,3, or 4
Replace      h2o_trips_cat=2 if h2o_trips=5,6,7,8, or 9
Replace      h2o_trips_cat=3 if h2o_trips≥10 and h2o_trips<97
Label values 0 "No trips"
              1 "1-4 trips"
              2 "5-9 trips"
              3 "10+ trips"
Label variable "Number of trips for drinking water, past 7 days
               (categorical)"
```

**Step 3.** After applying the household sampling weight, calculate the percentage of households that made 0, 1-4, 5-9, and 10 or more trips to collect water—overall and by residence (urban or rural).

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab h2o_trips_cat, perc ci
svy: tab h2o_trips_cat ahtype, col perc ci
```

### 8.3 Percent of households using a correct water treatment practice or technology

This indicator is the percentage of households using a water treatment practice or technology that effectively kills or removes pathogens. Effective practices and technologies include boiling the water; adding bleach or chlorine to the water; using a ceramic, sand, or composite water filter; and using solar disinfection. Practices such as straining the water through a cloth or letting the water stand and settle are not considered effective water treatment methods. Respondents may report using more than one practice or technology. A household is considered to be using a correct water treatment practice or technology if it uses at least one of the effective practices or technologies.

#### Definitions

Numerator	Number of households using at least one correct water treatment practice or technology
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Households missing information about water treatment, with a “don’t know” response, or with only an “other” treatment response are considered to be not using an effective water treatment method.
Survey variables used	v217a-v217e, hhea, wgt_hh, strata, ahtype
Analytic variables used	n/a
Analytic variables created	h2o_corrtrat

## Calculations

**Step 1.** Create a binary variable to flag households using an effective water treatment practice or technology (`h2o_corrtrtreat`).

```
Set          h2o_corrtrtreat=0
Replace      h2o_corrtrtreat=1 if (v217a=1 or v217b=1 or v217d=1
              or v217e=1)
```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that correctly treat their drinking water using the `h2o_corrtrtreat` variable. Repeat using the residence variable (`ahtype`) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab h2o_corrtrtreat
svy: tab h2o_corrtrtreat ahtype, col
```

## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

UNICEF & World Health Organization. (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/publications-detail-redirect/9789241516235>

World Health Organization & UNICEF. (2017). *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines*. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789241512893>

### 8.4 Percent of households with soap and water at a handwashing station on premises (Feed the Future phase two indicator)

This indicator, a Feed the Future phase two indicator, measures the percentage of households with soap and water at a handwashing station on premises. Enumerators collect information for this indicator by physically observing the handwashing station when they administer Module 2, *Dwelling characteristics and household assets*.

## Definitions

Numerator	Number of households in which soap and water are found at a handwashing station on premises
Denominator	Number of households where a place for handwashing was observed or where no handwashing facility exists. (Households that do not provide permission to see the handwashing facility or where it is not seen for some other reason are excluded.)
Unit of measure	Percentage
Level of data	Household

Sampling weight	Household
Disaggregation levels	Gendered household type* Residence (urban, rural)*
Treatment of missing data	Households with missing information on a place for washing or on the availability of water, soap, or other cleansing agent are assumed not to have the characteristic with the missing value.
Survey variables used	<b>v205, v206, v207, hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	<b>genhhstype_dj</b>
Analytic variables created	<b>handwash</b>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Create a binary variable to flag households in which both soap and water are found at a handwashing station (**handwash**). Include households where a fixed or mobile place for handwashing was observed and households where no handwashing facility exists in the denominator—but exclude households that do not provide permission to see the handwashing facility or where it is not seen for some other reason.

Set	<b>handwash=0 if v205=1 or v205=2 or v205=3</b>
Replace	<b>handwash=1 if (v205=1 or v205=2) and v206=1 and v207=1</b>
Label values	0 "No" 1 "Yes"
Label variable	<b>"HH has a handwashing station with soap and water"</b>

**Step 2.** After applying the household sampling weight, calculate the percentage of households in which soap and water are found at the commonly used handwashing station using the **handwash** variable. Repeat using the gendered household type variable constructed using de jure household members (**genhhstype\_dj**) to obtain results disaggregated for male and female adult, male adult-only, female adult-only, and children-only households and the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: tab handwash
svy: tab handwash genhhstype_dj, col
svy: tab handwash ahtype, col
```

## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

World Health Organization. (2018). *Global reference list of 100 core health indicators*. Available at: [https://www.who.int/publications/i/item/2018-global-reference-list-of-100-core-health-indicators-\(-plus-health-related-sdgs\)](https://www.who.int/publications/i/item/2018-global-reference-list-of-100-core-health-indicators-(-plus-health-related-sdgs))

World Health Organization. (2019). *Water sanitation hygiene. Monitoring hygiene*. Available at: [https://www.who.int/water\\_sanitation\\_health/monitoring/coverage/monitoring-hygiene/en/](https://www.who.int/water_sanitation_health/monitoring/coverage/monitoring-hygiene/en/)

World Health Organization & UNICEF. (2017). *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines*. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789241512893>

## 8.5 Percent of households using an improved but shared sanitation facility

This indicator is the percentage of households using an improved sanitation facility that is shared with other households. Improved sanitation facilities include flush or pour-flush toilets that empty into a piped sewer system, a septic tank, a pit latrine, or “don’t know where”; composting toilets; ventilated improved pit latrines; and pit latrines with a slab. Sanitation facilities that are not considered to be improved include flush or pour-flush toilets that empty into “somewhere else,” open pits or pit latrines without a slab, bucket toilets, and hanging toilets. Shared facilities may be less hygienic than unshared facilities, which could deter use of the facilities.

### Definitions

Numerator	Number of households using an improved sanitation facility that is shared with other households
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Households that report using an improved toilet facility, but with missing information about sharing the toilet facility, are considered to be using an improved but shared facility; they are included in the numerator and denominator. Households missing information about type of sanitation facility used or with an “other” type of sanitation facility are considered to be not using an improved sanitation facility and are included in the denominator but excluded from the numerator.
Survey variables used	v208, v209, hhea, wgt_hh, strata, ahtype
Analytic variables used	n/a
Analytic variables created	san_improved, san_shared, san_impshared

### Calculations

**Step 1.** Create a binary variable to flag households using an improved sanitation facility (**san\_improved**).

```
Set          san_improved=0
Replace      san_improved=1 if v208=11,12,13,15,21,22,or 31
Label values 0 "No"
              1 "Yes"
Label variable "HH's sanitation facility is improved"
```

**Step 2.** Create a binary variable to flag households that share their sanitation facility with other households (**san\_shared**).

```
Set                san_shared=missing
Replace            san_shared=0 if v209=0
Replace            san_shared=1 if v209=1
Label values       0 "No"
                   1 "Yes"
Label variable     "HH shares sanitation facility with other HHs"
```

**Step 3.** Create a binary variable to flag households using an improved sanitation facility that is shared with other households (**san\_impshared**).

```
Set                san_impshared=0
Replace            san_impshared=1 if san_improved=1 and
                   (san_shared=1 or san_shared=missing)
Label values       0 "No"
                   1 "Yes"
Label variable     "HH uses a shared improved sanitation facility"
```

**Step 4.** After applying the household sampling weight, calculate the percentage of households that use an improved sanitation facility that is shared with other households using the **san\_impshared** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab san_impshared
svy: tab san_impshared ahtype, col
```

## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

UNICEF & World Health Organization. (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/publications-detail-redirect/9789241516235>

World Health Organization & UNICEF. (2017). *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines*. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789241512893>

## 8.6 Percent of households using an improved facility that is not shared (basic sanitation service) (Feed the Future phase two indicator)

This indicator is the percentage of households using an improved sanitation facility that is not shared with other households. Improved sanitation facilities include flush or pour-flush toilets that empty into a piped sewer system, a septic tank, a pit latrine, or “don’t know where”; composting toilets; ventilated improved pit latrines; and pit latrines with a slab. Sanitation facilities that are not considered to be

improved include flush or pour-flush toilets that empty into “somewhere else,” open pits or a pit latrines without a slab, bucket toilets, and hanging toilets.

## Definitions

Numerator	Number of households using an improved sanitation facility that is not shared with other households
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Gendered household type* Residence (urban, rural)*
Treatment of missing data	Households that report using an improved toilet facility, but with missing information about sharing the toilet facility, are considered not to be using an improved, non-shared facility. They are included in the denominator and excluded from the numerator. Households missing information about the type of sanitation facility they use or with an “other” type of sanitation facility are considered to be not using an improved sanitation facility. They are included in the denominator but excluded from the numerator.
Survey variables used	<b>hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	<b>san_improved, san_shared</b>
Analytic variables created	<b>san_impnotshared</b>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Create a binary variable to flag households using an improved sanitation facility that is not shared with other households (**san\_impnotshared**).

```

Set          san_impnotshared=0
Replace      san_impnotshared=1 if san_improved=1 and
              san_shared=0
Label values 0 "No"
              1 "Yes"
Label variable "HH uses an improved, unshared sanitation
              facility"
```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that use an improved sanitation facility that is not shared with other households using the **san\_impnotshared** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas and the variable constructed using de jure household members (**genhhstype\_dj**) to obtain results disaggregated for male and female adult, male adult-only, female adult-only, and children-only households. (Sample code uses Stata syntax.)

```

svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab san_impnotshared
svy: tab san_impnotshared genhhstype_dj, col
svy: tab san_impnotshared ahtype, col
```



## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

UNICEF & World Health Organization. (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/publications-detail-redirect/9789241516235>

World Health Organization & UNICEF. (2017). *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines*. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789241512893>

World Health Organization & UNICEF. (2018). *JMP Methodology: 2017 Update & SDG Baselines*. Available at: <https://washdata.org/sites/default/files/documents/reports/2018-04/JMP-2017-update-methodology.pdf>

## 8.7 Percent of households using an unimproved sanitation facility

This indicator is the percentage of households using an unimproved sanitation facility. A sanitation facility is considered to be unimproved if human excreta is not adequately separated from human contact. Unimproved sanitation facilities include flush or pour flush toilets that empty into “somewhere else,” open pits or pit latrines without a slab, bucket toilets, and hanging toilets. Households that report having no sanitation facility or report using the bush or field are considered to be using an unimproved sanitation facility. “Other” responses are also considered to be unimproved.

### Definitions

Numerator	Number of households using an unimproved sanitation facility
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Households missing information on type of sanitation facility used are considered to be using an unimproved facility; they are included in the numerator and denominator. Households with an “other” type of sanitation facility are considered to be using an unimproved facility; they are included in both the numerator and denominator.
Survey variables used	v208, hhea, wgt_hh, strata, ahtype
Analytic variables used	n/a
Analytic variables created	san_notimp

### Calculations

**Step 1.** Create a binary variable to flag households using an unimproved sanitation facility (san\_notimp).

```

Set          san_notimp=0
Replace      san_notimp=1 if v208=14,23,41,51,61,96, or
              missing
Label values 0 "No"
              1 "Yes"
Label variable "HH uses an unimproved sanitation facility"

```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that use an unimproved sanitation facility using the `san_notimp` variable. Repeat using the residence variable (`ahtype`) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```

svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab san_notimp
svy: tab san_notimp ahtype, col

```

## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

UNICEF & World Health Organization. (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/publications-detail-redirect/9789241516235>

World Health Organization & UNICEF. (2017). *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines*. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789241512893>

World Health Organization (WHO) & UNICEF. (2018). *JMP Methodology: 2017 Update & SDG Baselines*. Available at: <https://washdata.org/sites/default/files/documents/reports/2018-04/JMP-2017-update-methodology.pdf>

## 8.8 Percent of households practicing open defecation

This indicator is the percentage of households practicing open defecation. Households that report having no sanitation facility or report using the bush or field are considered to be practicing open defecation.

### Definitions

Numerator	Number of households that report having no sanitation facility or using the bush or field for defecation
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)

Treatment of missing data	Households missing information on type of sanitation facility used are considered to be using an unimproved facility; they are included in the denominator but excluded from the numerator. Households with an “other” type of sanitation facility are considered to not be practicing open defecation; they are included in the denominator but excluded from the numerator.
Survey variables used	<b>v208, hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	n/a
Analytic variables created	<b>san_opendef</b>

## Calculations

**Step 1.** Create a binary variable to flag households practicing open defecation (**san\_opendef**).

```
Set                san_opendef=0
Replace            san_opendef=1 if v208=61
Label values       0 "No"
                   1 "Yes"
Label variable     "HH practices open defecation"
```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that report having no sanitation facility or using the bush or field for defecation using the **san\_opendef** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: tab san_opendef
svy: tab san_opendef ahtype, col
```

## References

UNICEF & World Health Organization. (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/publications-detail-redirect/9789241516235>

World Health Organization & UNICEF. (2017). *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines*. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789241512893>

## 8.9 Mean number of household members who regularly defecate in the open

This indicator is the mean number of household members who regularly defecate in the open while at home or at work. They may defecate, for example, in a field, bush, forest, ditches, or canals; on riverbanks, garbage piles, or railway tracks; or along the road. The question required to calculate this indicator is not included in the ZOI Baseline Survey questionnaire; it was added to the ZOI Midline Survey main survey questionnaire, so the indicator estimates cannot be compared between baseline and midline.

## Definitions

Numerator	Number of household members in surveyed households who are reported to regularly defecate in the open
Denominator	Number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban/rural) Whether the household has a toilet facility
Treatment of missing data	Households missing data for <b>v210a</b> are excluded from both the indicator numerator and denominator.
Survey variables used	<b>v210a, hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	n/a
Analytic variables created	<b>hhmem_open_def</b>

## Calculations

**Step 1.** Create a variable to capture the number of household members who defecate in the open regularly (**hhmem\_open\_def**)

```
Set                hhmem_open_def=v210a if v210a<97
Label variable    "Number of HH members who defecate in open
                  regularly"
```

**Step 2.** After applying the household sampling weight, calculate the mean number of household members who regularly defecate in the open both overall and by residence and whether the household mainly uses any type of toilet using the **hhmem\_open\_def** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight= wgt_hh], strata(strata)
svy: mean hhmem_open_def
svy: mean hhmem_open_def, over(ahtype)
```

## 8.10 HWISE-4 Scale indicator

The HWISE-4 Scale indicator estimates the percentage of households experiencing water insecurity in the ZOI. Household water insecurity is calculated by summing the frequency with which households experience challenges related to availability, accessibility, acceptability, and use of water in the household. The four questions are each scored on a 0 to 3 scale, such that the total score ranges from 0 to 12, with a score greater than or equal to 4 indicating that the household is water insecure. The information is collected in ZOI Surveys beginning in 2022. **Table 5** summarizes the HWISE-4 questions asked and associated categories and the variables used in the indicator calculation. Although HWISE-4 is not an official Feed the Future indicator, it will be reported going forward across ZOI Surveys.

**Table 5: HWISE-4 Indicator Questions and Variables**

<b>HWISE-4 question</b>	<b>Variables</b>
In the last 4 weeks, how frequently did you or anyone in your household worry you would not have enough water for all of your household needs?	<b>v218a</b>
In the last 4 weeks, how frequently have you or anyone in your household had to change schedules or plans due to problems with your water situation?	<b>v218b</b>
In the last 4 weeks, how frequently have you or anyone in your household had to go without washing hands after dirty activities (e.g., defecating or changing diapers, cleaning animal dung) because of problems with water?	<b>v218c</b>
In the last 4 weeks, how frequently has there not been as much water to drink as you would like for you or anyone in your household?	<b>v218d</b>

## Definitions

Numerator	Number of households that score at least 4 out of 12 points across the 4 HWISE-4 questions
Denominator	Number of households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Gendered household type Residence (urban, rural) Wealth quintile Poverty status Shock experience severity
Treatment of missing data	Households missing information about one or more of the four HWISE-4 questions are excluded from the numerator and denominator.
Survey variables used	<b>v218a-d, wgt_hh, hhea, strata, ahtype</b>
Analytic variables used	<b>genhhtype_dj, awiquint, poor190, shock_sev</b>
Analytic variables created	<b>bwise4, hwise4, miss_h4, v218ax, v218bx, v218cx, v218dx</b>

## Calculations

This section describes the detailed step-by-step procedures to calculate the HWISE-4 indicator. The step-by-step procedures follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax HWISE4.do* file.

**Step 1.** Ensure that all four variables of interest are not missing for any households, by generating a variable to capture households with missing values (**miss\_h4**).

```

Set          miss_h4=0
Replace      miss_h4=1 if v218a=missing or v218b=missing or
              v218c=missing or v218d=missing
Replace      miss_h4=missing if v218a=missing and
              v218b=missing and v218c=missing and v218d=missing

Label values 0 "No"
              1 "Yes"
```

```
Label variable "HH is missing 1+ variable for HWISE-4
calculation"
```

**Step 2.** Create analytic variables (**v218ax**, **v218bx**, **v218cx**, and **v218dx**) that recode the variables **v218a**, **v218b**, **v218c**, and **v218d** with values of 4 to be 3.

```
Set          v218ax=v218a
Replace      v218ax=3 if v218a=4
Set          v218bx=v218b
Replace      v218bx=3 if v218b=4
Set          v218cx=v218c
Set          v218dx=3 if v218c=4

For each     variable (var) in v218ax v218bx v218cx v218dx:
  Label values 0 "Never (0 times)"
               1 "Rarely (1-2 times)"
               2 "Sometimes (3-10 times)"
               3 "Often/Always (11+ times)"
  Label variable "Recode of `var' ([`var' label])"
```

**Step 3.** Create a variable (**bwise4**) that sums the four variables created in Step 2, excluding households with any missing data.

```
Set          bwise4=missing
Replace      bwise4=0 if miss_h4=0
Replace      bwise4=v218ax+v218bx+v218cx+v218dx if miss_h4=0
Label variable "HH's HWISE-4 score (0-12)"
```

**Step 4.** Create a binary variable that flags households facing water insecurity according the HWISE-4 indicator (**hwise4**).

```
Set          hwise4=1 if bwise4≥4 and bwise4≠missing
Replace      hwise4=0 if bwise4<4
Label values 0 "Not water insecure"
               1 "Water insecure"
Label variable "Brief Household Water Insecurity Experience
Scale (HWISE-4)"
```

**Step 5.** After applying the household weight (**wgt\_hh**), calculate the percentage of households that are facing water insecurity using the **hwise4** analytic variable. Repeat using the gendered household type variable constructed using de jure household members (**genhh\_type\_dj**) to obtain results disaggregated for male and female adult, male adult-only, female adult-only, and children-only households; the wealth quintile variable (**awiquint**) to obtain results disaggregated for households in each asset-based wealth quintile; the poverty status variable (**poor190**) to obtain results disaggregated for households that are poor and non-poor according the \$1.90 poverty threshold; the residence variable (**ah\_type**) to obtain results disaggregated for households in urban and rural areas; the shock exposure severity variable (**shock\_sev**) to obtain results disaggregated for households that experienced no shocks during the 12 months preceding the survey, and those that experienced low, moderate, and high shock exposure severity; and the drinking water collection variable (**h2o\_trips**)

to obtain results disaggregated for households by the number of trips they made to obtain drinking water during the 7 days preceding the survey. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab hwise4
svy: tab hwise4 genhhtype_dj, col
svy: tab hwise4 ahtype, col
svy: tab hwise4 awiquint, col
svy: tab hwise4 poor190, col
svy: tab hwise4 shock_sev, col
svy: tab hwise4 h2o_trips, col
```

## References

HWISE Research Coordination Network. (2019). *HWISE user manual*. Northwestern University. Available at: [https://arch.library.northwestern.edu/concern/generic\\_works/ww72bb80d](https://arch.library.northwestern.edu/concern/generic_works/ww72bb80d) [Google Scholar]

Young, S.L., Miller, J.D., Frongillo, E.A., Boateng, G.O., Jamaluddine, Z., Neilands, T.B., & HWISE Research Coordination Network. (2021). Validity of a four-item Household Water Insecurity Experiences scale for assessing water issues related to health and well-being. *Am J Trop Med Hyg*, 104(1), 391-394. Doi: 10.4269/ajtmh.20-0417. PMID: 33124535; PMCID: PMC7790094.

## 9. DWELLING AND HOUSEHOLD CHARACTERISTIC INDICATORS

In this chapter, indicators related to dwelling and household characteristics are defined and their calculation described. The following six indicators are included in the chapter:

- Percent of households using solid fuels for cooking
- Mean number of de jure household members per sleeping room
- Percent distribution of households by dwelling roof materials
- Percent distribution of households by dwelling exterior wall materials
- Percent distribution of households by dwelling floor materials
- Percent of households that have electricity

All indicators in this chapter are based on information reported in Module 1, *Household roster and demographics*, and Module 2, *Dwelling characteristics and household assets*. The step-by-step procedures for all but the last indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax WASH dwelling.do* file.

### 9.1 Percent of households using solid fuels for cooking

This indicator is the percentage of households using solid fuels (coal or lignite; charcoal, wood, straw, shrubs, or grass; agricultural crop residue; and animal dung) as their main source of cooking fuel.

#### Definitions

Numerator	Number of households that use solid fuels for cooking
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Households that report that food is not cooked in the house, that are missing information on their main cooking fuel, or that have a response of “other” are excluded from the numerator and denominator.
Survey variables used	<b>v219, hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	n/a
Analytic variables created	<b>dw_solidfuel</b>

#### Calculations

**Step 1.** Create a binary variable that flags households using solid fuels for cooking (**dw\_solidfuel**).

```
Set          dw_solidfuel=missing
Replace      dw_solidfuel=1 if v219=6,7,8,9,10,or 11
Replace      dw_solidfuel=0 if v219<6
Label values 0 "No"
              1 "Yes"
Label variable "HH uses solid cooking fuel"
```



**Step 2.** After applying the household sampling weight, calculate the percentage of households that use solid fuels for cooking using the `dw_solidfuel` variable. Repeat using the residence variable (`ahtype`) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab dw_solidfuel
svy: tab dw_solidfuel ahtype, col
```

## 9.2 Mean number of de jure household members per sleeping room

This indicator is the mean number of de jure households per sleeping room in the household's dwelling.

### Definitions

Numerator	Number of de jure household members per sleeping room across all surveyed households
Denominator	Number of surveyed households
Unit of measure	Number
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Household size cannot have any missing values. If the number of sleeping rooms is missing, it should be replaced with 1, assuming that every household has at least one room in which to sleep.
Survey variables used	<code>v204</code> , <code>hhea</code> , <code>wgt_hh</code> , <code>strata</code> , <code>ahtype</code>
Analytic variables used	<code>hhsizes_dj</code>
Analytic variables created	<code>roomsleep</code> , <code>memsleep_dj</code>

### Calculations

**Step 1.** Create a variable that indicates the number of rooms in each household's dwelling that are used for sleeping (`roomsleep`). If the number of sleeping rooms in the household-level data file is missing or equal to 0, set the number of sleeping rooms to be 1, assuming that every household has at least one room in which to sleep.

```
Set          roomsleep=v204 if v204<96
Replace      roomsleep=1 if v204=0 or v204=missing
```

**Step 2.** Create a variable that indicates the number of de jure household members per sleeping room (`memsleep_dj`).

```
Set          memsleep_dj=0
Replace      memsleep_dj=(hhsizes_dj÷roomsleep)
Label variable "Number of de jure HH members per sleeping room"
```

**Step 3.** After applying the household sampling weight, calculate the mean value of the `memsleep_dj` variable. Repeat using the residence variable (`ahtype`) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wt_hh], strata(strata)
svy: mean memsleep_dj
svy: mean memsleep_dj, over(ahtype)
```

### 9.3 Percent distribution of households by dwelling roof materials

This indicator is the percent distribution of households by specified categories of the main dwelling roof materials: (1) natural roofing (no roof, thatch, and sod); (2) rudimentary roofing (bamboo, wood planks, and cardboard); (3) finished roofing (metal, wood, ceramic tiles, cement, calamine or cement fiber, and roofing shingles); and (4) other.

#### Definitions

Numerator	Number of households in each specified category of main dwelling roof materials
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Information on dwelling characteristics is recorded based on interviewer's observations; therefore, the dwelling roof variable cannot have missing values. Dwelling characteristics classified as "other" are reported in a separate "other" category.
Survey variables used	v201, hhea, wt_hh, strata, ahtype
Analytic variables used	n/a
Analytic variables created	dw_roof

#### Calculations

**Step 1.** Create a variable that categorizes each household's main dwelling roof material as natural, rudimentary, finished, or other (**dw\_roof**). Note that the types of roof materials vary by country, so be sure to adapt the template syntax as needed to include all response options in the household data file.

```
Set          dw_roof=1 if v201=11,12,or 13
Replace      dw_roof=2 if v201=14,21,or 22
Replace      dw_roof=3 if v201=31,32,33,34,35,or 36
Replace      dw_roof=4 if v201=96
Label values 1 "Natural, including none"
              2 "Rudimentary"
              3 "Finished"
              4 "Other"
Label variable "Roof material of HH's dwelling"
```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that have roofs made of natural, rudimentary, and finished materials on their dwellings using the **dw\_roof** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab dw_roof
svy: tab dw_roof ahtype, col
```

## 9.4 Percent distribution of households by dwelling exterior wall materials

This indicator is the percent distribution of households by specified categories of main exterior dwelling wall materials: (1) natural wall materials (no walls, dirt, or cane, palm, or tree trunks); (2) rudimentary wall materials (bamboo with mud, stone with mud, cardboard, reused wood, plywood, and unbaked bricks); (3) finished wall materials (wood planks or shingles, unbaked bricks covered with plaster, bricks, cement blocks, cement, and stone with lime or cement); and (4) other.

### Definitions

Numerator	Number of households in each specified category of main exterior wall materials
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Information on dwelling characteristics is recorded based on interviewer's observations; therefore, the dwelling walls variable cannot have missing values. Dwelling characteristics classified as "other" are reported in a separate "other" category.
Survey variables used	<b>v203, hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	n/a
Analytic variables created	<b>dw_wall</b>

### Calculations

**Step 1.** Create a variable that categorizes each household's main wall material as natural, rudimentary, finished, or other (**dw\_wall**). Note that the types of wall materials vary by country, so be sure to adapt the template syntax as needed to capture all response options in the household data file.

```
Set          dw_wall=1 if v203=11,12,or 13
Replace      dw_wall=2 if v203=14,15,21,22,23,or 24
Replace      dw_wall=3 if v203=31,32,33,34,35,or 36
Replace      dw_wall=4 if v203=96
Label values 1 "Natural, including none"
              2 "Rudimentary"
              3 "Finished"
              4 "Other"
Label variable "Wall material of HH's dwelling"
```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that have walls made of natural, rudimentary, and finished materials on their dwellings using the **dw\_wall** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab dw_wall
svy: tab dw_wall ahtype, col
```

## 9.5 Percent distribution of households by dwelling floor materials

This indicator is the percent distribution of households by specified categories of dwelling floor materials: (1) natural floor materials (earth or sand, dung and palm leaves); (2) rudimentary floor materials (wood planks and bamboo slats); (3) finished floor materials (parquet or polished wood, vinyl, or asphalt strips; wall-to-wall carpet; ceramic tiles; and cement); and (4) other.

### Definitions

Numerator	Number of households in each specified category of floor materials
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Not applicable. Information on dwelling characteristics is recorded based on interviewer's observations; therefore, the dwelling floor variable cannot have missing values. Dwelling characteristics classified as "other" are reported in a separate "other" category.
Survey variables used	v202, hhea, wgt_hh, strata, ahtype
Analytic variables used	n/a
Analytic variables created	dw_floor

### Calculations

**Step 1.** Create a variable that categorizes each household's main floor material as natural, rudimentary, finished, or other (**dw\_floor**). Note that the types of floor materials vary by country, so be sure to adapt the template syntax as needed to capture all response options in the household data file.

```
Set          dw_floor=1 if v202=11,12,or 13
Replace      dw_floor=2 if v202=21 or 22
Replace      dw_floor=3 if v202=31,32,33,34,or 35
Replace      dw_floor=4 if v202=96
Label values 1 "Natural, including none"
              2 "Rudimentary"
              3 "Finished"
              4 "Other"
Label variable "Floor material of HH's dwelling"
```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that have floors made of natural, rudimentary, and finished materials in their dwellings using the **dw\_floor** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab dw_floor
svy: tab dw_floor ahtype, col
```

## 9.6 Percent of households that have electricity

This indicator is the percentage of households that have electricity.

### Definitions

Numerator	Number of households that have electricity
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Residence (urban, rural)
Treatment of missing data	Households that are missing information for <b>v222a</b> in the household data file are excluded from the numerator and denominator.
Survey variables used	<b>v222a, hhea, wgt_hh, strata, ahtype</b>
Analytic variables used	n/a
Analytic variables created	<b>dw_elec</b>

### Calculations

**Step 1.** Create a binary variable that flags households that have electricity (**dw\_elec**).

```
Set          dw_elec=missing
Set          dw_elec=1 if v222a=1
Set          dw_elec=0 if v222a=2
Label values 0 "No"
              1 "Yes"
Label variable "HH has electricity"
```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that have electricity using the **dw\_elec** variable. Repeat using the residence variable (**ahtype**) to obtain results disaggregated for households in urban and rural areas. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab dw_elec
svy: tab dw_elec ahtype, col
```

## 9.7 References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

## 10. THE CONSUMPTION AGGREGATE AND THE POVERTY INDICATORS

In this chapter, the Feed the Future poverty indicators are defined. In addition, their calculation is described when using consumption expenditure data collected in Module 8 of the core P2-ZOI Midline Parallel Survey questionnaire. This methodology, which was used to calculate the poverty indicators at baseline, is used at midline to validate the midline poverty indicator estimates calculated using survey-to-survey imputation.<sup>42</sup>

This chapter is divided into two sections: the first provides guidelines to construct the indicators, and the second provides the step-by-step procedures to calculate the indicators.

### 10.1 Guidelines to construct the consumption aggregate and the poverty indicators

This section includes five subsections. The first subsection provides an overview of the consumption aggregate, which is the basis to calculate expenditure-based poverty indicators, the second subsection provides information on handling missing values and outliers, the third subsection provides information on analyzing the data used to calculate the consumption aggregate, the fourth subsection provides information on calculating the consumption aggregate, and the final subsection provides information on how to construct the Feed the Future poverty indicators using the consumption aggregate.

#### 10.1.1 Overview of the consumption aggregate

Household welfare measures, such as poverty, are commonly based on a nominal consumption aggregate calculated using data on household expenditures. Poverty is more commonly estimated based on household consumption expenditures, rather than on income, because income is more difficult to measure, particularly in poor agrarian economies and in urban economies with large informal sectors. Income may be seasonal and erratic, making it difficult to estimate, particularly for agricultural households in which income might not be monetized. Households may also underreport (e.g., if people are engaged in more than one activity, they often forget to report income from all sources). Consumption-based metrics are more closely related to individual well-being in the sense of having enough to meet current basic needs, and households have strategies to smooth out consumption, making consumption less likely than income to vary from month to month.

Following guidelines developed by Deaton & Zaidi (2002) and Grosh & Muñoz (1996), a *consumption aggregate* is constructed by aggregating the total monetary value of the goods and services consumed by each household during a period of reference. In this approach, every household member is assumed to have an equal share of total consumption, regardless of age and other household member characteristics.

In the parallel survey questionnaire, consumption expenditures are grouped into four categories in seven submodules, based on characteristics and likely frequency of purchase or recall periods:

---

<sup>42</sup> Guidance, step-by-step instructions, and template syntax are forthcoming in an update to this edition of the *Guide to Feed the Future Statistics for P2-ZOI Midline Surveys*.

- Food consumption over the past 7 days
- Non-food expenditures over the past 7 days
- Non-food expenditures in the past 30 days
- Non-food expenditures in the past 3 months
- Non-food expenditures in the past 12 months
- Consumer durables
- Housing

To compute the consumption aggregate, all purchased and non-purchased items consumed must be converted to a daily, per capita measure. Per capita measures are computed using the number of de jure household members. There are some general procedures that apply to all categories, such as the treatment of outliers and missing values, and some considerations that are specific to each of the four categories.

### **10.1.2 Handling missing values and outliers**

The consumption expenditure dataset should be reviewed for missing values and potential outliers, or extreme values, even if standard consistency checks and verifications were in place during data collection and performed on the overall dataset. Even with a carefully executed interview, there can be missing observations and outliers remaining in the dataset. Respondents may not know or not recall the quantity or value of an item that has been consumed by the household. In addition, they may report a quantity or value that is flagged as an outlier during the analysis stage because it stands out from the overall pattern of observations in the dataset. These missing observations and outliers need to be replaced with an imputed value, rather than coded as missing, because we need to comprehensively capture and value everything consumed by the household.

The same general procedure is used to replace both outliers and missing values. In the following sections, we first describe how to impute a value to replace a missing observation or an outlier, and then provide general guidelines for detecting and verifying outliers.

Analysts need to carefully document all cases of missing or outlying values, the protocol that was followed to detect outliers, and how the replacement value was imputed. The frequency and geographical distribution of missing values and outliers should be noted, because they may reveal issues with data collection or data entry, indicating that further investigation and possible data correction is necessary. In addition, analysts must keep both the raw dataset as well as the “cleaned” dataset, because another user may want to treat missing and extreme values differently. In all cases, detailed, complete, and accessible documentation is essential.

#### **10.1.2.1 Procedures for imputing values**

If an observation is missing or determined to be an outlier (see Section 10.1.2.2, Treatment of outliers), it should be replaced (imputed) by the median<sup>43</sup> value within the smallest geographic unit that has at

---

<sup>43</sup> The median and not the mean should be used, because the mean is affected by extreme values. The median should be calculated over the per capita values available in the smallest geographical unit possible. Imputation should be done by multiplying the per capita median value by the number of de jure household members in the household for which the reported value is found to be an outlier or is missing.

least five observations, starting with the cluster. The median and not the mean should be used, because the mean is affected by extreme values.

The steps to do this are as follows:

- Identify and rank the administrative units available in the survey dataset that can be used as locational boundaries of sampled households, from the smallest (cluster) to the largest (the boundaries of the sample as a whole).
- Starting with the smallest local administrative unit (cluster), if there are five or more observations, calculate the median value at that level.
- If the total number of valid observations is less than five at the cluster level, but greater than or equal to five at the next smallest local administrative unit level (e.g., administration level 2), calculate the median at that level.
- If the total number of valid observations is less than five at that local administrative unit level, but greater than or equal to five at the next local administrative unit level (e.g., administration level 1), calculate the median at that level.
- If the total number of valid observations at that administrative level is less than five but greater than or equal to five at the sample (ZOI) level, calculate the median at the sample level.
- Multiply the per capita median value by the number of de jure household members in the household for which the observation is missing or determined to be an outlier and impute the result.

If the total number of observations in the entire sample is less than five, all values should be reviewed together to determine if these values are plausible. If they are plausible, calculate the median using the available values at the sample level. If they are not plausible, the median should not be calculated, and the observation (that is either missing or confirmed to be an outlier) should likely be dropped from the analysis.

It may be useful to also compare a median value calculated at the cluster level to the median value of the same item in other clusters, to make sure it is not markedly different. If the median value seems too large or too small compared to similar administrative units, it should be checked for plausibility. If it does not seem plausible, all observations should be reviewed carefully for outliers and implausible values, and the median should be calculated at the next administrative level.

#### **10.1.2.2 Treatment of outliers**

The consumption module data should be checked for outliers for each individual item and across all items at the household and per capita levels, because a value could fall within the normal range as a total but not per capita.

The general rule adopted here is that if a value is more than three standard deviations (SD) from the mean, it should be flagged as a potential outlier and further examined for plausibility. We check for plausibility because an extreme value may be legitimate depending on the circumstances (e.g., X and Y both bought mobile phones, but X spent \$500 on a smart phone and Y spent just \$30 on a basic flip phone). There is no established protocol for further examination, so analysts should develop their own protocol and document every step carefully. Other questions from the survey can be used to triangulate



a particular response or provide context. For instance, at the end of the Sub-module 8.1, the respondent is asked about non-members of the household eating meals in the household (questions V8108-V8111). This series of questions could indicate that larger than average food expenditures (and certain non-food expenditures as well) are plausible because of guests present in the household during the recall period. Note that even if a large value can be explained by particular circumstances, it may be adequate to treat it as an outlier, because the extreme value could have a large impact on the survey statistics.

If, after further examination, an observation is determined to be an outlier, it should be replaced with an imputed value—that is, the median calculated at the smallest administrative unit with enough valid observations, as described in Section 10.1.2.1.

### **10.1.3 Analyzing data comprising the consumption aggregate**

In this section, specific procedures for analyzing the data for each of the four expenditure categories used to compute the consumption aggregate (i.e., food, non-food, durables, and housing) are described.

#### **10.1.3.1 Food consumption**

The food item sub-module records consumption of a comprehensive list of country-specific foods. In this sub-module, both food volume and monetary values are recorded. The interviewer asks whether the item was consumed and, if so, what quantity the household consumed in total in the past 7 days. The interviewer then asks how much of the total quantity was purchased, how much came from home production, and how much was obtained as a gift or from other sources. For each source, the respondent is asked to give a corresponding value. The analyst should do the following across all items:

1. Verify that all quantities reported for a single food item are in a common unit (v8103b, v8104b, v8106b, and v8107b). If different units are reported, depending on the source, convert them to a common one.
2. Verify that the sum of the quantities consumed from various sources equals the total quantity consumed ( $v8103a = v8104a + v8106a + v8107a$ ).<sup>44</sup> If the sum is smaller than the total quantity consumed, v8103a is set equal to the sum. If the sum is larger, v8103a is compared to the quantity from each source and then one or more of the quantities is adjusted following a flagging protocol.
3. Compute the unit value for an item by dividing the amount spent on purchases by the quantity consumed from purchases (v8105 divided by v8104a), if they both have positive values. If either one is 0 or missing, any other source can be used in which both value and quantity are reported (quantity and value of own production consumed or quantity and value consumed from gifts or other sources).

#### ***Missing food item observations***

It is assumed that every household must have consumed some food in the 7 days preceding the survey. If there are no data on food consumption for a household, either because the respondent says 'no' or

---

<sup>44</sup> If computer-assisted personal interviewing is used, the program will automatically flag any inconsistency, and the problem should be resolved during data collection. It is useful, however, to also check this during data analysis.

‘don’t know’ to all food items listed in the questionnaire, the household is first dropped from the working dataset and then brought back at the end of the food consumption module analysis. A per capita daily (PCD) food consumption value is then imputed for the household, using the median value calculated at the lowest administrative unit possible, as described in Section 10.1.2.1.

### ***Additional testing for extreme values in the food consumption sub-module***

After missing observations and confirmed outliers have been replaced with imputed values as appropriate, the sample distribution should look relatively normal. It may be useful, however, to disaggregate the data to see if common expectations are met. For instance, we would not expect households to consume food obtained mainly from “gifts and other sources,” unless the sample population has been impacted by disaster or other hardships and food assistance is widely present. We would expect rural farm households to consume a significant proportion of their staple food crops from home production. We should also observe that, in general, the ratio of food purchased to food not purchased goes up as income goes up. If these expectations are not met, the data are not necessarily faulty, but further investigation is recommended.

#### **10.1.3.2 Non-food, non-durable goods**

Consumption expenditure data for non-food items are collected for four different recall periods: 7 days, 1 month, 3 months, and 12 months. Items such as charcoal or kerosene, candles, and public transportation are regular purchases and are better recalled on a weekly basis. Soap and cosmetics, petrol, wages to servants, utilities, and costs related to cell phones (e.g., monthly bills or airtime cards), which are also purchased regularly but less frequently, have a recall period of 1 month. Items purchased even less frequently, such as clothing, footwear, or cooking utensils, have a recall period of 3 months. Finally, purchases of goods such as furniture or school expenditures are infrequent events and thus are collected on a 12-month recall basis. Because of these different time periods, calculating total consumption of non-food items first requires converting the values reported to a common reference period.

We include all non-food items that contribute to household consumption or enhance household welfare. Expenditures on items such as clothing, footwear, beauty, or recreation increase household consumption and welfare and therefore are included in the consumption aggregate. However, expenditures on items such as debt payment, fines, or legal fees do not lead to higher consumption and welfare and thus are excluded. Expenditures that are made as productive investments or as part of household production activities, such as purchase of a bus, fertilizers, or agricultural seeds, are also excluded because they do not improve household consumption at present. They indirectly impact consumption when they result in increased home production consumed by the household or increased non-food expenses through higher income.

Finally, following the recommendations made by Deaton & Zaidi (2002), we exclude infrequent and lumpy expenditures, such as those for births and funerals, marriage ceremonies, dowries, and night lodging in a rest house or hotel, because including them would overestimate a household’s true level of consumption. Many of these lumpy, occasional expenditures are included in the questionnaire because collecting these separately helps with recall during the interview and ensures that these expenditures are not inadvertently bundled into others. These are as follows:

- Health expenditures (items 8228 to 8236 and 8304 to 8309)
- Night lodging in a rest house or hotel (item 8275)
- Fines, legal costs, marriage and funeral costs (items 8299 to 8303)

For items that could be gathered in the wild, such as wood poles, bamboo, or grass for thatching roofs (items 8319 to 8321), the interviewer asks the respondent to estimate the total quantity used in the past 12 months and then to estimate the value for any amount gathered.

### 10.1.3.3 Consumer durables

Purchases of durable goods (or household assets) represent large and relatively infrequent expenses. Almost all households incur these expenditures at some point in time, but only a small proportion of households would have made such expenditures during the preceding 12 months. Deaton & Zaidi (2002) argue that the value of services that the household receives from durable goods is a more appropriate measure of consumption than the price of the good, either contemporaneous or at the time of purchase. Hence, consumption of a durable good is calculated as the “user cost” or “annual rental equivalent” of owning the item and is approximated by multiplying the value of the item in its current shape by the sum of the real interest rate and the depreciation rate:

$$R^j = P_t^j(r_t + \delta^j)$$

Where:  $P_t^j$  is the current value of the item (j),  $r_t$  is the real rate of interest, and  $\delta^j$  is the depreciation rate for the durable good. Each of these components is computed separately as follows:

- The current value of the item ( $P_t^j$ ): This is the value of the item in its current shape (second-hand) (question V8705). If the household owns more than one item (in question V8703), the respondent is asked to report an average value.
- The real rate of interest ( $r_t$ ): In theory, this should be the specific real interest rate calculated for each durable good. In practice, a single average real rate of interest is used for all goods (see Deaton & Zaidi, 2002). Data on real interest rates by country are available from the World Bank Databank<sup>45</sup> and should be averaged over as many uninterrupted years as possible to minimize the effect of large fluctuations or any distortion in interest rates. However, real interest rates are not available for all countries. For the Feed the Future target countries,<sup>46</sup> time series are available for all but three countries.<sup>47</sup> If a continuous real interest rate series is not available for at least 10 years (and up to at least 2 years of the survey year), nominal interest rates and inflation rates published by the country’s Central Bank can be used. The average inflation rate

<sup>45</sup> Available at: <http://databank.worldbank.org/data/home.aspx>, under the World Development Indicators database and Real Interest Rate (%) series

<sup>46</sup> Bangladesh, Democratic Republic of Congo, Ethiopia, Ghana, Guatemala, Honduras, Kenya, Liberia, Madagascar, Malawi, Mali, Mozambique, Nepal, Niger, Nigeria, Rwanda, Senegal, Tanzania, Uganda, and Zambia

<sup>47</sup> At least 20 years of data are available for Bangladesh, Guatemala, Honduras, Kenya, Madagascar, Malawi, Mozambique, Nigeria, Rwanda, Tanzania, Uganda, and Zambia. There are more than 10 years of data available for the Democratic Republic of Congo. Data are available for Liberia from 2001 to 2017, and data are available for Mali, Niger, and Senegal from 2005 to 2017. There are no data for Ghana, Ethiopia, or Nepal.

(using the Consumer Price Index [CPI]) for each year should be subtracted from the nominal interest rate ( $I_t$ ) and then averaged over as many years as possible:

$$r_t = I_t - CPI_t$$

- The rate of depreciation ( $\delta^i$ ) is calculated for each item<sup>48</sup> as follows:

$$\delta^j = 1 - \left( \frac{P_t^j}{P_{t-T}^j} \right)^{\frac{1}{T}}$$

Where  $P_t^j$  is the current average price (average value as reported by the respondent) of the item,  $P_{t-T}^j$  is the average price of the item when purchased, and  $T$  is the average age of the item in years.<sup>49</sup>

Note that although some goods can gain value over time, such as some pieces of art or antiques, leading to a negative depreciation rate, this is unlikely to occur for household durable goods. If the data show that the current value exceeds the original price for an item, it is most likely due to inaccurate reporting of the selling (second-hand) value, the age of the goods, or the original price. Negative values as well as extreme values (outliers), after being examined for plausibility, should be replaced by the durable-specific median value at the smallest geographic unit possible (see Sections 10.1.1.2 and 10.1.1.3). In general, depreciation rates between 10 and 25 percent are considered reasonable.

An average annual rental equivalent can then be estimated for each durable good and multiplied by the number of units ( $S^j$ ) owned by the household, if more than one, as reported in question V8703.

The total value of annual consumption of durable goods ( $A$ ) consumed by household ( $i$ ) is calculated as follows:

$$A_i = \sum_{j=1}^n R^j \times S^j$$

The PCD rental equivalent of durable goods consumed by the household is calculated by dividing the annual value by 365 days and by the number of de jure household members.

#### 10.1.3.4 Housing

Housing is an important component of the total welfare of households and should be included in the estimation of the consumption aggregate. We are interested in measuring the flow of services accruing to the household from occupying the dwelling and not the expenditure for buying the house, which

<sup>48</sup> In this sub-module, “item” refers to a line in the questionnaire, because in some cases, there could be more than one good bundled together.

<sup>49</sup> If the household owns more than one item of the durable goods, the average value and average age should be used.

should be treated as an investment. However, imputing the value of housing services is not straightforward,<sup>50</sup> especially if the housing market is not well developed in the areas where the survey is implemented.

If the household pays a rent for its dwelling, rent payment (question V8605) is a good approximation and should be used as the consumption value from housing for these households. If the household either owns its dwelling or lives in it for free, a “rental equivalent” needs to be estimated.

Respondents who do not rent their dwelling are asked to provide an estimate of how much monthly rent could be charged, if the house were to be rented today (question V8604). This rent estimate should be checked for plausibility,<sup>51</sup> because it will be used to impute a value for housing consumption.

Questions V8606 to V8608 are directed to homeowners only, who are asked whether their household pays a mortgage on the house and, if so, how much and with what frequency. These questions should not be used as a proxy for housing services, because the mortgage depends on the terms of the contract, including the length of the contract and the interest rate charged. These questions are merely asked to help with the recall of all expenditures and avoid some expenses to be inadvertently included with others.

Using reported actual monthly rent paid (from renters) and estimated monthly rental value of dwellings provided by employers or occupied for free by the household, a hedonic regression model can be used to estimate a rental equivalent for households that are not reporting actual or estimated rent.<sup>52</sup> This model is developed by regressing available rental value data on a series of dwelling characteristics and then using the resulting equation to estimate a rental equivalent for the non-renting households that did not provide an estimate of rental value. A log-linear functional form is commonly used and often performs better than a linear form, although for some datasets, other functional forms could provide better predictability.<sup>53</sup>

$$\ln(R_i) = \beta_0 + \beta_i X_i + \varepsilon_i$$

Where:

$R_i$  is the rent paid or rent estimate of housing unit  $i$

$X_i$  is a set of characteristics or attributes of housing unit  $i$

$\beta_0$  and  $\beta_i$  are the coefficients to estimate in the hedonic model

$\varepsilon_i$  is the error term

---

<sup>50</sup> For a review of approaches in measuring the value of housing for the welfare aggregate, see: Balcazar, et al., 2014.

<sup>51</sup> In regions where there is an active rental market, respondents may have a reasonable idea of the rental value of their house. However, it is not always the case, and some evidence has shown that people tend to have an inflated estimation of the value of their home, so self-assessment of rental value should be carefully examined.

<sup>52</sup> If there are insufficient rental data to use hedonic regression, please contact the United States Agency for International Development to discuss and agree on an alternate approach.

<sup>53</sup> Other functional forms can be tested, such as a log-log function.

There are several characteristics or attributes (the set of independent variables  $X_i$ ) of the dwelling that can be fitted in the equation, such as the following:

- Structural attributes: material of the exterior walls (v203), floors (v202), and roof (v201); number of rooms for sleeping (v204<sup>54</sup>); type of toilet (v208); source of drinking water (v211); access to electricity (v222a)
- Location: district (c05) or region (c06), rural or urban (ah`type`)<sup>55</sup>
- If available: distance from an improved road, distance from a market<sup>56</sup>

Some of the attributes are categorical and will be coded as binary or dummy variables. A stepwise regression procedure can be used to find the best fit, with the objective of maximizing the  $R^2$ . Multi-collinearity is likely to be present, but because the purpose is imputing rent where it is missing (predicting the dependent variable), the contribution or significance of individual variables should not be much of a concern.

The equation with the estimated regression coefficients ( $\beta_i$ ) can be applied to the characteristics of non-renting households to impute their rent equivalent. These imputed rent equivalents should be examined for plausibility.

Question V8609 asks about expenditures on repairs and maintenance in the past month. These should be added to the imputed, estimated, or actual monthly rental values to come up with total monthly expenditures on housing for the household. However, repair and maintenance expenditures should first be examined to determine whether they could include large, lumpy expenses, such as renovation costs, that would otherwise be infrequent. In such a case, it may be better to replace the reported value with the median value of monthly expenditures on repairs and maintenance of the other households within the smallest geographical unit (see Sections 10.1.1.2 and 10.1.1.3).

#### 10.1.4 Computing the consumption aggregate

The PCD consumption aggregate is obtained by adding the PCD consumption expenditures across all food and non-food categories to obtain a total PCD consumption expenditure.

$$X_{pc_i} = \sum_{j=1}^m x_{pc_{ij}}$$

Where:

$x_{pc_{ij}}$  is the PCD consumption expenditure on item category  $j$  by household  $i$

$X_{pc_i}$  is the PCD consumption aggregate of household  $i$

<sup>54</sup> Note that v204 asked for the number of sleeping rooms, which is not the same as the total number of rooms in the house and may be more or less correlated with rent.

<sup>55</sup> This is not included in survey questionnaire; it is added during the post-processing stage using the country-specific definition of urban and rural.

<sup>56</sup> These observations are not available from the standard ZOI Midline Survey dataset but could potentially be obtained from secondary geographic information system sources.

### 10.1.5 Poverty indicators

After the PCD consumption aggregate is computed, the following poverty indicators used by Feed the Future can be estimated:

- Prevalence of poverty: Percent of people living on less than the international poverty line
- Depth of poverty of the poor: Mean percent shortfall of the poor, relative to the international poverty line
- Percent of people who are 'near-poor': living on 100 percent to 125 percent of the international poverty line

The official Feed the Future phase two poverty indicators are calculated using the \$1.90 per day at 2011 purchasing power parity (PPP) poverty line. However, in the midline indicator assessment reports, the poverty indicators measured at each of the three most recent international poverty lines (\$1.25 per day at 2005 PPP, \$1.90 per day at 2011 PPP, and \$2.15 per day at 2017 PPP) are presented.

In 1990, researchers in the World Bank and a group of independent researchers introduced the first dollar-a-day international poverty line measured in common currency by using PPP exchange rates. The idea is to use the same yardstick to measure poverty across countries. The international poverty line is calculated based on a methodology developed by the World Bank (Ravallion et al., 1991). It provides a population-level per capita consumption threshold for extreme poverty. Expressed in international dollars, using PPP conversion factors,<sup>57</sup> it provides a straightforward metric that can be used to compare prevalence of extreme poverty across countries. Since 1990, the international poverty line has been revised to \$1.25 at 2005 PPP, \$1.90 at 2011 PPP, and recently \$2.15 at 2017 PPP; that is, the methodology has been revised three times since it was first introduced. Revisions of the poverty line align with updates of the PPP global estimates by the International Comparison Program.

Sections 10.1.2.1-10.1.2.3 present how to calculate the three Feed the Future poverty indicators using the \$1.90 at 2011 PPP prices poverty line to facilitate the demonstration, but the same process applies with the other international poverty lines using the correct PPP conversion factor and CPI<sup>58</sup> to convert the poverty line expressed in international dollars into local currency of the year and month of the survey. The PPP and CPI should align with the year of the poverty line used (2005 CPI for the \$1.25 2005 PPP, 2010 CPI for the \$1.90 2011 PPP, and 2017 CPI for the \$2.15 2017 PPP).

---

<sup>57</sup> The PPP conversion factor is a spatial price deflator and currency converter that controls for price level differences between countries. It is used, among other things, to make spatial comparisons of price levels.

<sup>58</sup> CPIs measure changes over time in the general level of prices of goods and services that households acquire (use or pay for) for the purpose of consumption. CPIs are calculated as weighted averages of the percentage of price changes for a basket of goods and services of consumer products, with the weights reflecting their relative importance in household consumption during a specified period. CPI are usually calculated on the basis of periodic surveys of consumer prices and are expressed relative to a fixed base year. CPIs for Feed the Future target countries are available from the International Monetary Fund, International Financial Statistics website (Prices, Production, and Labor – IMF Data: <https://data.imf.org/regular.aspx?key=63087884>), or from individual countries' national bank websites or national statistics offices. The analyst should use official websites to get appropriate CPI measures. The monthly disaggregated CPI for the survey year is needed to deflate prices and conversion of PPP values in LCU.

PPP conversion factors for private consumption for all countries can be in the World Bank's Databank:

- 2017: <https://databank.worldbank.org/source/world-development-indicators>
- 2011: [https://databank.worldbank.org/source/international-comparison-program-\(icp\)-2011](https://databank.worldbank.org/source/international-comparison-program-(icp)-2011)
- 2005: [https://databank.worldbank.org/source/international-comparison-program-\(icp\)-2005](https://databank.worldbank.org/source/international-comparison-program-(icp)-2005)

For convenience, the 2005, 2011, and 2017 PPP conversion factors for the 20 Feed the Future target countries are provided **Table 6**.

**Table 6: PPP Conversion Factors (Local Currency Unit per International \$) for Private Consumption for Feed the Future Target Countries**

Country name	PPP 2005 <sup>a</sup>	PPP 2011 <sup>b</sup>	PPP 2017 <sup>c</sup>
Bangladesh	25.494	24.849	29.514
Democratic Republic of Congo	316.232	537.732	630.606
Ethiopia	2.751	5.439	8.496
Ghana <sup>d</sup>	4,475.822	0.788	1.751
Guatemala	4.54	3.873	4.403
Honduras	9.66	10.080	10.839
Kenya	32.684	35.430	41.635
Liberia	0.511	0.568	0.426
Madagascar	756.381	704.913	962.960
Malawi	56.922	78.017	241.931
Mali	289.679	221.868	205.273
Mozambique <sup>d</sup>	11,625.693	15.527	21.988
Nepal	26.467	25.759	30.513
Nigeria	78.583	79.531	112.098
Niger	267.331	228.753	245.160
Rwanda	236.745	246.834	293.705
Senegal	298.245	246.107	238.578
Tanzania	482.451	585.520	754.621
Uganda	744.618	946.890	1,221.088
Zambia <sup>e</sup>	2,830.326	2,505.341	4.224

<sup>a</sup> Data from World Development Indicators database: <https://databank.worldbank.org/source/world-development-indicators>

<sup>b</sup> Data from International Comparison Program 2011 databank: [https://databank.worldbank.org/source/international-comparison-program-\(icp\)-2011](https://databank.worldbank.org/source/international-comparison-program-(icp)-2011), using Global PPPs as Classification Name (Classification Code=PPPGlob); and Individual Consumption Expenditure by Households as Series Name: (Series Code: 110000)

<sup>c</sup> Data from International Comparison Program 2005 databank: [https://databank.worldbank.org/source/international-comparison-program-\(icp\)-2005](https://databank.worldbank.org/source/international-comparison-program-(icp)-2005), using Global PPPs as Classification Name (Classification Code=PPPGlob); and Individual Consumption Expenditure by Households as Series Name: (Series Code: 110000)

<sup>d</sup> Redenominated their currency between 2005 and 2011, which resulted in a dramatic drop in their PPP; taken from the July 2016 [Feed the Future Indicator Handbook Definition Sheets \(agrilinks.org\)](#)

<sup>e</sup> Redenominated their currency between 2011 and 2017, which resulted in a dramatic drop in their PPP; taken from the July 2016 [Feed the Future Indicator Handbook Definition Sheets \(agrilinks.org\)](#)



### 10.1.5.1 Prevalence of poverty: Percent of people living on less than the international poverty line in PPP value

The prevalence of poverty, or poverty headcount ratio, is the proportion of the population in the survey area living below the international poverty line. The steps to calculate the prevalence of poverty indicator are as follows. The \$1.90 at 2011 PPP prices poverty line is used, but the same process applies if another international poverty line is used.

1. Convert the \$1.90 poverty line into the local currency unit (LCU) using the 2011 PPP conversion factor for private consumption based on the 2011 International Comparison Program data.<sup>59</sup> See **Table 6** for PPP 2011 conversion factors for Feed the Future target countries.
2. Adjust the poverty line in LCU for inflation from 2011 to the year and month of the survey.<sup>60</sup> In all cases, the official source for the CPI should be used. The formula to adjust the poverty line for inflation is as follows:

$$(\$1.90)_{LCU_t} = 1.90 \times (PPP\ 2011) \times \left( \frac{CPI_t}{CPI_{2011}} \right)$$

Where the subscript  $t$  refers to the month and year of the ZOI Survey.

3. Calculate the prevalence of poverty:
  - Calculate the numerator as the sum of the sample-weighted number of de jure household members in sampled households in which the PCD consumption is less than the poverty line (less than  $(\$1.90)_{LCU_t}$ ).
  - Calculate the denominator as the sum of sample-weighted number of de jure household members in the sampled households with consumption data.
  - Multiply the ratio by 100 to obtain a percentage.

$$PrevPov = \left( \frac{1}{\left( \sum_{i=1}^n H_i * wt_i \right)} \left( \sum_{i=1}^n I_i (H_i * wt_i) \right) \right) * 100$$

Where:

$H_i$  is the number of de jure household members in household  $i$

$wt_i$  is the sample weight of household  $i$

$n$  is the number of households in the sample with consumption data

$I_i$  is a parameter that takes on the following values:

$$I_i = 1 \text{ if } X_{pci} < (\$1.90)_{LCU_t}$$

$$I_i = 0 \text{ if } X_{pci} \geq (\$1.90)_{LCU_t}$$

<sup>59</sup> [https://databank.worldbank.org/source/international-comparison-program-\(icp\)-2011](https://databank.worldbank.org/source/international-comparison-program-(icp)-2011)

<sup>60</sup> Because ZOI Surveys span multiple months, use the CPI for the month in which the bulk of data collection took place when adjusting the poverty line. If multiple sources are explored, and the CPI for this month is not available, determine an alternate approach (e.g., taking the geometric mean of the product of adjacent CPIs to estimate the unavailable CPI or using a quarterly CPI).

### 10.1.5.2 Depth of poverty of the poor: Mean percent shortfall of the poor relative to the international poverty line in PPP value

The depth of poverty of the poor measures how far individuals with a daily consumption expenditure less than \$1.90 are below the \$1.90 per day (2011 PPP) poverty line. This indicator differs from the commonly used depth of poverty indicator that was used in Feed the Future phase one in that this indicator only tracks the depth of poverty of households that are below the poverty line (i.e., the poor). In other words, it **does not include all households**; households for which the per capita consumption aggregate is equal to or greater than the poverty threshold are **not** included in this calculation.

The steps to calculate the depth of poverty of the poor indicator are as follows:

1. Using the poverty line converted to LCU and adjusted for inflation  $((\$1.90)_{LCU_t})$ , as was done for the prevalence of poverty, subtract per capita expenditure in LCU of each poor household in the sample from the poverty line, and then divide by the poverty line to obtain the proportional shortfall from the poverty line of each poor household in the sample.
2. Multiply each of these proportional shortfalls by the sample-weighted number of de jure household members and then sum across all poor households.
3. Sum the sample-weighted number of de jure household members in poor households.
4. Divide (2) by (3) and multiply by 100 to obtain the depth of poverty of the poor expressed as a percentage of the \$1.90 per person per day poverty line.

$$DepthPoor = \frac{\sum_i^k \delta_i * (H_i * wt_i)}{\sum_i^k (H_i * wt_i)} * 100$$

With:

$$\delta_i = 1 - \left( \frac{X_{pc_i}}{(\$1.90)_{LCU_t}} \right)$$

Where:

$\delta_i$  is the proportional shortfall of poor household  $i$

$X_{pc_i}$  is the daily per capita consumption aggregate (in LCU) of poor household  $i$

$H_i$  is the number of de jure household members in poor household  $i$

$wt_i$  is the weight assigned to household  $i$

$(\$1.90)_{LCU_t}$  is the \$1.90 poverty threshold converted to local currency at current prices for the year and month of the survey

$k$  is the number of poor households in the sample

### 10.1.5.3 Percent of people who are ‘near-poor’, living on 100 percent to 125 percent of the international poverty line \$1.90 per day (2011 PPP)

This indicator measures the proportion of the population that is near-poor, defined as an income marginally above the poverty line—that is, between the poverty line and 1.25 times the poverty line. In

other words, in the case of a poverty line equal to \$1.90 2011 PPP, near-poor individuals are those with daily consumption expenditures equal or greater than \$1.90 but less than \$2.375 per day at 2011 PPP.

The steps to calculate this indicator are essentially the same as for the prevalence of poverty (Section 10.1.5.1). In this indicator, however, the numerator is the sum of individuals who are defined as near-poor, and the upper limit of the near-poor range (e.g., \$2.375 [2011 PPP]) also needs to be converted to LCU and adjusted for inflation for the year and month of the survey

$$PrevNearPoor = \left( \frac{1}{\sum_{i=1}^n H_i * wt_i} \right) \left( \sum_{i=1}^n I_i(H_i * wt_i) \right) * 100$$

Where:

$H_i$  is the number of de jure household members in household  $i$

$wt_i$  is the sample weight of household  $i$

$n$  is the number of households in the sample

$I_i$  is a parameter that takes on the following values:

$I_i = 1$  if  $(\$1.90)_{LCU_t} \leq X_{pc_i} < (\$2.375)_{LCU_t}$

$I_i = 0$  if  $X_{pc_i} < (\$1.90)_{LCU_t}$  or  $X_{pc_i} \geq (\$2.375)_{LCU_t}$

When calculating this indicator using the \$1.25 (2005 PPP) poverty line, the 125 percent threshold is \$1.5625, and when calculating this indicator using the \$2.15 (2017 PPP) poverty line, the 125 percent threshold is \$2.6875.

## 10.2 Step-by-step procedure to calculate the poverty indicators

This section presents the step-by-step procedures to calculate the Feed the Future ZOI-level poverty indicators. The poverty indicators require the daily per capita consumption aggregate to first be calculated. Sections 10.2.1 through 10.2.7 present the steps to calculate the components of the daily per capita consumption aggregate, and Section 10.2.8 presents the steps to calculate the daily per capita consumption aggregate in local currency at ZOI Survey year prices. This then allows for the calculation of poverty indicators using the \$1.90 per day (2011 PPP) poverty line (Sections 10.2.12 through 10.2.14), the \$1.25 per day (2005 PPP) poverty line (Sections 10.2.15 through 10.2.17), and the \$2.15 per day (2017 PPP) poverty line (Sections 10.2.18 through 10.2.20).

Before beginning to calculate the consumption aggregate, the analyst must first review three protocols and the corresponding syntax:

**Protocol a:** to identify outlier values in the dataset

**Protocol b:** to replace identified outliers with an acceptable imputed median value

**Protocol c:** to change the dataset from wide to long format

These protocols are referred to in Sections 10.2.1-10.2.8, but in each section, '**var**' is replaced with the relevant variables from the section.

There are seven Stata syntax files available in the Midline Toolkit that are used to calculate components of the consumption aggregate; these files are numbered 1-7. The eighth syntax file is used to calculate the consumption aggregate and the poverty indicators. Each syntax file contains a set of Stata commands to identify outliers in a program called **FLAG\_OUTLIERS**, which is described in **Protocol a**, and a set of Stata commands to calculate the median values of the variables in a program called **CALC\_MEDIAN**, which is described in **Protocol b**. The programs are called multiple times with different parameters in the syntax files. The programs are included in the syntax file so that the files are self-contained. The analyst must carefully document every instance of flagged potential outliers, the protocol followed for examination, and decisions made.

### 10.2.1 Protocol a: to identify outlier values in the dataset

**Step a1.** Select the threshold and local administrative unit that will be used to identify potential outliers for an item.

As a **general rule**, a value is defined as a potential outlier if it is more than three SD from the mean in the selected local administrative unit. For specific situations, however, a different SD threshold may be considered (for instance, more than two SD if the spread of data is narrow or more than five SD if the spread of data is wide). The default local administrative unit set up in the template Stata syntax files is the cluster (**hhea**), the smallest local administrative unit available in the dataset. However, another—higher—local administrative unit should be used if more records are needed for the mean and SD calculations or if it makes more sense for the variable being examined. Local administrative units relevant to a survey are found in the household identification section of the country-customized questionnaire and generally have variables that begin with “c0,” such as **c05**, usually the country’s administrative level 2 (e.g., district, ward, or circle), and **c06**, usually the country’s administrative level 1 (e.g., province or region). The core P2-ZOI midline analytic datasets include district (variable **c05**) and region (variable **c06**). Administrative levels smaller than Admin 2 may be included in the questionnaire (e.g., **c04** [county]), but they are often not exported from the computer-assisted personal interviewing system for analysts to use. Consult the household-level analytic data file or discuss with the Data Processing Manager to determine what local administrative level variables are available.

**Step a2.** Calculate the mean and SD within the selected local administrative unit. Start with cluster (**hhea**). Create a variable to flag potential ‘var’ outlier values (**out\_`var’**) that is assigned 1 if the value is less than three SD from the mean, 2 if the value is more than three SD from the mean, and 0 otherwise.

```
By hhea: Set      mean_`var'=mean(`var')
Label variable   "mean of [`var']"

By hhea: Set      sd_`var'=SD(`var')
Label variable   "standard deviation of [`var']"
Set              out_`var'=0
Replace          out_`var'=1 if `var'<(mean_`var'-(3*sd_`var'))
Replace          out_`var'=2 if `var'>(mean_`var'+(3*sd_`var'))
Label values     0 "Value within 3 SD"
                  1 "Value < -3SD"
```

```

2 "Value > +3SD"
Label variables "Outlier flag for [`var']"

```

**Note:** If a threshold of three SD results in too many or too few flagged values for an item, repeat this step after adjusting the three SD threshold. For specific situations, a different SD threshold may be considered (for instance, more than two SD if the spread of data is narrow or more than five SD if the spread of data is wide).

**Step a3.** Save the flagged potential outliers (`out_`var'>0`) and relevant administrative unit identifiers to an Excel file: "``submodule' _`var'_outliers.xlsx`" where ``submodule'` is the Module 8 sub-module (e.g., 8.1 or 8.2) and ``var'` is the variable name. Open the Excel file and examine each outlier one by one for plausibility.

**Step a4.** If a potential outlier is determined not to be an outlier after inspection, set the `out_`var'` value to 0 in the data file. For confirmed outliers, set the value of ``var'` to missing.

```

Replace      out_`var'=0 if [Include in this IF statement the
               extreme values determined to be plausible]
Replace      `var'=missing if out_`var'>0 and [Include in this
               IF statement for extreme values determined to be
               implausible]

```

### 10.2.2 Protocol b: to replace identified outliers with an acceptable imputed median value

If the total number of valid observations of an item at the selected local administrative unit level is greater than or equal to five, calculate the median at this level. However, if the total number of valid observations is less than five at the selected local administrative unit level, calculate the median at the next smallest local administrative unit level (e.g., if the number is less than five at the cluster level, calculate the median at the next level [e.g., administrative 2]). If the total number of valid observations in the full sample is less than five, do not calculate the median. In this case, all observations should be reviewed together to determine whether to include the item in the calculation and, if so, what values are plausible. If any values are implausible, set them to missing and calculate the median with the available plausible values.

**Step b1.** Identify and rank the administrative units available from the survey that can be used as geographical boundaries of sampled households, from the smallest (cluster) to the largest (the boundaries of the ZOI). In the toolkit syntax, four local administrative unit levels are used to demonstrate the procedure: cluster (`hhea`), Admin level 2 (district, `c05`), Admin level 1 (region, `c06`), and ZOI. Adjust the levels as relevant to the survey.

**Step b2.** If the total number of valid observations is greater than or equal to five at the cluster level, calculate the median at the cluster level.

```

Set          med_`var'=missing
Label variable "median of [`var']"

By hhea: Replace med_`var'=median(`var') if number of cases≥5

```

**Step b3.** If the total number of valid observations is less than five at the cluster level but greater than or equal to five at Admin 2 level, calculate the median at the Admin 2 level.

By c05: Replace med\_`var`=median(`var`) if number of cases<5 in hhea and number of cases≥5 in c05

**Step b4.** If the total number of valid observations at the Admin 2 level is less than five but greater than or equal to 5 at the Admin 1 level, calculate the median at the Admin 1 level.

By c06: Replace med\_`var`=median(`var`) if number of cases<5 in c05 and number of cases≥5 in c06

**Step b5.** If the total number of valid observations at the Admin 1 level is less than five but greater than or equal to 5 at the ZOI level, calculate the median at the ZOI level.

Replace med\_`var`=median(`var`) if number of cases<5 in c06 and number of cases≥5 in the ZOI sample

**Step b6.** If the total number of valid observations in the sample is less than five, all observations should be reviewed together to determine whether the values are plausible. If they are plausible, calculate the median with the available values.

Review `var` if number of cases <5 in the ZOI sample  
 Replace `var`=missing if value is not plausible  
 Replace med\_`var`=median(`var`)

**Step b7.** Replace all confirmed outliers of `var` by the median.

Replace `var`=med\_`var` if out\_`var`>0 and med\_`var`≠missing

### 10.2.3 Protocol c: to change the dataset from wide to long format

**Protocol c** describes how to change the level of observation in the data from household level (flat or wide format) to item level (long or rectangular format).

The original consumption data included in the household analytic data file are in flat format, meaning that one record includes all consumption data per household (see **Table 7A**). At multiple points in the following sections, the data will be converted from flat format to rectangular format, meaning that there are multiple records for each household—one for each food item (see **Table 7B**).

**Table 7A: Example of Flat Format Data: Consumption Amount**

hhea	hhnum	Food item			Quantity consumed			Unit of quantity consumed		
		v8101_001	v8101_002	v8101_003	v8103a_001	v8103a_002	v8103a_003	v8103b_001	v8103b_002	v8103b_003
101	1	Maize	Rice	Millet	100	4	0	GRAM	KG	
101	2	Maize	Rice	Millet	0	5	1		KG	KG
101	3	Maize	Rice	Millet	2	2	0	KG	GRAM	

**Table 7B: Example of Rectangular Format Data: Consumption Amount**

hhea	hhnum	Food item (v8101)	Quantity consumed (v8103a)	Unit of quantity consumed (v8103b)	Index variable (j)
101	1	Maize	100	GRAM	1
101	1	Rice	4	KG	2
101	1	Millet	0		3
101	2	Maize	0		1
101	2	Rice	5	KG	2
101	2	Millet	1	KG	3
101	3	Maize	200	KG	1
101	3	Rice	2	GRAM	2
101	3	Millet	0		3

The pseudo syntax to reshape the data from wide to long format and rename the variables to remove the ‘\_’ suffix in their names in the item-level data used through this section of the Guide is as follows:

```

Reshape    wide to long
    variable stubs    [list of variable stubs]
    index variables   [list of index variables]
    new variable      j

Label variable j      "Item number"
Rename variable      `var' _ to be `var'

```

Where:

- List of stub variables: a list of the part of the variable name that is common across the variables that will be reshaped. For example, for the variables **v8101\_001**, **v8101\_002**, and **v8101\_003**, the variable stub is **v8101\_**.
- List of index variables: a list of variables that uniquely identify the household-level (wide) observation. For the consumption data, these variables are **hhea** and **hhnum**. In addition, any variables that should be maintained in the long version of the data file for all records should be included in this list (e.g., **c05**, **c06**, and **hhsize\_dj**).
- New variable: **j** is the new index variable for the items within each household that is created when the data are reshaped.

#### 10.2.4 Food consumption expenditures of the past 1 week (7 days)

The step-by-step procedures to calculate PCD food consumption expenditures in the 1 week (7 days) preceding the ZOI Survey follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov1\_food/1w.do*. Variables from Sub-module 8.1 are used for the calculations.

**Step 1.** Prepare the data.

**Step 1a.** Load the household-level analytic data file. Drop all variables except for cluster number (**hhea**), household number (**hhnum**), geographic administrative units (**c05** and **c06**), de jure household size (**hhsize\_dj**), all sub-module 8.1 variables (variables starting with **v810**), and the

outcome variables for Module 8 (**v8100r** and **v8700r**), and drop households that did not complete Sub-module 8.1.<sup>61</sup>

```
Load          "FTF ZOI Survey [Country] [Year] household data
               analytic"
Keep variables hhea hhnum hhsize_dj v8100r v8700r c05 c06, and
               all variables starting with 'v810'
Drop          record if v8100r#1
```

**Step 1b.** Create a variable (**v8102miss**) to identify households in which none of the food items included in Sub-module 8.1 were consumed in the 7 days preceding the survey. [XXX] should be equal to the number of food items included in Sub-module 8.1; adapt the template syntax to include all food items.

```
Set          v8102miss=0
Replace      v8102miss=v8102miss+1 if v8102_001≥2 or missing
Replace      v8102miss=v8102miss+1 if v8102_002≥2 or missing
Replace      v8102miss=v8102miss+1 if v8102_003≥2 or missing
...
Replace      v8102miss=v8102miss+1 if v8102_[XXX]≥2 or missing

Label variable "Number of v8102 items with consumption not
               reported or missing"
```

If no consumption was reported or consumption information is missing across all food items in Sub-module 8.1, **v8102miss** will be equal to the number of food items in Sub-module 8.1.

**Step 1c.** Drop all households from the working dataset that reported no consumption or that are missing consumption information across all food items. These dropped households are added back into the dataset in Step 10.<sup>62</sup>

```
Drop          record if v8102miss=[XXX]
```

**Step 1d.** Change the level of observation in the data from household level to food item level, as explained in **Protocol c**. Convert the variables that start with the following “stub” from wide format (household level) to long format (item level): **v8101\_**, **v8102\_**, **v8103a\_**, **v8103b\_**, **v8104a\_**, **v8104b\_**, **v8105\_**, **v8106a\_**, **v8106b\_**, **v8106c\_**, **v8107a\_**, **v8107b\_**, and **v8107c\_**, and then remove the ‘\_’ suffix from all food item variable names.

```
Reshape      wide to long
```

---

<sup>61</sup> In most P2-ZOI Midline Surveys, there are two outcome variables for Module 8: one for Sub-module 8.1 and another for Sub-modules 8.2-8.7 because different respondents may have been administered the two pieces of the module. However, some P2-ZOI Midline Surveys—especially earlier surveys—may have only one Module 8 outcome variable (**v8709**); if this is the case, adjust the syntax throughout the consumption aggregate calculation to use the one Module 8 outcome variable.

<sup>62</sup> It is assumed that every household must have consumed some food in the past 7 days. Households without recorded food consumption are dropped at this stage and brought back into the dataset later in Step 10. Food consumption of these dropped households is then replaced by the local median consumption.



variable stubs	v8101_ v8102_ v8103a_ v8103b_ v8104a_ v8104b_ v8105_ v8106a_ v8106b_ v8106c_ v8107a_ v8107b_ v8107c_
index variables	hhea hhnum c05 c06 hhsize_dj
new variable	j
Label variable j	"Item number"
Rename variable	v8101_ to be v8101
Rename variable	v8102_ to be v8102
Rename variable	v8103a_ to be v8103a
Rename variable	v8103b_ to be v8103b
Rename variable	v8104a_ to be v8104a
Rename variable	v8104b_ to be v8104b
Rename variable	v8105_ to be v8105
Rename variable	v8106a_ to be v8106a
Rename variable	v8106b_ to be v8106b
Rename variable	v8106c_ to be v8106c
Rename variable	v8107a_ to be v8107a
Rename variable	v8107b_ to be v8107b
Rename variable	v8107c_ to be v8107c

**Step Ie.** Drop food items that are not consumed by the household.

Drop	record if v8102≥2 or missing
------	------------------------------

**Step If.** Label and check all food item variables.

Label v8101	"Name of food item"
Label v8102	"Household consumed item, past week"
Label v8103a	"Quantity consumed, past week"
Label v8103b	"Unit of quantity consumed, past week"
Label v8104a	"Quantity consumed from purchases, past week"
Label v8104b	"Unit of quantity purchased, past week"
Label v8105	"Value of quantity purchased, past week"
Label v8106a	"Quantity consumed from production, past week"
Label v8106b	"Unit of quantity produced, past week"
Label v8106c	"Value of quantity produced, past week"
Label v8107a	"Quantity consumed from gifts or other sources, past week"
Label v8107b	"Unit of quantity received, past week"
Label v8107c	"Value of quantity received, past week"

**Step Ig.** Create analytic variables for each food item (*item*), quantity (*\_q*), unit (*\_u*), and monetary value (*\_v*) to use in the analysis for food consumed (*cons\_*), purchased (*purc\_*), produced (*prod\_*), and received as a gift or from other sources (*gift\_*).

Set	item=v8101
Label variable	"Food item"
Set	cons_q=v8103a
Label variable	"Quantity consumed by household"
Set	cons_u=v8103b
Label variable	"Unit of quantity consumed by household"
Set	purc_q=v8104a
Label variable	"Quantity purchased by household"
Set	purc_u=v8104b
Label variable	"Unit of quantity purchased by household"
Set	purc_v=v8105
Label variable	"Value of quantity purchased"
Set	prod_q=v8106a
Label variable	"Quantity produced"
Set	prod_u=v8106b
Label variable	"Unit of quantity produced by household"
Set	prod_v=v8106c
Label variable	"Value of quantity produced by household"
Set	gift_q=v8107a
Label variable	"Quantity received from gift and other sources"
Set	gift_u=v8107b
Label variable	"Unit of quantity received from gift and other sources"
Set	gift_v=v8107c
Label variable	"Value of quantity received from gifts and other sources"

**Step 1h.** Check the variables measuring quantities and monetary values for each food item and set refused, don't know, and missing values to missing.

Set the consumption (**cons\_q**), purchased (**purc\_q**), production (**prod\_q**), and gifted (**gift\_q**) quantities to missing if any of them have one of the following values: refused (9999997), don't know (9999998), or missing (9999999). **Note:** Ensure that the number of digits in the syntax for the refused, don't know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

Replace	cons_q=missing if cons_q≥9999997
Replace	purc_q=missing if purc_q≥9999997
Replace	prod_q=missing if prod_q≥9999997
Replace	gift_q=missing if gift_q≥9999997

Set the purchased (**purc\_v**), production (**prod\_v**), and gifted (**gift\_v**) monetary values to missing if any of them have one of the following values: refused (999999997), don't know (999999998), or missing (999999999). **Note:** Ensure that the number of digits in the syntax for the refused, don't know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

```
Replace      purc_v=missing if purc_q≥999999997
Replace      prod_v=missing if prod_q≥999999997
Replace      gift_v=missing if gift_q≥999999997
```

**Step Ii.** Verify that for each food item, units are the same for all quantities reported—that is, for total consumption, purchases, production, and gifts or other sources.

**Step Ii.1.** Create a variable to flag items that have multiple or inconsistent units (**prob\_unit**).

```
Set          prob_unit=0
Replace      prob_unit=1 if cons_u≠purc_u & purc_u≠missing
Replace      prob_unit=1 if cons_u≠prod_u & prod_u≠missing
Replace      prob_unit=1 if cons_u≠gift_u & gift_u≠missing
Label value  0 "No"
              1 "Yes"
Label variable "Item has multiple/inconsistent units"
```

**Step Ii.2.** Save the flagged records with relevant identifiers (e.g., **hhea**, **hhnum**, **c05**, and **c06**) to an Excel file: "**flag\_probunit.xlsx**." Open the Excel file and carefully examine each flagged record and, if possible, correct any issues. Corrections can only be made for issues with clear-cut resolutions. If the issue cannot be corrected, set the unit variable (i.e., **cons\_u**, **purc\_u**, **prod\_u**, **gift\_u**) that is causing the issue to be missing. For example, if the quantity consumed unit is kilograms, the purchased quantity unit is liters and relatively similar in quantity, and the item is something that is measured by weight, update the purchased units to be kilograms. As another example, if the units for quantity consumed are liters and the units for quantity produced are milliliters, look at the quantities consumed and produced and decide which units make sense and update accordingly. If the solution is not obvious, because of the purchased quantity reported, set the purchased quantity units to missing. Add syntax to update the quantity and unit values as needed per the review performed.

**Step Ij.** Create a variable to flag food items for which values of all seven variables measuring quantity and monetary values of the item are 0 or missing (**miss\_data=7**).

```
Set          miss_data=0
Replace      miss_data=miss_data+1 if cons_q=0 or missing
Replace      miss_data=miss_data+1 if purc_q=0 or missing
Replace      miss_data=miss_data+1 if prod_q=0 or missing
Replace      miss_data=miss_data+1 if gift_q=0 or missing
Replace      miss_data=miss_data+1 if purc_v=0 or missing
Replace      miss_data=miss_data+1 if prod_v=0 or missing
Replace      miss_data=miss_data+1 if gift_v=0 or missing
Label variable "Number of quantity/value variables missing
              values per food item"
```

**Step 1k.** Verify that for each food item, the quantity consumed equals the sum of quantities from the three sources (i.e., `cons_q=purc_q+prod_q+gift_q`). The sum will not equal the quantity consumed if, for example, there is repetition of quantities for multiple sources and the sum is two or three times the quantity consumed. In theory, because of the way the questions are formulated, it should be the same, but in practice, it may differ. A household may report quantities obtained from various sources that were not entirely consumed in the 7 days preceding the survey, or there could be recall or data entry errors. We treat an amount consumed greater than the sum (see Step 1k.2) differently from an amount consumed less than the sum (see Steps 1k.3-1k.7).

**Step 1k.1.** Create a variable that sums the quantity of each food item purchased, produced, and received (`sumfood1`) and verify that for each item, the amount consumed by the household equals the sum of the amounts purchased, produced, and received—that is, `cons_q=purc_q+prod_q+gift_q`.<sup>63</sup>

```
Set          sumfood1=purc_q+prod_q+gift_q
Label variable  "Sum of food quantities purchased, produced,
                received (initial)"
```

**Step 1k.2.** Compare `sumfood1` and `cons_q` (quantity consumed). If the quantity consumed is greater than the sum of the amounts purchased, produced, and received, replace the quantity consumed with `sumfood1`. That is, if `cons_q > (purc_q+prod_q+gift_q)`, adjust `cons_q` to equal the sum. This is because the quantity consumed cannot be greater than the amounts obtained from various sources.

```
Replace          cons_q=sumfood1 if cons_q>sumfood1
```

**Step 1k.3.** Create a flag (`flag_1`) equal to 1 if the quantity and units consumed equal the quantity and units purchased. Set the quantity for the produced and received to zero and set the units for the produced and received to the same unit as purchased. In this step, we are assuming that `purc_q` and `purc_u` are the values that should be used to calculate the household's food consumption.

```
Set          flag_1=1 if cons_q=purc_q and cons_u=purc_u
Label variable  "Consumed qty & units = purchased qty & units"

If flag_1=1:
  Replace      prod_q=0
  Replace      prod_u=purc_u
  Replace      prod_v=0
  Replace      gift_q=0
  Replace      gift_u=purc_u
  Replace      gift_v=0
```

**Step 1k.4.** Create a flag (`flag_2`) equal to 1 if the quantity and units consumed equal the quantity and units produced. Set the quantity for the purchased and received to zero and set the units for the

---

<sup>63</sup> If computer-assisted personal interviewing is used, the program will automatically flag this inconsistency, and the problem should be resolved during data collection. It is useful, however, to also check during data analysis.

purchased and received to the same unit as produced. In this step, we are assuming that **prod\_q** and **prod\_u** are the values that should be used to calculate the household's food consumption.

```
Set          flag_2=1 if cons_q=prod_q and cons_u=prod_u
Label variable  "Consumed qty & units = produced qty & units"

If flag_2=1:
  Replace      purc_q=0
  Replace      purc_u=prod_u
  Replace      purc_v=0
  Replace      gift_q=0
  Replace      gift_u=prod_u
  Replace      gift_v=0
```

**Step 1k.5.** Create a flag (**flag\_3**) equal to 1 if the quantity and units consumed equal the quantity and units received. Set the quantity for purchased and produced to zero and set the units for purchased and produced to the same unit as consumed. In this step, we are assuming that **gift\_q** and **gift\_u** are the values that should be used to calculate the household's food consumption.

```
Set          flag_3=1 if cons_q=gift_q and cons_u=gift_u
Label variable  "Consumed qty & units = Received qty & units"

If flag_3=1:
  Replace      purc_q=0
  Replace      purc_u=cons_u
  Replace      purc_v=0
  Replace      prod_q=0
  Replace      prod_u=cons_u
  Replace      prod_v=0
```

**Step 1k.6.** Create a variable (**sumfood2**) equal to the sum of the quantities from the three sources (purchased, produced, and received) after the values are adjusted in Steps 1k.2-1k.5, and then create a flag (**flag\_4**) equal to 1 if **sumfood2** is not equal to the quantity consumed.

```
Set          sumfood2=purc_q+prod_q+gift_q
Label variable  "Sum of food quantities purchased, produced,
               received (post-adjustment)"

Set          flag_4=1 if sumfood2≠cons_q
Label variable  "sumfood2 and cons_q are not equal"
```

**Step 1k.7.** Continue to flag and check inconsistencies where the sum of the quantities of the food items (**sumfood2**) is not equal to the quantity consumed (**cons\_q**). Save the flagged records with relevant administrative unit identifiers to an Excel file: "**flag4.xlsx**."

Open the Excel file and carefully examine each record one by one to determine what the issue may be (e.g., data entry errors in the quantity or value variables), whether a correction can be made, and what the correction should be.

**Step 1k.8.** Add syntax to update the records that can be corrected, per the examination performed in Step 1k.7.

**Step 2.** Convert units to the same numeraire to the extent possible to calculate unit price and reconcile units of measure. In addition, checking outliers requires all the units of an item to be the same.

In the syntax that follows, the units are those listed in Sub-module 8.1 of the core parallel survey questionnaire for the P2-ZOI Midline Survey: *Household consumption expenditure—food consumption over past 7 days*. The survey team should have obtained or developed a country-specific conversion table that should be used to adapt the syntax.

- Identify the type (i.e., weight, volume, or count), name, and number of measurement units collected in the country-specific survey.
- Check whether country-specific or local units have an equivalent in standard metric units they can be converted to by consulting the unit conversion table developed for the country-specific survey at the time of fieldwork.
- Customize the Midline Toolkit syntax to convert all measurement units collected for quantity consumed, purchased, produced, and gifted to standard metric units. For example, replace “xxxxx unit” with the metric/standard unit equivalent.

**Step 2a.** Create duplicate unit, value, and quantity variables as a backup or reference with ‘orig\_’ included at the start of their names.

```
Set          orig_cons_u=cons_u
             orig_cons_q=cons_q
             orig_cons_v=cons_v
             orig_prod_u=prod_u
             orig_prod_q=prod_q
             orig_prod_v=prod_v
             orig_purc_u=purc_u
             orig_purc_q=purc_q
             orig_purc_v=purc_v
             orig_gift_u=gift_u
             orig_gift_q=gift_q
             orig_gift_v=gift_v
```

**Step2b.** Create a weight unit conversion table to convert weight units used in the country-specific survey to the standard weight unit equivalent (gram [g]) (see **Table 8**).

Note that separate weight unit conversion table entries should be developed for food items measured using the same local non-standard weight units that have different conversion factors. For example, the standard unit equivalent of one small pail may not be the same for all food items, so separate entries are required.

**Table 8: Example Weight Unit Conversion Table**

Food item	Food item number*	Weight unit (cons_u)	Description	Standard unit equivalent (gram)
All	n/a	18	Grams	1
All	n/a	1	Kilograms	1,000
All	n/a	2	50-kg bag	50,000
Rice	006	4	Small pail	400
Cowpeas	043	4	Small pail	300
Brown beans	037	4	Small pail	300

\*Minus leading "8"

**Step 2c.** Convert all weight units included in the data to grams, the standard unit weight equivalent, using the weight conversion table prepared in Step 2b. The syntax that follows aligns with the table in Step 2b and shows how to convert two standard units (i.e., kilograms and 50-kg bags) for all food items and one non-standard unit (i.e., small pails) for three food items as an example. The syntax must be adapted for the weight units and food items included in the country-specific survey.

```

Replace      purc_q=purc_q*1000   if cons_u=1
Replace      purc_q=purc_q*50000  if cons_u=2
Replace      purc_q=purc_q*400    if cons_u=4 and v8101=006
Replace      purc_q=purc_q*300    if cons_u=4 and (v8101=37 or
                                043)

```

**Step 2d.** Create a volume unit conversion table to convert volume units used in the country-specific survey to the standard volume unit equivalent (milliliters).

Note that separate volume unit conversion table entries should be developed for food items measured using the same local non-standard volume units that have different conversion factors (see **Table 9**). For example, the standard unit equivalent of 1 sachet/tube may not be the same for all food items, so separate entries are required.

**Table 9: Example Volume Unit Conversion Table**

Food item	Food item number*	Volume unit (cons_u)	Description	Standard unit equivalent (milliliters)
All	n/a	19	Milliliters	1
All	n/a	15	Liters	1,000
All	n/a	20	Teaspoon	4.93
Fresh milk	109	22	Sachet/tube	25
Soured milk	113	22	Sachet/tube	25
Cooking oil	126	22	Sachet/tube	10

\*Minus leading "8"

**Step 2e.** Convert all volume units included in the data to milliliters, the standard volume unit equivalent, using the volume unit conversion table prepared in Step 2d. The syntax that follows aligns with the table in Step 2d and shows how to convert two standard volume units (i.e., liters and teaspoons) to the standard unit equivalent (milliliters) for all food items and one non-standard unit

(i.e., sachet/tube) for three food items as an example. The syntax must be adapted for the volume units and food items included in the country-specific survey.

```
Replace      purc_q=purc_q*1000      if cons_u=15
Replace      purc_q=purc_q*4.93      if cons_u=20
Replace      purc_q=purc_q*25        if cons_u=22 and (v8101=109
or 113)
Replace      purc_q=purc_q*10        if cons_u=126
```

**Step 2f.** Create a count unit conversion table to convert count units used in the country-specific survey to the standard count unit equivalent (pieces) (see **Table 10**).

**Table 10: Example Count Unit Conversion Table**

Food item	Food item number*	Original count units (cons_u)	Recoded count units (cons_u)	Description	Standard unit equivalent (piece)
All	n/a	09	09	Piece	1
All	n/a	96	24	Other: dozen	12
All	n/a	96	25	Other: pair	2

\*Minus leading “8”

**Step 2g.** Convert all count units included in the data to pieces using the count unit conversion table prepared in Step 2f. The syntax that follows aligns with the table in Step 2f and shows how to convert two count units to the standard count unit equivalent (i.e., pieces) for all food items as an example. The syntax must be adapted for the count units and food items included in the country-specific survey.

```
Replace      purc_q=purc_q*12 if cons_u=24
Replace      purc_q=purc_q*2  if cons_u=25
```

**Step 2h.** Calculate the unit price (**unitprice\_item**) for each food item using the following rules:

If (at least some of) the food item was purchased:

```
Set          unitprice_item=missing
Replace      unitprice_item=purc_v÷purc_q if purc_v>0 and
purc_q>0
Replace      unitprice_item=prod_v÷prod_q if prod_v>0 and
prod_q>0 & unitprice_item=missing
Replace      unitprice_item=gift_v÷gift_q if gift_v>0 and
gift_q>0 and unitprice_item=missing
Label variable "Unit price of item consumed in the past 1 week"
```

**Step 3.** Identify and examine potential unit price outliers of each item (**unitprice\_item**) and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ‘var’ with **unitprice\_item**. Variables **mean\_unitprice\_item**, **sd\_unitprice\_item**, and **out\_unitprice\_item** are created in this step.



**Step 4.** Calculate the local median unit price for each food item (**unitprice\_item**) and impute outliers by the median value. Follow **Protocol b** to calculate the median value, replacing ``var'` with **unitprice\_item**. Variable **med\_unitprice\_item** is created in this step.

**Step 5.** Calculate total household food consumption expenditure on each food item in the 7 days preceding the survey (**totfood\_item**).

**Step 5a.** For the following variables: **cons\_q**, **purc\_q**, **prod\_q**, **gift\_q**, **purc\_v**, **prod\_v**, **gift\_v**, and **med\_unitprice\_item**, replace missing values with “0.”

```
Replace      cons_q=0 if cons_q=missing
Replace      purc_q=0 if purc_q=missing
Replace      prod_q=0 if prod_q=missing
Replace      gift_q=0 if gift_q=missing

Replace      purc_v=0 if purc_v=missing
Replace      prod_v=0 if prod_v=missing
Replace      gift_v=0 if gift_v=missing
```

**Step 5b.** If the unit price for the food item is not missing, calculate total household food consumption expenditure (**totfood\_item**) as the item’s unit price multiplied by the quantity consumed.

```
Set          totfood_item=missing
Replace      totfood_item=unitprice_item*cons_q if
              unitprice_item≠missing
```

**Step 5c.** If the food item is purchased from a vendor (items 8167 through 8186) and **purc\_v**>0, set total household consumption expenditure equal to the purchase value of the item.

```
Replace      totfood_item=purc_v if purc_v>0 and v8101≥8167
              and v8101≤8186
```

**Step 5d.** If the unit price for the food item is missing, set total household food consumption expenditure of each item equal to the quantity consumed multiplied by the local median unit price for the food item (median unit price is created in Step 4).

```
Replace      totfood_item=(cons_q*med_unitprice_item) if
              totfood_item=missing
```

**Step 6.** Create a variable to capture the daily per capita food consumption of each food item during the 7 days preceding the survey (**pcdfood\_item**) by dividing the total household food consumption expenditure (**totfood\_item**) by the number of de jure members in the household (**hhsize\_dj**) and 7 as the number of days in the week.

```
Set          pcdfood_item=totfood_item÷hhsize_dj÷7
Label variable "Per capita daily consumption of food item in
               local currency"
```

**Step 7.** Identify and verify potential daily per capita food consumption of each food item (**pcdfood\_item**) outliers and set confirmed outliers to missing. Follow **Protocol a** to identify and

examine potential outliers, replacing `var` with `pcdfood_item`. Variables `mean_pcdfood_item`, `sd_pcdfood_item`, and `out_pcdfood_item` are created in this step.

**Step 8.** Calculate the local median of daily per capita food consumption of each food item during the 7 days preceding the survey (`pcdfood_item`) and replace verified outliers with the imputed local median. Follow **Protocol b** to calculate the median value, replacing `var` with `pcdfood_item`. Variable `med_pcdfood_item` is created in this step.

**Step 9.** Sum all `pcdfood_item` variables by household and create a variable (`pcdfood_hh`) that captures total daily per capita consumption expenditure by household in the 7 days preceding the survey. The variable `pcdfood_item` will have the same value for a household across all food item records. Keep only the first record for each household.

```
By hhea hhnum: Set pcdfood_hh=sum(pcdfood_item)
By hhea hhnum: Keep if _n=1
Label variable "Per capita daily food consumption expenditure,
               past 1 week"
```

**Step 10.** Add the households that were excluded in Step 1b because they did not report any food consumption in the 7 days preceding the survey back into the working dataset. Keep only the relevant variables, including cluster number (`hhea`), household number (`hhnum`), de jure household size (`hhsize_dj`), geographic/administrative units (variables beginning with "c0"), and the Module 8 outcome variables (`v8100r` and `v8700r`) for those households.

**Step 10a.** Merge the working data file with the household analytic data file using `hhea` and `hhnum` as key matching variables.

```
Merge          1 to 1
key variables   hhea hhnum
data file       "FTF ZOI Survey [COUNTRY] [YEAR] household
                data analytic"
keep variables   hhea hhnum hhsize_dj c05 c06 v8100r v8700r
```

**Step 10b.** Drop households that did not complete the consumption expenditure module (Module 8).

```
Drop          record if v8100r≠1
```

**Step 11.** Identify and verify potential `pcdfood_hh` outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing `var` with `pcdfood_hh`. Variables `mean_pcdfood_hh`, `sd_pcdfood_hh`, and `out_pcdfood_hh` are created in this step.

**Step 12.** Calculate local median of `pcdfood_hh` and replace confirmed outliers and missing values with the local median. Because every household must have consumed some food in the past week, all missing `pcdfood_hh` values are replaced by the median. Follow **Protocol b** to calculate the median value, replacing `var` with `pcdfood_hh`. Variable `med_pcdfood_hh` is created in this step.

**Step 13.** Rename the `pcdfood_hh` variable to be `pcdfood_1w`, keep only the renamed PCD food consumption expenditure (`pcdfood_1w`), cluster number (`hhea`), and household number (`hhnum`) variables and save the data file.

```
Rename      pcdfood_hh to be pcdfood_1w
Keep variables hhea hhnum pcdfood_1w
Save        "FTF ZOI Survey [COUNTRY] [YEAR] pov1_food"
```

### 10.2.5 Regular non-food, non-durable goods expenditures of the past 1 week (7 days)

The step-by-step procedures to calculate the non-food, non-durable goods consumption expenditures during the 1 week (7 days) preceding the survey follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov2\_nfood\_1w*. Pet expenditure variables (`v8112`, `v8112a`, `v8113`, `v8114`, `v8115`) from Sub-module 8.1 and variables from Sub-module 8.2 are used for these calculations.

**Step 1.** Prepare the data.

**Step 1a.** Create a temporary data file that holds the pet expenditures in the same format as other items included in this section. Pet expenditure variables are `v8112`, `v8112a`, `v8113`, `v8114`, and `v8115` in Sub-module 8.1.

Load the household-level analytic data file. Drop households that did not complete Module 8 and keep only necessary variables required for this section, namely cluster number (`hhea`), household number (`hhnum`), geographic administrative units (`c05` and `c06`), de jure household size (`hhsz_dj`), pet expenditure variables (`v8112`, `v8112a`, `v8113`, `v8114`, `v8115`), all Sub-module 8.2 variables (all variables starting with '`v820`'), and the outcome variables for Module 8 (`v8100r` and `v8700r`), and drop households that did not complete Sub-modules 8.2-8.7.

```
Load      "FTF ZOI Survey [COUNTRY] [YEAR] household data
          analytic"
Keep variables hhea hhnum hhsz_dj c05 c06 v8100r v8700r v8112-
          v8115, and all variables starting with 'v820'
Drop      record if v8700r#1
```

Create variables indicating whether the household bought any pet food in the past week and how much the household spent on pet food. Because questions on pet food expenditures are formulated differently from the other expenditures in the past 7 days, it is necessary to recode these variables to align them with the structure of Sub-module 8.2 (for example: `v8201_21`: item, `v8202_21`: purchased, `v8203_21`: amount paid). The suffix '`_21`' is used here because 21 is the first available item number based on the number of items included in Module 8.2 of the core questionnaire. (Items 8187-8206 are renumbered from 1-20 in Step 1d when the Module 8.2 data are transformed from wide to long format.) Adapt the suffix to be the first available number that does not overlap with an item number already used in the country-specific questionnaire.

```
Set      v8201_21="21"
Label variables "Pet food item code"
```

```

Set          v8202_21=v8112a
Replace      v8202_21=2 if v8112=2
Label variable "Bought pet food"

Set          v8203_21=v8113 if v8112=1
Replace      v8203_21=0 if v8112=2
Label variable "Amount spent on pet food"

```

Create variables indicating if the household had any other pet expenditures in the past week and how much was spent (v8201\_22: item, v8202\_22: purchased, v8203\_22: amount spent). Adapt the suffix to be the next available number that does not overlap with an item number already used in the country-specific questionnaire.

```

Set          v8201_22="22"
Label variable "Other pet expenditures code"

Set          v8202_22=v8114
Replace      v8202_22=2 if v8112=2
Label variable "Bought other pet expenditures"

Set          v8203_22=v8115 if v8114=1
Replace      v8203_22=0 if v8112=2
Label variable "Amount spent on other pet expenditures"

```

Save the data as a temporary data file.

```

Save          "FTF ZOI Survey [COUNTRY] [YEAR]
              temp_pet_expenditures"

```

**Step 1b.** Create a variable (v8202miss) to flag households if none of the items in Sub-module 8.2 were purchased in the 7 days preceding the survey. [XXX] should be equal to the number of items included in Sub-module 8.2; adapt the template syntax to include all items.

```

Set          v8202miss=0
Replace      v8202miss=v8202miss+1 if v8202_001≥2 or missing
Replace      v8202miss=v8202miss+1 if v8202_002≥2 or missing
Replace      v8202miss=v8202miss+1 if v8202_003≥2 or missing
...
Replace      v8202miss=v8202miss+1 if v8202_[XXX]≥2 or missing
Label variable "Number of v8202 items with consumption not
              reported or missing with consumption not reported
              or missing"

```

If no consumption was reported or consumption information is missing across all items included in Sub-module 8.2, v8202miss will be equal to the total number of items in Sub-module 8.2.

**Step 1c.** Drop all households from the working dataset that reported no consumption or that are missing consumption information across all items in Sub-Module 8.2. These dropped households will be added back into the dataset in Step 6.

```

Drop          record if v8202miss=[XXX]

```

**Step 1d.** Change the level of observation in the data from household level to item level, as explained in **Protocol c**. Convert the variables that start with the following “stubs” from wide format (household level) to long format (item level): `v8201_`, `v8202_`, and `v8203_` and then remove the ‘\_’ suffix from all item variable names.

```
Reshape          wide to long
  variable stubs v8201_ v8202_ v8203_
  index variables hhea hhnum c05 c06 hhsize_dj
  new variable    j

Label variable j  "Item number"
Rename           v8201_ to be v8201
Rename           v8202_ to be v8202
Rename           v8203_ to be v8203
```

**Step 1e.** Drop items that are not consumed.

```
Drop          record if v8202≥2 or missing
```

**Step 1f.** Create new variables to use in the analysis (`item` and `totnfood_item`).

```
Set          item=v8201
Label variable "Non-food item"

Set          totnfood_item=v8203
Label variable "Expenditure on non-food item, past 7 days"
```

**Step 1g.** Check the expenditure values for each item and set the values to missing if any of them have one of the following values: refused (999999997), don’t know (999999998), or missing (999999999). **Note:** Ensure that the number of digits in the syntax for the refused, don’t know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

```
Replace          totnfood_item=missing if totnfood_item≥999999997
```

**Step 2.** Calculate PCD expenditure on each non-food item purchased in the 7 days preceding the ZOI Survey (`pcdnfood_item`). Divide `totnfood_item` by the number of de jure household members and by 7 days.

```
Set          pcdnfood_item=totnfood_item÷hhsize_dj÷7
Label variable "PCD non-food consumption exp on item, past 7
              days"
```

**Step 3.** Identify and examine potential `pcdnfood_item` outliers and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ‘`var`’ with `pcdnfood_item`. Variables `mean_pcdnfood_item`, `sd_pcdnfood_item`, and `out_pcdnfood_item`, are created in this step.

**Step 4.** Calculate the local median of `pcdnfood_item` and replace verified outliers with the local median. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdnfood_item`. Variable `med_pcdnfood_item` is created in this step.

**Step 5.** Sum all `pcdnfood_item` variables by household and create a variable that captures the total household-level non-food PCD expenditure for the week preceding the survey (`pcdnfood_hh`) and keep only one record per household (the first record if there are multiple records).

```
By hhea hhnum: Set pcdnfood_hh=sum(pcdnfood_item)
By hhea hhnum: Keep if _n=1
Label variable "PCD total HH non-food consumption exp, past 1
week"
```

**Step 6.** Add the households that were dropped in Step 1 because they did not have expenditures on any Sub-module 8.2 items back into the working dataset by merging the working dataset with the household data analytic file using the cluster and household numbers (`hhea` and `hhnum`) as the key matching variables. When merging, keep only geographic administrative units (variables beginning with `c0`), and the Module 8 outcome variables (`v8100r` and `v8700r`) from the household data analytic file. Then drop households that did not complete Sub-modules 8.2-8.7.

```
Merge          1 to 1
key variables  hhea hhnum
data file      "FTF ZOI Survey [COUNTRY] [YEAR] household
data analytic"
keep variables hhsizedj c05 c06 v8100r v8700r
Drop          record if v8700r#1
```

**Step 7.** Identify and examine potential `pcdnfood_hh` outliers and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `pcdnfood_hh`. Variables `mean_pcdnfood_hh`, `sd_pcdnfood_hh`, and `out_pcdnfood_hh` are created in this step.

**Step 8.** Calculate the local median of `pcdnfood_hh` and replace `pcdnfood_hh` with the local median if the flagged value is a confirmed outlier. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdnfood_hh`. Variable `med_pcdnfood_hh` is created in this step.

**Step 9.** Set `pcdnfood_hh` to 0 if it is still missing. A missing value means that the household is missing Sub-module 8.2 consumption information and it could not be imputed. Households that legitimately did not make any purchases of regular non-food, non-durable goods items in the past 7 days already have their PCD expenditures set to zero (in Step 5).

```
Replace          pcdnfood_hh=0 if pcdnfood_hh=missing
```

**Step 10.** Rename the `pcdnfood_hh` variable to `pcdnfood_1w`. Keep only the renamed PCD non-food consumption expenditure variable, the cluster and household number variables, and save the data file.

```
Rename variable pcdnfood_hh to be pcdnfood_1w
Keep variables  hhea hhnum pcdnfood_1w
```

```
Save "FTF ZOI Survey [COUNTRY] [YEAR] pov2_nfood_1w"
```

### 10.2.6 Occasional non-food, non-durable goods expenditures of the past 1 month

The step-by-step procedures to calculate the PCD non-food, non-durable good consumption expenditures during the 1 month preceding the survey follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov3\_nfood\_1m*. Variables from Sub-module 8.3 are used for the calculations.

**Step 1.** Prepare the data.

**Step 1a.** Load the household-level analytic data file. Keep only necessary variables required for this section: cluster number (**hhea**), household number (**hhnum**), geographic administrative units (**c05** and **c06**), de jure household size (**hhsz\_dj**), all Sub-module 8.3 variables (variables starting with 'v830'), and the Module 8 outcome variables (**v8100r** and **v8700r**), and drop households that did not complete Sub-modules 8.2-8.7.

```
Load "FTF ZOI Survey [COUNTRY] [YEAR] household data analytic"
Keep variables hhea hhnum hhsz_dj c05 c06 v8100r v8700r, and all variables starting with 'v830'
Drop record if v8700r#1
```

**Step 1b.** Create a variable (**v8302miss**) to flag households if none of the items in Sub-module 8.3 were purchased in the month preceding the survey. [XXX] should be the number of items included in Sub-module 8.3; adapt the template syntax to include all items.

```
Set v8302miss=0
Replace v8302miss=v8302miss+1 if v8302_001≥2 or missing
Replace v8302miss=v8302miss+1 if v8302_002≥2 or missing
Replace v8302miss=v8302miss+1 if v8302_003≥2 or missing
...
Replace v8302miss=v8302miss+1 if v8302_[XXX]≥2 or missing

Label variable "Number of v8302 items with consumption not reported or missing"
```

If no consumption was reported or consumption information is missing across all items included in Sub-module 8.3, **v8302miss** will be equal to the total number of items included in Sub-module 8.3.

**Step 1c.** Drop all households from the working dataset that reported no consumption or that are missing consumption information across all items in Sub-Module 8.3. These dropped households will be added back into the dataset in Step 6.

```
Drop record if v8302miss=[XXX]
```

**Step 1c.** Change the level of observation in the data from household level to item level, as explained in **Protocol c**. Convert the variables that start with the following "stubs" from wide format (household level) to long format (item level): **v8301\_**, **v8302\_**, and **v8303\_** and then remove the '\_' suffix from all item variable names.

```

Reshape          wide to long
  variable stubs  v8301_ v8302_ v8303_
  index variables hhea hhnum c05 c06 hhsize_dj
  new variable    j

Label variable j  "Item number"

Rename variable   v8301_ to be v8301
Rename variable   v8302_ to be v8302
Rename variable   v8303_ to be v8303

```

**Step Id.** Drop items that are not purchased.

```

Drop          record if v8302≥2 or missing

```

**Step Ie.** Create new variables to use in the analysis (*item* and *totnfood\_item*).

```

Set          item=v8301
Label variable "Non-food item, past 1 month"

Set          totnfood_item=v8303
Label variable "Total expenditure on non-food item, past 1
month"

```

**Step If.** Drop items related to health, which do not contribute to household consumption, even if survey respondents report expenditures on these items.<sup>64</sup> In the parallel survey questionnaire template, these items are numbered 8228 through 8236, but the variable range should be adjusted to the country-customized questionnaire.

```

Drop          record if item=8228, 8229, 8230, 8231, 8232,
8233, 8234, 8235, or 8236

```

**Step Ig.** Check the expenditure values for each item and set the values to missing if any of them have one of the following values: refused (999999997), don't know (999999998), or missing (999999999). **Note:** Ensure that the number of digits in the syntax for the refused, don't know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

```

Replace          totnfood_item=missing if totnfood_item≥999999997

```

**Step 2.** Calculate the PCD non-food consumption expenditure on each item purchased in the past 1 month (*pcdnfood\_item*). Divide *totnfood\_item* by the number of de jure household members and by 30 days.

```

Set          pcdnfood_item=(totnfood_item÷hhsize_dj÷30) if
totnfood_item≠missing
Label variable "PCD non-food consumption expenditure on item,
past 1 month"

```

<sup>64</sup> See Section 10.1 for a discussion about these items and why they are included in the questionnaire.



**Step 3:** Identify and examine potential outliers of `pcdnfood_item` and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `pcdnfood_item`. Variables `mean_pcdnfood_item`, `sd_pcdnfood_item`, and `out_pcdnfood_item` are created in this step.

**Step 4:** Calculate the local median of `pcdnfood_item` (`med_pcdnfood_item`) and replace outliers with the median. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdnfood_item`. Variable `med_pcdnfood_item` is created in this step.

**Step 5.** Sum across all `pcdnfood_item` variables by household and create a variable that captures total household-level PCD non-food expenditure for the month preceding the survey (`pcdnfood_hh`). Keep only one record per household.

```
By hhea hhnum:   Set pcdnfood_hh=sum(pcdnfood_item)
By hhea hhnum:   Keep if _n=1
Label variable   "PCD total HH non-food consumption exp, past 1
                  month"
```

**Step 6.** Add the households that were dropped in Step 1 because they did not have expenditures on any of the items listed in Sub-module 8.3 back into the working dataset by merging the working dataset with the household data analytic file using cluster number and household number (`hhea` and `hhnum`) as the key matching variables. When merging, keep only the geographic administrative units (`c05` and `c06`), de jure household size (`hhsizes_dj`), and the Module 8 outcome variables (`v8100r` and `v8700r`) from the household data analytic file. Drop households that did not complete Sub-modules 8.2-8.7.

```
Merge           1 to 1
  key variables  hhea hhnum
  data file      "FTF ZOI Survey [COUNTRY] [YEAR] household
                  data analytic"
  keep variables hhsizes_dj c05 c06 v8100r v8700r
Drop            record if v8700r#1
```

**Step 7.** Identify and examine potential outliers of `pcdnfood_hh` and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `pcdnfood_hh`. Variables `mean_pcdnfood_hh`, `sd_pcdnfood_hh`, and `out_pcdnfood_hh` are created in this step.

**Step 8.** Calculate the local median of `pcdnfood_hh` and replace confirmed outliers with the median. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdnfood_hh`. Variable `med_pcdnfood_hh` is created in this step.

**Step 9.** Set `pcdnfood_hh` to 0 if `pcdnfood_hh` is missing. A missing value means that the household either did not purchase any of these items in the past month or the household is missing consumption information for Sub-module 8.3. Households that legitimately did not make any purchases

of regular non-food, non-durable goods items in the past month also have their PCD expenditures set to zero.

```
Replace          pcdnfood_hh=0 if pcdnfood_hh=missing
```

**Step 10.** Rename the `pcdnfood_hh` variable as `pcdnfood_1m`. Keep only the renamed PCD non-food consumption expenditure (`pcdnfood_1m`), cluster number (`hhea`), and household number (`hhnum`) variables, and save the data file.

```
Rename variable  pcdnfood_hh to be pcdnfood_1m
Label variable   "PCD nonfood consumption expenditure on item,
                  past 1 month"
Keep variables    hhea hhnum pcdnfood_1m
Save              "FTF ZOI Survey [COUNTRY] [YEAR] pov3_nfood_1m"
```

### 10.2.7 Occasional non-food, non-durable goods of the past 3 months

The step-by-step procedures to calculate the PCD non-food, non-durable good consumption expenditures during the 3 months preceding the survey follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov4\_nfood\_3m*. Variables from Sub-module 8.4 are used for the calculations.

**Step 1.** Prepare the data.

**Step 1a.** Load the household-level analytic data file, keep relevant variables required for this section, namely cluster number (`hhea`), household number (`hhnum`), geographic administrative units (`c05` and `c06`), de jure household size (`hhsz_dj`), all Sub-module 8.4 (variables starting with '`v840`'), and the Module 8 outcome variables (`v8100r` and `v8700r`), and drop households that did not complete Sub-modules 8.2-8.7.

```
Load              "FTF ZOI Survey [COUNTRY] [YEAR] household data
                  analytic"
Keep variables     hhea hhnum c05 c06 hhsz_dj v8100r v8700r, and
                  all variables starting with 'v840'

Drop              record if v8700r≠1
```

**Step 1b.** Create a variable (`v8402miss`) to flag households that did not consume any of the items included in Sub-module 8.4 in the 3 months preceding the survey. `[XXX]` should be the number of items included in Sub-module 8.4; adapt the template syntax to include all items.

```
Set               v8402miss=0
Replace           v8402miss=v8402miss+1 if v8402_001≥2 or missing
Replace           v8402miss=v8402miss+1 if v8402_002≥2 or missing
Replace           v8402miss=v8402miss+1 if v8402_003≥2 or missing
...
Replace           v8402miss=v8402miss+1 if v8402_[XXX]≥2 or missing
```

```
Label variable    "Number of v8402 items with consumption not
                  reported or missing"
```

If no consumption was reported or consumption information is missing across all items included in Sub-module 8.4, `v8402miss` will be equal to the total number of items in Sub-module 8.4.

**Step Ic.** Drop all households from the working dataset that reported no consumption or that are missing consumption information across all food items. These dropped households are added back into the working dataset in Step 6.

```
Drop              record if v8402miss=[XXX]
```

**Step Id.** Change the level of observation in the data from household level to item level, as explained in **Protocol c**. Convert the variables that start with the following “stubs” from wide format (household level) to long format (item level): `v8401_`, `v8402_`, and `v8403_` and then remove the ‘\_’ suffix from all item variable names.

```
Reshape           wide to long
  variable stubs   v8401_ v8402_ v8403_
  index variables  hhea hhnum c05 c06 hhsizes_dj
  new variable     j

Label variable j    "Item number"
Rename variable     v8401_ to be v8401
Rename variable     v8402_ to be v8402
Rename variable     v8403_ to be v8403
```

**Step Ie.** Drop items that are not purchased.

```
Drop              record if v8402≥2 or missing
```

**Step If.** Create new variables to use in the analysis (`item` and `totnfood_item`).

```
Set               item=v8401
Label variable     "Non-food item, past 3 months"

Set               totnfood_item=v8403
Label variable     "Total expenditure on non-food item, past 3
                  months"
```

**Step Ig.** Drop expenditures on lodging in a rest house or hotel in the past 3 months, even if the survey respondents report expenditures on this item.<sup>65</sup> In the parallel survey questionnaire template, this item is numbered 8275. Adjust this item number to reflect the ZOI country-customized questionnaire.

```
Drop              record if item=8275
```

---

<sup>65</sup> See Section 10.1.3.2 for a discussion about these items and why they are included in the ZOI Survey questionnaire.

**Step 1h.** Check the expenditure values for each item and set the values to missing if any of them have one of the following values: refused (999999997), don't know (999999998), or missing (999999999). **Note:** Ensure that the number of digits in the syntax for the refused, don't know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

```
Set totnfood_item=missing if totnfood_item≥999999997
```

**Step 2.** Calculate the PCD expenditure on each non-food item purchased in the past 3 months (`pcdnfood_item`) by dividing `totnfood_item` by the number of de jure household members and by 90 days.

```
Set pcdnfood_item=(totnfood_item÷hhsize_dj÷90)
Label variable "PCD non-food consumption expenditure on item,
past 3 months"
```

**Step 3.** Identify and examine potential outliers of `pcdnfood_item` and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `pcdnfood_item`. Variables `mean_pcdnfood_item`, `sd_pcdnfood_item`, and `out_pcdnfood_item` are created in this step.

**Step 4.** Calculate the local median of `pcdnfood_item` and replace outliers with the median. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdnfood_item`. Variable `med_pcdnfood_item` is created in this step.

**Step 5.** Sum across all `pcdnfood_item` variables by household to create a variable that captures total household-level PCD non-food expenditure for the 3 months preceding the survey (`pcdnfood_hh`). Keep only one record per household.

```
By hhea hhnum: Set pcdnfood_hh=sum(pcdnfood_item)
By hhea hhnum: Keep if _n=1

Label variable "PCD total HH non-food consumption exp, past 3
month"
```

**Step 6.** Add the households that were dropped in Step 1 because they did not have expenditures on any of the items listed in Sub-module 8.4 back into the working dataset by merging the working dataset with the household data analytic file using cluster number and household number (`hhea` and `hhnum`) as the key matching variables. Include only geographic administrative units (`c05` and `c06`), de jure household size (`hhsize_dj`), and the Module 8 outcome variables (`v8100r` and `v8700r`) from the household data analytic file. Drop households that did not complete Module 8.

```
Merge 1 to 1
key variables hhea hhnum
data file "FTF ZOI Survey [COUNTRY] [YEAR] household
data analytic"
keep variables hhea hhnum c05 c06 hhsize_dj v8100r v8700r
Drop record if v8100r≠1 or v8700r≠1
```

**Step 7.** Identify and examine potential outliers of `pcdnfood_hh` and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `pcdnfood_hh`. Variables `mean_pcdnfood_hh`, `sd_pcdnfood_hh`, and `out_pcdnfood_hh` are created in this step.

**Step 8.** Calculate the local median of `pcdnfood_hh` and replace confirmed outliers with the median. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdnfood_hh`. Variable `med_pcdnfood_hh` is created in this step.

**Step 9.** Set `pcdnfood_hh` to 0 if `pcdnfood_hh` is still missing. A missing value means that the household either did not purchase any of these items in the 3 months preceding the survey or the household is missing consumption information. Households that legitimately did not make any purchases of regular non-food, non-durable goods items in the past 3 months also have their PCD expenditures set to zero.

```
Replace          pcdnfood_hh=0 if pcdnfood_hh=missing
```

**Step 10.** Rename `pcdnfood_hh` to `pcd_nfood_3m`, keep necessary variables, and save the data file.

```
Rename variable  pcdnfood_hh to be pcd_nfood_3m
Keep variables   hhea hhnum pcd_nfood_3m
Label variable   "PCD non-food consumption expenditure on item,
                  past 3 months"
Save             "FTF ZOI Survey [Country] [Year] pov4_nfood_3m"
```

### 10.2.8 Occasional non-food, non-durable goods of the past 1 year (12 months)

The step-by-step procedures to calculate the PCD non-food, non-durable goods consumption expenditures during the 1 year (12 months) preceding the survey follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov5\_nfood\_1y*. Variables from Sub-module 8.5 are used for the calculations.

**Step 1.** Prepare the data.

**Step 1a.** Load the household-level analytic data file, keep relevant variables required for this section, namely cluster number (`hhea`), household number (`hhnum`), geographic administrative units (`c05` and `c06`), de jure household size (`hhsz_dj`), all variables from Sub-module 8.5 (variables beginning with ``v8501_'`, ``v8502_'` and ``v8503_'`), and the Module 8 outcome variables (`v8100r` and `v8700r`), and drop households that did not complete Sub-modules 8.2-8.7. Flag households if none of the items were purchased in the 1 year preceding the survey.

```
Load            "FTF ZOI Survey [COUNTRY] [YEAR] household data
                  analytic"
Keep variables   hhea hhnum c05 c06 hhsz_dj v8100r v8700r, and
                  all variables starting with `v8501_' `v8502_' and
                  `v8503_'
```

```
Drop record if v8700r#1
```

**Step 1b.** Create a variable (**v8502miss**) to flag households if none of the items in Sub-module 8.5 were consumed in the 1 year preceding the survey. [XXX] should be the equal to the number of items included in Sub-module 8.5; adapt the template syntax to include all items.

```
Set v8502miss=0
Replace v8502miss=v8502miss+1 if v8502_001≥2 or missing
Replace v8502miss=v8502miss+1 if v8502_002≥2 or missing
Replace v8502miss=v8502miss+1 if v8502_003≥2 or missing
...
Replace v8502miss=v8502miss+1 if v8502_[XXX]≥2 or missing

Label variable "Number of v8502 items with consumption not
reported or missing"
```

If no consumption was reported or consumption information is missing across all items included in Sub-module 8.5, **v8502miss** will be equal to the total number of items in Sub-module 8.5.

Drop all households from the working dataset that reported no consumption or that are missing consumption information across all items. These dropped households are added back into the working dataset in Step 6.

```
Drop record if v8502miss=[XXX]
```

**Step 1c.** Change the level of observation in the data from household level to item level, as explained in **Protocol c**. Convert the variables that start with the following “stubs” from wide format (household level) to long format (item level): **v8501\_**, **v8502\_**, and **v8503\_** and then remove the ‘\_’ suffix from all item variable names.

```
Reshape wide to long
variable stubs v8501_ v8502_ v8503_
index variables hhea hhnum c05 c06 hhsize_dj
new variable j
Rename variable v8501_ to be v8501
Rename variable v8502_ to be v8502
Rename variable v8503_ to be v8503
```

**Step 1d.** Drop items that are not consumed.

```
Drop record if v8502≥2 or missing
```

**Step 1e.** Create new variables to use in the analysis (**item** and **totnfood\_item**).

```
Set item=v8501
Label variable "Non-food item"

Set totnfood_item=v8503
```

**Label variable** "Expenditure on non-food item, past 1 year"

**Step 1f.** Drop items that do not contribute to household consumption or are lumpy and related to rare events, such as legal fees, marriages, funerals, and hospitalization, even if survey respondents report expenditures on these items.<sup>66</sup> In the parallel survey questionnaire, these items are numbered 8299 through 8309. The item numbers should be adjusted to the country-specific questionnaire.

**Drop** record if item=8299 or 8300 or 8301 or 8302 or  
8303 or 8304 or 8305 or 8306 or 8307 or 8308 or  
8309

**Step 1g.** Save the data to a temporary data file as *FTF ZOI Survey [COUNTRY] [YEAR] temp\_pov5*, while consumption expenditures from Sub-module 8.5(2) are calculated. The temporary data file will be merged back into the working dataset in Step 1h.6.

**Save** "FTF ZOI Survey [COUNTRY] [YEAR] temp\_pov5"

**Step 1h.** Prepare the Sub-module 8.5(2) data to calculate consumption expenditures for items that may or may not have been purchased in the 12 months preceding the survey. In the core parallel survey questionnaire for the P2-ZOI Midline Survey, these items are numbered 8319 to 8321. The item numbers should be adjusted to the country-customized questionnaire.

**Step 1h.1.** Load the household-level analytic data file, keep relevant variables, and drop households that did not complete Sub-modules 8.2-8.7.

**Load** "FTF ZOI Survey [COUNTRY] [YEAR] household data  
analytic"  
**Keep variables** hhea hhnum c05 c06 hhsize\_dj v8504\_1-v8508\_3  
v8100r v8700r  
**Drop** record if v8700r≠1

Flag households if none of the items were purchased or gathered in the past 12 months.

**Set** v8505miss=0  
**Replace** v8505miss=v8505miss+1 if v8505\_001≥2 or missing  
**Replace** v8505miss=v8505miss+1 if v8505\_002≥2 or missing  
**Replace** v8505miss=v8505miss+1 if v8505\_003≥2 or missing

**Label variable** "No consumption reported or missing across all  
v8505 variables"

If no consumption was reported or consumption information is missing across all items included in Sub-module 8.5(2), **v8505miss** will be equal to 3, assuming that the three items in this sub-module of the core questionnaire are included in the country-customized questionnaire.

---

<sup>66</sup> See Section 10.1.3.2, Non-food, non-durable goods.

**Step 1h.2.** Drop all flagged households from the working dataset.

```
Drop          record if v8505miss=total number of Sub-module
                8.5(2) items
```

**Step 1h.3.** Change the level of observation in the data from household level to item level, as explained in **Protocol c**. Convert the variables that start with the following “stubs” from wide format (household level) to long format (item level): `v8504_`, `v8505_`, `v8505a_`, `v8505b_`, `v8506_`, `v8507_`, `v8508_` and then remove the ‘\_’ suffix from all item variable names.

```
Reshape          wide to long
  variable stubs  v8504_ v8505_ v8505a_ v8505b_ v8506_ v8507_
                  v8508_
  index variables hhea hhnum c05 c06 hhsize_dj
  new variable    j
  Label variable  j      "Item number"

Rename variable  v8504_  to be v8504
Rename variable  v8505_  to be v8505
Rename variable  v8505a_ to be v8505a
Rename variable  v8505b_ to be v8505b
Rename variable  v8506_  to be v8506
Rename variable  v8507_  to be v8507
Rename variable  v8508_  to be v8508
```

**Step 1h.4.** Drop items that are not consumed.

```
Drop          record if v8505=2 or missing
```

**Step 1h.5.** Create new variables to use in analysis (`item` and `totnfood_item`).

```
Set          item=v8504
Label variable "Non-food item"

Set          totnfood_item=v8507 if v8506=1
Replace      totnfood_item=v8508 if v8506=2
Label variable "Expenditure on non-food item, past 1 year"
```

**Note:** Question V8506 in the core parallel survey questionnaire for the P2-ZOI Midline Survey does not allow for an item to be both gathered and purchased. However, if the country-customized questionnaire does allow for items to be both gathered and purchased, add a line of syntax to replace `totnfood_item` by the sum of `v8507` and `v8508`.

**Step 1h.6.** Drop unnecessary variables and append *FTF ZOI Survey [COUNTRY] [YEAR] temp\_pov5* data file to the working dataset.

```
Drop variables  v8505a v8505b v8506 v8507 v8508
```



**Append**                      **"FTF ZOI Survey [COUNTRY] [YEAR] temp\_pov5" to working dataset**

**Step 1i.** Check `totnfood_item` values for each item and set the values to missing if any of them have one of the following values: refused (999999997), don't know (999999998), or missing (999999999). **Note:** Ensure that the number of digits in the syntax for the refused, don't know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

**Set**                      `totnfood_item=missing if totnfood_item≥999999997`

**Step 2.** Calculate PCD expenditure on each item in the 12 months preceding the survey (`pcdnfood_item`) by dividing `totnfood_item` by the number of de jure household members and by 365 days.

**Set**                      `pcdnfood_item=(totnfood_item÷hhsize_dj÷365)`  
**Label variable**      `"PCD non-food consumption expenditure on item, past 1 year"`

**Step 3.** Identify and examine potential `pcdnfood_item` outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `pcdnfood_item`. Variables `mean_pcdnfood_item`, `sd_pcdnfood_item`, and `out_pcdnfood_item` are created in this step.

**Step 4.** Calculate the local median of `pcdnfood_item` and replace verified outliers with the local median. Follow **Protocol b** for calculating median values, replacing ``var'` with `pcdnfood_item`. Variable `med_pcdnfood_item` is created in this step.

**Step 5.** Sum across the `pcdnfood_item` values by household to create a variable that captures total household-level PCD non-food, non-durable expenditure during the 1 year preceding the survey (`pcdnfood_hh`), and keep only one record per household (the first record if there are multiple records).

**By hhea hhnum:**      `Set pcdnfood_hh=sum(pcdnfood_item)`  
**By hhea hhnum:**      `Keep if _n=1`  
**Label variable**      `"PCD total HH non-food consumption expenditure, past 1 year"`

**Step 6.** Add the households that were dropped in Step 1 because they did not have expenditures on any of the items listed in the sub-module back into the working dataset by merging the working dataset with the household data analytic file using cluster number and household number (`hhea` and `hhnum`) as the key matching variable. Keep only de jure household size (`hhsize_dj`), geographic administrative units (`c05` and `c06`), and the Module 8 outcome variables (`v8100r` and `v8700r`) from the household data analytic file. Drop households that did not complete Sub-modules 8.2-8.7.

**Merge**                      `1 to 1`  
**key variables**            `hhea hhnum`

```

data file      "FTF ZOI Survey [COUNTRY] [YEAR] household
               data analytic"
keep variables hhsiz_dj c05 c06 v8100r v8700r
Drop          record if v8700r≠1

```

**Step 7.** Identify and verify potential `pcdnfood_hh` outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `pcdnfood_hh`.

Variables `mean_pcdnfood_hh`, `sd_pcdnfood_hh`, and `out_pcdnfood_hh` are created in this step.

**Step 8.** Calculate the local median of `pcdnfood_hh` and replace with the local median if the flagged value is a confirmed outlier. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdnfood_item`. Variable `med_pcdnfood_hh` is created in this step.

**Step 9.** Set `pcdnfood_hh` to 0 if it is still missing. A missing value means that the household is missing Sub-module 8.5 consumption information and it could not be imputed. Households that legitimately did not make any purchases on regular non-food, non-durable goods items in in the past 12 months also have their PCD expenditures set to zero.

```

Replace      pcdnfood_hh=0 if pcdnfood_hh=missing

```

**Step 10.** Rename the `pcdnfood_hh` variable to be `pcdnfood_1y`. Keep only the renamed PCD non-food consumption expenditure, cluster number, and household number variables, and save the data.

```

Rename variable pcdnfood_hh to be pcdnfood_1y
Keep variables  hhea hhnum pcdnfood_1y
Save           "FTF ZOI Survey [COUNTRY] [YEAR] pov5_nfood_1y"

```

### 10.2.9 Consumption expenditures on durable goods owned by the household

The step-by-step procedures to calculate the PCD durable good consumption expenditures follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov6\_asset*. Variables from Sub-module 8.7 are used for the calculations.

**Step 1.** Prepare the data.

**Step 1a.** Load the household-level analytic data file, keep only the variables required for this section, namely cluster number (`hhea`), household number (`hhnum`), geographic administrative units (`c05` and `c06`), de jure household size (`hhsiz_dj`), all Sub-module 8.7 variables (variables starting with `'v870'`), and the Module 8 outcome variables (`v8100r` and `v8700r`), and drop households that did not complete Sub-modules 8.2-8.7.

```

Load        "FTF ZOI Survey [COUNTRY] [YEAR] household data
            analytic"
Keep        hhea hhnum c05 c06 hhsiz_dj v8100r v8700r, and
            all variables starting with `v870'
Drop        record if v8700r≠1

```

**Step 1b.** Create a variable (**v8702miss**) to flag households if the household owned none of the durable goods in Sub-module 8.7. [XXX] should be the equal to the number of items included in Sub-module 8.5; adapt the template syntax to include all items.

```
Set          v8702miss=0
Replace      v8702miss=v8702miss+1 if v8702_001≥2 or missing
Replace      v8702miss=v8702miss+1 if v8702_002≥2 or missing
Replace      v8702miss=v8702miss+1 if v8702_003≥2 or missing
...
Replace      v8702miss=v8702miss+1 if v8702_[XXX]≥2 or missing

Label variable "Number of v8702 items with consumption not
               reported or missing"
```

If none of the durable goods in Sub-module 8.7 were owned or information is missing across all durable goods, **v8702miss** will be equal to the number of durable goods in Sub-module 8.7.

Drop all flagged households from the working dataset. These dropped households will be added back into the dataset in Step 13.

```
Drop          record if v8702miss=[XXX]
```

**Step 1c.** Change the level of observation in the data from household level to item level, as explained in **Protocol c**. Convert the variables that start with the following “stubs” from wide format (household level) to long format (item level): **v8701\_**, **v8702\_**, **v8703\_**, **v8704\_**, **v8705\_**, and **v8706\_** and then remove the ‘\_’ suffix from all item variable names.

```
Reshape      wide to long
variable stubs v8701_ v8702_ v8703_ v8704_ v8705_ v8706_
index variables hhea hhnum c05 c06 hhsize_dj
new variable j
Rename variable v8701_ to be v8701
Rename variable v8702_ to be v8702
Rename variable v8703_ to be v8703
Rename variable v8704_ to be v8704
Rename variable v8705_ to be v8705
Rename variable v8706_ to be v8706
```

**Step 1d.** Drop durable goods that are not owned.

```
Drop          record if v8702≥2 or missing
```

**Step 1e.** Create new variables for analysis (**item**, **currvalue**, **purcvalue**, **assetage**, and **number\_item**). If more than one durable good of a kind is owned, **currvalue**, **purcvalue**, and **assetage** are set equal to the average across the multiple durable goods. If a durable good has been owned for less than a year, its **assetage** is set equal to 1 year.

```

Set          item=v8701
Label variable "Asset"

Set          currvalue=v8705
Label variable "Current value of asset, average if multiple"

Set          purcvalue=v8706
Label variable "Amount paid for asset, average if multiple"

Set          assetage=v8704
Replace      assetage=1 if v8703≠0 and v8706>0 & v8704=0
Label variable "Asset age, average if multiple"

Set          number_item=v8703
Label variable "Number of assets owned"

```

**Step 1f.** Check current values, purchase values, and durable good age values, and set the following values to missing: refused (997 or 999999997), don't know (998 or 999999998), and missing (999 or 999999999). **Note:** Ensure that the number of digits in the syntax for the refused, don't know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

```

Replace      assetage=missing if assetage≥997
Replace      currvalue=missing if currvalue≥999999997
Replace      purcvalue=missing if purcvalue≥999999997

```

**Step 2.** Identify and verify potential outliers of asset age (**assetage**), current value (**currvalue**), and purchase value (**purcvalue**) and set confirmed outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing **`var'** with **assetage**, **currvalue**, and **purcvalue**. Variables **mean\_assetage**, **mean\_currvalue**, **mean\_purcvalue**, **sd\_assetage**, **sd\_currvalue**, **sd\_purcvalue**, **out\_assetage**, **out\_currvalue**, and **out\_purcvalue** are created in this step.

**Step 3.** Calculate the local median of current value (**currvalue**) and purchase value (**purcvalue**) and replace the confirmed outliers with the local median. Follow **Protocol b** to calculate median values, replacing **`var'** with **currvalue** and with **purcvalue**. Variables **med\_currvalue** and **med\_purcvalue** are created in this step.

**Step 4.** Calculate the depreciation rate (**dep**) of each durable good. Investigate depreciation values that are less than 0 or very high. The depreciation rate should not be negative (i.e., the current value of a durable good should not exceed the purchase price, as may happen with an investment asset). Set **dep** to missing if **currvalue** is greater than **purcvalue**. It will be replaced by the median in Step 6.

```

Set          dep=missing
Replace      dep=1-((currvalue÷purcvalue)^(1÷assetage)) if
              assetage≠0
Replace      dep=missing if currvalue>purcvalue

```

Label variable "Depreciation rate of each item owned (average)"

**Step 5.** Identify and verify potential depreciation rate (**dep**) outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with **dep**. Variables **mean\_dep**, **sd\_dep**, and **out\_dep** are created in this step.

**Step 6.** Calculate the local median of the **dep** variable and replace with the local median if the flagged value is a confirmed outlier. Follow **Protocol b** to calculate median values, replacing ``var'` with **dep**. Variable **med\_dep** is created in this step.

**Step 7.** Calculate the annual "rental price" of each durable good (**totrent\_item**) by multiplying **currvalue** by the depreciation rate and the average real interest rate (**avgintrate**). The average interest rate is country-specific and should be calculated over 10 to 20 years, depending on data availability. The World Bank's World Development Indicators database<sup>67</sup> has a real interest rate series for most Feed the Future countries. If no information for the country of the parallel survey is available from the World Bank, a real interest rate can be calculated from data on nominal interest rate and CPI from the country's central government. Use as many years as possible to calculate an average. Keep detailed records of the procedures followed.

Set **avgintrate=XXX** [Survey country's average real interest rate calculated using available data]  
Label variable "Average interest rate, [Country] over [X] years"

Set **totrent\_item1=currvalue\*(avgintrate+dep)**  
Label variable "Annual rental price of durable goods"

**Step 8.** Calculate the annual rental equivalent of the durable good (**totrent\_item2**) by multiplying **totrent\_item1** by the number of durable goods of the same kind the household owns (**v8703**).

Set **totrent\_item2=totrent\_item1\*v8703**  
Label variable "Annual rental equivalent of durable goods"

**Step 9.** Calculate the PCD rental equivalent of durable goods (**pcdrent\_item**) by dividing **totrent\_item2** by the number of de jure household members and by 365 days.

Set **pcdrent\_item=totrent\_item2÷hhsizedj÷365**

**Step 10.** Identify and examine potential **pcdrent\_item** outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with **pcdrent\_item**. Variables **mean\_pcdrent\_item**, **sd\_pcdrent\_item**, and **out\_pcdrent\_item** are created in this step.

---

<sup>67</sup> <https://datatopics.worldbank.org/world-development-indicators/>

**Step 11.** Calculate the local median of the `pcdrent_item` variable and replace with the local median if the flagged value is a confirmed outlier. Follow **Protocol b** to calculate median values, replacing ``var'` with `pcdrent_item`. Variable `med_pcdrent_item` is created in this step.

**Step 12.** Sum across all `pcdrent_item` values by household to create a variable that captures total household-level PCD durable good rental equivalent expenditure (`pcdrent_hh`). Keep only one record per household (the first record if there are multiple records).

```
By hhea hhnum:   Set pcdrent_hh=sum(pcdrent_item)
By hhea hhnum:   Keep if _n=1
Label variable   "Per capita daily household asset rental
                  equivalent"
```

**Step 13.** Add the households that were dropped in Step 1 because they reported to not own any durable goods by merging the working dataset with the household data analytic file using cluster number (`hhea`) and household number (`hhnum`) as the key matching variables. Keep only household size (`hhsizes_dj`), and geographic administrative units (`c05` and `c06`), and the Module 8 outcome variables (`v8100r` and `v8700r`) from the household data analytic file. Drop households that did not complete Sub-modules 8.2-8.7 and set `pcdrent_hh` to 0 for all households added back into the dataset.

```
Merge           1 to 1
  key variables  hhea hhnum
  data file      "FTF ZOI Survey [COUNTRY] [YEAR] household
                  data analytic"
  keep variables hhsizes_dj c05 c06 v8100r v8700r
Drop            record if v8700r≠1
Replace         pcdrent_hh=0 if the household is added back
                  into the dataset
```

**Step 14.** Rename `pcdrent_hh` to be `pcd_asset`. Keep only the renamed asset consumption expenditure, cluster number, and household number variables, and save the data file.

```
Rename variable pcdrent_hh to be pcd_asset
Keep variables  hhea hhnum pcd_asset
Save            "FTF ZOI Survey [COUNTRY] [YEAR] pov6_asset"
```

### 10.2.10 Consumption expenditures on housing

The step-by-step procedures to calculate the PCD consumption expenditures on housing follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov7\_housing*. Variables from Sub-module 8.6 are used for the calculations.

**Step 1.** Prepare the data.

**Step 1a.** Load the household-level analytic data file, keep relevant variables required for this section, namely cluster number (`hhea`), household number (`hhnum`), geographic administrative units (`c05` and `c06`), de jure household size (`hhsizes_dj`), all sub-module 8.6 variables (variables beginning with

'v860'), and the Module 8 outcome variables (v8100r and v8700r), and drop households that did not complete Sub-modules 8.2-8.7.

```
Load          "FTF ZOI Survey [COUNTRY] [YEAR] household data
              analytic"
Keep          hhea hhnum c05 c06 hhsize_dj v8100r v8700r, and
              all variables starting with 'v860'
Drop          record if v8700r#1
```

**Step 1b.** Create analytic variables v8602x, v8603x, v8604ax, v8605ax, v8608x, and v8609x that are the same as their corresponding survey variables but with the following values set to missing: refused (999999997), don't know (999999998), or missing (999999999). **Note:** Ensure that the number of digits in the syntax for the refused, don't know, and missing values is appropriate given the country-customized questionnaire; their length can vary.

```
Set          v8602x=v8602
Replace      v8602x=missing if v8602>999999996
Label variable "Value of dwelling - selling"

Set          v8604ax=v8604a
Replace      v8604ax=missing if v8604a>999999996
Label variable "Value of dwelling - renting (number)"

Set          v8605ax=v8605a
Replace      v8605ax=missing if v8605a>999999996
Label variable "Amount of rent payments (number)"

Set          v8608x=v8608
Replace      v8608x=missing if v8608>999999996
Label variable "Amount of mortgage payments"

Set          v8609x=v8609
Replace      v8609x=missing if v8609>999999996
Label variable "Amount spent on house maintenance in past month"

Set          v8603x=v8603
Replace      v8603x=missing if v8603>996
Label variable "Age of house in years"
```

**Step 2.** Calculate the PCD rental equivalent (pcdhouse1) for households living in rented homes (v8601=5). Divide the amount paid for rent (v8605ax) by the number of de jure household members (hhsize\_dj) and by the number of days for which the rent applies.

```
Set          pcdhouse1=missing
Replace      pcdhouse1=v8605ax÷hhsize_dj if v8601=5 and
              v8605b=1 and v8605ax#missing [per day]
```

```

Replace      pcdhouse1=v8605ax÷hhsizе_dj÷7 if v8601=5 and
              v8605b=2 and v8605ax≠missing [per week]
Replace      pcdhouse1=v8605ax÷hhsizе_dj÷30.4 if v8601=5 and
              v8605b=3 and v8605ax≠missing [per month]
Replace      pcdhouse1=v8605ax÷hhsizе_dj÷365 if v8601=5 and
              v8605b=4 and v8605ax≠missing [per year]
Label variable "Per capita daily rent, rented home"

```

**Step 3.** Calculate the PCD rental equivalent (**pcdhouse2**) for households living in employer-provided or free homes (**v8601=3** or **4**). Divide the rent value (**v8604ax**) by the number of de jure household members (**hhsizе\_dj**) and by the number of days for which the rent value applies.

```

Set          pcdhouse2=missing
Replace      pcdhouse2=v8604ax÷hhsizе_dj if v8604b=1 and
              (v8601=3 or v8601=4) [per day]
Replace      pcdhouse2=v8604ax÷hhsizе_dj÷7 if v8604b=2 and
              (v8601=3 or v8601=4) [per week]
Replace      pcdhouse2=v8604ax÷hhsizе_dj÷30.4 if v8604b=3 and
              (v8601=3 or v8601=4) [per month]
Replace      pcdhouse2=v8604ax÷hhsizе_dj÷365 if v8604b=4 and
              (v8601=3 or v8601=4) [per year]
Label variable "Per capita daily rental equivalent,
              free/provided home"

```

**Step 4.** Create a new variable to capture total PCD housing expenditure (**pcdhouse3**) by adding **pcdhouse1** and **pcdhouse2**.

```

Set          pcdhouse3=pcdhouse1+pcdhouse2
Replace      pcdhouse3=missing if pcdhouse1=missing and
              pcdhouse2=missing
Label variable "PCD housing consumption expenditure for rented &
              provided/free homes"

```

**Step 5.** Identify and examine potential **pcdhouse3** outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with **pcdhouse3**. Variables **mean\_pcdhouse3**, **sd\_pcdhouse3**, and **out\_pcdhouse3** are created in this step.

**Step 6.** Calculate the local median of **pcdhouse3** and replace with the local median if the flagged value is a confirmed outlier. Follow **Protocol b** to calculate median values, replacing ``var'` with **pcdhouse3**. Variable **med\_pcdhouse3** is created in this step.

**Step 6a.** Set **pcdhouse3** to missing if **pcdhouse3** is 0 to avoid indeterminate value for the hedonic regression (in log linear form).

```

Replace      pcdhouse3=missing if pcdhouse3=0

```



**Step 7.** Use a hedonic regression model to impute values for households that are missing a PCD housing expenditure value (see Section 10.1.1.7).<sup>68</sup>

A hedonic regression model is used to impute a rental value of owner-occupied housing and for missing rental values. The step-wise, log-linear model described in Section 10.1.1.7 is one of the possible specifications for the hedonic regression, and the data analysis team may want to use a different specification, such as a double-log, to better fit the data. A common list of explanatory variables is used in this syntax, but other variables, including external ones, could be used to improve the predictability of the model.

**Step 7a.** Keep only variables relevant for the hedonic regression in the working dataset and merge these with the household data analytic dataset. The hedonic regression model uses household-level variables as explanatory variables

```
Keep variables      hhea hhnum pcdhouse1 pcdhouse2 pcdhouse3
Merge              1 to 1
  key variables      hhea hhnum
  data file          "FTF ZOI Survey [COUNTRY] [YEAR] household
                    data analytic"
  keep variables     all variables
```

Drop households that did not complete Module 8.

```
Drop              record if v8100r#1 or v8700r#1
```

**Step 7b.** Create a dependent variable for the log linear model (`log_house`).

```
Set              log_house=log(pcdhouse3)
```

**Step 7c.** Create independent variables to include in the regression model using the household data analytic data file. The choice of independent variables will be to some extent country-specific. In this step, `dw_finishedroof`, `dw_finished_floor`, `dw_finished_wall`, and `access_electric` are created as example independent variables.

```
Set              dw_finishedroof=0 if dw_finishedroof#missing
Replace          dw_finishedroof=1 if dw_roof=3
Label values     0 "No"
                1 "Yes"
Label variable    "HH has finished roofing"

Set              dw_finishedfloor=0 if dw_finishedfloor#missing
Replace          dw_finishedfloor=1 if dw_floor=3
Label values     0 "No"
                1 "Yes"
```

<sup>68</sup> If rental data are too deficient to do a hedonic regression, consult with the United States Agency for International Development to determine an alternative approach.

```

Label variable  "HH has finished flooring"

Set            dw_finishedwall=0 if dw_finishedwall#missing
Replace        dw_finishedwall=1 if dw_wall=3
Label values   0 "No"
               1 "Yes"

Label variable  "HH has finished exterior walls"

Set            access_electric=0 if access_electric#missing
Replace        access_electric=1 if v222a=1
Label values   0 "No"
               1 "Yes"

Label variable  "HH has electricity"

```

The following two variables are already created in the household analytic file:

```

roomsleep      "Number of sleeping rooms"
h2o_improved    "Household uses an improved water source"

```

**Step 7d.** Keep the variables needed for the stepwise regression in Step 7e and save the data as a temporary data file, *FTF ZOI Survey [COUNTRY] [YEAR] temp\_house.dta*.

```

Keep variables  hhea hhnum c05 c06 hhsize_dj log_house, all
                variables beginning with 'pcdhouse' and 'v860,'
                all variables created in Step 7c, and all other
                variables needed for the stepwise regression

Save            "FTF ZOI Survey [COUNTRY] [YEAR] temp_house"

```

**Step 7e.** Run a stepwise regression to find the best fit with the objective of maximizing the  $R^2$ . Multi-collinearity is likely to be present, but because the purpose of the model is to predict the dependent variable (to impute a rent-equivalent), the contribution or significance of individual independent variables should not be much of a concern.

```

Load           "FTF ZOI Survey [COUNTRY] [YEAR]
                temp_house"

Run             OLS regression
  dependent variable  log_house
  independent variables  dw_finishedroof dw_finishedfloor
                        dw_finishedwall access_electric
                        roomsleep h2o_improved

Predict        rstudent rstudent
Detect         outliers
Set            outlier=1 if (rstudent≥3 and
                rstudent<missing) or rstudent≤3

```

Run	OLS regression again, excluding outliers (if outlier=missing)
Predict	log_model
Save	predicted values as "log_model"

**Step 7f.** Create a variable of the predicted values (**model**) by taking the inverse of the **log\_model** values.

Set	<b>model=exp(log_model)</b>
-----	-----------------------------

**Step 8.** Create a new variable to capture PCD use-value of housing (**pcdhouse**). Set the value equal to the predicted value of housing from the regression model if **pcdhouse** is missing.

Set	<b>pcdhouse=pcdhouse3</b>
Replace	<b>pcdhouse=model if pcdhouse3=missing and (v8601=1,2, or 8) or (pcdhouse1=missing and v8601=5) or (pcdhouse2=missing and v8601=3 or 4)</b>
Label variable	<b>"Per capita daily use-value of housing, plus predicted"</b>

**Step 9.** Calculate PCD expenditure on repairs and maintenance to the house.

Reported expenditures on repairs and maintenance should be examined carefully to ensure that they are realistic, based on the PCD use-value of the dwelling and compared to other various household expenditures, so that they do not unduly inflate housing PCD expenditures.

**Step 9a.** Create a variable to capture expenditures on dwelling repairs and maintenance.

Set	<b>repair= v8609 if v8609&lt;999999996</b>
-----	--

**Step 9b.** Create a variable for PCD expenditures on repairs.

Set	<b>pcdhouse4=(repair ÷ hysize_dj ÷ 30.4)</b>
Label variable	<b>"PCD expenditure house repairs and maintenance"</b>

**Step 10.** Identify and verify potential **pcdhouse4** outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing **`var'** with **pcdhouse4**. Variables **mean\_pcdhouse4**, **sd\_pcdhouse4**, and **out\_pcdhouse4** are created in this step.

**Step 11.** Calculate the local median of the **pcdhouse4** variable and replace with the local median if the flagged value is a confirmed outlier. Follow **Protocol b** to calculate median values, replacing **`var'** with **pcdhouse4**. Variable **med\_pcdhouse4** is created in this step.

**Step 12.** Create a variable to calculate total daily per capita expenditure on housing (**pcd\_house**). Add together the PCD use-value of housing **pcdhouse** from Step 8 and PCD expenditure on repairs/maintenance **pcdhouse4**.

Set	<b>pcd_house=(pcdhouse+pcdhouse4)</b>
-----	---------------------------------------

```

Replace          pcd_house=missing if pcdhouse=missing and
                  pcdhouse4=missing
Label variable   "Per capita daily consumption expenditure on
                  housing"

```

**Step 13.** Keep only the PCD consumption expenditure on housing, cluster number, and household number variables, and save the data file.

```

Keep variables   hhea hhnum pcd_house
Save             "FTF ZOI Survey [COUNTRY] [YEAR] pov7_house"

```

### 10.2.11 Calculating the consumption aggregate

The step-by-step procedures to calculate the consumption aggregate follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. Variables created in Sections 10.2.1 to 10.2.7 are used for the calculations. **Note:** Throughout the rest of the section, “20xx” refers to survey year and “ccc” refers to the survey country.

This section describes the procedures to combine the results of the seven previous consumption expenditure sections to create a single household-level variable, the PCD consumption expenditure, or PCD consumption aggregate, which is then used to calculate the poverty indicators.

It is important to make sure that there are sufficient data to produce the aggregate variable, even if the Module 8 outcome variables `v8100x` and `v8700x` are coded as 'Complete' (01). If the PCD expenditure of a sub-module is missing or equal to 0 for a household, further investigation is required.

There are two PCD expenditure variables that cannot be missing or equal to zero.

- If `pcd_food` was missing for a household, it should have been replaced by the local median in Section 10.2.1, Step 12 because every household must have consumed food in the 7 days preceding the survey.
- If `pcd_house` was missing for a household, it should have been estimated using the hedonic regression model in Section 10.2.7, Step 8, because every household must have a dwelling.

For the other non-food, non-housing sub-modules, it may be legitimate for a household to not have had expenditures over the recall periods, but some scrutiny is advised. If the lack of expenditures appears legitimate and there are otherwise sufficient data for the household, the PCD expenditure for that sub-module should be left equal to 0. If the lack of expenditures appears questionable, especially if there are other data issues with that household, it may be preferable to drop the household from the poverty analysis.

**Step 1.** Combine the seven consumption expenditure data files created in the previous sections into one working dataset and calculate the consumption aggregate for each household. This consumption aggregate variable is expressed in local currency at ZOI Survey year prices.

**Step 1a.** Merge all seven consumption expenditure data files created in Section 10.2.1 through Section 10.2.7 with the household analytic data file using the cluster number and household number (hhea and hhnum) as matching variables. Drop households that did not complete Module 8.

```
Load "FTF ZOI Survey [COUNTRY] [YEAR] household data analytic"

Merge      1 to 1
  key variables    hhea hhnum
  data file        "FTF ZOI Survey [COUNTRY] [YEAR] pov1_food"
  keep variables   all variables
Merge      1 to 1
  key variables    hhea hhnum
  data file        "FTF ZOI Survey [COUNTRY] [YEAR]
pov2_nfood_1w"

  keep variables   all variables

Merge      1 to 1
  key variables    hhea hhnum
  data file        "FTF ZOI Survey [COUNTRY] [YEAR]
pov3_nfood_1m"
  keep variables   all variables
Merge      1 to 1
  key variables    hhea hhnum
  data file        "FTF ZOI Survey [COUNTRY] [YEAR]
pov4_nfood_3m"
  keep variables   all variables
Merge      1 to 1
  key variables    hhea hhnum
  data file        "FTF ZOI Survey [COUNTRY] [YEAR]
pov5_nfood_1y"
  keep variables   all variables
Merge      1 to 1
  key variables    hhea hhnum
  data file        "FTF ZOI Survey [COUNTRY] [YEAR]
pov6_asset"
  keep variables   all variables
Merge      1 to 1
  key variables    hhea hhnum
  data file        "FTF ZOI Survey [COUNTRY] [YEAR]
pov7_house"
  keep variables   all variables
Drop
record if v8700r#1
```

**Step 1b.** Create a total PCD non-food consumption expenditure variable (pcd\_nfood) in LCU by summing across the consumption expenditure variables that capture non-food consumption expenditures in the 1 week, 1 month, 3 months, and 1 year preceding the survey.

```
Set      pcd_nfood=pcd_nfood_1w+pcd_nfood_1m+
pcd_nfood_3m+pcd_nfood_1y
```

```
Label variable    "Per capita daily non-food consumption
                  expenditure in local currency, 20xx"
```

**Step 1c.** Create a total PCD consumption expenditure variable (`Xpc_20xx_LCU`) in LCU for the survey year by adding consumption expenditures on food, non-food, assets, and housing.

```
Set              Xpc_20xx_LCU=pcd_food_1w+pcd_nfood+
                  pcd_asset+pcd_house
Label variable    "Per capita daily consumption expenditure in
                  local currency, 20xx"
```

**Step 2.** Identify and examine potential `Xpc_20xx_LCU` outliers and set verified outliers to missing. Follow **Protocol a** to identify and examine potential outliers, replacing ``var'` with `Xpc_20xx_LCU`. Variables `mean_Xpc_20xx_LCU`, `sd_Xpc_20xx_LCU`, and `out_Xpc_20xx_LCU` are created in this step.

**Step 3.** Calculate the local median of `Xpc_20xx_LCU` (`med_Xpc_20xx_LCU`) and replace confirmed outliers with the local median. Follow **Protocol b** to calculate median values, replacing ``var'` with `Xpc_20xx_LCU`. Variable `med_Xpc_20xx_LCU` is created in this step.

Replace confirmed outliers with the local median. **Note:** There should not be any `Xpc_20xx_LCU` equal to missing. These households should have been dropped in Step 1 from the indicator analysis because of insufficient data.

```
Replace          Xpc_20xx_LCU=med_Xpc_20xx_LCU if
                  out_Xpc_20xx_LCU>0 and med_Xpc_20xx_LCU!=missing
```

### 10.2.12 Prevalence of poverty (\$1.90 per day 2011 PPP) indicator

The step-by-step procedures to calculate the prevalence of poverty, or the percentage of people living on less than \$1.90 per day 2011 PPP, follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Number of de jure household members living on less than \$1.90 per day (2011 PPP) in surveyed households
Denominator	Total number of de jure household members in surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type* Wealth quintile Shock exposure severity
Treatment of missing data	Missing observations are replaced with imputed values in the calculation of the consumption aggregate. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.

Survey variables used	hhea, wgt_hh, strata
Analytic variables used	ppp2011_ccc, cpi20xx_ccc, cpi2011_ccc, Xpc_20xx_LCU, hhsz_dj, genhhtype_dj, awiquint, shock_sev
Analytic variables created	povline190_LCU, poor190, wgt_hhmem

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Obtain the values of the CPI and PPP parameters used in the calculation of the poverty indicators at \$1.90 2011 PPP and create three variables: **cpi20xx\_ccc**, which is the survey country's CPI for the month and year in which the bulk of survey data collection occurred, **cpi2011\_ccc**, which is the survey country's CPI for 2011, and **ppp2011\_ccc**, which is the survey country's 2011 PPP conversion factor for private consumption. See **Table 6** in Section 10.1.5 for the 2011 PPP conversion factors. Document the sources of the parameter values.

```
Set          cpi20xx_ccc=xx.xx
Label variable "[Country] CPI [Month] [Year]"

Set          cpi2011_ccc=xx.xx
Label variable "[Country] 2011 CPI"

Set          ppp2011_ccc=xx.xx
Label variable "[Country] 2011 PPP conversion factor, private
consumption"
```

**Step 2.** Convert the \$1.90 per day (2011 PPP) poverty line into local, survey year currency (**povline190\_LCU**).

Multiply the \$1.90 per day (2011 PPP) poverty line by the 2011 PPP conversion rate of the ZOI Survey country (**ppp2011\_ccc**) and adjust the resulting figure for cumulative price inflation since 2011 by multiplying it by the ratio of the ZOI Survey country's CPI for the year and month of the survey (**cpi20xx\_ccc**) to the 2011 CPI (**cpi2011\_ccc**).

```
Set          povline190_LCU=(1.90*ppp2011_ccc)*
(cpi20xx_ccc÷cpi2011_ccc)
Label variable "$1.90 poverty line in local currency at time of
survey"
```

**Step 3.** Create a variable (**poor190**) that flags households that have a PCD consumption aggregate below the \$1.90 poverty line at 2011 PPP expressed in local currency.

```
Set          poor190=0
Replace      poor190=1 if Xpc_20xx_LCU<povline190_LCU
Label values 0 "HH members living on or above $1.90 poverty
line (2011 PPP)"
1 "HH members living below $1.90 poverty line
(2011 PPP)"
```

```
Label variable "HH members living below $1.90 poverty line (2011 PPP)"
```

**Step 4.** Create a de jure household member weight and use it to calculate the prevalence of poverty estimates.

**Step 4a.** Create the de jure household member weight (**wgt\_hhmem**) variable, using the household weight variable (**wgt\_hh**) and the number of de jure household members variable (**hhsiz\_dj**), and apply it to the data, accounting for the complex survey design. (Sample code uses Stata syntax.)

```
Set          wgt_hhmem=wgt_hh*hhsiz_dj
svyset       hhea [pw=wgt_hhmem], strata(strata)
```

**Step 4b.** Calculate the prevalence of poverty in the ZOI population—that is, the percentage of the ZOI population that lives below the \$1.90 poverty line—using the **poor190** analytic variable. Repeat the calculation using the gendered household type, wealth quintile, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svy: tab poor190
svy: tab poor190 genhhstype_dj, col
svy: tab poor190 awiquint, col
svy: tab poor190 shock_sev, col
```

### 10.2.13 Depth of poverty of the poor (\$1.90 per day 2011 PPP) indicator

The step-by-step procedures to calculate the depth of poverty of the poor using the \$1.90 per day 2011 PPP poverty line follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Total shortfall of de jure household members living below the \$1.90 per day (2011 PPP) poverty line relative to the poverty line
Denominator	Total number of de jure household members living below the \$1.90 per day (2011 PPP) poverty line in surveyed households
Unit of measure	Mean percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type* Wealth quintile Shock exposure severity
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<b>hhea, wgt_hh, strata</b>



Analytic variables used	<code>Xpc_20xx_LCU</code> , <code>povline190_LCU</code> , <code>genhhstype_dj</code> , <code>awiquint</code> , <code>shock_sev</code>
Analytic variables created	<code>povdepth190</code>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** See Step 2 in Section 10.2.12 to ensure that the `povline190_LCU` variable is created

**Step 2.** Create a variable measuring an individual's proportional shortfall from the poverty line by subtracting the PCD consumption expenditure in local currency (`Xpc_20xx_LCU`) from the \$1.90 poverty line expressed in local currency in the month and year of the survey (`povline190_LCU`) and then dividing by the same poverty line (`povline190_LCU`). Multiply the result by 100 to obtain the depth of poverty expressed as a percentage of the poverty line (`povdepth190`).

```
Set          povdepth190=(povline190_LCU-Xpc_20xx_LCU) ÷
              povline190_LCU*100
```

**Step 3.** Exclude households if their household PCD consumption expenditure is equal to or exceeds the \$1.90 poverty threshold (`povline190_LCU`) because this indicator measures the depth of poverty of only the poor.

```
Replace      povdepth190=missing if
              Xpc_20xx_LCU ≥ povline190_LCU
Label variable "Depth of poverty of the poor, $1.90 poverty
              line"
```

**Step 4.** Calculate the depth of poverty of the poor in the ZOI as the mean of the `povdepth190` analytic variable using the de jure household member weight in the complex survey design specification (see Section 10.2.12, Step 4). Repeat the calculation using the gendered household type, wealth quintile, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svy: mean povdepth190
svy: mean povdepth190, over(genhhstype_dj)
svy: mean povdepth190, over(awiquint)
svy: mean povdepth190, over(shock_sev)
```

### 10.2.14 Percent of people who are 'near-poor' (\$1.90 per day 2011 PPP) indicator

The step-by-step procedures to calculate the percent of people who are near-poor based on the \$1.90 per day (2011 PPP) poverty line follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

## Definitions

Numerator	Number of de jure household members living on 100 percent to less than 125 percent of the \$1.90 (2011 PPP) per day poverty line
Denominator	Total number of de jure household members in surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type* Wealth quintile Shock exposure severity
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<code>hhea, wgt_hh, strata</code>
Analytic variables used	<code>xpc_20xx_LCU, povline190_LCU, genhhtype_dj, awiquint, shock_sev</code>
Analytic variables created	<code>nearpoor190</code>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Create a variable (`nearpoor190`) to flag households living on 100 percent to less than 125 percent of the \$1.90 per day (2011 PPP) poverty line expressed in local currency in the year and month of the survey (`povline190_LCU`, which was created in Section 10.2.12, Step 2).

```

Set          nearpoor190=0
Replace      nearpoor190=1 if (xpc_20xx_LCU≥povline190_LCU)
              and (xpc_20xx_LCU<povline190_LCU*1.25)
Label values 0 "HH members living below $1.90 poverty line or
              at or above 125% of $1.90 poverty line"
              1 "HH members living on 100% to less than 125% of
              $1.90 poverty line"
Label variable "HH members living on 100% to less than 125% of
              the $1.90 poverty line (2011 PPP)"

```

**Step 2.** Calculate the percentage of the ZOI population who are near-poor using the `nearpoor190` analytic variable and the de jure household member weight in the complex survey design specification (see Section 10.2.12, Step 4). Repeat the calculation using the gendered household type, wealth quintile, and shock severity disaggregates. (Sample code uses Stata syntax.)

```

svy: tab nearpoor190
svy: tab nearpoor190 genhhtype_dj, col
svy: tab nearpoor190 awiquint, col
svy: tab nearpoor190 shock_sev, col

```

### 10.2.15 Prevalence of poverty (\$1.25 per day 2005 PPP) indicator

The step-by-step procedures to calculate the prevalence of poverty, or the percentage of people living on less than \$1.25 per day 2005 PPP, follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8\_final*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Number of de jure household members living on less than \$1.25 per day (2005 PPP) in surveyed households
Denominator	Total number of de jure household members in surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type* Wealth quintile Shock exposure severity
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>ppp2005_ccc, cpi20xx_ccc, cpi2005_ccc, xpc_20xx_LCU, genhhtype_dj, awiquint, shock_sev</b>
Analytic variables created	<b>povline125_LCU, poor125</b>

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Obtain the values of the CPI and PPP parameters used in the calculation of the poverty indicators at \$1.25 2005 PPP and create two variables (**cpi20xx\_ccc**, was already created in Section 10.2.12, Step 1): **cpi2005\_ccc**, which is the survey country's CPI for 2005, and **ppp2005\_ccc**, which is the survey country's 2005 PPP conversion factor for private consumption. See **Table 6** in Section 10.1.5 for the 2005 PPP conversion factors. Document the sources of the parameter values.

```

Set          cpi2005_ccc=xx.xx
Label variable "[Country] 2005 CPI"

Set          ppp2005_ccc=xx.xx
Label variable "[Country] 2005 PPP conversion factor, private
               consumption"
```

**Step 2.** Convert the \$1.25 per day (2005 PPP) poverty line into local, survey year currency (**povline125\_LCU**).

First, multiply the \$1.25 per day (2005 PPP) poverty line by the 2005 PPP conversion rate of the ZOI Survey country (`ppp2005_ccc`). Then adjust the resulting figure for the cumulative price inflation since 2005 by multiplying by ratio of the ZOI Survey country's CPI for the survey month (`cpi20xx_ccc`) to the 2005 CPI (`cpi2005_ccc`).

```
Set                povline125_LCU=1.25*ppp2005_ccc*
                  (cpi20xx_ccc÷cpi2005_ccc)
Label variable     "$1.25 poverty line in local currency at the time
                  of the survey"
```

**Step 3.** Create a variable (`poor125`) that flags households that have a PCD consumption aggregate below the \$1.25 per day poverty line at 2005 PPP expressed in local currency.

```
Set                poor125=0

Replace            poor125=1 if Xpc_20xx_LCU<povline125_LCU
Label values       0 "HH members living on or above the $1.25
                  poverty line (2005 PPP)"
                  1 "HH members living below the $1.25 poverty line
                  (2005 PPP)"
Label variable     "HH living below $1.25 poverty line (2005 PPP)"
```

**Step 4.** Calculate the prevalence of poverty in the ZOI—that is, the percentage of the ZOI population that lives below the \$1.25 poverty line—using the `poor125` analytic variable and the de jure household member weight in the complex survey design specification (see Section 10.2.12, Step 3). Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

```
svy: tab poor125
svy: tab poor125 genhhtype_dj, col
svy: tab poor125 awiquint, col
svy: tab poor125 shock_sev, col
```

### 10.2.16 Depth of poverty of the poor (\$1.25 per day 2005 PPP) indicator

The step-by-step procedures to calculate the depth of poverty of the poor using the \$1.25 per day 2005 PPP poverty line follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Total shortfall of de jure household members living below the \$1.25 per day (2005 PPP) poverty line relative to the poverty line
Denominator	Total number of de jure household members living below the \$1.25 per day (2005 PPP) poverty line in surveyed households
Unit of measure	Mean percentage
Level of data	Household

Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type* Wealth quintile Shock exposure severity
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>xpc_20xx_LCU, povline125_LCU, genhhstype_dj, awiquint, shock_sev</b>
Analytic variables created	<b>povdepth125</b>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** See Step 2 in Section 10.2.12 to ensure that the **povline190\_LCU** variable is created **Step 2.** Create a variable measuring an individual's proportional shortfall from the poverty line by subtracting the PCD consumption expenditure in local currency (**xpc\_20xx\_LCU**) from the \$1.25 poverty line expressed in local currency in the month and year of the survey (**povline125\_LCU**) and then dividing by the same poverty line (**povline125\_LCU**). Multiply the result by 100 to obtain the depth of poverty expressed as a percentage of the poverty line (**povdepth125**).

**Set** 
$$\text{povdepth125} = (\text{povline125\_LCU} - \text{xpc\_20xx\_LCU}) \div \text{povline125\_LCU} * 100$$

**Step 3.** Exclude households if their PCD consumption expenditure is equal to or exceeds the \$1.25 per day (2005 PPP) poverty threshold.

**Replace** 
$$\text{povdepth125} = \text{missing if } (\text{xpc\_20xx\_LCU} \geq \text{povline125\_LCU})$$
  
**Label variable** "Depth of poverty of the poor, \$1.25 poverty line (2005 PPP)"

**Step 4.** Calculate the depth of poverty of the poor in ZOI—that is, on average, how far the poor in the ZOI are below the \$1.25 per day (2005 PPP) poverty threshold—using the **poor125** analytic variable and the de jure household member weight in the complex survey design specification (see Section 10.2.12, Step 4). Repeat using the gendered household type, wealth quintile, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svy: mean povdepth125
svy: mean povdepth125, over(genhhstype_dj)
svy: mean povdepth125, over(awiquint)
svy: mean povdepth125, over(shock_sev)
```

### 10.2.17 Percent of people who are ‘near-poor’ (\$1.25 per day 2005 PPP) indicator

The step-by-step procedures to calculate the percent of people who are near-poor based on the \$1.25 per day (2005 PPP) poverty line follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Number of de jure household members living on 100 percent to less than 125 percent of the \$1.25 (2005 PPP) per day poverty line
Denominator	Total number of de jure household members in surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type* Wealth quintile Shock exposure severity
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<code>hhea, wgt_hh, strata</code>
Analytic variables used	<code>Xpc_20xx_LCU, povline125_LCU, genhhstype_dj, awiquint, shock_sev</code>
Analytic variables created	<code>nearpoor125</code>

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create a variable to flag households living on 100 percent to less than 125 percent of the \$1.25 per day (2005 PPP) poverty line using variable `povline125_LCU`, which was created in Section 10.2.15, Step 2

```

Set          nearpoor125=0
Replace      nearpoor125=1 if (Xpc_20xx_LCU>=povline125_LCU)
              and (Xpc_20xx_LCU<povline125_LCU*1.25)
Label values 0 "HH members living below $1.25 poverty line or
              at or above 125% of $1.25 poverty line"
              1 "HH living on 100% and up to less than 125% of
              $1.25 poverty line"
Label variable "HH members living on 100% to less than 125% of
              the $1.25 poverty line (2005 PPP)"

```

**Step 2.** Calculate the percentage of people who are near-poor in ZOI—that is, the percentage of the ZOI population that lives on 100 percent to less than 125 percent of the \$1.25 per day (2005 PPP) poverty threshold—using the `nearpoor125` analytic variable and the de jure household member

weight in the complex survey design specification (see Section 10.2.12, Step 4). Repeat using the gendered household type, wealth quintile, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svy: tab nearpoor125
svy: tab nearpoor125 genhhtype_dj, col
svy: tab nearpoor125 awiquint, col
svy: tab nearpoor125 shock_sev, col
```

### 10.2.18 Prevalence of poverty (\$2.15 per day 2017 PPP) indicator

The step-by-step procedures to calculate the prevalence of poverty, or the percentage of people living on less than \$2.15 per day 2017 PPP, follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Number of de jure household members living on less than \$2.15 per day (2017 PPP) in surveyed households
Denominator	Total number of de jure household members in surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type*
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<b>hhea, wgt_hh, strata</b>
Analytic variables used	<b>ppp2017_ccc, cpi20xx_ccc, cpi2017_ccc, xpc_20xx_LCU, genhhtype_dj, awiquint, shock_sev</b>
Analytic variables created	<b>povline215_LCU, poor215</b>

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Obtain the values of the CPI and PPP parameters used in the calculation of the poverty indicators at \$2.15 2017 PPP and create two variables (**cpi20xx\_ccc**, was already created in Section 10.2.12, Step 1): **cpi2017\_ccc**, which is the survey country's CPI for 2017, and **ppp2017\_ccc**, which is the survey country's 2017 PPP conversion factor for private consumption. See **Table 6** in Section 10.1.5 for the 2017 PPP conversion factors. Document the sources of the parameter values.

```
Set          cpi2017_ccc=xx.xx
Label variable "[Country] 2017 CPI"

Set          ppp2017_ccc=xx.xx
```

```
Label variable "[Country] 2017 PPP conversion factor, private
consumption"
```

**Step 2.** Convert the \$2.15 per day (2017 PPP) poverty line into local survey year currency (`povline215_LCU`).

First, multiply the \$2.15 per day (2017 PPP) poverty line by the 2017 PPP conversion rate of the ZOI Survey country (`ppp2017_ccc`). Then adjust the resulting figure for the cumulative price inflation since 2017 by multiplying by the ratio of the ZOI Survey country's CPI for the survey month (`cpi20xx_ccc`) to the 2017 CPI (`cpi2017_ccc`).

```
Set povline215_LCU=(2.15*ppp2017_ccc)*(cpi20xx_ccc÷
cpi2017_ccc)
Label variable "$2.15 poverty line in local currency at the time
of the survey"
```

**Step 3.** Create a variable (`poor215`) that flags households living below the \$2.15 per day (2017 PPP) poverty line.

```
Set poor215=0
Replace poor215=1 if (Xpc_20xx_LCU<povline215_LCU)
Label values 0 "HH members living on or above the $2.15
poverty line (2017 PPP)"
1 "HH members living below the $2.15 poverty line
(2017 PPP)"
Label variable "HH living below $2.15 poverty line (2017 PPP)"
```

**Step 4.** Calculate the prevalence of poverty in ZOI—that is, the percentage of the ZOI population that lives below the \$2.15 poverty line—using the `poor215` analytic variable and the de jure household member weight in the complex survey design specification (see Section 10.2.12, Step 4). Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

```
svy: tab poor215
svy: tab poor215 genhhtype_dj, col
svy: tab poor215 awiquint, col
svy: tab poor215 shock_sev, col
```

### 10.2.19 Depth of poverty of the poor (\$2.15 per day 2017 PPP) indicator

The step-by-step procedures to calculate the depth of poverty of the poor using the \$2.15 per day 2017 PPP poverty line follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Total shortfall of de jure household members living below the \$2.15 per day
-----------	--



	(2017 PPP) poverty line relative to the poverty line
Denominator	Total number of de jure household members living below the \$2.15 per day (2017 PPP) poverty line in surveyed households
Unit of measure	Mean percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type*
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<code>hhea, wgt_hh, strata</code>
Analytic variables used	<code>xpc_20xx_LCU, povline215_LCU, genhhtype_dj, awiquint, shock_sev</code>
Analytic variables created	<code>povdepth215</code>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** See Step 2 in Section 10.2.18 to ensure that the `povline215_LCU` variable is created

**Step 2.** Create a variable measuring an individual's proportional shortfall from the poverty line by subtracting the PCD consumption expenditure in local currency (`xpc_20xx_LCU`) from the \$2.15 poverty line expressed in local currency in the month and year of the survey (`povline215_LCU`) and then dividing by the same poverty line (`povline215_LCU`). Multiply the result by 100 to obtain the depth of poverty expressed as a percentage of the poverty line (`povdepth215`).

```
Set          povdepth215=(povline215_LCU-xpc_20xx_LCU) ÷
              povline215_LCU*100
```

**Step 3.** Exclude households if their PCD consumption expenditure exceeds the \$2.15 per day (2017 PPP) poverty threshold.

```
Replace      povdepth215=missing if xpc_20xx_LCU>
              povline215_LCU
Label variable "Depth of poverty of the poor, $2.15 poverty
              line"
```

**Step 4.** Calculate the depth of poverty of the poor in ZOI—that is, on average, how far the poor in the ZOI are below the \$2.15 per day (2017 PPP) poverty threshold—using the `poor215` analytic variable and the de jure household member weight in the complex survey design specification (see Section 10.2.12, Step 4). Repeat using the gendered household type, wealth quintile, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svy: mean povdepth215
svy: mean povdepth215, over(genhhtype_dj)
svy: mean povdepth215, over(awiquint)
svy: mean povdepth215, over(shock_sev)
```

### 10.2.20 Percent of people who are ‘near-poor’ (\$2.15 per day 2017 PPP) indicator

The step-by-step procedures to calculate the percentage of people who are near-poor based on the \$2.15 per day (2017 PPP) poverty line follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax pov8*. The PCD consumption expenditure in local currency at ZOI Survey year prices created in Section 10.2.11 is used for the calculations.

#### Definitions

Numerator	Number of de jure household members living on 100 percent to less than 125 percent of the \$2.15 (2017 PPP) per day poverty line
Denominator	Total number of de jure household members in surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Individual (de jure household member)
Disaggregation levels	Gendered household type*
Treatment of missing data	Missing observations included in the calculation of the consumption aggregate are replaced with imputed values. Households that did not complete the consumption expenditure module (Module 8) are excluded from the indicator calculation.
Survey variables used	<code>hhea, wgt_hh, strata</code>
Analytic variables used	<code>Xpc_20xx_LCU, povline215_LCU, genhhstype_dj, awiquint, shock_sev</code>
Analytic variables created	<code>nearpoor215</code>

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create a variable (`nearpoor215`) to flag households living on 100 percent to less than 125 percent of the \$2.15 per day (2017 PPP) poverty line using variable `povline215_LCU`, which was created in Section 10.2.18, Step 2

```

Set          nearpoor215=0
Replace      nearpoor215=1 if Xpc_20xx_LCU>=povline215_LCU
              and Xpc_20xx_LCU<povline215_LCU*1.25
Label values 0 "HH members living below $2.15 poverty line or
              at or above 125% of $2.15 poverty line"
              1 "HH members living on 100% to less than 125% of
              $2.15 poverty line"
Label variable "HH members living on 100% to less than 125% of
              the $2.15 poverty line (2017 PPP)"

```

**Step 2.** Calculate the percentage of people who are near-poor in ZOI—that is, the percentage of the ZOI population that lives on 100 percent to less than 125 percent of the \$2.15 per day (2017 PPP) poverty threshold—using the `nearpoor215` analytic variable and the de jure household member weight in the complex survey design specification (see Section 10.2.12, Step 4). Repeat using the

gendered household type, wealth quintile, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svy: tab nearpoor215
svy: tab nearpoor215 genhhtype_dj, col
svy: tab nearpoor215 awiquint, col
svy: tab nearpoor215 shock_sev, col
```

### 10.3 References

Balcazar, C.F., Ceriani, L., Olivieri, S., & Ranzani, M. (2014). *Rent imputation for welfare measurement: A review of methodologies and empirical findings (English)*. Policy research working paper 7103. Washington, DC: World Bank Group. Available at:

<http://documents.worldbank.org/curated/en/313651468152726645/Rent-imputation-for-welfare-measurement-a-review-of-methodologies-and-empirical-findings>

Deaton, A., & Zaidi, S. (2002). *Guidelines for constructing consumption aggregates for welfare analysis (English)*. Living standards measurement study working paper, no. LSM 135. Washington, DC: The World Bank. Available at: <http://documents.worldbank.org/curated/en/206561468781153320/Guidelines-for-constructing-consumption-aggregates-for-welfare-analysis>

Grosh, M.E., & Munoz, J.E. (1996). *A manual for planning and implementing the living standards measurement study survey (English)*. Living standards measurement study working paper; no. LSM 126. Washington DC: The World Bank. Available at: <http://documents.worldbank.org/curated/en/363321467990016291/A-manual-for-planning-and-implementing-the-living-standards-measurement-study-survey>

Ravallion, M., Chen, S., Sangraula, P. (2009). Dollar a day revisited. *The World Bank Economic Review*, 23(2), 163–184. <https://doi.org/10.1093/wber/lhp007>

Ravallion, M., Datt, G., van de Walle, D. (1991). Quantifying absolute poverty in the developing world. *Review of Income and Wealth*, 37(4): 345–361.

## II. THE COMPARATIVE WEALTH INDEX

Although the Demographic and Health Surveys (DHS) Wealth Index, an asset-based wealth index (AWI), is useful for studying economic inequalities in a country at a given time, it cannot be directly compared across countries or over time. Hence, Rutstein and Staveteig (2014) developed a new methodology to calculate a comparative wealth index (CWI) that allows for direct comparison of economic status across countries and over time. Feed the Future adopted the CWI methodology to develop its ZOI-level indicator: the percent of households below the comparative threshold for the poorest quintile of the asset-based CWI.

This indicator reflects the percentage of households in the Feed the Future ZOI whose ownership (or lack thereof) of selected assets places the household below a fixed threshold that defines the poorest quintile (bottom 20 percent) in the comparative baseline wealth index that was used to create a cross-nationally, cross-temporally comparable AWI. The use of a fixed threshold across ZOIs is possible because the CWI is an index with a value that is relative to the baseline wealth index that is used for comparison. This means that the index score and thresholds can be compared across ZOI Surveys and over time.

This chapter is divided into two sections; the first section describes the general guidelines to construct Feed the Future's CWI indicator, and the second section outlines the step-by-step procedures to calculate the indicator.

### II.1 Guidelines to construct Feed the Future's CWI indicator

Constructing the CWI indicator involves seven key steps: (1) selection of a reference survey to serve as the point for comparison across all Feed the Future ZOI Surveys, (2) calculation of the AWI for the selected reference survey, (3) calculation of a set of anchoring points for the reference survey, (4) calculation of the AWI for the ZOI Survey being analyzed, (5) calculation of a set of anchoring points for the ZOI Survey being analyzed, (6) conversion of the AWI scores for all sampled households in the ZOI Survey being analyzed into comparable scores using the anchoring points calculated in Steps 3 and 4, and (7) determination of the percentage of households below the comparative threshold for the poorest quintile of the reference survey.

Following are detailed descriptions of these seven steps. Note that Steps 1 through 3 have already been performed and do not need to be performed again. The results already obtained will be used across all Feed the Future ZOI Surveys to calculate the CWI indicator.

**I. Selection of the reference survey:** To make survey-specific AWIs comparable with each other, reference data are required. Selection of the reference data is somewhat arbitrary, similar to the base year in a price index. For the Feed the Future CWI indicator, ZOI Survey data were not yet available, so a recently conducted DHS survey was chosen as the reference because recent DHS surveys collect the same asset and dwelling information that are collected in Feed the Future ZOI Surveys and similar household member information, and can, therefore, be used to construct the reference wealth index and anchoring points that are required to calculate the Feed the Future CWI indicator.

In the CWI study that DHS conducted, DHS selected the 2002 Vietnam survey as the reference because Vietnam's per capita income fell in the middle of the per capita incomes of DHS countries included in the study and because the survey fell in the middle of the time period being analyzed—that is, the years 2000–2010. Following the DHS approach, the Bureau for Resilience and Food Security (RFS) examined the World Bank's 2018 gross national income per capita (GNI/p) estimates in current U.S. dollars calculated using the Atlas method (World Bank Group, n.d.a) for countries classified as low-income or lower-middle-income countries by the World Bank (World Bank Group, n.d.b). The median GNI/p (Atlas method, current USD) for these countries was determined to be \$1,580. RFS then determined when countries with a GNI/p close to this median GNI/p last conducted a DHS survey. RFS chose the 2017 Senegal DHS as the reference survey because the country's 2018 GNI/p (Atlas method, current USD) was \$1,410—not far below the median—and because the survey was one of the most recent DHS surveys for which data were publicly available. The 2017 Senegal DHS was initially used as the reference survey to calculate the Feed the Future CWI indicator across all ZOI Surveys.

In 2023, the anchoring points for the reference survey were adjusted to align with the Joint Monitoring Programme for Water Supply, Sanitation and Hygiene definitions of improved drinking water sources and sanitation facilities and to account for an additional personal car variable in the 2017 Senegal DHS dataset that had not been previously included. After the updates were made and the anchoring points revised, results from some ZOI Surveys indicated that no households were in the poorest CWI quintile, even at baseline. Starting with no households in the bottom CWI quintile reduces the sensitivity and thus utility of the indicator for tracking changes in household asset wealth over time. RFS then explored the possibility of using a different reference survey; both the 2018 Zambia DHS and 2017–18 Bangladesh DHS were considered. Using the 2018 Zambia DHS as the reference survey yielded results similar to those obtained using the 2017 Senegal DHS as the reference survey. However, using the 2017–18 Bangladesh DHS as the reference survey yielded larger percentages of households in the poorest quintile across surveys, likely tied to Bangladesh's higher GNI/p in 2018 (\$1,750), compared to Senegal (\$1,410) and Zambia (\$1,430). The 2017–18 Bangladesh DHS was, therefore, chosen as the new reference survey, and the anchoring points and CWI quintile cutoffs were updated accordingly.

**2. Calculation of the AWI for the reference survey:** The AWI for the 2017–18 Bangladesh DHS calculated by DHS and included in the recode dataset available through The DHS Program's website was used (Rutstein, 2014).<sup>69</sup> All surveyed households were assigned a standardized score, generated through factor analysis, for each asset included in the index, depending on whether the household owned that asset, or in the case of sleeping arrangements, the number of de jure household members per room. These scores were then used to create the thresholds that define wealth quintiles using household weights.<sup>70</sup> The steps to calculate the AWI for the reference survey are the same steps used to calculate

---

<sup>69</sup> Note that the reference country wealth scores and quintiles were calculated using the wealth asset data file and syntax obtained directly from DHS. The reference country wealth scores and quintiles cannot be re-created exactly using survey data files available through the DHS website. In addition, decisions made about how and when to combine asset and dwelling characteristic category variables with a small number of observations also affect the wealth scores and quintiles calculated.

<sup>70</sup> These household-weighted quintiles used for the CWI indicator differ from the household member-weighted quintiles usually used for AWI quintiles. The AWI quintile estimates and disaggregate included in the Feed the Future ZOI Survey reports are generated using household member weights. Only the CWI indicator is generated using household weights.

the AWI for the ZOI Surveys, which are described in the step-by-step instructions in the next section. The reference survey AWI is calculated only one time, following the selection of the reference survey. The same reference survey AWI quintile cutpoints will be used in all Feed the Future CWI indicator calculations across all ZOI Surveys. The reference survey AWI scores are presented below for each quintile. Quintile cutoffs are also presented below. The quintile cutoffs were calculated by averaging the highest wealth score in the lower quintile and the lowest wealth score in the higher quintile.

Quintile	Reference country AWI scores	Quintile cutoffs
Lowest	-1.63452 to -0.83577	-0.835740
Second	-0.83571 to -0.55395	-0.553915
Middle	-0.55388 to -0.04444	-0.044190
Fourth	-0.04394 to 0.86411	0.864125
Highest	0.86414 to 3.9354	

**3. Calculation of anchoring points for the reference survey:** Following Rutstein and Staveteig's (2014) recommendations, two distinct methods are combined to obtain adequate anchoring points for the CWI. The first method uses information on the ownership of four assets included in DHS surveys—that is, car or truck, refrigerator, computer,<sup>71</sup> and television. The second method is based on the Unsatisfied Basic Needs (UBN) index (Feres & Mancero, 2001).

The asset anchoring points are obtained using logistic regression. The anchoring points are equal to the predicted wealth index score at which the probability of a household owning the respective item is equal to 50 percent. To obtain these scores, four logistic regression models, one for each item, were estimated separately for the reference survey.

$$\text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = \alpha + \beta(\text{AWI})$$

Where:

- p probability that the household owns the respective item,  $p = \text{prob}(Y = 1)$
- AWI wealth index score
- Y binary outcome variable indicating whether a household owns the respective item: yes (1) or no (0).

Each asset anchoring point is the predicted value of the AWI, where  $p=0.5$ , which can be calculated from the estimated parameter values for  $\alpha$  and  $\beta$ :

$$\ln\left(\frac{0.5}{1-0.5}\right) = \alpha + \beta(\text{AWI})$$

$$\ln(1) = \alpha + \beta(\text{AWI})$$

$$0 = \alpha + \beta(\text{AWI})$$

<sup>71</sup> In the DHS study, fixed telephone was used as an anchoring point, but because fixed telephones are becoming obsolete, fixed telephone is being replaced with computer for the Feed the Future CWI indicator.

$$AWI = \alpha/\beta$$

The UBN anchoring points are obtained using the UBN framework, which defines a set of basic needs and points to households that do not satisfy them. A household can have a UBN score ranging from 0 (no unmet basic needs) to 4 (unmet basic needs in all four UBN categories). **Table 11** shows the four UBN items suggested by Rutstein and Staveteig (2014) and considered in the ZOI Surveys, with some modifications to accommodate differences between DHS and ZOI Survey content.

**Table 11: Unsatisfied Basic Needs Categories and Criteria**

UBN category	UBN point assigned to household if:
<b>Inadequate dwelling construction</b>	The household resides in a dwelling with inadequate walls or floors. Inadequate walls are no walls or walls made of natural or rustic materials (e.g., dirt; cane, palm, or tree trunks; bamboo with mud; stone with mud), and inadequate floors are made of natural materials (e.g., earth, sand, dung, or palm leaves).
<b>Inadequate sanitation or water source</b>	The household lacks access to an adequate toilet facility, an adequate source of drinking water, or both. Inadequate toilet facilities include having no facility, a pit latrine without a slab, a bucket, a hanging toilet, a flush toilet to “somewhere else,” or sharing a facility—even an improved facility—with other households. Inadequate drinking water sources include unprotected wells, unprotected springs, and surface water.
<b>Overcrowded housing</b>	There are more than three de jure household members per sleeping room.
<b>High economic dependency<sup>a</sup></b>	(a) There are no de jure working-age adults (15-64 years of age) who have completed primary education in the household; OR (b) the only de jure working-age adults in the household are still in school; OR (c) there are no working-age adults in the household, and a primary adult decision-maker <sup>b</sup> has not completed primary education.

<sup>a</sup> Rutstein and Staveteig (2014) also included another condition: “if households have more than three household members per worker.” However, because the ZOI Surveys do not collect the employment status of all household members, this condition has been dropped from both the reference survey and ZOI Surveys.

<sup>b</sup> Rutstein and Staveteig (2014) used data on the head of household, but the ZOI Surveys do not collect head of household information. The ZOI Surveys do, however, collect education data on the primary adult male decision-maker and the primary adult female decision-maker in each household, so these data were used instead. Note that a household may have only a female primary adult decision-maker, only a male primary adult decision-maker, both, or neither.

Eight anchoring points—four asset and four UBN—were calculated using the reference survey data. These anchoring points, shown in **Table 12**, will be used in the Feed the Future CWI indicator calculation across all ZOI Surveys.

**Table 12: Reference Survey Anchoring Point Values**

Type	Anchoring point	Value for reference survey
Asset	Television	-0.0751299
	Refrigerator	0.5593175
	Computer	2.308742
	Car or truck	3.664623
UBN	1+ UBN	0.72244
	2+ UBNs	-0.34796
	3+ UBNs	-0.84296
	4 UBNs	-1.15814



**4. Calculation of the AWI for the ZOI Survey being analyzed:** The AWI for the ZOI Survey being analyzed is calculated following the method developed by the DHS, with a few modifications (Rutstein, 2014).<sup>72</sup> The step-by-step instructions to calculate the AWI for the ZOI Survey being analyzed are included in the step-by-step procedures.

**5. Calculation of anchoring points for the ZOI Survey being analyzed:** The four asset anchoring points and the four UBN anchoring points for the ZOI Survey being analyzed are calculated using the approach described in Step 3. The step-by-step instructions to calculate the ZOI Survey anchoring points are included in the step-by-step procedures.

**6. Conversion of the ZOI Survey AWI scores into the CWI scores:** CWI scores are calculated by first running a linear regression with the reference survey anchoring points (*basecut<sub>i</sub>*) as the dependent variables and the ZOI Survey anchoring points as the independent variables (*compcut<sub>i</sub>*) to obtain  $\alpha$  and  $\beta$ :

$$basecut_i = \alpha + \beta * compcut_i$$

Where *basecut<sub>i</sub>* is the value of the anchoring point on the baseline AWI of item *i*, and *compcut<sub>i</sub>* is the value of the anchoring point on the ZOI Survey AWI.

The constant  $\alpha$  represents the amount of adjustment of the ZOI Survey AWI relative to the Baseline Survey AWI, and  $\beta$  represents the dispersion of the ZOI Survey AWI relative to the Baseline Survey AWI.

The regression coefficient  $\beta$  and constant  $\alpha$  are then used to convert each household's AWI score into a CWI score. The CWI score for a household is equal to the AWI score for that household multiplied by the regression coefficient  $\beta$ , plus the constant:

$$CWI = \alpha + \beta * AWI$$

**7. Determination of the percentage of households below the comparative threshold for the poorest quintile of the reference survey:** Using the CWI scores calculated in the previous step, determine the percentage of households that are below the comparative threshold for the poorest quintile of the CWI. This indicator reflects the percentage of households in the ZOI whose ownership (or lack thereof) of selected assets places the household below a fixed threshold that defines the poorest quintile (bottom 20 percent) of the reference survey AWI. This fixed threshold is -0.86276845.

---

<sup>72</sup> In the ZOI Survey AWI calculations: (1) the land variable is created using only a household-level question, whereas in the DHS calculations, the land variable is created using questions in the household, men's, and women's questionnaires; (2) the house variable is created using only a household-level question included in the consumption expenditure module, whereas in the DHS calculations, the house variable is created using questions in the men's and women's questionnaires.



## 11.2 Step-by-step procedures to calculate the CWI indicator

The step-by-step procedures to calculate the AWI follows the Stata syntax in the following do files: *FTF ZOI Survey [COUNTRY] [YEAR] syntax wealthindex AWI.do* and *FTF ZOI Survey [COUNTRY] [YEAR] syntax wealthindex CWI.do*. Variables from Modules 1 and 2 are used for the calculations.

### Definitions

Numerator	Number of households below the comparative threshold for the poorest quintile of the asset-based CWI generated using household weights
Denominator	Total number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Gendered household type * Shock exposure severity
Treatment of missing data	Individual variables created from questions with missing responses are set to 0 “No” for the AWI and anchoring point calculations.
Survey variables used	See variables below—listed separately for Parts 1 and 2.
Analytic variables used	See variables below—listed separately for Parts 1 and 2.
Analytic variables created	See variables below—listed separately for Parts 1 and 2.

\*Standard Feed the Future disaggregate

### Calculations

Following is a detailed description of the steps involved in calculating the CWI indicator for the ZOI Survey being analyzed, using reference survey information that is already available. The steps are divided into two parts: AWI (Part 1) and CWI (Part 2). The step-by-step procedures for Part 1 follow the Stata syntax in the following do files: *FTF ZOI Survey [COUNTRY] [YEAR] syntax wealthindex AWI.do*.

#### Part 1. AWI

Survey variables used	<code>hhea, hhnum, ahtype, v103, v201, v202, v203, v208, v209, v211, v219, v222a-v222f, v223a-v223g, v224a, v224b<sup>+</sup>, v225, v226a-v226g, v240a, v240b, wgt_hh</code>
Analytic variables used	<code>memsleep_dj</code>
Analytic variables created	<code>memsleep, land, landarea, landarea_median_rural, landarea_median_urban, house, domestic, bankacct, water_11-water_96, toilet_61, toilet_111-toilet_962, floor_11-floor_96, roof_11-roof_96, wall_11-wall_96, cookfuel_1-cookfuel_96, num_cow, num_cattle, num_horse, num_goat, num_sheep, num_poultry, num_fish, cat_cow1_4, cat_cow5_9, cat_cow10, cat_cattle1_4, cat_cattle5_9, cat_cattle10, cat_horse1_4, cat_horse5_9, cat_horse10, cat_goat1_4, cat_goat5_9, cat_goat10, cat_sheep1_4, cat_sheep5_9, cat_sheep10, cat_poultry1_9, cat_poultry10_29, cat_poultry30, cat_fish1_49, cat_fish50_99, cat_fish100, v222ax-v222fx, v223ax-v223gx, com, urb, rur, rur_const, rur_coeff, urb_const, urb_coeff, awi, hhmemwgt, awiquint, null</code>

<sup>†</sup>Variable **v224b**, which captures household ownership, was variable **v8601** in the Baseline Survey dataset. Because the household consumption module (Module 8) is not included in the main survey questionnaire at midline, the question was added to Module 2, *Dwelling characteristics and household assets*.

**Step 1.** Create the variables needed to construct the AWI in the household-level analytic data file.

**Step 1.1.** Load the analytic household data file.

```
Load          "FTF ZOI Survey [COUNTRY] [YEAR] household data  
              analytic"
```

**Step 1.2.** Create an integer version of the **memsleep\_dj** analytic variable that captures the number of de jure household members per sleeping room in the household (**memsleep**).<sup>73</sup>

```
Set          memsleep=integer(memsleep_dj)  
Label variable "Number of HH members per sleeping room,  
              truncated"
```

**Step 1.3.** Create a variable that indicates whether the household owns agricultural land (**land**). Set missing responses to "no."<sup>74</sup>

```
Set          land=0  
Replace      land=1 if v240a=1  
Label values 0 "No"  
              1 "Yes"  
Label variable "HH owns agricultural land"
```

**Step 1.4.** Create a variable that indicates the amount of agricultural land owned by the household (**landarea**). Households that own 95 or more hectares of land are coded as having 95 hectares in the dataset; ensure that this is true. Set missing and "don't know" responses to missing; they will be replaced by the median in the next step.

```
Set          landarea=0 if land=0  
Replace      landarea=v240b if v240b≠missing and  
              landarea=missing  
Replace      landarea=missing if v240b>95 and v240b≠missing
```

**Step 1.5.** Determine the median land area value by residence (urban: **landarea\_median\_urban**, rural: **landarea\_median\_rural**) and substitute the median value for any missing land area values.

```
Set          landarea_median_urban=median(landarea) if  
              ahtype=1
```

---

<sup>73</sup> Per DHS protocol, the number of de jure household members per sleeping room variable is captured as a whole number by truncating the value, rather than rounding to the nearest whole number.

<sup>74</sup> In the reference survey AWI calculations, the land variable was created using the variables **v745b** (land ownership) and **v135** (de jure household member) from the women's questionnaire, the variables **mv745b** (land ownership) and **mv135** (de jure household member) from the men's questionnaire, and the variable **hv244** (agriculture land ownership) from the household questionnaire. Land was set to 1 if **v745b** or **mv745b** was 1 (owns alone), 2 (owns jointly), or 3 (owns both alone and jointly) for any de jure household member or if **hv244** was 1 (household owned agricultural land), and otherwise set to 0.

```

Replace      landarea=landarea_median_urban if ahtype=1 and
              landarea=missing

Set          landarea_median_rural=median(landarea) if
              ahtype=2

Replace      landarea=landarea_median_rural if ahtype=2 and
              landarea=missing

```

**Step 1.6.** Create a variable that indicates whether the household owns its dwelling (**house**). If a household is missing this information, consider the household to not own its dwelling.<sup>75</sup><sup>76</sup>

```

Set          house=0
Replace      house=1 if v224b=1
Label values 0 "No"
              1 "Yes"
Label variable "HH owns dwelling"

```

**Step 1.7.** Create a binary variable for each drinking water source response category, including an “other” category (15 ‘**water\_**’ variables with the response category number appended to the end of the name). Set the indicator variable to 1 if the household reported getting its water from that source and 0 otherwise.

```

For each      value (i) of 11-14 21 31 32 41 42 51 61 71 81 91
              96:
Set          water_`i`=0
Replace      water_`i`=1 if v211=`i`
Label values 0 "No"
              1 "Yes"

Label variable water_11 "Piped into dwelling"
Label variable water_12 "Piped into yard/plot"
Label variable water_13 "Piped to neighbor"
Label variable water_14 "Piped to tap/standpipe"
Label variable water_21 "Tubewell or borehole"
Label variable water_31 "Protected well"
Label variable water_32 "Unprotected well"
Label variable water_41 "Protected spring"
Label variable water_42 "Unprotected spring"
Label variable water_51 "Rainwater"
Label variable water_61 "Tanker truck"
Label variable water_71 "Cart with small tank"
Label variable water_81 "Surface water"

```

<sup>75</sup> In the reference survey AWI calculations, the house variable was created using the variables **v745a** (house ownership) and **v135** (de jure household member) from the women’s questionnaire, and the variables **mv745a** (house ownership) and **mv135** (de jure household member) from the men’s questionnaire. House was set to 1 if **v745a** or **mv745a** was 1 (owns alone), 2 (owns jointly), or 3 (owns both alone and jointly) for any de jure household member, and otherwise set to 0.

<sup>76</sup> In Baseline Surveys, household ownership was captured in Module 8, *Household consumption expenditure*, but in Midline Surveys, the data are captured in Module 2, *Dwelling characteristics and household assets*.

```

Label variable   water_91 "Bottled water"
Label variable   water_96 "Other water source"

```

**Step 1.8.** Create two binary variables for each response category of the variable that indicates the household's main sanitation facility (12 'toilet\_' variables with the response category number and a shared [1] unshared [2] indication appended to the end of the name). The first variable is for sanitation facilities that are used by only one household, and the second variable is for sanitation facilities that are shared by multiple households. The exception is for the no toilet/bush response category. Open defecators are not asked question 209 about sharing status, so there is only one indicator for this response category (toilet\_61). Set the indicator variable to 1 if the household had that type of sanitation facility and 0 otherwise.

```

For each          value (i) of 11-15 21-23 31 41 51 96:
  For each        value (l) of 1 2:
    Set           toilet_`i'`_l'=0
    Replace       toilet_`i'`_l'=1 if v208=`i' and v209=`l'
    Label values  0 "No"
                  1 "Yes"

```

```

Set               toilet_61=0
Replace           toilet_61=1 if v208=61

```

```

Label variable    toilet_61 "No toilet, open bush"
Label variable    toilet_112 "Flush piped toilet, not shared"
Label variable    toilet_122 "Flush septic tank, not shared"
Label variable    toilet_132 "Flush pit latrine, not shared"
Label variable    toilet_142 "Flush elsewhere, not shared"
Label variable    toilet_152 "Flush DK where, not shared"
Label variable    toilet_212 "Ventilated pit, not shared"
Label variable    toilet_222 "Pit with slab, not shared"
Label variable    toilet_232 "Open pit, not shared"
Label variable    toilet_312 "Composting toilet, not shared"
Label variable    toilet_412 "Bucket toilet, not shared"
Label variable    toilet_512 "Hanging toilet, not shared"
Label variable    toilet_962 "Other toilet, not shared"
Label variable    toilet_111 "Flush piped toilet, shared"
Label variable    toilet_121 "Flush septic tank, shared"
Label variable    toilet_131 "Flush pit latrine, shared"
Label variable    toilet_141 "Flush elsewhere, shared"
Label variable    toilet_151 "Flush don't know where, shared"
Label variable    toilet_211 "Ventilated pit, shared"
Label variable    toilet_221 "Pit with slab, shared"
Label variable    toilet_231 "Open pit, shared"
Label variable    toilet_311 "Composting toilet, shared"
Label variable    toilet_411 "Bucket toilet, shared"
Label variable    toilet_511 "Hanging toilet, shared"
Label variable    toilet_961 "Other toilet, shared"

```

**Step 1.9.** Create a binary variable for each response category of the variable that indicates the primary flooring material of the household's dwelling. Create 11 'floor\_' variables with the response category

number appended to the end of the name. Set the indicator variable to 1 if the household had that type of flooring material and 0 otherwise.

```

For each      value (i) of 11-13 21 22 31-35 96:
  Set          floor_`i`=0
  Replace      floor_`i`=1 if v202=`i'
  Label values 0 "No"
               1 "Yes"

Label variable floor_11 "Floor - earth/sand"
Label variable floor_12 "Floor - dung"
Label variable floor_13 "Floor - palm leaves"
Label variable floor_21 "Floor - wood planks"
Label variable floor_22 "Floor - bamboo slats"
Label variable floor_31 "Floor - vinyl or asphalt strips"
Label variable floor_32 "Floor - wall-to-wall carpet"
Label variable floor_33 "Floor - cement"
Label variable floor_34 "Floor - parquet or polished wood"
Label variable floor_35 "Floor - ceramic tiles"
Label variable floor_96 "Floor - other"

```

**Step 1.10.** Create a binary variable for each response category of the variable that indicates the primary roof material of the household's dwelling. Create 13 'roof\_' variables with the response category number appended to the end of the name. Set the indicator variable to 1 if the household had that type of roof material and 0 otherwise.

```

For each      value (i) of 11-14 21 22 31-36 96:
  Set          roof_`i`=0
  Replace      roof_`i`=1 if v201=`i'
  Label values 0 "No"
               1 "Yes"

Label variable roof_11 "Roof - no roof"
Label variable roof_12 "Roof - thatch"
Label variable roof_13 "Roof - sod"
Label variable roof_14 "Roof - bamboo"
Label variable roof_21 "Roof - wood planks"
Label variable roof_22 "Roof - cardboard"
Label variable roof_31 "Roof - metal"
Label variable roof_32 "Roof - wood"
Label variable roof_33 "Roof - calamine/cement fiber"
Label variable roof_34 "Roof - ceramic tiles"
Label variable roof_35 "Roof - cement"
Label variable roof_36 "Roof - roofing shingles"
Label variable roof_96 "Roof - other"

```

**Step 1.11.** Create a binary variable for each response category of the variable that indicates the primary exterior wall material of the household's dwelling. Create 16 'wall\_' variables with the response category number appended to the end of the name. Set the indicator variable to 1 if the household had that type of exterior wall material and 0 otherwise.

```

For each      value (i) of 11-15 21-24 31-36 96:
  Set         wall_`i`=0
  Replace     wall_`i`=1 if v203=`i`
  Label values 0 "No"
              1 "Yes"

Label variable wall_11 "Wall - no walls"
Label variable wall_12 "Wall - dirt"
Label variable wall_13 "Wall - cane/palm/tree trunks"
Label variable wall_14 "Wall - bamboo with mud"
Label variable wall_15 "Wall - stone with mud"
Label variable wall_21 "Wall - cardboard"
Label variable wall_22 "Wall - reused wood"
Label variable wall_23 "Wall - plywood"
Label variable wall_24 "Wall - unbaked bricks"
Label variable wall_31 "Wall - wood planks/shingles"
Label variable wall_32 "Wall - unbaked bricks w plaster"
Label variable wall_33 "Wall - bricks"
Label variable wall_34 "Wall - cement blocks"
Label variable wall_35 "Wall - cement"
Label variable wall_36 "Wall - stone with lime/cement"
Label variable wall_96 "Wall - other"

```

**Step 1.12.** Create a binary variable for each response category of the variable that indicates the primary type of cooking fuel the household uses (13 'cookfuel\_' variables with the response category number appended to the end of the name). Set the indicator variable to 1 if the household used that type of fuel for cooking and 0 otherwise.

```

For each      value (i) of 1-11 95 96:
  Set         cookfuel_`i`=0
  Replace     cookfuel_`i`=1 if v219=`i`
  Label values 0 "No"
              1 "Yes"

Label variable cookfuel_1 "Cooking fuel - electricity"
Label variable cookfuel_2 "Cooking fuel - liquid propane gas"
Label variable cookfuel_3 "Cooking fuel - natural gas"
Label variable cookfuel_4 "Cooking fuel - biogas"
Label variable cookfuel_5 "Cooking fuel - kerosene"
Label variable cookfuel_6 "Cooking fuel - coal"
Label variable cookfuel_7 "Cooking fuel - charcoal"
Label variable cookfuel_8 "Cooking fuel - wood"
Label variable cookfuel_9 "Cooking fuel - straw/shrubs/grass"
Label variable cookfuel_10 "Cooking fuel - agri crop residue"
Label variable cookfuel_11 "Cooking fuel - animal dung"
Label variable cookfuel_95 "Cooking fuel - food not cooked in
house"
Label variable cookfuel_96 "Cooking fuel - other"

```

**Step 1.13.** Create a continuous variable for each farm animal equal to the number that the household owns, setting missing values to 0 and leaving "don't know" responses as 98 (num\_[animal]). Then

create a set of categorical variables for each farm animal that indicates the number that the household owns. In the template syntax, the following variables are created:

- Large animals: `cat_[animal]_1-4` (household owns 1-4), `cat_[animal]_5_9` (household owns 5-9), and `cat_[animal]_10` (household owns 10 or more or the respondent reported not knowing the number owned)
- Small animals: `cat_[animal]_1-9` (household owns 1-9), `cat_[animal]_10_29` (household owns 10-29), and `cat_[animal]_30` (household owns 30 or more or the respondent reported not knowing the number owned)
- Fish: `cat_fish_1-49` (household owns 1-49), `cat_fish_50_99` (household owns 50-99), and `cat_fish_100` (household owns 100 or more or the respondent reported not knowing the number owned)

In most cases, these categories work, but it is important to review carefully to ensure that there are no categories with a small number of households (i.e., categories have at least 3-5 households in them).

Review the analytic categorical variable alongside the survey variable frequencies for each farm animal to make sure that the categories make sense for the dataset. If there are not many households in certain categories, then re-create the variable using different categories. For example, if only 6 households own more than 5 cows, and only 1 household owns more than 10 cows, the 1-4, 5-9, and 10+ categories can be combined into a 1+ category. Or if there are many households that own more than 30 goats, the 10+ category can be split into two categories: 10-29 and 30+. For consistency, it is recommended that one of the following category options be used: (a) two categories: 0, 1+; (b) three categories: 0, 1-4, 5+; (c) four categories, generally for large animals: 0, 1-4, 5-9, 10+; or (c) four categories, generally for small animals: 0, 1-9, 10-29, 30+. Use categories that work well with the data.

```
Set          num_cow=0 if v225=2 or v226a=missing
Replace      num_cow=v226a if v226a≤98
Label variable "Number of cows or bulls HH owns"
```

```
Set          cat_cow1_4=0
Replace      cat_cow1_4=1 if num_cow≥1 and num_cow≤4
Label variable "HH owns 1-4 cows or bulls"
Set          cat_cow5_9=0
Replace      cat_cow5_9=1 if num_cow≥5 and num_cow≤9
Label variable "HH owns 5-9 cows or bulls"
Set          cat_cow10=0
Replace      cat_cow10=1 if num_cow≥10 and num_cow≤98
Label variable "HH owns 10+ cows or bulls"
Label values 0 "No"
              1 "Yes"
```

```
Set          num_cattle=0 if v225=2 or v226b=missing
Replace      num_cattle=v226b if v226b≤98
Label variable "Number of other cattle HH owns"
```

```
Set          cat_cattle1_4=0
Replace      cat_cattle1_4=1 if num_cattle≥1 and num_cattle≤4
```

Label variable	"HH owns 1-4 other cattle"
Set	cat_cattle5_9=0
Replace	cat_cattle5_9=1 if num_cattle≥5 and num_cattle≤9
Label variable	"HH owns 5-9 other cattle"
Set	cat_cattle10=0
Replace	cat_cattle10=1 if num_cattle≥10 and num_cattle≤98
Label variable	"HH owns 10+ other cattle"
Label values	0 "No"
	1 "Yes"
Set	num_horse=0 if v225=2 or v226c=missing
Replace	num_horse=v226c if v226c≤98
Label variable	"Number of horses, donkeys, mules HH owns"
Set	cat_horse1_4=0
Replace	cat_horse1_4=1 if num_horse≥1 and num_horse≤4
Label variable	"HH owns 1-4 horses, donkeys, mules"
Set	cat_horse5_9=0
Replace	cat_horse5_9=1 if num_horse≥5 and num_horse≤9
Label variable	"HH owns 5-9 horses, donkeys, mules"
Set	cat_horse10=0
Replace	cat_horse10=1 if num_horse≥10 and num_horse≤98
Label variable	"HH owns 10+ horses, donkeys, mules"
Label values	0 "No"
	1 "Yes"
Set	num_goat=0 if v225=2 or v226d=missing
Replace	num_goat=v226d if v226d≤98
Label variable	"Number of goats owned by HH"
Set	cat_goat1_4=0
Replace	cat_goat1_4=1 if num_goat≥1 and num_goat≤4
Label variable	"HH owns 1-4 goats"
Set	cat_goat5_9=0
Replace	cat_goat5_9=1 if num_goat≥5 and num_goat≤9
Label variable	"HH owns 5-9 goats"
Set	cat_goat10=0
Replace	cat_goat10=1 if num_goat≥10 and num_goat≤98
Label variable	"HH owns 10+ goats"
Label values	0 "No"
	1 "Yes"
Set	num_sheep=0 if v225=2 or v226e=missing
Replace	num_sheep=v226e if v226e≤98
Label variable	"Number of sheep owned by HH"
Set	cat_sheep1_4=0
Replace	cat_sheep1_4=1 if num_sheep≥1 and num_sheep≤4
Label variable	"HH owns 1-4 sheep"
Set	cat_sheep5_9=0
Replace	cat_sheep5_9=1 if num_sheep≥5 and num_sheep≤9
Label variable	"HH owns 5-9 sheep"



```

Set          cat_sheep10=0
Replace      cat_sheep10=1 if num_sheep≥10 and num_sheep≤98
Label variable "HH owns 10+ sheep"
Label values 0 "No"
              1 "Yes"

```

```

Set          num_poultry=0 if v225=2 or v226f=missing
Replace      num_poultry=v226f if v226f≤98
Label variable "Number of poultry owned by HH"

```

```

Set          cat_poultry1_9=0
Replace      cat_poultry1_9=1 if num_poultry≥1 and
              num_poultry≤9
Label variable "HH owns 1-9 poultry"

```

```

Set          cat_poultry10_29=0
Replace      cat_poultry10_29=1 if num_poultry≥10 and
              num_poultry≤29
Label variable "HH owns 10-29 poultry"
Set          cat_poultry30=0
Replace      cat_poultry30=1 if num_poultry≥30 and
              num_poultry≤98
Label variable "HH owns 30+ poultry"
Label values 0 "No"
              1 "Yes"

```

```

Set          num_fish=0 if v225=2 or v226g=missing
Replace      num_fish=v226g if v226g≤9998
Label variable "Number of fish owned by HH"

```

```

Set          cat_fish1_9=0
Replace      cat_fish1_9=1 if num_fish≥1 and num_fish≤49
Label variable "HH owns 1-49 fish"
Set          cat_fish50_49=0
Replace      cat_fish50_99=1 if num_fish≥50 and num_fish≤99
Label variable "HH owns 50-99 fish"
Set          cat_fish100=0
Replace      cat_fish100=1 if num_fish≥100 and num_fish≤9998
Label variable "HH owns 100+ fish"
Label values 0 "No"
              1 "Yes"

```

**Step 1.14.** Create binary variables for each asset included in the survey (v222ax–v222fx and v223ax–v223gx) so that “no” responses (2) and “missing” responses (9), and missing values all have a value of 0.

```

For each      variable (var) in variable list v222a-v222f
              v223a-v223g:
  Set          `var'x=0
  Replace      `var'x=1 if `var'=1

```

```

Label values 0 "No"
              1 "Yes"

Label variable v222ax "HH has electricity"
Label variable v222bx "HH has a radio"
Label variable v222cx "HH has a television"
Label variable v222dx "HH has a non-mobile phone"
Label variable v222ex "HH has a computer"
Label variable v222fx "HH has a refrigerator"
Label variable v223ax "HH member has a watch"
Label variable v223bx "HH member has a mobile phone"
Label variable v223cx "HH member has a bicycle"
Label variable v223dx "HH member has a motorcycle or motor
                    scooter"
Label variable v223ex "HH member has an animal-drawn cart"
Label variable v223fx "HH member has a car or truck"
Label variable v223gx "HH member has a boat with motor"

```

**Step 1.15.** Create a binary variable that indicates whether any household member has a bank account (bankacct) so that “no” responses (2), “missing” responses (9), and missing values all have a value of 0.

```

Set          bankacct=0 if v224a=2 or 9 or missing
Replace      bankacct=1 if v224a=1
Label values 0 "No"
              1 "Yes"
Label variable "HH member has bank account"

```

**Step 2.** Create a binary variable that indicates whether the household had a maid or servant (domestic).<sup>77</sup>

**Step 2.1.** Save the current version of the household-level analytic data file as a temporary data file and load the persons-level analytic data file.

```

Save          "temp_awi"
Load          "FTF ZOI Survey [COUNTRY] [YEAR] persons data
              analytic"

```

**Step 2.2.** Create a binary variable that indicates whether any household members in the roster are maids or servants (domestic).

```

Set          domestic=0
Replace      domestic=1 if v103=14
Label values 0 "No"
              1 "Yes"
Label variable "HH has a maid or servant"

```

---

<sup>77</sup> The relationship information necessary to create the domestic variable was not included in the reference country survey, so it was not included in the wealth index calculations for the reference survey.

**Step 2.3.** Create a file that has one record per household indicating whether there is a maid or servant in that household, and save the file as a temporary data file.

```
By hhea and hhnum:    set domestic=1 if domestic=1 for any HH
                      members
By hhead and hhnum:   replace domestic=0 if domestic=0 for all HH
                      members
Save                  "temp_domestic"
```

**Step 2.4.** Load the temporary AWI data file created in Step 2.1 and add the **domestic** variable from the temporary data file created in Step 2.3.

```
Load data file  "temp_awi"
Merge          1 to 1
key variables   hhea hhnum
data file       "temp_domestic"
keep variable   domestic
```

**Step 3.** Determine which variables created in Steps 1 and 2 do not have any variation so that they can be excluded from the factor analysis or have little variation so that they can be combined with other variables.

**Step 3.1.** Define a variable list (**vars**) that includes all the variables created to calculate the wealth index.

```
Set          vars=domestic memsleep_dj land landarea house
bankacct cat_cow1_4 cat_cow5_9 cat_cow10
cat_cattle1_4 cat_cattle5_9 cat_cattle10
cat_horse1_4 cat_horse5_9 cat_horse10 cat_goat1_4
cat_goat5_9 cat_goat10 cat_sheep1_4 cat_sheep5_9
cat_sheep10 cat_poultry1_9 cat_poultry10_29
cat_poultry30 cat_fish1_49 cat_fish50_99
cat_fish100 water_11 water_12 water_13 water_14
water_21 water_31 water_32 water_41 water_42
water_51 water_61 water_71 water_81 water_91
water_96 toilet_111 toilet_112 toilet_121
toilet_122 toilet_131 toilet_132 toilet_141
toilet_142 toilet_151 toilet_152 toilet_211
toilet_212 toilet_221 toilet_222 toilet_231
toilet_232 toilet_311 toilet_312 toilet_411
toilet_412 toilet_511 toilet_512 toilet_961
toilet_962 toilet_61 floor_11 floor_12 floor_13
floor_21 floor_22 floor_31 floor_32 floor_33
floor_34 floor_35 floor_96 roof_11 roof_12
roof_13 roof_14 roof_21 roof_22 roof_31 roof_32
roof_33 roof_34 roof_35 roof_36 roof_96 wall_11
wall_12 wall_13 wall_14 wall_15 wall_21 wall_22
wall_23 wall_24 wall_31 wall_32 wall_33 wall_34
wall_35 wall_36 wall_96 cookfuel_1 cookfuel_2
cookfuel_3 cookfuel_4 cookfuel_5 cookfuel_6
cookfuel_7 cookfuel_8 cookfuel_9 cookfuel_10
```

```
cookfuel_11 cookfuel_95 cookfuel_96 v222ax v222bx
v222cx v222dx v222ex v222fx v223ax v223bx v223cx
v223dx v223ex v223fx v223g'
```

**Step 3.2.** Run frequencies for variables in the **vars** variable list and flag any variables that have no variation.<sup>78</sup> All variables identified as having no variation—that is, all binary variables that have a value of 0 for all households or that have a value of 1 for all households—will be excluded from the factor analysis in Steps 4 and 5. For example, if the variable for unprotected spring (**water\_42**) has no ‘Yes’ responses, exclude **water\_42** and document this as a comment in the analysis syntax file.

```
For each      variable (var) of variable list vars:
  summarize   `var' [examine number of observations, mean,
                    standard deviation, minimum, maximum]
  tabulate     `var' [examine number of observations and
                    variable values]
```

**Step 3.3.** Using the frequency results obtained in Step 3.2, examine all binary variables that have ‘Yes’ values for fewer than five households. Determine whether the variable is similar enough to another variable that the two can be combined. If there is not a similar variable, and the variable has only one or two ‘Yes’ values, combine the ‘Yes’ values with the “other” variable if the “other” category already has any ‘Yes’ responses. If the “other” category does not have any ‘Yes’ responses, and the variable cannot be combined with another category, leave it as is. For example, if the variable unprotected spring (**water\_42**) has only one ‘Yes’ value, and the variable unprotected well (**water\_32**) has any ‘Yes’ values, combine the two categories (case 1 in **Table 13**), or if there are no ‘Yes’ values for **water\_32**, combine **water\_42** with the “other” variable (**water\_96**) if **water\_96** has ‘Yes’ values (case 2 in **Table 13**). Recode the variables according to your findings and document any recoding as comments in the analysis syntax file.

**Table 13: Example Cases that Show Variables with 0 or 1 ‘Yes’ Responses**

Case 1		Case 2	
Variable	# ‘Yes’ responses	Variable	# ‘Yes’ responses
<b>water_32</b> (unprotected well )	1	<b>water_32</b> (unprotected well )	0
<b>water_42</b> (unprotected spring)	1	<b>water_42</b> (unprotected spring)	1
<b>water_96</b> (other)	0	<b>water_96</b> (other)	1
Action: combine <b>water_32</b> and <b>water_42</b>		Action: combine <b>water_42</b> and <b>water_96</b>	

**Step 4.** Create a list of all variables that will be included in the wealth index (**vars2**), after excluding or combining variables in Step 3.

```
Set          var2=[variables in variable list vars minus any
                    flagged for dropping or that were combined and
                    are no longer needed]
```

<sup>78</sup> Note that ‘**summarize**’ and ‘**tabulate**’ are Stata commands used to perform this step. If using a different software package, be sure to use the commands available in that package.

**Step 5.** Create a list that includes all variables thought to have the same relationship with the underlying economic status in both urban and rural areas (**varsc**)—that is, remove those variables that do not apply to either rural or urban areas or that are thought to indicate different levels of wealth in urban and rural areas from the list of variables created in Step 4. The resulting list will be the variables to be included in the common factor analysis. This list is based on one’s understanding of and experience with the surveyed ZOI. For example, the land area and animal variables are usually removed for the common factor analysis.

**Step 6.** Run the common factor analysis using the variable list created in Step 5 (**varsc**) and save the component scores as a new variable (**com**).<sup>79</sup> (Note that ‘**factor**’ and ‘**predict**’ are Stata commands used in the template syntax file. Specifying the ‘**pcf**’ option indicates that a principal-component factor model is used—rather than a principal factor, iterated principal factor, or maximum-likelihood factor model. Specifying ‘**factors (1)**’ indicates that only the first factor is retained. Specifying ‘**no rotation**’ suppresses display of the rotation matrix.)

```
factor [variable list varsc], pcf factors(1)
predict com, no rotation
```

**Step 7.** Create a variable list that includes all variables relevant to the economic status in urban areas (**varsu**). The selection of which variables to include in the urban factor analysis is again based on one’s understanding and experience. Variables important in rural areas may be relevant in urban areas as well, but with a different relationship to wealth. If a variable has no variation in urban areas, omit it from the urban factor analysis. The resulting list will be the variables to be included in the urban factor analysis.

```
For each      variable (var) in variable list vars2:
  Summarize   `var' if ahtype=1
  Tabulate    `var' if ahtype=1
Set           varsu=[variables in variable list vars2 minus any
                  that have no variation in urban areas]
```

**Step 8.** Run the urban factor analysis for households in urban areas using the variable list created in Step 7 (**varsu**) and save the component scores as a new variable (**urb**).<sup>80</sup> (Note that ‘**factor**’ and ‘**predict**’ are Stata commands used in the template syntax file. Additional background information can be found in Step 6.)

```
factor [variable list varsu] if ahtype=1, pcf factors(1)
predict urb if ahtype=1, no rotation
```

**Step 9.** Create a variable list that includes all variables relevant to the economic status in rural areas (**varsr**). The selection of which variables to include in the rural factor analysis is again based on one’s understanding and experience. Variables important in urban areas may be relevant in rural areas as well,

<sup>79</sup> All factor analyses should be run using unweighted data with the following options: principal components extraction using correlation method with one factor extracted, substitution of mean for missing values (if missing values have not yet been imputed), and estimation of the factor scores using the regression method.

<sup>80</sup> All factor analyses should be run using unweighted data with the following options: principal components extraction using correlation method with one factor extracted, substitution of mean for missing values (if missing values have not yet been imputed), and estimation of the factor scores using the regression method.

but with a different relationship to wealth. If a variable has no variation in rural areas, omit it from the rural factor analysis. The resulting list will be the variables to be included in the rural factor analysis.

```
For each      variables in variable list vars2:
  Summarize   `var' if ahtype=2
  Tabulate    `var' if ahtype=2
Set           varsr=[variables in variable list vars2 minus any
               that have no variation in rural areas]
```

**Step 10.** Run the rural factor analysis for households in rural areas using the variable list created in Step 9 (**varsr**), and save the component scores as a new variable (**rur**).<sup>81</sup> (Note that **'factor'** and **'predict'** are Stata commands used in the template syntax file. Additional background information can be found in Step 6.)

```
factor [variable list varsr] if ahtype=2, pcf factors(1)
predict rur if ahtype=2, no rotation
```

**Step 11.** Run a regression with the common factor score (**com**) as the dependent variable and the urban area factor score (**urb**) as the independent variable. Save the constant term (**urb\_const**) and the coefficient (**urb\_coeff**).

```
Regress      com urb if ahtype=1
Set          urb_const=_b[_cons]
Set          urb_coeff=_b[urb]
```

**Step 12.** Run a regression with the common factor score (**com**) as the dependent variable and the rural area factor score (**rur**) as the independent variable. Save the constant term (**rur\_const**) and the coefficient (**rur\_coeff**).

```
Regress      com rur if ahtype=2
Set          rur_const=_b[_cons]
Set          rur_coeff=_b[urb]
```

**Step 13.** Create a variable for the combined score (**awi**) equal to 0. Then calculate the combined score using the appropriate urban or rural factor scores, constant, and coefficient obtained in Steps 7-12.

```
Set          awi=0
Replace      awi=urb_const+(urb_coeff*urb) if ahtype=1
Replace      awi=rur_const+(rur_coeff*rur) if ahtype=2
Label variable `Household's asset-based wealth index score'
```

**Step 14.** Create a variable for the household member weight (**hmemwgt**), if not already created, by multiplying the number of de jure household members (**hhsz\_dj**) by the household weight (**wgt\_hh**).

---

<sup>81</sup> All factor analyses should be run using unweighted data with the following options: principal components extraction using correlation method with one factor extracted, substitution of mean for missing values (if missing values have not yet been imputed), and estimation of the factor scores using the regression method.

```

Set                hhmemwgt=(hhsize_dj*wgt_hh)
Label variable     "Household member weight, de jure"

```

**Step 15.** Create wealth quintiles (**awiquint**) using the wealth index scores (**awi**) and applying the household member weight.<sup>82</sup> Stata has a command (**xtile**) that will do this in one line of code, and a similar command is available in SPSS (**RANK**). Otherwise, this can be done generally following the instructions in Step 5 of Rutstein’s Steps to Constructing the New DHS Wealth Index guidance document (2014). Note that in most cases, the cumulative distribution will not be smooth at the quintile cutpoints (e.g., 20 percent, 40 percent, 60 percent) because a single AWI score may increase the cumulative percentage by several percentage points.

```

Set                awiquint=1 if awi<awi cut point that indicates
                   poorest 20% of population
Replace            awiquint=2 if awiquint=missing and awi<awi cut
                   point that indicates poorest 40% of population
Replace            awiquint=3 if wquint=missing and awi<awi cut
                   point that indicates poorest 60% of population
Replace            awiquint=4 if awiquint=missing and awi<awi cut
                   point that indicates poorest 80% of population
Replace            awiquint=5 if wquint=missing and awi>=awi cut
                   point that indicates poorest 80% of population
Label values       1 "Lowest (poorest)"
                   2 "Second"
                   3 "Middle"
                   4 "Fourth"
                   5 "Highest (wealthiest)"
Label variable     "HH's asset-based wealth index quintile"

```

**Step 16.** Create a null variable (**null**) equal to 1 for all records to use for adding variables in a later step and save the data file to use in Part 2.

```

Set                null=1
Save data           "FTF ZOI Survey [COUNTRY] [YEAR] wealthindex AWI"

```

## Part 2. CWI

Survey variables used	v110, v111a, v111b, v202, v203, v208, v209, v211, v222c, v222e, v222f, v223f, hhea, hhnum, wgt_hh, strata
Analytic variables used	memsleep, age, awi, genhhstype_dj, shock_sev
Analytic variables created	ubn, ubn1, ubn2, ubn3, ubn4, wadult_dj, wadult_noprim_dj, wadult_att_dj, edu_prim_dj, edu_prim_pdm_dj, null, ptile1, ptile2, ptile3, ptile4, sumwts, tot_wgt, cut1, cut2, cut3, cut4, freq, cumfreq, percent, cumpercent, compcut1, compcut2, compcut3, compcut4, basecut1, basecut2, basecut3, basecut4, tv, computer, fridge, car, const, coeff, cw, cwiquint, comp_poor

<sup>82</sup> Note that for the wealth quintiles generated for the CWI reference survey, household weights were used rather than household member weights.

**Step 1.** Create the first three UBN variables (**ubn1**, **ubn2**, and **ubn3**) using household-level data.

**Step 1.1.** Load the data file created in Part I.

```
Load "FTF ZOI Survey [COUNTRY] [YEAR] wealthindex AWI"
```

**Step 1.2.** Create the first UBN variable (**ubn1**) that indicates whether the household has inadequate dwelling construction—that is, dirt floors or natural or rustic walls.

```
Set ubn1=0
Replace ubn1=1 if v202=11,12,13,or 96 [inadequate floor]
or v203=11,12,13,14,15,or 96 [inadequate wall]
Label values 0 "No"
1 "Yes"
Label variable "Inadequate dwelling construction (floors/walls)"
```

**Step 1.3.** Create the second UBN variable (**ubn2**) that indicates inadequate sanitation or drinking water source.

```
Set ubn2=0
Replace ubn2=1 if v208=14,23,41,51,61,or 96 or v209=1
[inadequate sanitation] or v211=32,42,81,or 96
[inadequate water source]
Label values 0 "No"
1 "Yes"
Label variable "Inadequate sanitation (toilet/drinking water)"
```

**Step 1.4.** Create the third UBN variable (**ubn3**) that indicates household crowding—more than three de jure household members per sleeping room.<sup>83</sup>

```
Set ubn3=0
Replace ubn3=1 if memsleep>3 and memsleep#missing
Label values 0 "No"
1 "Yes"
Label variable "HH crowding: >3 de jure HH members/sleeping
room"
```

**Step 2.** Create the fourth UBN variable (**ubn4**) that indicates high economic dependency using individual-level data.

**Step 2.1.** Save the current data in a temporary file and load the individual-level analytic data file to create the final UBN variable (**ubn4**).

```
Save "temp_ubn_hh"
Load "FTF ZOI Survey [COUNTRY] [YEAR] persons analytic
data"
```

---

<sup>83</sup> Per DHS protocol, the number of de jure household members per sleeping room variable is captured as a whole number by truncating the value, rather than rounding to the nearest whole number. The *memsleep* variable was created in Step 1.2 of the Part I, AWI step-by-step procedures.



**Step 2.2.** Create a variable to flag working-age adults who are de jure household members (`wadult_dj`).

```
Set          wadult_dj=0
Replace      wadult_dj=1 if (age≥15 and age≤64) and hhmem_dj=1
Label values 0 "No"
              1 "Yes"
Label variable "Working-age adult (15-64), de jure HH member"
```

**Step 2.3.** Create a variable to flag working-age adults who are de jure household members but did not complete primary education (`wadult_noprimary_dj`), but first create a variable that captures whether household members completed primary education (`edu_prim_dj`). Be sure to adapt the syntax to appropriately account for primary education in the survey country; [A], [X], [B], and [Y] are placeholders in the pseudo code.

```
Set          edu_prim_dj=0 if v111a≠missing and hhmem_dj=1
Replace      edu_prim_dj=1 if v111a≥[A] and v111a≤[X] and
              v111b≥[B] and v111a≤[Y] and hhmem_dj=1
Label values 0 "No"
              1 "Yes"
Label variable "HH member completed primary education"

Set          wadult_noprimary_dj=missing
Replace      wadult_noprimary_dj=0 if wadult_dj=1
Replace      wadult_noprimary_dj=1 if wadult_dj=1 and
              edu_prim_dj=0
Label values 0 "No"
              1 "Yes"
Label variable "Working-age adult didn't complete primary
              school, de jure HH member"
```

**Step 2.4.** Create a variable to flag working-age adults who are de jure household members and were attending school at the time of the survey (`wadult_att_dj`).

```
Set          wadult_att_dj=missing
Replace      wadult_att_dj =0 if wadult_dj=1
Replace      wadult_att_dj =1 if v110=1 and wadult_dj=1
Label values 0 "No"
              1 "Yes"
Label variable "Working-age adult attending school at time of
              survey, de jure HH member"
```

**Step 2.5.** Create a variable to flag whether a primary adult decision-maker in the household completed primary school (`edu_prim_pdm_dj`).

```
Set          edu_prim_pdm_dj=0
Replace      edu_prim_pdm_dj=1 if (edu_prim_dj=1 and fdm_dj=1)
              or (edu_prim_dj=1 and mdm_dj=1)
Label values 0 "No"
              1 "Yes"
```

```
Label variable    "Primary adult decision-maker in HH completed
                  primary school"
```

**Step 2.6.** Create household-level variables by transforming the individual-level variables created so far in Step 2 so that there is one record for each household. Sum the individual-level variables `wadult_dj`, `wadult_noprim_dj`, and `wadult_att_dj` across household members and select the maximum value of `edu_prim_pdm_dj` (i.e., 0 or 1) for the primary adult decision-makers in the household. In this step, the working data file is transformed from an individual-level data file to a household-level data file.

```
By hhea hhnum:    set wadult_dj=sum(wadult_dj)
Label variable    "Number of working-age adults in HH"

By hhea hhnum:    set wadult_noprim_dj=sum(wadult_noprim_dj)
Label variable    "Number of working-age adults in HH who didn't
                  complete primary school"

By hhea hhnum:    set wadult_att_dj=sum(wadult_att_dj)
Label variable    "Number of working-age adults attending school at
                  time of survey"

By hhea hhnum:    set edu_prim_pdm_dj=max(edu_prim_pdm_dj)
Label variable    "De jure primary adult decisionmaker (male or
                  female) completed primary school"
```

**Step 2.7.** Create the fourth UBN variable (`ubn4`), which indicates that (a) there are no working-age adults who have completed primary education in the household, OR (b) the only working-age adults in the household are still in school, OR (c) there are no working-age adults in the household or neither a primary male adult decision-maker nor a primary female adult decision-maker has completed primary education.

```
Set              ubn4=0
Replace          ubn4=1 if wadult_dj>0 and wadult_dj≠missing and
                  wadult_dj=wadult_noprim_dj
Replace          ubn4=1 if wadult_dj>0 and wadult_dj≠missing and
                  wadult_dj=wadult_att_dj
Replace          ubn4=1 if wadult_dj=0 and edu_prim_pdm_dj≠1
Label values     0 "No"
                  1 "Yes"
Label variable    "High economic dependency"
```

**Step 2.8.** Drop all variables except cluster (`hhea`), household number (`hhnum`), and `ubn4` and add the `ubn4` variable to the `temp_ubn_hh` data file created in Step 2.1 using `hhea` and `hhnum` as the key matching variables.

```
Keep variables    hhea hhnum ubn4
Merge            1 to 1
                key variables hhea hhnum
                data file     "temp_ubn_hh"
                keep variable ubn4
```

**Step 2.9.** Calculate the UBN score (**ubn**), create a null variable (**null**) equal to 1 to enable merging data into this file in Step 3c, and save in a temporary data file, *temp\_ubn*.

```
Set                ubn=ubn1+ubn2+ubn3+ubn4
Label variable    "Unmet basic needs score (0-4)"
Set                null=1
Save              "temp_ubn"
```

**Step 3.** Calculate the UBN cutpoint values.

**Step 3.1.** Create a new dataset that includes the frequency (**freq**), cumulative frequency (**cumfreq**), percentage (**percent**), and cumulative percentage (**cumpercent**) for each UBN score, with a structure as shown in **Table 14**. (In Stata, the **'contract'** command will produce the desired dataset of frequencies and percentages.)

```
contract ubn, freq(freq) cfreq(cumfreq) percent(percent)
          cpercent(cumpercent)
```

**Table 14: Frequencies and Percentages by UBN Score**

UBN score	Frequency (freq)	Cumulative frequency (cumfreq)	Percentage (percent)	Cumulative percentage (cumpercent)
0				
1				
2				
3				
4				

**Step 3.2.** Create four percentile variables (**ptile1**, **ptile2**, **ptile3**, and **ptile4**) such that **ptile1** is a constant equal to the percentage of households with one or more unmet basic needs, **ptile2** is a constant equal to the percentage of households with two or more unmet basic needs, **ptile3** is a constant equal to the percentage of households with three or more unmet basic needs, and **ptile4** is a constant equal to the percentage of households with all four unmet basic needs.

```
Set                ptile1=0
Replace            ptile1=(100-cumpercent) if ubn=0
Label variable    "% of HHs with 1+ UBNs, constant"

Set                ptile2=0
Replace            ptile2=(100-cumpercent) if ubn=1
Label variable    "% of HHs with 2+ UBNs, constant"

Set                ptile3=0
Replace            ptile3=(100-cumpercent) if ubn=2
Label variable    "% of HHs with 3+ UBNs, constant"

Set                ptile4=0
Replace            ptile4=(100-cumpercent) if ubn=3
Label variable    "% of HHs with 4 UBNs, constant"
```

**Step 3.3.** Save the `ptile1-ptile4` variables created in the previous step in the first row of the working data file. Also save in the first row the sum of the `freq` variable across the five rows to get the number of households included in the `ptile` variable calculations. (In Stata, this can be done with the ‘`collapse (sum)`’ command.) Drop all other rows in the working data file. Create a null variable (`null`) equal to 1 and save the variables to a temporary data file, `temp_cwi_ptile`.

```
collapse      (sum) ptile1 ptile2 ptile3 ptile4 freq
Set           null=1
Save          "temp_cwi_ptile"
```

**Step 3.4.** Load the *FTF ZOI Survey [Country] [Year] wealthindex AWI* data file and add the variables in the `temp_cwi_ptile` data file to the file with the AWI results in it. Use the `null` variable, which was created in both data files, as the matching variable so that the variables, which are constants, are added to all household records, and then save the file.

```
Use           "FTF ZOI Survey [COUNTRY] [YEAR] wealthindex AWI"
Merge         many to 1
  key variable null
  data file    "temp_cwi_ptile"
  keep variables ptile1 ptile2 ptile3 ptile4 freq
```

**Step 3.5.** Sort the data by households’ AWI score, lowest to highest. Then create a variable (`sumwts`) to indicate cumulative household sampling weight for all households and a variable (`tot_wgt`) to indicate the weighted total number of households, or the sum of household sampling weights across all households. (‘`Case number`’ refers to the system index used in the software package.)

```
Sort          awi
Set           sumwts=0
Replace       sumwts=wgt_hh if case number=1
Replace       sumwts=wgt_hh of previous case+wgt_hh if case
              number>1
Label variable "Cumulative HH sampling weight when HHs are
              sorted by awi score"

Set           tot_wgt=missing
Replace       tot_wgt=total(wgt_hh)
Label variable "Sum of HH sampling weights, constant"
```

**Step 3.6.** Create UBN cutpoint values for each household (`cut1`, `cut2`, `cut3`, `cut4`) using the ZOI Survey AWI scores (`awi`), the percentile variables (`ptile1-4`), the cumulative sampling weight variables (`sumwts`), and the total household sampling weights variable (`tot_wgt`).

```
Set           cut1=awi if (ptile1÷100)≤(sumwts÷tot_wgt)
Set           cut2=awi if (ptile2÷100)≤(sumwts÷tot_wgt)
Set           cut3=awi if (ptile3÷100)≤(sumwts÷tot_wgt)
Set           cut4=awi if (ptile4÷100)≤(sumwts÷tot_wgt)
```

**Step 3.7.** Create UBN cutpoints by determining the minimum value of each of the four UBN cutpoint values (`compcut1`, `compcut2`, `compcut3`, `compcut4`).

```

Set          compcut1=minimum(cut1)
Set          compcut2=minimum(cut2)
Set          compcut3=minimum(cut3)
Set          compcut4=minimum(cut4)

```

**Step 3.8.** Drop all observations except the first one and drop all variables except variables `compcut1–compcut4`. The data file now contains only one observation and four variables: `compcut1, compcut2, compcut3, compcut4`.

```

Keep records   if row=1
Keep variables compcut1 compcut2 compcut3 compcut4

```

**Step 4.** Create the UBN cutpoint variables for the reference survey. (Note that the values of these variables were determined previously and are included in the template syntax file. They will be used in the CWI indicator calculation across all ZOI Surveys.)

```

Set          basecut1=1.261171
Set          basecut2=0.6508898
Set          basecut3=(-0.6808249)
Set          basecut4=(-1.330052)

```

The data file now contains only one observation and eight variables: `compcut1, compcut2, compcut3, compcut4, basecut1, basecut2, basecut3, and basecut4`.

**Step 5.** Rearrange the data that are currently all in one row so that the `compcut1–compcut4` values, the UBN cutpoint values for the ZOI Survey, are in one column, and the `basecut1–basecut4` values, the UBN cutpoint values for the reference survey, are in a second column, and save the data to a temporary file, `temp_UBNcutpoint`.<sup>84</sup>

```

Reshape       wide to long
variable stubs compcut basecut
index variable id
new variable num
Label variables:
id            "ID, constant=1"
num          "Observation number"
compcut      "UBN cutpoint values for the ZOI Survey"
basecut      "UBN cutpoint values for the reference survey"
Save         "temp_UBNcutpoint"

```

**Step 6.** Calculate asset cutpoint values.

**Step 6.1.** Load household-level analytic data file.

```

Load          "FTF ZOI Survey [COUNTRY] [YEAR] wealthindex awi"

```

**Step 6.2.** Create binary variables for the four asset items to be used as anchoring points (`tv`, `computer`, `fridge`, and `car`).

<sup>84</sup> In Stata, this can be done with the “reshape long” command.

```

Set          tv=0
Replace      tv=1 if v222c=1
Label values 0 "No"
              1 "Yes"
Label variable "HH has a TV"

Set          computer=0
Replace      computer=1 if v222e=1
Label values 0 "No"
              1 "Yes"
Label variable "HH has a computer"

Set          fridge=0
Replace      fridge=1 if v222f=1
Label values 0 "No"
              1 "Yes"
Label variable "HH has a refrigerator"

Set          car=0
Replace      car=1 if v223f=1
Label values 0 "No"
              1 "Yes"
Label variable "HH has a car or truck"

```

**Step 6.3.** Run a logistic regression of each asset item on `awi` using the household survey weight, and then calculate the cutpoint value for each asset item (`compcut1`, `compcut2`, `compcut3`, `compcut4`), which is equal to the negative value of the regression constant divided by the AWI score coefficient.

```

svyset hhea [pweight=wgt_hh], strata(strata)
svy: logit tv awi
Set compcut1=(-constant÷coeff) [tv]
Label variable "Asset anchoring point: TV"

svy: logit fridge awi
Set compcut2=(-constant÷coeff) [fridge]
Label variable "Asset anchoring point: Refrigerator"

svy: logit computer awi
Set compcut3=(-constant÷coeff) [computer]
Label variable "Asset anchoring point: Computer"

svy: logit car awi
Set compcut4=(-constant÷coeff) [car]
Label variable "Asset anchoring point: Car/truck"

```

**Step 6.4.** Drop all observations except the first observation and drop all variables except the `compcut` variables. The data file now has one observation and four variables.

```

Keep records    if row=1

```

```
Keep variables   compcut1 compcut2 compcut3 compcut4
```

**Step 7.** Create variables for the four asset **basecut** variables for the reference survey. (The values for these variables were calculated previously and included in the syntax file. They will be used in the CWI indicator calculation across all ZOI Surveys.) The data file will now contain only one observation and eight variables: **compcut1**, **compcut2**, **compcut3**, **compcut4**, **basecut1**, **basecut2**, **basecut3**, and **basecut4**.

```
Set           basecut1=0.1166113   [television]
Set           basecut2=1.107843    [fridge]
Set           basecut3=1.591285    [computer]
Set           basecut4=2.05812     [car]
```

**Step 8.** Reshape the data the same as the UBN data in Step 4 and save the asset cutpoint value as a temporary data file, *temp\_ASSETcutpoint*.

```
Reshape        wide to long
variable stubs compcut basecut
index variable      id
new variable  num2
Label variables:
id            "ID, constant=1"
num           "Observation number"
compcut       "UBN cutpoint values for the ZOI Survey"
basecut       "UBN cutpoint values for the reference survey"
Save          "temp_ASSETcutpoint"
```

**Step 9.** Append the UBN cutpoint data file to the asset cutpoint data file. The appended dataset will have two variables of interest (**basecut** and **compcut**) and eight observations (four for assets and four for UBN).

```
Append "temp_UBNcutpoint" records to "temp_ASSETcutpoint" records
```

**Step 10.** Run an ordinary least squares regression of the ZOI Survey cutpoint variables (**compcut**) on the Baseline Survey variables (**basecut**) and save the regression constant (**const**) and coefficient (**coeff**).

```
Regress        basecut compcut
Set            const=regression constant
Set            coeff=regression coefficient
```

**Step 11.** Drop all observations except the first observation and drop all variables except the regression constant (**const**) and coefficient (**coeff**). Create a null variable (**null**) equal to 1 that can be used to add the constant and coefficient variables to all household records in the data file created in Part I. Save the data to a temporary file, *temp\_regress\_result*. Then add these two variables to the data file created in the last step of Part I.

```
Keep records   if row=1
Keep variables const coeff
Set            null=1
```

```

Save          "temp_regress_result"
Load          "FTF_ZOI_Survey [COUNTRY] [YEAR] wealthindex AWI"
Merge         many to 1
    key variable null
    data file  "temp_regress_result"
    keep variables const coeff

```

**Step 12.** Calculate the CWI scores for each household (*cwi*) using the ZOI Survey wealth index scores (*awi*) and the regression constant and coefficient obtained in Step 10.

```

Set          cwi=0
Replace      cwi=const+(coeff*awi)
Replace      cwi=missing if awi=missing
Label variable "Comparative wealth index score"

```

**Step 13.** Create a variable that flags households in the ZOI Survey sample that fall below the threshold for the poorest quintile of the asset-based CWI (*comp\_poor*).

```

Set          comp_poor=0 if cwi!=missing
Replace      comp_poor=1 if cwi<-0.835740
Label values 0 "No"
              1 "Yes"
Label variable "HH is in poorest CWI quintile"

```

**Step 14.** After applying the household sampling weight, calculate the CWI indicator using the *comp\_poor* analytic variable. Repeat using the gendered household type and shock exposure severity disaggregate variables. (Sample code uses Stata syntax.).

```

svyset hhea [pweight=wgt_hh] strata(strata)
svy: tab comp_poor
svy: tab comp_poor genhhtype_dj, col
svy: tab comp_poor shock_sev, col

```

**Step 15.** Create a variable that assigns all surveyed households to a CWI quintile (*cwiquint*).

```

Set          cwiquint=missing
Replace      cwiquint=1 if cwi<(-0.835740)
Replace      cwiquint=2 if cwi<-0.553915 and cwiquint=missing
Replace      cwiquint=3 if cwi<-0.044190 and cwiquint=missing
Replace      cwiquint=4 if cwi<0.864125 and cwiquint=missing
Replace      cwiquint=5 if cwi>=0.864125 and cwiquint=missing
              and cwi!=missing
Label values 1 "Lowest (poorest)"
              2 "Second"
              3 "Middle"
              4 "Fourth"
              5 "Highest (wealthiest)"
Label variable "HH's CWI quintile"

```



### 11.3 References

Feres, J.C., & Mancero, X. (2001). El método de las necesidades básicas insatisfechas y sus aplicaciones en América Latina. Serie Estudios Estadísticos y Prospectivos. División de Estadística y Proyecciones Económicas. CEPAL-ECLAC: Santiago de Chile. Febrero.

The Demographic and Health Surveys Program. *Wealth index*. Available at:

<https://dhsprogram.com/topics/wealth-index/Index.cfm>

Rutstein, S.O. (2014). *Steps to constructing the new DHS wealth index*. Rockville, MD: ICF International. Available at:

[https://dhsprogram.com/programming/wealth%20index/Steps\\_to\\_constructing\\_the\\_new\\_DHS\\_Wealth\\_Index.pdf](https://dhsprogram.com/programming/wealth%20index/Steps_to_constructing_the_new_DHS_Wealth_Index.pdf)

Rutstein, S., & Staveteig, S. (2014). *Making the Demographic and Health Surveys Wealth Index comparable*. DHS Methodological Reports No. 9. Rockville, Maryland, USA: ICF International. Available at:

<https://www.dhsprogram.com/pubs/pdf/MR9/MR9.pdf>

The World Bank Group. (n.d.a). The World Bank Atlas method - detailed methodology.

<https://datahelpdesk.worldbank.org/knowledgebase/articles/378832-what-is-the-world-bank-atlas-method>. Accessed on July 17, 2019.

The World Bank Group. (n.d.b). World Bank Country and Lending Groups.

<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>. Accessed on July 17, 2019.

## 12. RESILIENCE INDICATORS

In this chapter, the Feed the Future resilience indicators are defined and their calculation described. Resilience refers to a set of capacities that enable households and communities to effectively function in the face of shocks and stresses and still meet a set of well-being outcomes. No single indicator measures resilience. The ZOI Survey collects data specifically to measure four resilience indicators: (1) the ability to recover from shocks and stresses index (ARSSI); (2) an index of social capital at the household level; (3) the proportion of households that believe local government will respond effectively to future shocks and stresses; and (4) the percent of households participating in group-based savings, micro-finance, or lending programs.

This chapter is divided into two subsections; the first describes the four resilience indicators and the guidelines to construct the indicators in general, and the second describes the step-by-step procedures to calculate each indicator.

### 12.1 Guidelines to construct the resilience indicators

#### 12.1.1 Ability to recover from shocks and stresses index

The ARSSI measures the ability of households to recover from the typical types of shocks and stressors, such as loss of a family member, loss of income, hunger, drought, flood, or conflict, that households experienced during the reference period.

The indicator is calculated using data on the number and severity of the shocks and stressors that households experienced in the year preceding the ZOI Survey, the households' recovery from the shocks and stressors that they experienced in the year preceding the survey, and their perceived ability to meet their food needs in the year following the survey.

The first step in calculating the ARSSI is to calculate the base ability to recover (ATR) index using the responses to two ZOI Survey questions posed to households that experienced at least one shock or stress during the year before the survey, after asking about all households' exposure to and the severity of the types of shocks and stressors the household might have experienced during the year before the survey:

Question V359: "Would you say that right now, your household's ability to meet your food needs is better than before these difficult times, the same as before these difficult times, or worse than before these difficult times?"

Question V360: "Looking ahead over the next year, do you believe your household's ability to meet your food needs will be better than before these difficult times, the same as before these difficult times, or worse than before these difficult times?"

The response options for both questions are as follows:

- Better than before these difficult times=1
- The same as before these difficult times=2
- Worse than before these difficult times=3

In the questionnaire, response options for questions V359 and V360 are ordered from positive to negative; a higher number means a lower ability to recover from shocks. To construct the indicator, first recode the variables, creating new variables for both questions so that a higher value means a greater ability to recover: Better than before these difficult times is 3, The same as before these difficult times is 2, and Worse than before these difficult times is 1.

Then add the recoded variables to a new variable to create the base ATR index, which has a minimum value of 2 and a maximum value of 6.

ATR=recoded response to question V359+recoded response to question V360

The next step in calculating the ARSSI is to calculate the shock exposure index (SEI) using the responses to ZOI Survey questions posed about households' exposure to the types of shocks and stressors and the severity that the household might have experienced during the year before the survey. The 16 shocks and stresses included in the core ZOI Survey questionnaire and their question numbers are as follows: too much rain (V309); too little rain (V312); erosion of land (V315); loss of land (V318); sharp increase in the price of food (V321); someone stealing or destroying belongings (V324); not being able to access inputs for crops (V328); disease affecting crops (V331); pests affecting crops (V334); theft of crops (V337); not being able to access inputs for livestock (V340); disease affecting livestock (V343); someone stealing animals (V346); not being able to sell crops, livestock, or other products at a fair price (V349); severe illness in the family (V352); and death in the household (V355).<sup>85</sup>

Because each surveyed household did not experience the same types of shocks and stresses of the same severity, it is necessary to create the SEI as an adjustment factor that is applied to the ATR index that accounts for the shocks and stresses to which a household was exposed out of the total number of shocks or stresses included in the ZOI Survey (i.e., 16 in the core questionnaire), and the perceived severity of the shock or stressor on household income and food consumption.

Perceived severity is measured using the impact of the shock or stress on food consumption. These variables are based on respondents' answers to the questions for each shock experienced:

“How severe was the impact on household food consumption?”

The possible responses are as follows:

Not severe=1

Somewhat severe=2

Severe=3

Extremely severe=4

The responses are combined into one severity variable that has a minimum value of 1 and a maximum value of 4 for each shock and stressor experienced.

---

<sup>85</sup> Both the number of shocks and stresses and the shocks and stresses themselves included in a country-customized questionnaire may differ from those included in the core main survey questionnaire used for Midline Surveys. Be sure to review the customized questionnaire carefully to ensure that all shocks and stresses are included in the analysis.

The SEI is then calculated as the sum of the incidence of experience of each shock or stress (a variable equal to 1 if the shock or stress was experienced and 0 otherwise) multiplied by the perceived severity of the shock or stress. The SEI ranges from 0 (if the household did not experience any shocks or stresses) to 64 (if the household experienced all 16 shocks or stresses at the highest level of severity).

Finally, the shock exposure-adjusted ARSSI is calculated to create a comparable measure of ability to recover that corrects for any differences among households in their shock and stress exposure. To create the ARSSI, a linear regression of the base ATR index on the SEI is run across all households with ATR and SEI values, yielding the amount by which an increase of 1 in the SEI can be expected to change the ATR index.

The estimated empirical equation is:

$$ATR = a + b(SEI)$$

The SEI coefficient ( $b$ ) is expected to be a negative number such that the higher the shock or stress exposure, the lower the ability to recover.

The coefficient  $b$  is then used to calculate the ARSSI for each household using the following equation:

$$ARSSI = ATR + b(Y - SEI)$$

Where  $Y$  is the mean SEI across households. As such, the ATR index value of a household with shock exposure below the mean would have a downward adjustment of its value, and the opposite would be true for a household with shock exposure above the mean.

### **12.1.2 Index of social capital at the household level**

This indicator measures the ability of households to draw on social networks to obtain support to reduce the impact of shocks and stressors on their households. It measures both the degree of bonding among households within their own communities (bonding social capital) and the degree of bridging between households in the area to households outside their own community (bridging social capital).

If the household responses indicate that they have reciprocal, mutually reinforcing relationships through which they could receive and provide support during times of need, they are considered to have social capital.

The indicator is constructed from two sub-indices, one measuring bonding social capital (support for people within the community) and one measuring bridging social capital (support from people in other communities). These two indices are based on the following eight questions in the household questionnaire:

Question V361: “Now I will ask you some questions about whether your household will be able to lean on others for financial or food support during difficult times. By difficult times I mean times when there is loss of a family member, loss of income, hunger, drought, flood, conflict or similar events.”

Question V361A: “During difficult times, will your household be able to lean on relatives living in your community?”

The response options are Yes=1 and No=2.

If the respondent answers “YES,” the subsequent question is asked:

Question V361B: “Will the same relatives living in your community that your household will be able to lean on during your difficult times also be able to lean on your household for financial or food support during their difficult times?”

The responses include: Yes=1, No, they won’t need to=2, No, they won’t be able to=3.

The above questions (V361A and V361B) are repeated for:

Relatives living outside your community? (V361C and V361D)

Non-relatives living in your community? (V361E and V361F)

Non-relatives living outside your community? (V361G and V361H)

The bonding social capital index considers responses to questions V361A, V361B, V361E, and V361F. The bridging social capital index considers responses to questions V361C, V361D, V361G, and V361H.

The index of social capital is the average of the two sub-indices. For both sub-components, the households need to report both being able to lean on and being able to be leaned on from a group to receive a score of 1 for that group; if they report being able to lean on others without others being able to lean on them for a group, they will receive a score of 0 for that group (i.e., that specific bonding or bridging sub-component). For bonding social capital, responses to both V361A and V361B need to be “Yes” to receive a score of 1, and then the same for V361E and V361F. For bridging social capital, responses to both V361C and V361D need to be “Yes” to receive a score of 1, and then the same for V361G and V361H.

During analysis, variables for questions that are skipped because of skip patterns in the questionnaire are also set to 0. That is, the analytic variable for V361B is set to 0 if the response to V361A is “No,” the analytic variable for V361D is set to 0 if the response to V361C is “No,” the analytic variable for V361F is set to 0 if the response to V361E is “No,” and the analytic variable for V361H is set to 0 if the response to V361G is “No.”

The indicator is calculated in two steps. First, the individual bonding social capital sub-index and the bridging social capital sub-index are calculated and standardized to a 0-100 scale as follows:

Bonding sub-index for each household= $([1 \text{ if yes to both V361A and V361B, 0 otherwise}] + [1 \text{ if yes to both V361E and V361F, 0 otherwise}]) \div 2 \times 100$

Bridging sub-index for each household= $([1 \text{ if yes to both V361C and V361D, 0 otherwise}] + [1 \text{ if yes to both V361G and V361H, 0 otherwise}]) \div 2 \times 100$

The second step is to average the two sub-indices:

Index of social capital score for each household=(Bonding sub-index+Bridging sub-index)÷2

### **12.1.3 Percent of households that believe local government will respond effectively to future shocks and stresses**

This indicator measures a household's perception of local government responsiveness in the face of future shocks and stresses. Local government responsiveness can refer to either local leaders or institutions. Believing in the ability of one's local government to respond to shocks and stresses is a proxy for trust, legitimacy, and effectiveness of local institutions and leadership. Such belief and trust contribute to transformative resilience capacity, or the enabling environment that supports—or limits—people's ability to prevent or mitigate the impact of shocks and stresses, and to deal with it and recover from it.

This indicator is based on the following household-level question (V362):

“Do you believe your local government will help the community cope with difficult times in the future, for example during [INSERT COUNTRY-SPECIFIC SHOCK]?”

The options are:    Yes=1  
                              No, will not be able to=2  
                              No, support not needed=3

Households that report that they will not need support are excluded from the indicator calculation.

### **12.1.4 Percent of households participating in group-based savings, micro-finance, or lending programs**

This indicator helps track the financial inclusion of households in the ZOI. The benefits of financial inclusion include lower transaction costs of day-to-day interactions (e.g., Mobile Money), ability to grow savings to smooth consumption and mitigate against shocks, and access to credit to invest in micro, small, and medium enterprises.

Group-based savings programs are formal or informal community programs that serve as a mechanism for people in poor communities, with otherwise limited access to financial services, to pool their savings. The specific composition and function of the savings groups vary and can include rotating disbursement as well as accumulating savings models.

According to the World Bank, micro-finance can be defined as approaches to provide financial services to households and micro-enterprises that are excluded from traditional commercial banking services. Typically, participants are low-income, self-employed, or informally employed individuals, with no formalized ownership titles on their assets and with limited formal identification papers.

A household is considered to be participating in a group-based savings, micro-finance, or lending program if any member of the household saved money with or took a loan or borrowed cash or in-kind from a group-based savings, micro-finance, or lending program in the past 12 months.

The information collected is on whether the household accessed credit through a group-based source or is an active member of a credit or micro-finance group—not whether the household saved money.

The data required to calculate this indicator were collected in Module 6, *Women's empowerment in agriculture*, of the core ZOI Survey questionnaire in surveys with fieldwork prior to 2022. Data, therefore, were not collected in households without a primary adult female decision-maker or households composed of only children. In 2022, the questions required to calculate this indicator were moved from Module 6 to Module 3, *Food security and resilience*, to ensure that data would be collected for all households. In Section 12.2.4, two approaches to calculating the indicator are presented:

- (1) Using only data collected from primary adult female decision-makers in Module 6 at both points in time (for baseline/midline comparative analyses)
- (2) Using only data collected in Module 3, which is available only in surveys with fieldwork in or after 2022 (for descriptive midline analyses)

To compare baseline and midline indicator estimates, approach (1) must be used because data for this indicator were collected from only households with primary adult female decision-makers in surveys with fieldwork before 2022. However, in cases in which baseline-midline comparison is not required or not possible, approach (2) can be used.

## 12.2 Step-by-step procedures to calculate resilience indicators

This section describes the detailed procedures to calculate each resilience indicator. The step-by-step procedures to calculate the four Feed the Future resilience indicators follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax resilience.do*.

### 12.2.1 Ability to recover from shocks and stresses index

This section describes the step-by-step procedures to calculate the Feed the Future ARSSI indicator. It is based on information collected in the core ZOI Survey questionnaire Module 3, *Food security and resilience*.

#### Definitions

Numerator	Sum of the ARSSI scores for all surveyed households that experienced at least one shock or stress during the 12 months preceding the survey
Denominator	Number of surveyed households that experienced at least one shock or stress during the 12 months preceding the survey
Unit of measure	Ability to recover score (range: 2 to 6) adjusted for the number and severity of shocks or stresses the household experienced (range varies across surveys) <sup>a</sup>
Level of data	Household
Sampling weight	Household
Disaggregation levels	Gendered household type* Wealth quintile
Treatment of missing data	Households with any missing or refused responses are excluded from the numerator and denominator.
Survey variables used	v309-v357, v359, v360, hhea, wgt_hh, strata
Analytic variables used	genhhstype_dj, awiquint, shock_sev
Analytic variables created	v309x-v357x, v359x, v360x, atr, perceived_sev1-perceived_sev16, sei, anymiss_sei, arssi, b_atr

\*Standard Feed the Future disaggregate

<sup>a</sup> Surveys analyzed to date have minimum ARSSI scores that fall between 1 and 2 and maximum ARSSI scores between 6 and 7.

## Calculations

**Step 1.** Calculate the base ATR index.

**Step 1.1.** Identify all the variables needed to construct the ARSSI indicator and check their frequencies to ensure that none of the questions have a large number of missing or refused responses. These variables are **v309-v360** in the core ZOI Survey questionnaire, but they may vary across countries, depending on the number of shocks included in the final country questionnaire.

**Step 1.2.** Create two analytic variables: **v359x** and **v360x**. Recode the response options of variables **v359** and **v360** so that a higher value indicates a greater ability to recover from shocks, and set any refused responses to missing.

```
Set          v359x=missing if v359>3
Replace      v359x=1 if v359=3
Replace      v359x=2 if v359=2
Replace      v359x=3 if v359=1
Label values 1 "Worse"
              2 "Same"
              3 "Better"
Label variable "HH's current recovery status"

Set          v360x=missing if v360>3
Replace      v360x=1 if v360=3
Replace      v360x=2 if v360=2
Replace      v360x=3 if v360=1
Label values 1 "Worse"
              2 "Same"
              3 "Better"
Label variable "HH's recovery status in 1 year"
```

**Step 1.3.** Sum the recoded variables **v359x** and **v360x** into one variable (**atr**) that has a minimum value of 2 and a maximum value of 6, and set **atr** to missing if **v359x** or **v360x** is missing.

```
Set          atr=v359x+v360x
Replace      atr=missing if v359x=missing or v360x=missing
Label variable "Ability to recover index (2-6)"
```

**Step 2.** Calculate the SEI (range 0-64).

**Step 2.1.** Create an analytic variable for each shock capturing how the shock affected the household's food consumption (**v311x**, **v314x**, **v317x**, **v320x**, **v323x**, **v326x**, **v330x**, **v333x**, **v336x**, **v339x**, **v342x**, **v345x**, **v348x**, **v351x**, **v354x**, **v357x**), recoding values greater than 4 to missing for variables **v311**, **v314**, **v317**, **v320**, **v323**, **v326**, **v330**, **v333**, **v336**, **v339**, **v342**, **v345**, **v348**, **v351**, **v354**, and **v357**.



```

For each      variable (var) in variable list v311 v314 v317
              v320 v323 v326 v330 v333 v336 v339 v342 v345 v348
              v351 v354 v357:
Set           `var'x=`var'
Replace      `var'x=missing if `var'>4
Label values 1 "Not severe"
              2 "Somewhat severe"
              3 "Severe"
              4 "Extremely severe"
Label variable "Recode of `var' ([`var' label])"

```

**Step 2.2.** Create analytic variables for each question asking whether the household experienced a certain shock during the year prior to the survey, recoding “no” responses from 2 to 0 (v309x, v312x, v315x, v318x, v321x, v324x, v328x, v331x, v334x, v337x, v340x, v343x, v346x, v349x, v352x, v355x).

```

For each      variable (var) in variable list v309 v312 v315
              v318 v321 v324 v328 v331 v334 v337 v340 v343 v346
              v349 v352 v355:
Set           `var'x=1 if `var'=1
Replace      `var'x=0 if `var'=2
Replace      `var'x=missing if `var'>2
Label values 0 "No"
              1 "Yes"
Label variable "Recode of `var' ([`var' label])"

```

**Step 2.3.** For shocks that were only asked of households that cultivated crops or that owned livestock, recode any variables that have missing values because they were skipped to be 0.

```

Replace      v328x=0 if v328x=missing and v327=2
Replace      v331x=0 if v331x=missing and v327=2
Replace      v334x=0 if v334x=missing and v327=2
Replace      v337x=0 if v337x=missing and v327=2
Replace      v340x=0 if v340x=missing and v340a=2
Replace      v343x=0 if v343x=missing and v340a=2
Replace      v346x=0 if v346x=missing and v340a=2

```

**Step 2.4.** Create analytic variables for the perceived impact of each shock on the household’s food consumption that set missing values to 0 so that they can be summed in Step 2.6 (perceived\_sev1-perceived\_sev16).

```

Set          perceived_sev1=v311x
Replace      perceived_sev1=0 if perceived_sev1=missing
Label variable "Severity on food consumption: too much rain"

Set          perceived_sev2=v314x
Replace      perceived_sev2=0 if perceived_sev2=missing
Label variable "Severity on food consumption: too little rain"

Set          perceived_sev3=v317x
Replace      perceived_sev3=0 if perceived_sev3=missing

```

Label variable	"Severity on food consumption: land erosion"
Set	perceived_sev4=v320x
Replace	perceived_sev4=0 if perceived_sev4=missing
Label variable	"Severity on food consumption: lost land"
Set	perceived_sev5=v323x
Replace	perceived_sev5=0 if perceived_sev5=missing
Label variable	"Severity on food consumption: sharp increase in food prices"
Set	perceived_sev6=v326x
Replace	perceived_sev6=0 if perceived_sev6=missing
Label variable	"Severity on food consumption: stolen/destroyed belongings"
Set	perceived_sev7=v330x
Replace	perceived_sev7=0 if perceived_sev7=missing
Label variable	"Severity on food consumption: access to crop inputs"
Set	perceived_sev8=v333x
Replace	perceived_sev8=0 if perceived_sev8=missing
Label variable	"Severity on food consumption: crop disease"
Set	perceived_sev9=v336x
Replace	perceived_sev9=0 if perceived_sev9=missing
Label variable	"Severity on food consumption: crop pests"
Set	perceived_sev10=v339x
Replace	perceived_sev10=0 if perceived_sev10=missing
Label variable	"Severity on food consumption: stolen crops"
Set	perceived_sev11=v342x
Replace	perceived_sev11=0 if perceived_sev11=missing
Label variable	"Severity on food consumption: access to livestock inputs"
Set	perceived_sev12=v345x
Replace	perceived_sev12=0 if perceived_sev12=missing
Label variable	"Severity on food consumption: livestock disease"
Set	perceived_sev13=v348x
Replace	perceived_sev13=0 if perceived_sev13=missing
Label variable	"Severity on food consumption: stolen animals"
Set	perceived_sev14=v351x
Replace	perceived_sev14=0 if perceived_sev14=missing
Label variable	"Severity on food consumption: unable to sell products"

```

Set           perceived_sev15=v354x
Replace       perceived_sev15=0 if perceived_sev15=missing
Label variable "Severity on food consumption: severe illness in
              HH"

Set           perceived_sev16=v357x
Replace       perceived_sev16=0 if perceived_sev16=missing
Label variable "Severity on food consumption: family member's
              death"

For each      variable (var) in variable list perceived_sev1-
              perceived_sev16:
  Label values 1 "Not severe"
               2 "Somewhat severe"
               3 "Severe"
               4 "Extremely severe"

```

**Step 2.5.** Create a variable that flags households that have a missing or refused response to a question included in the calculation of the SEI variable (`anymiss_sei`).

```

Set           anymiss_sei=0
Replace       anymiss_sei=1 if v309x=missing or (v309x=1 and
              v311x=missing)
Replace       anymiss_sei=1 if v312x=missing or (v312x=1 and
              v314x=missing)
Replace       anymiss_sei=1 if v315x=missing or (v315x=1 and
              v317x=missing)
Replace       anymiss_sei=1 if v318x=missing or (v318x=1 and
              v320x=missing)
Replace       anymiss_sei=1 if v321x=missing or (v321x=1 and
              v323x=missing)
Replace       anymiss_sei=1 if v324x=missing or (v324x=1 and
              v326x=missing)
Replace       anymiss_sei=1 if v328x=missing or (v328x=1 and
              v330x=missing)
Replace       anymiss_sei=1 if v331x=missing or (v331x=1 and
              v333x=missing)
Replace       anymiss_sei=1 if v334x=missing or (v334x=1 and
              v336x=missing)
Replace       anymiss_sei=1 if v337x=missing or (v337x=1 and
              v339x=missing)
Replace       anymiss_sei=1 if v340x=missing or (v340x=1 and
              v342x=missing)
Replace       anymiss_sei=1 if v343x=missing or (v343x=1 and
              v345x=missing)
Replace       anymiss_sei=1 if v346x=missing or (v346x=1 and
              v348x=missing)
Replace       anymiss_sei=1 if v349x=missing or (v349x=1 and
              v351x=missing)
Replace       anymiss_sei=1 if v352x=missing or (v352x=1 and
              v354x=missing)

```

```

Replace          anymiss_sei=1 if v355x=missing or (v355x=1 and
                  v357x=missing)
Label values      0 "No"
                  1 "Yes"
Label variable    "HH is missing any variables in SEI"

```

**Step 2.6.** Multiply the analytic variables indicating whether a household experienced a shock (those created in Steps 2.2 and 2.3) by the analytic variables indicating the perceived severity of each shock experienced (those created in Step 2.4) and sum into one variable to create the weighted SEI variable (**sei**). Then set **sei** to missing if **anymissing**, created in the previous step, is equal to 1. The **sei** can range from 0 (household experienced no shocks) to 64 (household experienced all 16 shocks/stresses at the highest level of severity).

```

Set              sei=
                  (perceived_sev1*v309x)+(perceived_sev2*v312x)+
                  (perceived_sev3*v315x)+(perceived_sev4*v318x)+
                  (perceived_sev5*v321x)+(perceived_sev6*v324x)+
                  (perceived_sev7*v328x)+(perceived_sev8*v331x)+
                  (perceived_sev9*v334x)+(perceived_sev10*v337x)+
                  (perceived_sev11*v340x)+(perceived_sev12*v343x)+
                  (perceived_sev13*v346x)+(perceived_sev14*v349x)+
                  (perceived_sev15*v352x)+(perceived_sev16*v355x)
Replace          sei=missing if anymissing=1
Label variable    "Shock exposure index"

```

**Step 2.7.** Create a variable equal to the average SEI across households (**mean\_sei**).

```

Set              mean_sei=mean(sei)
Label variable    "Mean SEI for surveyed HHs"

```

**Step 3.** Calculate shock-exposure adjusted ARSSI.

**Step 3.1.** Run a linear regression of the base ATR index on the SEI to obtain 'b' (**b\_atr**) the SEI coefficient.

```

Regress          atr sei
Set              b_atr=coefficient of variable sei
Label variable    "SEI regression coefficient"

```

**Step 3.2.** Calculate the ARSSI (**arssi**) for each household.

```

Set              arssi=atr+b_atr*(mean_sei-sei)
Label variable    "Ability to recover from shocks and stresses
                  index (ARSSI)"

```

**Step 4.** Create a variable that assigns households that reported not experiencing shocks into their own category and then divide the remaining households by their SEI values so that they comprise three roughly equal groups (terciles): low, moderate, and high shock exposure severity (**shock\_sev**). Reassign the categories so that they have values 1-4. (Sample code using Stata syntax.) This variable is used as a disaggregate for many indicators across the survey report.

```

xtile shock_sev=sei if sei >0 & sei!=., nq(3)
replace shock_sev=0 if sei==0
label variable shock_sev "Shock exposure severity"
recode shock_sev (0=1) (1=2) (2=3) (3=4)
label define shocklevel 1 "Did not experience any shock" 2
                        "Low"
                        3 "Moderate" 4 "High"
label values shock_sev shock_level

```

**Step 5.** After applying the household sampling weight, calculate the mean ARSSI using the **arssi** analytic variable. Repeat using the gendered household type and wealth quintile disaggregates. (Sample code uses Stata syntax.)

```

svyset hhea [pweight=wgt_hh], strata(strata)
svy: mean arssi
svy: mean arssi, over(genhhtype_dj)
svy: mean arssi, over(awiquint)

```

### 12.2.2 Index of social capital at the household level

This section describes the step-by-step procedures to calculate the index of social capital at the household level. It is based on information collected in the core ZOI Survey questionnaire Module 3, *Food security and resilience*.

#### Definitions

Numerator	Sum of social capital index scores normalized to 100 for all households
Denominator	Number of surveyed households
Unit of measure	Score ranging from 0 to 100 (higher score means more social capital)
Level of data	Household
Sampling weight	Household
Disaggregation levels	First level: Social capital component (overall, bonding, and bridging social capital)*  Second level: Gendered household type* Wealth quintile Poverty status Shock exposure severity
Treatment of missing data	If <b>v361b</b> has a value of missing because <b>v361a</b> has a value of no, set <b>v361b</b> to no. If <b>v361d</b> has a value of missing because <b>v361c</b> has a value of no, set <b>v361d</b> to no. If <b>v361f</b> has a value of missing because <b>v361e</b> has a value of no, set <b>v361f</b> to no.  If <b>v361h</b> has a value of missing because <b>v361g</b> has a value of no, set <b>v361h</b> to no. After recoding the values for <b>v361b</b> , <b>v361d</b> , <b>v361f</b> , and <b>v361h</b> , households with any missing responses are excluded from the numerator and denominator.
Survey variables used	<b>v361a</b> , <b>v361b</b> , <b>v361c</b> , <b>v361d</b> , <b>v361e</b> , <b>v361f</b> , <b>v361g</b> , <b>v361h</b> , <b>hhea</b> , <b>wgt_hh</b> , <b>strata</b>

Analytic variables used	genhhtype_dj, awiquint, poor190, shock_sev
Analytic variables created	v361ax, v361bx, v361cx, v361dx, v361ex, v361fx, v361gx, v361hx, scap_index, scap_bond, scap_bridge, i_scap_index, i_scap_bond, i_scap_bridge

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Create the required intermediate analytic variables (v361ax, v361bx, v361cx, v361dx, v361ex, v361fx, v361gx, v361hx) in which “no” responses are recoded from 2 to 0. Also set variables v361bx, v361dx, v361fx, and v361hx to 0 if the variable is missing a value because the question was skipped because the answer to the question preceding it was no.

```

Set          v361ax=1 if v361a=1
Replace      v361ax=0 if v361a=2
Replace      v361ax=missing if v300r#1

Set          v361cx=1 if v361c=1
Replace      v361cx=0 if v361c=2
Replace      v361cx=missing if v300r#1

Set          v361ex=1 if v361e=1
Replace      v361ex=0 if v361e=2
Replace      v361ex=missing if v300r#1

Set          v361gx=1 if v361g=1
Replace      v361gx=0 if v361g=2
Replace      v361gx=missing if v300r#1

Set          v361bx=1 if v361b=1
Replace      v361bx=0 if v361b#1 or v361a=2
Replace      v361bx=missing if v300r#1

Set          v361dx=1 if v361d=1
Replace      v361dx=0 if v361d#1 or v361c=2
Replace      v361dx=missing if v300r#1

Set          v361fx=1 if v361f=1
Replace      v361fx=0 if v361f#1 or v361e=2
Replace      v361fx=missing if v300r#1

Set          v361hx=1 if v361h=1
Replace      v361hx=0 if v361h#1 or v361h=2
Replace      v361hx=missing if v300r#1

For each     variable (var) in variable list v361ax v361bx
              v361cx v361dx v361ex v361fx v361gx v361hx:
Label values 0 "No"

```

```

1 "Yes"
Label variable "Recode of `var' ([`var' label])"

```

**Step 2.** Calculate bonding social capital (*scap\_bond*).

```

Set          scap_bond=0
Replace      scap_bond=1 if v361ax=1 and v361bx=1
Replace      scap_bond=scap_bond+1 if v361ex=1 and v361fx=1
Replace      scap_bond=missing if v361ax=missing or
361bx=missing or v361ex=missing or v361fx=missing
Label variable "Bonding social capital (0-2)"

```

**Step 3.** Calculate bridging social capital (*scap\_bridge*).

```

Set          scap_bridge=0
Replace      scap_bridge=1 if v361cx=1 and v361dx=1
Replace      scap_bridge=scap_bridge+1 if v361gx=1 and
v361hx=1
Replace      scap_bridge=missing if v361cx=missing or
361dx=missing or v361gx=missing or v361hx=missing
Label variable "Bridging social capital (0-2)"

```

**Step 4.** Calculate index of social capital (*scap\_index*) by averaging the bonding and bridging social capital.

```

Set          scap_index=missing
Replace      scap_index=(scap_bond+scap_bridge)÷2
Label variable "Index of social capital (0-2)"

```

**Step 5.** Rescale the index and the two sub-indices to be out of 100 (*i\_scap\_index*, *i\_scap\_bond*, *i\_scap\_bridge*).

```

Set          i_scap_index=(scap_index÷2)*100
Label variable "Social capital index (rescaled 0-100)"

Set          i_scap_bond=(scap_bond÷2)*100
Label variable "Bonding social capital (rescaled 0-100)"

Set          i_scap_bridge=(scap_bridge÷2)*100
Label variable "Bridging social capital (rescaled 0-100)"

```

**Step 6.** Calculate the sample-weighted mean index of social capital after applying the household weight to the data using the analytic variable created in Step 5 (*i\_scap\_index*). Repeat using the gendered household type, wealth quintile, poverty status, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```

svyset hhea [pweight=wgt_hh], strata(strata)
svy: mean i_scap_index
svy: mean i_scap_index, over(genhhtype_dj)
svy: mean i_scap_index, over(awiquint)

```

```
svy: mean i_scap_index, over (poor190)
svy: mean i_scap_index, over (shock_sev)
```

**Step 7.** Repeat Step 6 for the rescaled bonding and bridging social capital sub-indices.

```
svy: mean i_scap_bond
svy: mean i_scap_bond, over (genhhstype_dj)
svy: mean i_scap_bond, over (awiquint)
svy: mean i_scap_bond, over (poor190)
svy: mean i_scap_bond, over (shock_sev)

svy: mean i_scap_bridge
svy: mean i_scap_bridge, over (genhhstype_dj)
svy: mean i_scap_bridge, over (awiquint)
svy: mean i_scap_bridge, over (poor190)
svy: mean i_scap_bridge, over (shock_sev)
```

### 12.2.3 Percent of households that believe local government will respond effectively to future shocks and stresses

This section describes the step-by-step procedures to calculate the percent of households that believe local government will respond effectively to future shocks and stresses. It is based on information collected in the core ZOI Survey questionnaire Module 3, *Food security and resilience*.

#### Definitions

Numerator	Number of households that believe local government will respond effectively to future shocks and stresses
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Gendered household type* Wealth quintile Poverty status Shock exposure severity
Treatment of missing data	Households with a “No, support not needed” response or that are missing a response are excluded from the numerator and denominator.
Survey variables used	v362, hhea, wgt_hh, strata
Analytic variables used	genhhstype_dj, awiquint, poor190, shock_sev
Analytic variables created	locgov_resp

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create the indicator variable (**locgov\_resp**), dropping any households for which a response is missing or with a “No, support not needed” response.

```
Set          locgov_resp=missing
Replace      locgov_resp=1 if v362=1
Replace      locgov_resp=0 if v362=2
```



**Step 2.** After applying the household sampling weight, calculate the percentage of households that believe local government will respond effectively to future shocks and stresses using the `locgov_resp` analytic variable. Repeat using the gendered household type, wealth quintile, poverty status, and shock severity exposure disaggregates. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab locgov_resp
svy: tab locgov_resp genhhtype_dj, col
svy: tab locgov_resp awiquint, col
svy: tab locgov_resp poor190, col
svy: tab locgov_resp shock_sev, col
```

#### 12.2.4 Percent of households participating in group-based savings, micro-finance, or lending programs

This section describes the step-by-step procedures to calculate the percent of households participating in group-based savings, micro-finance, or lending programs. Because the data for this indicator are collected in different places in surveys with fieldwork before 2022 and in surveys with fieldwork in 2022 or later (see Section 12.1.4), two approaches to calculating the indicator are presented:

- (1) Using only data collected from primary adult female decision-makers in Module 6 (for baseline/midline comparative analyses)
- (2) Using only data collected in Module 3, which is available only in surveys with fieldwork in or after 2022 (for descriptive midline analyses)

To compare baseline and midline indicator estimates, approach (1) must be used because data for this indicator were collected from only households with primary adult female decision-makers in surveys with fieldwork before 2022. However, in cases in which baseline-midline comparison is not required or not possible, approach (2) can be used.

#### Approach 1: Using data from only primary female adult decision-makers collected in Module 6 (for comparative analyses)

##### Definitions, approach 1

Numerator	Number of households with a primary adult female decision-maker who participated in a group-based savings, micro-finance, or lending program in the 12 months preceding the survey
Denominator	Number of surveyed households with a primary adult female decision-maker
Unit of measure	Percentage
Level of data	Household
Sampling weight	Primary adult female decision-maker
Disaggregation levels	Gendered household type* Wealth quintile Poverty status Shock exposure severity
Treatment of missing data	All “don’t know” and missing responses are considered to be “no” responses when calculating this indicator.

Survey variables used	v600r, v6308_5, v6308_6, v6405_04, hhea, wgt_hh, strata
Analytic variables used	genhhtype_dj, awiquint, poor190, shock_sev
Analytic variables created	access_finance1

\*Standard Feed the Future disaggregate

## Calculations, approach 1

**Step 1.** Create a binary variable that indicates whether primary female adult decision-makers who completed Module 6 reported that at least one person from their household participated in a group-based savings, micro-finance, or lending program in the 12 months preceding the survey (**access\_finance1**). The variable indicates whether anyone in the household took a loan from a group-based micro-finance or lending program (v6308\_5) or from an informal credit or savings group (v6308\_6), or whether the primary adult female decision-maker herself is an active member of a savings, credit, or micro-finance group (v6405\_05).

```

Set          access_finance1=missing
Replace      access_finance1=0 if v600r=1
Replace      access_finance1=1 if v6308_5 ≤3
Replace      access_finance1=1 if v6308_6 ≤3
Replace      access_finance1=1 if v6405_05=1
Label values 0 "No"
              1 "Yes"
Label variable "PAFDM responded that HH participated in group-
               based savings, microfinance, or lending"

```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that participated in a group-based savings, micro-finance, or lending program in the 12 months preceding the survey using **access\_finance1**. Repeat using the gendered household type, wealth index, poverty status, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```

svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab access_finance1
svy: tab access_finance1 genhhtype_dj, col
svy: tab access_finance1 awiquint, col
svy: tab access_finance1 poor190, col
svy: tab access_finance1 shock_sev, col

```

**Approach 2: Using data collected in Module 3, which is available only in surveys with fieldwork in or after 2022 (for descriptive analyses at one point in time)**

## Definitions, approach 2

Numerator	Number of households with at least one household member who participated in a group-based savings, micro-finance, or lending program in the 12 months preceding the survey
Denominator	Number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household

Disaggregation levels	Gendered household type* Wealth quintile Poverty status Shock exposure severity
Treatment of missing data	All “don’t know” and missing responses are considered to be “no” responses when calculating this indicator.
Survey variables used	v300r, v364, v365, v366, v367, hhea, wgt_hh, strata
Analytic variables used	genhhtype_dj, awiquint, poor190, shock_sev
Analytic variables created	access_finance2

\*Standard Feed the Future disaggregate

## Calculations, approach 2

**Step 1.** Create a binary variable that indicates whether respondents to Module 3 reported that at least one person from their household participated in a group-based savings, micro-finance, or lending program in the 12 months preceding the survey (**access\_finance2**). The variable indicates whether anyone in the household took a loan from a group-based micro-finance or lending program (**v364**) or from an informal credit or savings group (**v365**), or whether a primary adult decision-maker in the household is an active member of a savings, credit, or micro-finance group (**v366**, **v367**).

```

Set          access_finance2=missing
Replace      access_finance2=0 if v300r=1
Replace      access_finance2=1 if v364≤3
Replace      access_finance2=1 if v365≤3
Replace      access_finance2=1 if v367=1 and v366=1
Label values 0 "No"
              1 "Yes"
Label variable "HH participated in group-based savings,
               microfinance, or lending"

```

**Step 2.** After applying the household sampling weight, calculate the percentage of households that participated in a group-based savings, micro-finance, or lending program in the 12 months preceding the survey using **access\_finance2**. Repeat using the gendered household type, wealth quintile, poverty status, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```

svyset hhea [pweight=wgt_hh], strata(strata)
svy: tab access_finance2
svy: tab access_finance2 genhhtype_dj, col
svy: tab access_finance2 awiquint, col
svy: tab access_finance2 poor190, col
svy: tab access_finance2 shock_sev, col

```

## 12.3 References

Frankenberger, T., Mueller, M., Spangler, T., & Alexander, S. (2013). *Community resilience: Conceptual framework and measurement Feed the Future learning agenda*. Rockville, MD: Westat. Available at: [https://www.agrilinks.org/sites/default/files/resource/files/FTF%20Learning\\_Agenda\\_Community\\_Resilience\\_Oct%202013.pdf](https://www.agrilinks.org/sites/default/files/resource/files/FTF%20Learning_Agenda_Community_Resilience_Oct%202013.pdf)

Sagara, B. (2018). *Resilience measurement practical guidance note series 2: Measuring shocks and stresses*. Produced by Mercy Corps as part of the Resilience Evaluation, Analysis and Learning (REAL) Associate Award. Available at:

[https://reliefweb.int/sites/reliefweb.int/files/resources/GN02\\_ShocksandStressesMsmnt\\_Final.pdf](https://reliefweb.int/sites/reliefweb.int/files/resources/GN02_ShocksandStressesMsmnt_Final.pdf)

## 13. THE FIVE DOMAINS OF EMPOWERMENT

### 13.1 Background

The Abbreviated Women's Empowerment in Agriculture Index (A-WEAI) is a composite index developed by the United States Agency for International Development, the International Food Policy and Research Institute, and the Oxford Poverty and Human Development Initiative to monitor changes in women's empowerment and gender equality resulting from United States Agency for International Development programs. For Midline Surveys, the A-WEAI indicator will not be calculated because data are not collected from primary adult male decision-makers, which are required to calculate one of the two A-WEAI sub-indices: the gender parity index. However, data collected from primary adult female decision-makers allow for the calculation of the second A-WEAI sub-index: the five domains of empowerment (5DE) and its components (the percentage of women achieving empowerment and the average adequacy score of disempowered women), as well as the percentage of women who are adequate in each of the six A-WEAI indicators, among three populations: (1) all primary adult female decision-makers, (2) all primary adult female decision-makers with complete 5DE data (that is, a value for each of the six A-WEAI indicators), and (3) female primary adult decision-makers who are disempowered.

The 5DE captures empowerment, agency, and inclusion of women in agriculture in five domains: (1) decision-making over production, (2) access to resources, (3) control over income, (4) group participation and leadership, and (5) time allocation. The domains of empowerment contribute to improved economic status of poor households, improved nutrition and health, and increased agricultural productivity at the household level (Hillesland, 2016). Six indicators are used to calculate the 5DE; each domain has one indicator, with the exception of the access to resources domain, which has two indicators (see **Table 15**).

**Table 15: A-WEAI Domains, Indicators, and Indicator Weights**

Domain	Indicator	Weight
1. Decision-making over production	1. Input in productive decisions	1/5
2. Access to resources	2. Ownership of assets	1/15
	3. Access to credit and decisions on it	2/15
3. Control over income	4. Control over use of income	1/5
4. Group participation and leadership	5. Group membership	1/5
5. Time allocation	6. Workload	1/5

The 5DE is calculated using data collected from primary adult female decision-makers. The primary adult female decision-maker is the adult female, 18 years of age or older, who makes more social and economic decisions than other women in the household. A household can have only one primary female decision-maker. A household will not have a primary adult female decision-maker if there are no adult females 18 years of age or older in the household, or if there is a woman 18 years of age or older, but she is too unwell or too cognitively impaired to be considered a primary decision-maker. Note that a household that has a woman 18 years of age or older should have a primary female decision-maker, even if she is not involved in many decisions made for the household. It is an extremely rare situation in

which there is a woman 18 years of age or older in a household, but the household lacks a primary adult female decision-maker.

## 13.2 Guidelines to construct the 5DE

This section provides instructions for constructing the 5DE and builds on existing material (see Alkire et al., 2012; Alkire et al., 2013; Malapit et al., 2015) by providing step-by-step guidance for calculating the 5DE. Part 1 describes the steps for data preparation and the calculation of the six A-WEAI indicators used to calculate the 5DE. Part 2 describes the steps for computing the 5DE, including how to calculate the percentage of women who are empowered. Part 3 summarizes measures of adequate achievement in each of the six A-WEAI indicators that are also reported in midline indicator assessment reports.

The 5DE is computed using data collected from primary adult female decision-makers who are de jure household members in Module 6, *Women's empowerment in agriculture*, in ZOI Surveys. The sample for calculating the 5DE includes only respondents who have complete 5DE data—that is, a value for all six A-WEAI indicators. Appendix A, **Table AI** summarizes the 5DE domains and A-WEAI indicators, relevant survey questions and ZOI Survey variables, the adequacy and inadequacy criteria, and weight assigned to each A-WEAI indicator to compute the 5DE.<sup>86</sup>

### 13.2.1 Part 1: Preparing the data and calculating the A-WEAI indicators

The data preparation steps include performing checks for consistent skip patterns, out-of-range values, outliers (extreme or implausible responses), missing responses, and duplicate observations, and then using the data to create the six A-WEAI indicators, which indicate achievement in that aspect of empowerment. The indicators are dichotomized and coded as 0 (inadequate) or 1 (adequate). Respondents receive a score of 0 or 1 to reflect whether they have adequate achievement in each indicator.

#### Domain 1: Decisionmaking over production

To capture the individual's decision-making over production, one indicator is used: input into production decisions. Empowerment in this domain is measured using two sub-indicators: (a) respondents' input into decisions related to four types of agriculture-related productive activities: food crop farming, cash crop farming, livestock raising, and fishing or fishpond culture; and (b) respondents' autonomy in making their own decisions about the same four activities.

For the first sub-indicator, each activity is scored 0 if the respondent has input into no decisions or few decisions made about the activity, or 1 if the respondent has input into at least some decisions about the activity. For the second sub-indicator, each activity is scored 0 if the respondent feels that she cannot make her own decisions about the activity to a medium or high extent, or 1 if the respondent feels that she can make her own decisions about the activity to a medium or high extent. Each sub-indicator has a possible score between 0 and 4. The two sub-indicators are aggregated to obtain a total score between 0 and 8 for the domain. Respondents are considered to have adequate achievement in this domain if they have a total score of 1 or higher. Respondents are asked:

---

<sup>86</sup> Appendix A, **Table AI** is adapted from Malapit et al., 2015; Alkire et al., 2012; and Alkire et al., 2013.

“Did you yourself participate in [ACTIVITY] in the past 12 months?” (question V6201)

The activities are food crop farming (6.2\_1), cash crop farming (6.2\_2), livestock raising (6.2\_3), and fishing or fishpond culture (6.2\_6).

Respondents who participated in an activity are then asked:

“When decisions are made regarding [ACTIVITY], who is it that normally takes the decision?” (question V6202)

The five options are (A) *self*, (B) *partner/spouse*, (C) *other household member*, (D) *other non-household member*, and (X) *not applicable*. Respondents may list more than one person.

Respondents who do not decide alone—those who report making decisions with others—are asked two follow-up questions:

“How much input do you have in making decisions about [ACTIVITY]?” (question V6203)

The options are (01) *no input or input into very few decisions*, (02) *input into some decisions*, (03) *input into most or all decisions*, and (93) *no decision made*.

“To what extent do you feel you can make your own personal decisions regarding [ACTIVITY] if you want(ed) to?” (question V6204)

The options are: (1) *not at all*, (2) *small extent*, (3) *medium extent*, and (4) *to a high extent*.

## **Domain 2: Access to resources**

To capture the individual’s control over productive resources, two indicators are used: (1) ownership of assets and (2) access to credit and decisions related to credit. Respondents are considered adequate in asset ownership if they own at least two types of small assets or one type of large asset,<sup>87</sup> jointly or alone. Respondents are considered adequate in access to credit and decisions on credit if they participate in at least one decision for at least one source of credit.

### **Ownership of assets**

An asset ownership score is calculated to reflect whether the respondent owns—alone or jointly—the assets included in the survey. For each asset that a household owns, 1 is assigned if respondents report that they own it solely or jointly, and 0 is assigned if they report that someone else owns it.

Respondents are considered to have adequate achievement in asset ownership if they own alone or jointly at least two types of small assets or one type of large asset. Small asset types include chickens, ducks, turkeys, and pigeons; hand tools, such as trowel, hoe, shovel, or machete; non-mechanized farming equipment, such as animal-drawn plough, cart, or wheelbarrow; and small consumer durables, such as a radio or cookware. Large asset types include agricultural land; other land not used for

---

<sup>87</sup> Small assets include chickens, ducks, turkeys, or pigeons; hand tools for farming; non-mechanized farming equipment; and small consumer durables. Large assets include agricultural land, other land not used for agriculture purposes, small and large livestock, fishpond or fishing equipment, mechanized farm equipment, non-farm business equipment, house or other structure, large consumer durables, cell phone, and means of transport.

agricultural purposes; a house or other structures; livestock, such as oxen, cattle, goats, sheep, or pigs; fishponds or fishing equipment; mechanized farm equipment, such as a tractor-drawn plough, power tiller, or treadle pump; non-farm business equipment, such as solar panels, a sewing machine, brewing equipment, or fryers; large consumer durables, such as a couch, refrigerator, or television; a cell phone; or a means of transportation, such as a bicycle, motorcycle, or car. Respondents are considered to be inadequate in this indicator if they do not own at least one type of large asset or two types of small assets, or if the household owns no assets.

Respondents are asked:

“Does anyone in your household currently have any [ITEM]?” (question V6301)

The options are (1) Yes and (2) No. For each asset the households own, respondents are then asked:

“Do you own any of the [ITEM] either by yourself or jointly with someone else?” (question V6303)

The options are (A) *self*, (B) *partner/spouse*, (C) *other household member*, (D) *other non-household member*, and (X) *no*. Respondents may list more one response if more than one person owns the asset.

### **Access to credit and decisions about credit**

This indicator on access to credit and decision-making about credit is measured using two sub-indicators: (a) the respondent provided input into the decision to borrow from at least one lending source, and (b) the respondent provided input into the decision about what to do with the money or item borrowed from at least one lending source. The survey collects information on access to credit from six different types of sources: non-government organizations; formal lenders, such as a bank or financial institution; informal lenders; friends or relatives; group-based micro-finance groups, such as village associations and saving groups; and informal credit or saving groups, such as merry-go-rounds, tontines, and funeral societies.

For the first sub-indicator, each lending source is scored 0 if anyone in the household borrowed from the source but the respondent did not participate in the decision to borrow, or 1 if someone in the household borrowed from the source and the respondent participated in the decision to borrow. For the second sub-indicator, each lending source is scored 0 if anyone in the household borrowed from the source but the respondent did not participate in the decision about what to do with the money or item borrowed, or 1 if someone in the household borrowed from the source and the respondent participated in the decision about what to do with the money or item borrowed. The sub-indicator is set to missing if the household borrowed from the source but no decisions regarding how to use the loan were made in the 12 months preceding the survey (i.e., the response was recorded as “Not applicable”).

Respondents are considered to have adequate achievement in this indicator if they made decisions alone or jointly to borrow from at least one lending source or if they made decisions alone or jointly about how to use the money or item borrowed from at least one lending source during the year preceding the survey. Respondents who did not participate in any of these decisions related to credit or who live in households that did not access credit are considered inadequate on access to credit and decisions on credit and receive a score of 0.



Respondents are asked:

“Has anyone in your household taken any loans or borrowed cash or in-kind from [SOURCE] in the past 12 months?” (question V6308)

The options are (1) *yes, cash*; (2) *yes, in-kind*; (3) *yes, cash and in-kind*; (4) *no*; and (8) *don’t know*.

The sources are (6.3b\_1) *Non-governmental organization*, (6.3b\_2) *Informal lender*, (6.3b\_3) *Formal lender*, (6.3b\_4) *Friends or relatives*, (6.3b\_5) *Group-based micro-finance or lending*, and (6.3b\_6) *Informal credit or savings group*.

Respondents who live in households that accessed credit are asked two follow-up questions:

“Who made the decision to borrow from [SOURCE]?” (question V6309)

“Who makes the decision about what to do with the money/item borrowed from [SOURCE]?” (question V6310)

Response options for both questions are (A) *self*, (B) *partner or spouse*, (C) *other household member*, (D) *other non-household member*, and (X) *not applicable*. Respondents can provide more than one response if more than one person makes the decision.

### **Domain 3: Control over income**

To capture the individual’s control over income and expenditures, one indicator is used, control over use of income based on economic activity participation. Empowerment in this domain is measured with two sub-indicators: (1) respondents’ input into decisions on the use of income from economic activities; and (2) respondents’ autonomy in making decisions about non-farm economic activities, salary or wage employment, major household expenditures, and minor household expenditures.

For the first sub-indicator, each activity is scored 0 if the respondent participated in the activity and has input into no decisions or few decisions made about the income from that activity, or 1 if the respondent has input into at least some decisions made about the income from activity. There are six relevant activities: (6.2\_1) food crop farming, (6.2\_2) cash crop farming, (6.2\_3) livestock production, (6.2\_4) non-farm activities, (6.2\_5) wage and salary work, and (6.2\_6) fishing or fishpond aquaculture. The first sub-indicator, therefore, has a possible score between 0 and 6. For the second sub-indicator, each relevant activity is scored 0 if the respondent feels that she cannot make her own decisions about the activity to a medium or high extent, or 1 if the respondent feels that she can make her own decisions about the activity to a medium or high extent. There are four relevant activities for the second sub-indicator: (6.2\_4) non-farm economic activities, (6.2\_5) wage and salary employment, (6.2\_7) major household expenditures, and (6.2\_8) minor household expenditures. The second sub-indicator, therefore, has a possible score between 0 and 4. The two sub-indicators are aggregated for a total score of between 0 and 10 for the indicator. Respondents are considered to have adequate achievement in this domain if they have a total score of 1 or higher, except in one instance: if respondents have a total score of 1 because they had input only into at least some decisions regarding minor household expenditures, they are considered to have inadequate achievement in this domain.

Three of the same questions used to calculate the Domain I indicator are also used to calculate this indicator—questions V6201, V6202, and V604—although additional activities are considered for this indicator. One additional question is used to calculate this indicator:

“How much input did you have in decisions on the use of income generated from [ACTIVITY]?”  
(question V6205)

The options are (1) *no input or input into very few decisions*, (2) *input into some decisions*, (3) *input into most or all decisions*, and (93) *no decision made*.

#### **Domain 4: Leadership in the community**

One indicator is used to capture the individual’s potential for leadership and influence in her community: membership in an economic or social group.

Respondents are asked:

“Is there a [GROUP] in your community?” (question V6404)

If respondents indicate that a certain group exists, they are asked a follow-up question:

“Are you an active member of this [GROUP]?” (question V6405)

Options are membership in the following types of groups: (6.4\_01) agricultural, livestock, or fisheries producers’ groups; (6.4\_02) water users’ group; (6.4\_03) forest users’ group; (6.4\_04) credit or micro-finance group; (6.4\_05) mutual help or insurance group; (6.4\_06) trade or business association; (6.4\_07) civic group; (6.4\_08) local government; (6.4\_09) religious group; (6.4\_10) other women’s group; and (6.4\_11) other formal or informal group.

For each type of group, a score of 1 is assigned if respondents report that they are active members of that group, and a score of 0 is assigned if they report that they are not active members of that group, if that group does not exist in their community, or if they do not know if the group exists in their community. Respondents are considered to have adequate achievement in leadership in the community if they are an active member of at least one group.

#### **Domain 5: Time allocation**

One indicator, workload, is used to capture the individual’s time allocation. Information is collected using a detailed 24-hour time allocation module in which respondents are asked to recall the time that they spent on activities in the 24 hours before the interview, starting at 4:00 a.m. on the day before the interview. Information is collected for primary activities only and recorded for 15-minute intervals (question V6601). The number of hours worked is defined as the sum of the time the individual reported spending on work-related tasks as the primary activity. Work and non-work activity categories are shown in **Table 16**.

**Table 16: Activities Included in the A-WEAI Time Use Module**

Work activity categories	Non-work activity categories
School, including homework	Commuting to or from work or school
Work as employed	Sleeping and resting
Own business work	Eating and drinking
Food crop farming	Personal care
Cash crop farming	Traveling, but not for work or school
Livestock raising	Watching TV, listening to radio, or reading
Shopping or getting service (including health services)	Exercising
Fishing or fishpond culture	Social activities and hobbies
Weaving, sewing, textile care	Religious activities
Cooking	Other
Domestic work, including collecting fuel <sup>a</sup>	
Fetching water <sup>a</sup>	
Caring for children and adults, including sick or elderly	

<sup>a</sup> Fetching water was included in the domestic work category in surveys implemented prior to 2022. Fetching water is defined as a separate work activity category as of 2022, and the domestic work category was updated accordingly.

The individual is defined as adequate on workload if the number of hours she worked the day preceding the survey was less than the time poverty line of 10.5 hours in 24 hours.

Respondents are also asked:

“In the past 24 hours, did you work, either at home or outside the home, more than usual, about the same amount as usual, or less than usual?” (question V6602)

This question is not used in the A-WEAI indicator calculation, but it can be used to compare 5DE results among all women, those that had a typical amount of work the day preceding the survey, and those that had an atypical amount of work the day preceding the survey.

### 13.2.2 Part 2: Computing the 5DE

The 5DE is computed using the following steps:

1. Calculate individual adequacy scores ( $\alpha$ ) for women who have complete 5DE data.
2. Calculate the percentage of women who are empowered ( $H_e$ ) and the percentage of women who are disempowered ( $H_d$ ).
3. Calculate the average adequacy score among the disempowered ( $A_d$ ).
4. Compute the 5DE ( $EA$ ).

**Step 1.** Calculate weighted adequacy scores for women who have complete 5DE data.

Having complete 5DE data means that the woman has a score (0 [inadequate] or 1 [adequate]) for each of the six A-WEAI indicators; she is not missing a score for any of the A-WEAI indicators.<sup>88</sup> A weighted

<sup>88</sup> Women who are not active in at least one agricultural activity (i.e., food crop farming, cash crop farming, raising livestock, or fishing or fishpond culture) are always excluded because they will always be missing a value for input into productive decision-making.

adequacy score ( $c_a$ ) is calculated for each respondent with complete 5DE data by summing the weighted scores for the six indicators of adequacy calculated in Part I, as shown in the following calculation:

$$c_{ai} = w_1I_1 + w_2I_2 + w_3I_3 + w_4I_4 + w_5I_5 + w_6I_6$$

Where:

$c_{ai}$  = weighted adequacy score of respondent  $i$

$w_i$  = weight assigned to indicator  $i$  so that the sum of the weights across all indicators equals 1

The weighted adequacy score can have a value between 0 and 1 and reflects the percentage of weighted indicators in which respondents have adequate achievement. A score of 0 indicates that the respondent is adequate in none of the indicators. A score of 1 indicates that the respondent is adequate in all indicators. A respondent with a score of 0.4 has adequate achievement in 40 percent of the weighted indicators, which is equal to two of five domains.

**Step 2.** Calculate the number of women who are empowered and the number of women who are disempowered.

**Step 2.1.** Apply an empowerment cutoff—greater than or equal to 0.8—to identify the empowered. The cutoff is set to be able to see improvement over time. A cutoff that is too low could overstate the prevalence of empowered respondents and result in an indicator lacking sensitivity. A cutoff that is too high could result in too few respondents being categorized as empowered and leave little room for improvement. Based on different sensitivity analyses, the empowerment cutoff is set to greater than or equal to 0.8, so that respondents with a weighted adequacy score greater than or equal to 0.8—those who are adequate in 80 percent or more of the weighted indicators—are considered to be empowered. Individuals with a weighted adequacy score of less than 0.8—those who are adequate in less than 80 percent of the weighted indicators—are considered to be disempowered.<sup>89</sup>

**Step 3.** Calculate the average weighted adequacy score among the disempowered ( $A_d$ ) by obtaining the average of the weighted adequacy scores among those who are disempowered—that is, among those who have a weighted adequacy score of less than 0.8.

**Step 4.** Compute the 5DE ( $EA$ ), as shown in the following calculation:

$$EA = H_e + H_d(A_d)$$

Where:

$H_e$  = Percent of women who are empowered (respondents whose average adequacy score is greater than or equal to 0.8)

<sup>89</sup> The sample for the calculation of 5DE excludes cases with missing values for any of the indicators of adequacy in the 5DE.

$H_d$  = Percent of women who are disempowered (respondents whose average adequacy score is less than 0.8)

$A_d$  = Average weighted adequacy score among disempowered women

### 13.2.3 Part 3: Calculating the 5DE standard error (jackknife approach) and performing a test of difference

Because the 5DE is calculated by summing two sample-weighted estimates, there is no variation across respondents. Without variation, standard errors and confidence intervals (CIs) cannot be calculated using the same approach that is used for other indicators. Due to the complicated nature of the non-linear 5DE, non-standard methods must be used to compute the CIs and design effect (DEFF). This section provides guidelines for calculating the 95 percent CI and DEFF for the 5DE.

#### Confidence interval estimation

To compute the CI for the 5DE, standard error must be estimated using a replication method such as the jackknife repeated replication (JRR) method. In the JRR method, the standard error is the square root of the replicate variance among the indicator estimates calculated based on several sample replicates of the full sample (Heeringa, et al 2010; Wolter, 2007). In each replicate, all but one sample cluster are considered in estimation of the indicator. Where  $C$  is the total number of sample clusters,  $C$  sample replicates should be created, and the variance of the estimated indicator  $\hat{\theta}$  (A-WEAI or A-WEAI context) can be calculated as follows:

$$var(\hat{\theta}) = \frac{1}{C(C-1)} \sum_{c=1}^C (\hat{\theta}_c - \hat{\theta})^2$$

in which

$$\hat{\theta}_c = C\hat{\theta} - (C-1)\hat{\theta}_{(c)}$$

where  $\hat{\theta}$  is the estimate computed from the full sample of  $C$  clusters and  $\hat{\theta}_{(c)}$  is the estimate computed from the reduced sample of  $(C-1)$  clusters ( $c^{\text{th}}$  cluster excluded)

To estimate the CI for the 5DE, the algorithm for the JRR method follows a sequence of six steps:

1. The 5DE is estimated based on the full sample as  $\hat{\theta}$ .
2. A total of  $C$  sample replicates are defined, where replicate  $c$  is a reduced sample of  $(C-1)$  clusters where  $c^{\text{th}}$  cluster is excluded.
3. For each sample replicate  $c$ , a replicate weight is calculated by multiplying the full sample survey weight by a strata-level adjustment factor  $\frac{W_h}{W_{h(c)}}$ , where  $W_h$  is the summation of the survey weight for all units within stratum  $h$  and  $W_{h(c)}$  is the same summation but after cluster  $c$  is excluded.

4. From each sample replicate  $c$ , the 5DE is calculated as  $\hat{\theta}_{(c)}$  using the replicate weights.
5. The JRR replicated variance estimation formula is used to compute the standard error based on the estimates from the full sample and  $C$  replicates, computed in Steps 1 and 4.
6. 95 percent CIs are estimated based on the estimated indicators and standard error as  $CI(\hat{\theta}) = \hat{\theta} \pm 1.96 \sqrt{var(\hat{\theta})}$ .

### Design effect estimation

Due to the reasons mentioned previously, there is not a constructed method to calculate the DEFF for the 5DE. Because sets of linear variables are used in the computation of both indicators, the average of DEFFs for these variables can be used as approximates for DEFFs as the average of DEFFs for the variables that are involved in the calculation of the main components: the sample-weighted percent of women who are empowered (and disempowered) and the sample-weighted average adequacy score.

### Differences of indicators and test of significance

The 5DE can be compared between two points in time—for instance, between baseline and midline—and hypothesis testing can be used to determine whether the difference between the estimated values is statistically significant or not.

Assuming that the 5DE follows the standard normal distribution, let  $\hat{\theta}_1$  denote the estimate computed for the Baseline Survey, and  $\hat{\theta}_2$  denote the estimate computed for the Midline Survey. To test the null hypothesis,  $H_0: \hat{\theta}_1 = \hat{\theta}_2$ , versus the alternative hypothesis,  $H_a: \hat{\theta}_1 \neq \hat{\theta}_2$ , the test statistic (z-score) can be calculated as follows:

$$Z = \frac{|\hat{\theta}_1 - \hat{\theta}_2|}{\sqrt{SE_1^2 + SE_2^2}}$$

where  $SE_1$  is the standard error for  $\hat{\theta}_1$

$SE_2$  is the standard error for  $\hat{\theta}_2$

Using the standard normal tables, the associated two-tailed p-value of the Z-score can be found as  $2 * P(Z > Z_\alpha)$  where  $Z_\alpha$  is the critical Z value for the chosen  $\alpha$  level significance (e.g.,  $Z_\alpha = 1.96$  for the 95 percent confidence level and  $Z_\alpha = 1.645$  for the 90 percent confidence level). The p-value is then assessed in comparison with the significance level (0.05 for 95 percent confidence level, or 0.10 for 90 percent confidence level), where the null hypothesis is accepted if the p-value is larger than the significance level, and rejected otherwise. If the null hypothesis is accepted, any observed difference is due to sampling or experimental error, but if the null hypothesis is rejected, it is likely that the estimated indicator for the two groups being compared are truly different.

### 13.2.4 Part 4: Calculating adequate achievement in each A-WEAI indicator

In addition to calculating the 5DE, the percentage of women who are empowered, and the average weighted adequacy score, calculate the percentage of women with adequate achievement in each A-WEAI indicator for three populations:

- Women with data for all six A-WEAI indicators
- Disempowered women
- All women

These results provide insight into which A-WEAI indicators women are most able and least able to achieve adequacy in—perhaps highlighting areas that programs can target.

## 13.3 Step-by-step instructions for computing the 5DE

The step-by-step procedures to calculate the 5DE follows the Stata syntax in the following do files: *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_1 prep.do* and *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_2 5DEcalc.do*. Variables from Module 6, *Women's empowerment in agriculture*, are used for the calculations.

### Definitions

Numerator	The sample-weighted 5DE among primary adult female decision-makers
Denominator	N/A
Unit of measure	Scale from 0 to 1 (higher value indicates more empowerment)
Level of data	Individual-level
Sampling weight	Primary adult female decision-maker
Disaggregation levels	Age category (18-29 years, 30+ years)*
Treatment of missing data	Primary adult female decision-makers who do not have complete data for all six A-WEAI indicators are excluded from the 5DE calculations.
Survey variables used	All variables that begin with 'v6', <i>hhea</i> , <i>hhnum</i> , <i>wgt_fpd</i>
Analytic variables used	<i>age</i> , <i>hhmem_dj</i>
Analytic variables created	<b>Part I:</b> <i>v6201_1x-v6201_8x, v6203_1x-v6203_8x, v6204_1x-v6204_8x, v6205_1x-v6205_6x, partact_1-partact_8, partact, partactagr, inputdec_1-inputdec_8, feelmakedec_1-feelmakedec_8, feelinputdecagr_sum, feelinputdecagr, v6301_01x-v6301_15x, own_01-own_15, own_sum, ownagr_sum, selfjointown_01-selfjointown_15, selfjointown_sum, jown_count, creditaccess_1-creditaccess_6, creditaccess, creditselfjointborrow_1-creditselfjointborrow_6, creditselfjointuse_1-creditselfjointuse_6, creditselfjointanydec_1-creditselfjointanydec_6, credjanydec_any, incomedec_1-incomedec_6, incomedec_sum, incdec_count, groupmember_01-groupmember_11, groupmember_any, num_time_incr, work, num_avail, work_hours, z105, npoor_z105, v6601p_15_01x-v6601p_15_24x, v6601p_30_01x-v6601p_30_24x, v6601p_45_01x-</i>

v6601p\_45\_24x,v6601p\_60\_01x-v6601p\_60\_24x

**Part 2:** w\_feelininputdecagr, w\_jown\_count, w\_credjanydec\_any, w\_incdec\_count, w\_groupmember\_any, w\_npoor\_z105, wg0\_feelininputdecagr, wg0\_jown\_count, wg0\_credjanydec\_any, wg0\_incdec\_count, wg0\_groupmember\_any, wg0\_npoor\_z105, feelininputdecagr\_miss, jown\_count\_miss, credjanydec\_any\_miss, incdec\_count\_miss, groupmember\_any\_miss, npoor\_z105\_miss, ca, n\_missing, miss\_any, emp\_80p, dis\_80p, ad\_80p, swp\_emp\_80p, swp\_dis\_80p, swm\_ad\_80p, EA\_80p

**Part 3:** fdm\_agecat, id, ID, alloc, alloc\_AGE, mean\_0\_emp\_80p, mean\_1\_emp\_80p, mean\_2\_emp\_80p, deff\_0\_emp\_80p, deff\_1\_emp\_80p, deff\_2\_emp\_80p, mean\_0\_dis\_80p, mean\_1\_dis\_80p, mean\_2\_dis\_80p, deff\_0\_dis\_80p, deff\_1\_dis\_80p, deff\_2\_dis\_80p, mean\_0\_ad\_80p, mean\_1\_ad\_80p, mean\_2\_ad\_80p, deff\_0\_ad\_80p, deff\_1\_ad\_80p, deff\_2\_ad\_80p, EA\_80p0, EA\_80p1, EA\_80p2, deff0, deff1, deff2, replicate, PS2, allocj, allocj\_AGE, wtadj, wtadj\_AGE, i, A\_EA20p0, A\_EA20p1, A\_EA20p2, B\_EA20p0, B\_EA20p1, B\_EA20p2, JKSE\_EA\_80p0, JKSE\_EA\_80p1, JKSE\_EA\_80p2, SE\_EA20p0, SE\_EA\_80p1, SE\_EA\_80p2, SE0, SE1, SE2, LCI, UCI, \_j, Z, P, age\_cat

**Part I Step 7 (additional variables created for report but not needed for 5DE calculations):** partact\_any, partact\_agr\_any, inputdec\_any, inputdec\_agr\_any, selfjointown\_01x-selfjointown\_15x, hh\_cash\_1-hh\_cash\_6, hh\_inkind\_1-hh\_inkind\_6, hh\_cash\_any, hh\_inkind\_any, hh\_loan\_any, creditdecborrow\_any, creditdecuse\_any, creditdec\_any, incomedec\_any, incomedec\_agr\_any, groupmember\_01x-groupmember\_11x, work\_any, partook\_A-partook\_X, partook\_OP, partook\_GHIJ, partook\_QR, time\_A-time\_X, time\_OP, time\_GHIJ, time\_QR, hours\_A-hours\_X, hours\_OP, hours\_GHIJ, hours\_QR

\*Standard Feed the Future disaggregate

## Calculations

### **Part I: Preparing the data and calculating the A-WEAI indicators**

The following steps in Part I describe how to create the six A-WEAI indicators indicating adequacy across the five domains and follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_1 prep.do*.

First, load the individual-level data file with the Module 6, *Women's empowerment in agriculture*, data in it and keep only records for women who are de jure household members and completed Module 6.



```

Load          "FTF ZOI Survey [COUNTRY] [YEAR] persons
              analytic"
Keep record   if v600r=1 and hhmem_dj=1

```

## Step 1. A-WEAI indicator 1: Decisionmaking over production

Note that only activities 1, 2, 3, and 6 in the core ZOI Survey questionnaire are used to calculate the indicator for this domain; however, all activities will be checked, and variables will be created for them because they are used in the A-WEAI indicator 4 calculation.

### Step 1.1. Review and prepare the data.

**Step 1.1.A.** Review frequency distributions and the number of missing cases for variables **v6201**, **v6202**, **v6203**, and **v6204** for activities 1 to 6 and also variables **v6202**, **v6203**, and **v6204** for minor and major household expenditures, activities 7 and 8. (Variable **v6201** does not exist for activity 7 or 8.)

**Step 1.1.B.** Ensure that skip patterns are correct; if the respondent did not participate in the activity (**v6201**), she should not have a response for **v6202**, **v6202aa**, **v6203**, and **v6204**. Also check who usually makes the decisions for each activity (**v6202aa**); if the respondent is usually the sole decision-maker for an activity, she should not have responses for **v6203** and **v6204**.

**Step 1.1.C.** Create new variables for **v6201** (activities 1-6), and **v6203** and **v6204** (activities 1-8) that set any variables coded as missing during data processing (9 or 99) and "no decision made" responses (93) to missing (**v6201\_1x-v6201\_6x**, **v6203\_1x-v6203\_8x**, and **v6204\_1x-v6204\_8x**).

```

For each      activity (x) of 1-6:
              Set   v6201_`x'x=v6201_`x'
              Replace v6201_`x'x=missing if v6201_`x'=9

For each      activity (x) of 1-8:
  Set         v6203_`x'x=v6203_`x'
  Replace     v6203_`x'x=missing if v6203_`x'=93 or
              v6203_`x'=99

Set          v6204_`x'x=v6204_`x'
Replace     v6204_`x'x=missing if v6204_`x'=9

```

**Step 1.1.D.** Create analytic variables **v6201\_7x** and **v6201\_8x** (activities 7 and 8). Set the variables equal to 1 if the respondent has data for the corresponding **v6202** variable so that the new variables can be used in a later step.

```

For each      economic activity (x) of 7 and 8:
  Set         v6201_`x'x=missing
  Replace     v6201_`x'x=1 if v6202_`x'≠missing
  Label values 0 "No"
              1 "Yes"
  Label variable "[`x' activity]: Participated, past 12 months"

```

**Step 1.2.** Create binary variables indicating participation in each activity. Use the variables **v6201\_1x-v6201\_8x** created in the previous two steps to generate eight binary participation variables (**partact\_1-partact\_8**). If the respondent participated in the activity, the variable is coded as 1. If the respondent did not participate in the activity, the variable is coded as 0. If information is missing, the variable is coded as missing.

```

For each      economic activity (x) of 1-8:
  Set          partact_`x`=1 if v6201_`x`x=1
  Replace      partact_`x`=0 if v6201_`x`x=2
  Replace      partact_`x`=missing if v6201_`x`=missing
  Label values 0 "No"
               1 "Yes"
  Label variable "[`x' activity]: Participated, past 12 months"

```

**Step 1.3.** Create a variable that sums the number of productive activities (**partact\_1-partact\_6**) in which the respondent participated (**partact**). The variable is coded as missing if all six variables being summed are coded as missing. Ensure that **partact** includes records that are missing values for only a subset of the variables being summed.

```

Set          partact=partact_1+partact_2+partact_3+partact_4+
              partact_5+partact_6
Replace      partact=missing if partact_1=missing and
              partact_2=missing and partact_3=missing and
              partact_4=missing and partact_5=missing and
              partact_6=missing
Label variable "Number of productive activities participated in
              (0-6)"

```

**Step 1.4.** Create a variable that counts the number of agriculture-related productive activities (**partact\_1-partact\_3** and **partact\_6**) in which the respondent participated (**partact\_agr**). The variable is coded as missing if all four variables being summed are coded as missing. Ensure that **partact\_agr** includes records that are missing values for only a subset of the variables being summed.

```

Set          partactagr=partact_1+partact_2+partact_3+
              partact_6
Replace      partactagr=missing if partact_1=missing and
              partact_2=missing and partact_3=missing and
              partact_6=missing
Label variable "Number of productive ag activities participated
              in (0-4)"

```

**Step 1.5.** Create a binary variable for each activity to indicate whether the respondent has adequate participation in decision-making. Use **partact\_1-partact\_8** created in Step 1.2, **v6203\_1x-v6203\_8x** created in Step 1.1.C, and the **v6202aa\_1-v6202aa\_8** survey variables to generate eight binary variables on input into decision-making (**inputdec\_1-inputdec\_8**). Respondents who report that they alone are usually responsible for making decisions about an activity in **v6202\_1-v6202\_8** (i.e., **v6202=A [SELF]** only) have a value of 1 for the corresponding decision-making filter variable

(v6202aa\_1-v6202aa\_8), indicating that they must have input into all decisions about the activity. Respondents who report that they make all of the decisions by themselves (v6202aa=1) or that they have input in some, most, or all decisions made with others (v6203=2 or 3) are coded as 1, and respondents who report that they have no input or input in a few decisions are coded as 0. Respondents who did not participate in the activity, respondents who reported “no decisions made,” and respondents who are missing input into decision-making information for v6202aa and v6203 are coded as missing.

```

For each      economic activity (x) of 1-8:
  Set         inputdec_`x`=1 if (v6202aa_`x`=1 or v6203_`x`x=2
              or v6203_`x`x=3) and partact_`x`=1
  Replace     inputdec_`x`=0 if (v6202aa_`x`=2 and v6203_`x`x≠2
              and v6203_`x`x≠3) and partact_`x`=1
  Replace     inputdec_`x`=missing if (v6202aa_`x`=missing or
              (v6202aa_`x`=2 and v6203_`x`x=missing)) and
              partact_`x`=1
  Replace     inputdec_`x`=missing if partact_`x`≠1
  Label values 0 "No"
              1 "Yes"
  Label variable "[`x' activity]: Has input into at least some
              decisions"

```

**Step 1.6.** Create binary variables for each activity to indicate whether the respondent feels she has autonomy in decision-making. Use partact\_1-partact\_8 created in Step 1.2, v6204\_1x-v6204\_8x created in Step 1.1.C, and the v6202aa\_1-v6202aa\_8 survey variables to generate eight binary variables on autonomy in decision-making (feelmakedec\_1-feelmakedec\_8). Respondents who report that they alone are usually responsible for making decisions about an activity in v6202\_1-v6202\_8 (i.e., v6202=A [SELF] only) have a value of 1 for the decision-making filter survey variable (v6202aa\_1-v6202aa\_8), indicating that they must have autonomy in decision-making about the activity. Respondents who report that they make all of the decisions by themselves (v6202aa=1) or that they feel they can make decisions to a “medium” or “high extent” if they want to are coded as 1. Responses of “not at all” or “small extent” are coded as 0. Respondents who did not participate in the activity and respondents who are missing input into decision-making information for v6202aa and v6204 are coded as missing.

```

For each      activity (x) of 1-8:
  Set         feelmakedec_`x`=1 if (v6202aa_`x`=1 or
              v6204_`x`x=3 or v6204_`x`x=4) and partact_`x`=1
  Replace     feelmakedec_`x`=0 if (v6202aa_`x`=2 and
              v6204_`x`x≠3 and v6204_`x`x≠4) and partact_`x`=1
  Replace     feelmakedec_`x`=missing if (v6202aa_`x`=missing
              or (v6202aa_`x`=2 and v6204_`x`x=missing)) and
              partact_`x`=1
  Replace     feelmakedec_`x`=missing if partact_`x`≠1
  Label values 0 "No"
              1 "Yes"
  Label variable "[`x' activity]: Feels can make decisions to at
              least medium extent"

```

**Step 1.7.** Create a variable that counts the number of agricultural economic activities for which the respondent has input into at least some decisions, plus the number of agriculture economic activities for which the respondent feels that she can make decisions to a medium or high extent if she wanted to. This variable (**feelinputdecagr\_sum**) is a count of the variables related to agricultural economic activities created in Steps 1.5 and 1.6 that have a value of 1 (meaning 'YES'). The variable can have a value between 0 and 8. The variable is coded as missing if all eight variables being summed are coded as missing. Ensure that **feelinoutdecagr\_sum** includes records that are missing values for only a subset of the variables being summed. Variables created in Steps 1.5 and 1.6 related to the following economic activities are not included in the count because they are not related to agriculture: non-farm activity, wage and salary employment, and major and minor household expenditures.

Set	<b>feelinputdecagr_sum=inputdec_1+inputdec_2+inputdec_3+inputdec_6+feelmakedec_1+feelmakedec_2+feelmakedec_3+feelmakedec_6</b>
Replace	<b>feelinputdecagr_sum=missing if inputdec_1=missing and inputdec_2=missing and inputdec_3=missing and inputdec_6=missing and feelmakedec_1=missing and feelmakedec_2=missing and feelmakedec_3=missing and feelmakedec_6=missing</b>
Label variable	<b>"Number of areas makes decisions or feels can, ag activities (0-8)"</b>

**Step 1.8.** Create the indicator of adequate achievement in decision-making about agricultural production, A-WEAI indicator 1 (**feelinputdecagr**). The indicator is coded as 1 if the variable created in Step 1.7, **feelinputdecagr\_sum**, has a value of 1 or more and is coded as 0 if **feelinputdecagr\_sum** has a value of 0. The indicator is coded as missing if **feelinputdecagr\_sum** has a value of missing. Note that if the respondent did not engage in any agricultural activities, this indicator will be coded as missing.

Set	<b>feelinputdecagr=1 if feelinputdecagr_sum≥1</b>
Replace	<b>feeliinputdecagr=0 if feelinputdecagr_sum=0</b>
Replace	<b>feelinputdecagr=missing if feelinputdecagr_sum=missing</b>
Label values	<b>0 "No" 1 "Yes"</b>
Label variable	<b>"Adequate in decision-making, women active in ag (A-WEAI Indicator 1)"</b>

## **Step 2. A-WEAI indicator 2: Ownership of assets**

**Step 2.1.** Review and prepare the data.

**Step 2.1A.** Review frequency distributions and number of missing cases for assets 1-15 for variables **v6301** and **v6303**.

**Step 2.1B.** Ensure that skip patterns are correct; if a respondent's household did not own an asset (**v6301**), the respondent should not have a response for **v6303**.

**Step 2.1C.** Create new variables for **v6301** that have ‘no’ responses (2) recoded as 0 and any variables coded as missing during data processing (9) recoded as missing (**v6301\_01x-v6301\_15x**).

```
For each      asset (x) of 1-15:
  Set         v6301_`x'x=v6301_`x'
  Replace     v6301_`x'x=0 if v6301_`x'=2
  Replace     v6301_`x'x=missing if v6301_`x'=9
```

**Step 2.2.** Create binary variables indicating whether someone in the respondent’s household owns each asset. The variables created in the previous step are used to create 15 binary variables on household asset ownership (**own\_01-own\_15**), which are coded as 1 to indicate that someone in the household owns the asset and as 0 to indicate that no one in the household owns the asset. The variable is coded as missing if the corresponding variable created for the asset in Step 2.1C is coded as missing.

```
For each      asset of (x) of 01-15:
  Set         own_`x'=1 if v6301_`x'x=1
  Replace     own_`x'=0 if v6301_`x'x=0
  Replace     own_`x'=missing if v6301_`x'x=missing
  Label values 0 "No"
               1 "Yes"
  Label variable "Household owns [`x' asset]"
```

**Step 2.3.** Create variables that count the number of total assets the respondent’s household owns (**own\_sum**) and the number of agricultural assets the respondent’s household owns (**ownagr\_sum**). The variable is coded as missing if all variables being summed are coded as missing. Ensure that **own\_sum** and **ownagr\_sum** include records that are missing values for only a subset of the variables being summed.

```
Set           own_sum=own_01+own_02+own_03+own_04+own_05+
              own_06+own_07+own_08+own_09+own_10+own_11+
              own_12+own_13+own_14+own_15
  Replace     own_sum=missing if own_01 through own_15 are all
              missing values
  Label variable "Number of asset types household owns (0-15)"

Set           ownagr_sum=own_01+own_02+own_03+own_04+own_05+
              own_06+own_07+own_08
  Replace     ownagr_sum=missing if own_01 through own_08 are
              all missing values
  Label variable "Number of agricultural asset types household
              owns (0-8)"
```

**Step 2.4.** Create binary variables for each asset indicating whether the respondent owns the asset alone or jointly. The variables created in Step 2.3 are used to determine whether the respondent owns each asset—either alone or jointly. Fifteen binary self-or-joint-ownership variables (**selfjointown\_01-selfjointown\_15**) are created that are coded as 1 if the respondent alone or jointly owns the asset and as 0 if someone else owns the asset. If the respondent’s household does not own the asset or if the household owns the asset but the woman is missing a value for the corresponding **v6303** variable, the variable is coded as missing.

```

For each      asset (x) of 0-15:
  Set          selfjointown_`x`=1 if v6303_`x` includes "A" and
              own_`x`=1
  Replace      selfjointown_`x`=0 if v6303_`x` does not include
              "A" and own_`x`=1
  Replace      selfjointown_`x`=missing if own_`x`≠1 or
              (own_`x`=1 and v6303_`x`=missing)
  Label values 0 "No"
              1 "Yes"
  Label variable "Owns alone or jointly [`x' asset]"

```

**Step 2.5.** Create a variable that counts the number of assets owned by the respondent. This variable (**selfjointown\_sum**) is a count of the number of assets owned solely or jointly by the respondent. The variable is coded as missing if all 15 variables being summed are coded as missing. Ensure that **selfjointown\_sum** includes records that are missing values for only a subset of the variables being summed.

```

Set          selfjointownsum=selfjointown_01+selfjointown_02+
              selfjointown_03+selfjointown_04+selfjointown_05+
              selfjointown_06+selfjointown_07+selfjointown_08+
              selfjointown_09+selfjointown_10+selfjointown_11+
              selfjointown_12+selfjointown_13+selfjointown_14+
              selfjointown_15
  Replace      selfjointownsum=missing if selfjointownsum_01
              through selfjointownsum_15 are all missing values
  Label variable "No. of assets the respondent owns alone or
              jointly (0-15)"

```

**Step 2.6.** Create the indicator for adequate achievement in asset ownership, A-WEAI indicator 2 (**jown\_count**). This variable is coded as 1 (adequate) if the respondent owns at least one large asset alone or jointly or if the respondent owns at least two small assets alone or jointly. This is done by setting **jown\_count** to be 1 if the respondent owns at least one asset alone or jointly and then recoding **jown\_count** to 0 if the respondent does not own any assets that the household owns, if the respondent owns only one asset alone or jointly that is considered to be a small asset (i.e., chickens [4], hand tools for farming [6], non-mechanized farm equipment [7], or small consumer durables [12]), or if the respondent's household does not own any of the assets. This indicator is coded as missing if **selfjointownsum** and **own\_sum** both have values of missing.

```

Set          jown_count=missing
  Replace      jown_count=1 if selfjointownsum≥1 and
              selfjointownsum≠missing
  Replace      jown_count=0 if jown_count=1 and
              selfjointownsum=1
              and (selfjointown_04=1 or selfjointown_06=1 or
              selfjointown_07=1 or selfjointown_12=1)
  Replace      jown_count=0 if selfjointownsum=0 or own_sum=0
  Label values 0 "No"
              1 "Yes"

```

Label variable "Adequate in asset ownership, all women (AWEAI Indicator 2)"

### Step 3. A-WEAI indicator 3: Access to credit

**Step 3.1.** Review and prepare the data.

**Step 3.1A.** Review frequency distributions and number of missing cases for credit sources 1-6 for variables **v6308**, **v6309**, and **v6310**.

**Step 3.1B.** Ensure that skip patterns are correct; if the respondent's household did not borrow from the credit source (**v6308**), they should not have responses for **v6309** and **v6310**.

**Step 3.2.** Create binary variables indicating whether the respondent's household borrowed from each credit source. The survey collects information on access to credit from six credit sources (**v6308\_1-v6308\_6**). This information is used to generate six binary variables on households' credit access (**creditaccess\_1-creditaccess\_6**), one for each source of credit. If the household borrowed cash or in-kind or a combination of the two (cash and in-kind), the variable is coded as 1. If the household did not borrow from that credit source, the variable is coded as 0. If the respondent does not know, a response is missing for the **v6308** variable, or if the **v6308** variable was set to missing during data processing (i.e., 9), the variable is coded as missing.

```
For each      credit source (x) of 1-6:
  Set         creditaccess_`x`=1 if v6308_`x`<4
  Replace     creditaccess_`x`=0 v6308_`x`=4
  Replace     creditaccess_`x`=missing if v6308_`x`=missing or
              v6308_`x`=8 or v6308_`x`=9
  Label values 0 "No"
              1 "Yes"
  Label variable "Household borrowed from [`x` credit source]"
```

**Step 3.3.** Create a variable that indicates the number of credit sources the respondent's household used (**creditaccess**). The variable is coded as missing if all six variables being summed are coded as missing. Ensure that **creditaccess** includes records that are missing values for only a subset of the variables being summed.

```
Set          creditaccess=creditaccess_1+ creditaccess_2+
              creditaccess_3+ creditaccess_4+ creditaccess_5+
              creditaccess_6
  Replace     creditaccess=missing if creditaccess_1=missing
              and creditaccess_2=missing and
              creditaccess_3=missing and creditaccess_4=missing
              and creditaccess_5=missing and
              creditaccess_6=missing
  Label variable "Number of credit sources household used (0-6)"
```

**Step 3.4.** For each credit source, create variables indicating whether the respondent participated in decisions related to borrowing (Step 3.4.1), how to use the borrowed money or item (Step 3.4.2), or either borrowing or how to use the borrowed money or item (Step 3.4.3).



**Step 3.4A.** For each credit source, create a variable that indicates whether the respondent participated in the decision to borrow from that source (**creditselfjointborrow\_1-creditselfjointborrow\_6**). The variables are coded as 1 if the respondent alone or jointly made the decision to borrow from the credit source and 0 if the respondent was not involved in the decision. The variables are coded as missing if the respondent's household did not borrow from the credit source or if the corresponding **v6309** variable has a missing value or a value of "not applicable" (X).<sup>90</sup>

```

For each      credit source (x) of 1-6:
Set           creditselfjointborrow_`x`=1 if v6309_`x` includes
              "A" and creditaccess_`x`=1
Replace       creditselfjointborrow_`x`=0 if v6309_`x` does not
              include "A" and creditaccess_`x`=1
Replace       creditselfjointborrow_`x`=missing if
              (v6309_`x`=missing or v6309_`x`="X") and
              creditaccess_`x`=1
Label values  0 "No"
              1 "Yes"
Label variable "Participated in decision to borrow from [`x'
              credit source]"

```

**Step 3.4B.** For each credit source, create a variable that indicates whether the respondent participated in the decision on how to use the money or item borrowed from that source (**creditselfjointuse\_1-creditselfjointuse\_6**). The variables are coded as 1 if the respondent alone or jointly made the decision on how to use the money or item borrowed from the credit source and 0 if the respondent was not involved in the decision. The variables are coded as missing if the respondent's household did not borrow from the credit source or if the corresponding **v6310** variable has a missing value or a value of "not applicable" (X).

```

Set           creditselfjointuse_`x`=1 if v6310_`x` includes "A"
              and creditaccess_`x`=1
Replace       creditselfjointuse_`x`=0 if v6310_`x` includes "A"
              and creditaccess_`x`=1
Replace       creditselfjointuse_`x`=missing if
              (v6310_`x`=missing or v6310_`x`="X") and
              creditaccess_`x`=1
Label values  0 "No"
              1 "Yes"
Label variable "Participated in the decision to use credit from
              [`x' credit source]"

```

**Step 3.4C.** For each credit source, create a variable indicating whether the respondent participated in decisions to borrow from that credit source or how to use the borrowed money or item (**creditselfjointanydec\_1-creditselfjointanydec\_6**). The variables are coded as 1 if either the corresponding **creditselfjointborrow** variable or the corresponding **creditselfjointuse** variable has a value of 1. The variables are coded as 0 if the corresponding

<sup>90</sup> "X" has been removed from Module 6 in the core P2-ZOI Midline Survey questionnaire as a response option, but it remains in the template syntax because it was included in some Midline Surveys.



**creditselfjointborrow** variable and the corresponding **creditselfjointuse** variable both have values of 0. The variables are also coded as 0 if either the corresponding **creditselfjointborrow** variable or the corresponding **creditselfjointuse** variable has a value of 0 and the other has a missing value. The variables are coded as missing if both of the corresponding **creditselfjointborrow** and **creditselfjointuse** variables are missing values.

```

Set          creditselfjointanydec_`x`=0 if
              creditselfjointborrow_`x`#missing or
              creditselfjointuse_`x`#missing
Replace      creditselfjointanydec_`x`=1 if
              creditselfjointborrow_`x`=1 or
              creditselfjointuse_`x`=1
Replace      creditselfjointanydec_`x`=missing if
              creditselfjointborrow_`x`=missing and
              creditselfjointuse_`x`=missing
Label values 0 "No"
              1 "Yes"
Label variable "Participated in 1+ credit decision from [`x'
              credit source]"

```

**Step 3.5.** Create the indicator for adequate achievement in access to credit (**credjanydec\_any**). If the respondent made the decision to borrow from at least one source of credit alone or jointly or the decision on how to use the money or item borrowed from at least one source of credit alone or jointly, the indicator is coded as 1. If the respondent did not participate in a decision to borrow from any credit source or a decision on how to use the money or item borrowed from any source of credit, the indicator is coded as 0. If a respondent's household did not borrow from any credit source, the indicator is coded as 0. The indicator is coded as missing if each of the six **creditselfjointanydec** variables has a value of missing.

```

Set          credjanydec_any=0 if creditaccess#missing
Replace      credjanydec_any=1 if creditselfjointanydec_1=1 or
              creditselfjointanydec_2=1 or
              creditselfjointanydec_3=1 or
              creditselfjointanydec_4=1 or
              creditselfjointanydec_5=1 or
              creditselfjointanydec_6=1
Replace      credjanydec_any=missing if
              creditselfjointanydec_1=missing and
              creditselfjointanydec_2=missing and
              creditselfjointanydec_3=missing and
              creditselfjointanydec_4=missing and
              creditselfjointanydec_5=missing and
              creditselfjointanydec_6=missing
Label values 0 "No"
              1 "Yes"
Label variable "Adequate in credit decision-making, all women
              (A-WEAI Indicator 3)"

```

## Step 4. A-WEAI indicator 4: Control over income

**Step 4.1.** Review and prepare the data.

**Step 4.1A.** Review frequency distributions and number of missing cases.

**Step 4.1B.** Ensure that skip patterns are correct; if the respondent did not participate in the activity (v6201), they should not have a response for v6205. Also check who usually makes the decisions for each activity (v6202aa); if the respondent is usually the sole decision-maker for an activity, they should not have a response for v6205.

**Step 4.1C.** Create new variables for v6205 that set any variables recoded to missing during data processing (i.e., 99) and “no decisions made” responses (i.e., 93) to missing (v6205\_1x-v6205\_6x).

```
For each      activity (x) of 1-6:
  Set         v6205_`x'x=v6205_`x'
  Replace     v6205_`x'x=missing if v6205_`x'=99 or
              v6205_`x'=93
```

**Step 4.2.** Create binary variables to indicate whether the respondent had input into decisions made on the use of income from each activity. Use partact\_1-partact\_8 created in Step 1.2, v6205\_1x-v6205\_6x created in the previous step, and the v6202aa\_1-v6202aa\_8 survey variables. The survey collects data on decision-making for the use of income generated from food crop farming, cash crop farming, livestock raising, non-farm activities, wage and salary employment, and fishing (v6205\_1-v6205\_6). This information is used to generate six income decision-making variables (incomedec\_1-incomedec\_6), one for each activity. Respondents who report that they alone are usually responsible for making decisions about an activity in v6202\_1-v6202\_8 (i.e., v6202=A [SELF] only) have a value of 1 for the corresponding decision-making filter variable (v6202aa\_1-v6202aa\_8), indicating that they must have input into all decisions about the activity. Respondents who indicate that they make all of the decisions by themselves (v6202aa=1) or that they have input into some or most or all decisions about income from the activity (v6205=2 or 3) are coded as 1, and respondents who have no input or input into very few decisions are coded as 0. Respondents who do not participate in the activity, respondents who reported “no decisions made,” and respondents who are missing input into decision-making information for v6202aa and v6205 are coded as missing.

```
For each      economic activity (x) of 1-6:
  Set         incomedec_`x'=1 if (v6205_`x'x=2 or v6205_`x'x=3
              or v6202aa_`x'=1) and partact_`x'=1
  Replace     incomedec_`x'=0 if v6205_`x'x≠2 and v6205_`x'x≠3
              and v6202aa_`x'=2 and partact_`x'=1
  Replace     incomedec_`x'=missing if v6202aa_`x'=2 and
              v6205_`x'x=missing and partact_`x'=1
  Replace     incomedec_`x'=missing if partact_`x'≠1
  Label values 0 "No"
              1 "Yes"
  Label variable["`x] activity]: Has input into at least some
              decisions about income"
```

**Step 4.3.** Create a variable that counts the number of activities for which the respondent has input into decisions on the use of income or feels that she could make decisions if she wanted. The variable (**incdec\_sum**) is a sum of the six binary variables created in the previous step and also four of the variables created in Step 1.6 (i.e., those related to non-farm economic activities, wage and salary work, minor household expenditures, and major household expenditures). The variable is coded as missing if all 10 variables being summed are coded as missing. Ensure that **incdec\_sum** includes records that are missing values for only a subset of the variables being summed. The variable can have a value between 0 and 10, depending on the number of variables with a value of 1.

```

Set          incmedec_sum=missing
Replace      incmedec_sum=incmedec_1+incmedec_2+
              incmedec_3+ incmedec_4+incmedec_5+incmedec_6+
              feelmakedec_4+feelmakedec_5+feelmakedec_7+
              feelmakedec_8
Replace      incmedec_sum=missing if incmedec_1=missing and
              incmedec_2=missing and incmedec_3=missing and
              incmedec_4=missing and incmedec_5=missing and
              incmedec_6=missing and feelmakedec_4=missing and
              feelmakedec_5=missing and feelmakedec_7=missing
              and feelmakedec_8=missing and
Label variable "No. of areas in which has income decision-making
              input/power (0-10)"

```

**Step 4.4.** Create the indicator for adequate achievement in control over income (**incdec\_count**). Responses are coded as 1 if the respondent has input into some or most or all decisions on the use of income for at least one activity or feels that she can make decisions to a medium or high extent for at least one activity. If the respondent does not have input into decisions on the use of income or does not feel that she can make decisions to a medium or high extent for at least one activity, the indicator is coded as 0. The indicator is also coded as 0 if the respondent can make decisions only about minor household expenditures. The indicator is coded as missing if the variable **incmedec\_sum** has a value of missing.

```

Set          incdec_count=1 if incmedec_sum>0 and
              incmedec_sum≠missing
Replace      incdec_count=0 if incdec_count=1 and
              incmedec_sum=1 and feelmakedec_8=1
Replace      incdec_count=missing if incmedec_sum=missing
Label values 0 "No"
              1 "yes"
Label variable "Adequate in income decision-making, all women
              (A-WEAI Indicator 4)"

```

## **Step 5. A-WEAI indicator 5: Leadership (group membership)**

**Step 5.1.** Review and prepare the data.

**Step 5.1A.** Review frequency distributions and the number of missing cases for variables **v6404** and **v6405**.

**Step 5.1B.** Ensure that skip patterns are correct; if the respondent reported that a group does not exist in their community (v6404), they should not have a response for v6405.

**Step 5.2.** Create binary variables to indicate whether the respondent is an active member of each type of community group. The survey collects information about the respondent's participation in 11 types of community groups (v6405\_01-v6405\_11). This information is used to generate 11 binary participation variables (groupmember\_01-groupmember\_11), 1 for each activity. If a group exists in the community and the respondent is an active member, the variable is coded as 1. If a group exists in the community and the respondent is not an active member, the variable is coded as 0. If a group does not exist in the community, the respondent does not know whether it exists, or information on whether there is a group in the community is missing, the variable is coded as 0. If a group exists in the community but the respondent refuses to answer the question about membership, or the variable has a value of missing or was coded as missing during data processing, the variable is coded as missing.

```
For each      group (x) of 01-11:
  Set         groupmember_`x`=1 if v6404_`x`=1 and v6405_`x`=1
  Replace     groupmember_`x`=0 if (v6404_`x`=1 and
              v6405_`x`=2) or v6404_`x`=2 or v6404_`x`=8 or
              v6404_`x`=9 or v6404_`x`=missing
  Replace     groupmember_`x`=missing if v6404_`x`=1 and
              (v6405_`x`=missing or v6405_`x`=7 or v6405_`x`=9)
  Label values 0 "No"
               1 "Yes"
  Label variable "[`x' group]: Active member"
```

**5.3.** Create the indicator for adequate achievement in group membership (groupmember\_any), which is coded as 1 if the respondent is an active member of at least one group and coded as 0 if the respondent is not an active member of any group. Ensure that groupmember\_any includes records that are missing values for a subset of the variables being summed. The indicator is coded as missing only if the respondent is missing data for the entire sub-module 6.4B.

```
Set          groupmember_any=0
Replace      groupmember_any=1 if groupmember_01=1 or
              groupmember_02=1 or groupmember_03=1 or
              groupmember_04=1 or groupmember_05=1 or
              groupmember_06=1 or groupmember_07=1 or
              groupmember_08=1 or groupmember_09=1 or
              groupmember_10=1 or groupmember_11=1
Replace      groupmember_any=missing if v6404_01-v6404_11 and
              v6405_01-v6405_11 are all missing
Label values 0 "No"
               1 "Yes"
Label variable "Adequate in group membership, all women (AWEAI
              indicator 5)"
```

## **Step 6. A-WEAI indicator 6: Time allocation (workload)**

**Step 6.1.** Review and prepare the data. Review frequency distributions and the number of missing cases for all 15-minute time increment activity variables, which begin with 'v6601p\_'. Create a variable equal

to the number of these variables that have a value for each respondent. Respondents should have non-missing values for all 96 of these variables (24 hours\*[4\*15minutes]). If any respondents are missing values, investigate their records; they may have to be dropped from the workload indicator calculation (see Step 6.3).

```
Set                num_time_incr=count(all variables that start with
                  'v6601p_')
Label variable    "Number of 15-min time increments with data"
```

**Step 6.2.** Create a variable to count the number of 15-minute increments that respondents spent performing work activities as their primary activities (**work**). The 15-minute time increments (**x**) are coded as 15, 30, 45, and 60. Time allocation is reported for 1 day, or 24 hours—01 to 24 (**i**). Fifteen activities (D through R) (**z**) are considered to be work activities.<sup>91</sup> At the same time, also create **v6601p\_`x'`i'x** variables for which any missing values assigned during data processing (i.e., variables with a value="?" in Stata) are set to missing so that they can be used in Step 6.3.

```
Set                work=0
For each           value (x) of 15 30 45 60:
  For each         value (i) in 01 02 03 04 05 06 07 08 09 10 11 12
                  13 14 15 16 17 18 19 20 21 22 23 24:
    Set v6601p_`x'`i'x=v6601p_`x'`i'
    Replace v6601p_`x'`i'x=missing if v6601p_`x'`i'=="?"
    For each value (z) in D E F G H I J K L M N O P Q R:
      Replace      work=work+1 if v6601p_`x'`i'=="z"

Label variable    "Number of 15-min. increments worked day
                  preceding survey"
```

**Step 6.3.** Create a variable similar to the one created in Step 6.1 to count the number of time increments with data (**num\_avail**)—but this time for the **v6601p\_`x'`i'x** variables created in the previous step. Set **work** to missing for respondents missing all time increment data (i.e., **num\_avail**=0).

**Note:** If there are women who have at least some time increment data but who do not have data for all 96 time increments, determine if **work** also should be set to missing for them.

```
Set                num_avail=count(all variables that start with
                  'v6601p_')
Label variable    "Number of 15-min recoded time increments with
                  data"
Replace          Work=missing if num_avail=0
```

**Step 6.4.** Create a variable that converts the total number of 15-minute increments spent working into hours (**work\_hours**).

```
Set                work_hours=work÷4
```

<sup>91</sup> In surveys conducted prior to 2022, 14 activities (D through Q) were considered to be work activities. In these surveys, "Fetching water" was included in activity Q, "Domestic work." Be sure to account for this if conducting comparative analysis over time.

```
Label variable    "Number of hours worked day prior to survey"
```

**Step 6.5.** Create a variable that defines the time poverty line at 10.5 hours per day (**z105**), which implies that a workload in excess of 10.5 hours in a 24-hour period is considered excessive.

```
Set              z105=10.5
Label variable    "Time poverty line (10.5 hours per day)"
```

**Step 6.6.** Create the indicator of adequate achievement in productive and domestic workload (**npoor\_z105**). If the respondents' time in work and domestic activities (**work\_hours**) exceeds 10.5 hours, the respondents are time poor; set the indicator to 0. If the productive and domestic workload is 10.5 hours or less (630 minutes), the respondents are not time poor; set the indicator to 1. The indicator is coded as missing if the variable **work\_hours** has a value of missing.

```
Set              npoor_z105=0 if work_hours>z105 &
                  work_hours#missing
Replace          npoor_z105=1 if work_hours≤z105
Label values     0 "No"
                  1 "Yes"
Label variable    "Adequate in work hours, all women (A-WEAI
                  Indicator 6)"
```

**Step 6.7.** Save the data file.

```
Save              "FTF ZOI Survey [COUNTRY] [YEAR] aweai prep"
```

## Step 7. Additional variables needed for midline indicator assessment tables

The variables created in this step are not used to calculate the 5DE, but they are tabulated in results tables in Section 6.4 of the midline indicator report template and thus used in the 6.4 Tables section of the "FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_3 BL\_ML\_compare" Stata do file.

**Step 7.1.** Create a variable that indicates if the respondent participated in at least one of the six economic activities: food crop farming, cash crop farming, livestock raising, fishing or fishpond culture, non-farm work, or wage/salaried employment (**partact\_any**). The variable is coded as missing if **partact** has a value of missing.

```
Set              partact_any=0 if partact#missing
Replace          partact_any=1 if partact_1=1 or partact_2=1 or
                  partact_3=1 or partact_4=1 or partact_5=1 or
                  partact_6=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Participated in 1+ econ activity"
```

**Step 7.2.** Create a variable that indicates if the respondent participated in at least one of the four agriculture economic activities: food crop farming, cash crop farming, livestock raising, or fishing or fishpond culture (**partact\_agr\_any**). The variable is coded as missing if **partact\_agr** has a value of missing.

```

Set          partact_agr_any=0 if partact_agr#missing
Replace      partact_agr_any=1 if partact_1=1 or partact_2=1
              or partact_3=1 or partact_6=1
Label values 0 "No"
              1 "Yes"
Label variable "Participated in 1+ ag econ activity"

```

**Step 7.3.** Create a variable that indicates if the respondent participated in at least some decisions for at least one of six economic activities: food crop farming, cash crop farming, livestock raising, fishing or fishpond culture, non-farm work, or wage/salaried employment (`inputdec_any`). The variable is coded as missing if all six variables being summed are coded as missing.

```

Set          inputdec_any=0
Replace      inputdec_any=1 if inputdec_1=1 or inputdec_2=1 or
              inputdec_3=1 or inputdec_4=1 or inputdec_5=1 or
              inputdec_6=1
Replace      inputdec_any=missing if inputdec_1=missing and
              inputdec_2=missing and inputdec_3=missing and
              inputdec_4=missing and inputdec_5=missing and
              inputdec_6=missing
Label values 0 "No"
              1 "Yes"
Label variable "Had input into 1+ econ activity"

```

**Step 7.4.** Create a variable that indicates if the respondent participated in at least some decisions for at least one of four agriculture economic activities: food crop farming, cash crop farming, livestock raising, or fishing or fishpond culture (`inputdec_agr_any`). The variable is coded as missing if all four variables being summed are coded as missing.

```

Set          inputdec_agr_any=0
Replace      inputdec_agr_any=1 if inputdec_1=1 or
              inputdec_2=1 or inputdec_3=1 or inputdec_6=1
Replace      inputdec_agr_any=missing if inputdec_1=missing
              and inputdec_2=missing and inputdec_3=missing and
              inputdec_6=missing
Label values 0 "No"
              1 "Yes"
Label variable "Had input into 1+ ag econ activity"

```

**Step 7.5.** Create a variable for each asset that replaces missing `selfjointown` values for respondents whose household does not own the asset with 0s so that all tabulated estimates have the same denominator (i.e., all respondents) (`selfjointown_01x-selfjointown_11x`). These variables should not be coded as missing unless data are missing for the entire Sub-module 6.3A.

```

For each      asset (x) of 01 02 03 04 05 06 07 08 09 10 11 12
              13 14 15:
Set          selfjointown_`x`x=selfjointown_`x`
Replace      selfjointown_`x`x=0 if 'own_`x`'#missing and
              selfjointown_`x`=missing

```

```

replace selfj'intown_`x'x=missing if v6301_01-v6301_15
and v6303_01-v6303_15 all=missing

```

**Step 7.6.** Create variables to capture whether the respondent's household borrowed cash, in-kind, or either cash or in-kind from any credit source.

**Step 7.6.1.** Create two variables for each credit source—one indicating whether the household took a cash loan from that source and the other indicator whether the household took an in-kind loan from that source. Set the cash loan variables to 1 if the respondent's household borrowed only cash or both cash and in-kind (**hh\_cash\_1-hh\_cash\_6**), and set the in-kind variables to 1 if the respondent's household borrowed only in-kind or both cash and in-kind (**hh\_in\_kind\_1- hh\_inkind\_6**). The variables are coded as missing if the value of the corresponding **v6308** variable is don't know (8) or missing (9 or missing).

```

For each credit source (x) of 1 2 3 4 5 6:
  Set hh_cash_`x'=0
  Replace hh_cash_`x'=1 if v6308_`x'=1 or v6308_`x'=3
  Replace hh_cash_`x'=missing if v6308_`x'>4
  Label values 0 "No"
               1 "Yes"
  Label variable "HH borrowed cash from [`x' credit source]"

  Set hh_inkind_`x'=0
  Replace hh_inkind_`x'=1 if v6308_`x'=2 or v6308_`x'=3
  Replace hh_inkind_`x'=missing if v6308_`x'>4
  Label values 0 "No"
               1 "Yes"
  Label variable "HH borrowed in-kind from [`x' credit source]"

```

**Step 7.6.2.** Create variables indicating if the respondent's household borrowed cash (**hh\_cash\_any**), in-kind (**hh\_inkind\_any**), or either cash or in-kind (**hh\_loan\_any**) from any credit source. The variables are coded as missing if all six variables being summed have a value of missing.

```

Set hh_cash_any=0
Replace hh_cash_any=1 if hh_cash_1=1 or hh_cash_2=1 or
hh_cash_3=1 or hh_cash_4=1 or hh_cash_5=1 or
hh_cash_6=1

Replace hh_cash_any=missing if hh_cash_1=missing and
hh_cash_2=missing and hh_cash_3=missing and
hh_cash_4= missing and hh_cash_5= missing and
hh_cash_6=missing

Label values 0 "No"
             1 "Yes"
Label variable "HH took a cash loan, any source"

Set hh_inkind_any=0
Replace hh_inkind_any=1 if hh_inkind_1=1 or hh_inkind_2=1
or hh_inkind_3=1 or hh_inkind_4=1 or
hh_inkind_5=1 or hh_inkind_6=1

```



```

Replace          hh_inkind_any=missing if hh_inkind_1=missing and
                  hh_inkind_2=missing and hh_inkind_3=missing and
                  hh_inkind_4=missing and hh_inkind_5=missing and
                  hh_inkind_6=missing
Label values      0 "No"
                  1 "Yes"
Label variable    "HH took an in-kind loan, any source"

Set              hh_loan_any=0
Replace          hh_loan_any=1 if creditaccess_1=1 or
                  creditaccess_2=1 or creditaccess_3=1 or
                  creditaccess_4=1 or creditaccess_5=1 or
                  creditaccess_6=1
Replace          hh_loan_any=missing if creditaccess_1=missing and
                  creditaccess_2=missing and creditaccess_3=missing
                  and creditaccess_4=missing and
                  creditaccess_5=missing and creditaccess_6=missing
Label values      0 "No"
                  1 "Yes"
Label variable    "HH took a cash or in-kind loan, any source"

```

**Step 7.6.3.** Create variables indicating if the respondent participated in decisions related to borrowing from any credit source (`creditdecborrow_any`), using the credit from any credit source (`creditdecuse_any`), or either borrowing from or using the credit from any credit source (`creditdec_any`). The variables are coded as missing if `hh_loan_any` has a value of missing.

```

Set              creditdecborrow_any=missing
Replace          creditdecborrow_any=0 if hh_loan_any≠0
Replace          creditdecborrow_any=1 if creditdecborrow_1=1 or
                  creditdecborrow_2=1 or creditdecborrow_3=1 or
                  creditdecborrow_3=1 or creditdecborrow_4=1 or
                  creditdecborrow_5=1 or creditdecborrow_6=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Participated in decision to borrow, any source"

Set              creditdecuse_any=missing
Replace          creditdecuse_any=0 if hh_loan_any≠0
Replace          creditdecuse_any=1 if creditdecuse_1=1 or
                  creditdecuse_2=1 or creditdecuse_3=1 or
                  creditdecuse_3=1 or creditdecuse_4=1 or
                  creditdecuse_5=1 or creditdecuse_6=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Participated in decision to use credit, any
                  source"

Set              creditdec_any=0 if hh_loan_any=1
Replace          creditdec_any=1 if creditdecuse_any=1 or
                  creditdecborrow_any=1

```

Label values	0 "No"
	1 "Yes"
Label variable	"Participated in decision to borrow or use credit, any source"

**Step 7.7.** Create a variable that indicates if the respondent participated in at least some decisions related to using income or household expenditures: food crop farming, cash crop farming, livestock raising, fishing or fishpond culture, non-farm work, work or salaried employment, minor household expenditures, and major household expenditures (`incomedec_any`). The variable is coded as missing if all eight variables being summed are coded as missing. Note that the variables for making decisions about household expenditures are `inputdec_`x`` rather than `incomedec_`x``.

Set	<code>incomedec_any=0</code>
Replace	<code>incomedec_any=1 if incomedec_1=1 or incomedec_2=1 or incomedec_3=1 or incomedec_4=1 or incomedec_5=1 or incomedec_6=1 or inputdec_7=1 or inputdec_8=1</code>
Replace	<code>incomedec_any=missing if incomedec_1=missing and incomedec_2=missing and incomedec_3=missing and incomedec_4=missing and incomedec_5=missing and incomedec_6=missing and inputdec_7=missing and inputdec_8=missing</code>
Label values	0 "No"
	1 "Yes"
Label variable	"Had input into income decisions on 1+ activity, incl. HH expenditures"

**Step 7.8.** Create a variable that indicates if the respondent participated in at least some decisions related to using income from at least one of four agriculture activities: food crop farming, cash crop farming, livestock raising, and fishing or fishpond culture (`incomedec_agr_any`). The variable is coded as missing if all four variables being summed are coded as missing.

Set	<code>incomedec_agr_any=0 if incomedec#missing</code>
Replace	<code>incomedec_agr_any=1 if incomedec_1=1 or incomedec_2=1 or incomedec_3=1 or incomedec_6=1</code>
Replace	<code>incomedec_agr_any=missing if incomedec_1=missing and incomedec_2=missing and incomedec_3=missing and incomedec_6=missing</code>
Label values	0 "No"
	1 "Yes"
Label variable	"Had input into income decisions on 1+ ag activity, incl. HH expenditures"

**Step 7.9.** Create a variable for each group that replaces missing `groupmember` values for respondents who report that they do not have that type of group in their community with 0s so that all tabulated estimates have the same denominator (i.e., all respondents) (`groupmember_01x-groupmember_11x`). These variables should not be coded as missing unless data are missing for the entire Sub-module 6.4B.

For each	group (x) of 01 02 03 04 05 06 07 08 09 10 11:
----------	--

```

Set          groupmember_`x`x= groupmember_`x`
Replace      groupmember_`x`x=0 if groupmember_`x`=missing and
              groupmember_any#missing
Replace      groupmember_`x`x=missing if v6404_01-v6404_11 and
              v6405_01-v6405_11 all=missing

```

**Step 7.10.** Create a variable that indicates if the respondent did any work during the day preceding the survey (**work\_any**). The variable is coded as missing if **work\_hours** is missing a value.

```

Set          work_any=0 if work_hours#missing
Replace      work_any=1 if work_hours>0 and work_hours#missing
Label values 0 "No"
              1 "Yes"
Label variable "Performed any work activities, previous day"

```

**Step 7.11.** Create variables that count the number of 15-minute time increments spent performing each activity (**time\_A-time\_X**).

```

For each      letter (z) in A B C D E F G H I J K L M N O P Q R
              S T U V W X:
Set          time_`z`=0

For each      number (x) in 15 30 45 60:
  For each    number (i) in 01 02 03 04 05 06 07 08 09 10 11 12
              13 14 15 16 17 18 19 20 21 22 23 24:
    For each  letter(z) in A B C D E F G H I J K L M N O P Q R
              S T U V W X:
      Replace  time_`z`=time_`z`+1 if v6601p_`x`_`i`="`z`"

```

Label variables:

```

time_A      "Sleeping/resting (# 15-min periods)"
time_B      "Eating/drinking (# 15-min periods)"
time_C      "Personal care (# 15-min periods)"
time_D      "School/homework (# 15-min periods)"
time_E      "Work as employed (# 15-min periods)"
time_F      "Own business work (# 15-min periods)"
time_G      "Food crop farming (# 15-min periods)"
time_H      "Cash crop farming (# 15-min periods)"
time_I      "Livestock raising (# 15-min periods)"
time_J      "Fishing/fishpond culture (# 15-min periods)"
time_K      "Commuting (work/school) (# 15-min periods)"
time_L      "Shopping/getting services (# 15-min periods)"
time_M      "Weaving/sewing/textile care (# 15-min periods)"
time_N      "Cooking (# 15-min periods)"
time_O      "Domestic work (# 15-min periods)"
time_P      "Fetching water (# 15-min periods)"
time_Q      "Caring for children (# 15-min periods)"
time_R      "Caring for sick/elderly adults (# 15-min
              periods)"
time_S      "Traveling, not for work/school (# 15-min
              periods)"

```

```

time_T      "Watching TV/listening to radio/reading (# 15-min
             periods)"
time_U      "Exercise (# 15-min periods)"
time_V      "Social activities/hobbies (# 15-min periods)"
time_W      "Religious activities (# 15-min periods)"
time_X      "Other (# 15-min periods)"

```

**Step 7.12.** Create a variable that combines the number of 15-minute time increments spent performing domestic work excluding fetching water and the time spent fetching water as a stand-alone activity to enable comparison to baseline when the two activities were combined (**time\_OP**).

```

Set          time_OP=0
Replace      time_OP=time_O+time_P
Label variable "Domestic work, incl. fetching water (# 15-min
              periods)"

```

**Step 7.13.** Create a variable combines the number of 15-minute time increments spent on all agriculture work activities (i.e., food crop farming, cash crop farming, livestock raising, and fishing or fishpond aquaculture) to align with how the estimates are included in the results tables (**time\_GHIJ**).

```

Set          time_GHIJ=0
Replace      time_GHIJ=time_G+time_H+time_I+time_J
Label variable "Farming/livestock/fish work (# 15-min periods)"

```

**Step 7.14.** Create a variable that combines the number of 15-minute time increments spent on caring for children and caring for the elderly or sick to align with how the estimates are included in the results tables (**time\_QR**).

```

Set          time_QR=0
Replace      time_QR=time_Q+time_R
Label variable "Caring for children/elderly/sick (# 15-min
              periods)"

```

**Step 7.15.** Create variables that indicate if the respondent partook in each activity (**partook\_A-partook\_X**). These variables are coded as missing if the respondent is excluded from the workload indicator calculation (i.e., **work\_hours**=missing). **Note:** Fetching water, activity P, was separated from domestic work in the core Midline Survey questionnaire. If planning to compare with baseline, be sure that you name the baseline variables to enable comparison; **partook\_P** will have a value for midline only.

```

For each      activity (z) of A B C D E F G H I J K L M N O P Q
              R S T U V W X:
Set          partook_`z`=0 if work_hours=missing
Replace      partook_`z`=1 if time_`z`>0 & time_`z`≠missing
Label values  0 "No"
              1 "Yes"

Label variables:
  partook_A  "Sleeping/resting, previous day"
  partook_B  "Eating/drinking, previous day"

```

partook_C	"Personal care, previous day"
partook_D	"School/homework, previous day"
partook_E	"Work as employed, previous day"
partook_F	"Own business work, previous day"
partook_G	"Food crop farming, previous day"
partook_H	"Cash crop farming, previous day"
partook_I	"Livestock raising, previous day"
partook_J	"Fishing/fishpond culture, previous day"
partook_K	"Commuting (work/school), previous day"
partook_L	"Shopping/getting services, previous day"
partook_M	"Weaving/sewing/textile care, previous day"
partook_N	"Cooking, previous day"
partook_O	"Domestic work excl. fetching water, previous day"
partook_P	"Fetching water, previous day"
partook_Q	"Caring for children, previous day"
partook_R	"Caring for sick/elderly adults, previous day"
partook_S	"Traveling (not for work/school), previous day"
partook_T	"Watching TV/listening to radio/reading, previous day"
partook_U	"Exercise, previous day"
partook_V	"Social activities/hobbies, previous day"
partook_W	"Religious activities, previous day"
partook_X	"Other, previous day"

**Step 7.16.** Create a variable that combines domestic work excluding fetching water and fetching water as a stand-alone activity to enable comparison to baseline when the two activities were combined (**partook\_OP**). The variable is coded as missing if the respondent is excluded from the workload indicator calculation (i.e., **work\_hours=missing**).

Set	partook_OP=0 if work_hours≠missing
Replace	partook_OP=1 if partook_P=1 or partook_O=1
Label values	0 "No"
	1 "Yes"
Label variable	"Domestic work incl. fetching water, previous day"

**Step 7.17.** Create a variable that combines all agriculture work activities to align with how the estimates are included in the results tables (**partook\_GHIJ**). The variable is coded as missing if the respondent is excluded from the workload indicator calculation (i.e., **work\_hours=missing**).

Set	partook_GHIJ=0 if work_hours≠missing
Replace	partook_GHIJ=1 if partook_G=1 or partook_H=1 or partook_I=1 or partook_J=1
Label values	0 "No"
	1 "Yes"
Label variable	"Farming/livestock/fish work, previous day"

**Step 7.18.** Create a variable that combines caring for children and caring for elderly/sick to align with how the estimates are included in the results tables (**partook\_QR**). The variable is coded as missing if the respondent is excluded from the workload indicator calculation (i.e., **work\_hours=missing**).

```

Set          partook_QR=0 if work_hours≠missing
Replace      partook_QR=1 if partook_Q=1 or partook_R=1
Label values 0 "No"
              1 "Yes"
Label variable "Caring for children/elderly/sick, previous day"

```

**Step 7.19.** Create variables that the amount of time in hours that a respondent partook in each activity (hours\_A-hours\_X), assigning respondents who did not partake an activity a value of '0.' The variables are coded as missing if the respondent is excluded from the workload indicator calculation (i.e., work\_hours=missing).

**Note:** Fetching water, activity P, was separated from domestic work in the core Midline Survey questionnaire. If planning to compare with baseline, be sure that you name the baseline variables to enable comparison; hours\_P will have a value for midline only.

```

For each      activity (z) of A B C D E F G H I J K L M N O P Q
              R S T U V W X:
Set           hours_`z`=0 if work_hours≠missing
Replace       hours_`z`=time_`z`*15/60 if time_`z`≠missing

```

Label variables:

```

hours_A      "Sleeping/resting, time spent (hours)"
hours_B      "Eating/drinking, time spent (hours)"
hours_C      "Personal care, time spent (hours)"
hours_D      "School/homework, time spent (hours)"
hours_E      "Work as employed, time spent (hours)"
hours_F      "Own business work, time spent (hours)"
hours_G      "Food crop farming, time spent (hours)"
hours_H      "Cash crop farming, time spent (hours)"
hours_I      "Livestock raising, time spent (hours)"
hours_J      "Fishing/fishpond culture, time spent (hours)"
hours_K      "Commuting (work/school), time spent (hours)"
hours_L      "Shopping/getting services, time spent (hours)"
hours_M      "Weaving/sewing/textile care, time spent (hours)"
hours_N      "Cooking, time spent (hours)"
hours_O      "Domestic work, time spent (hours)"
hours_P      "Fetching water, time spent (hours)"
hours_Q      "Caring for children, time spent (hours)"
hours_R      "Caring for sick/elderly adults, time spent
              (hours)"
hours_S      "Traveling (not for work/school), time spent
              (hours)"
hours_T      "Watching TV/listening to radio/reading, time spent
              (hours)"
hours_U      "Exercise, time spent (hours)"
hours_V      "Social activities/hobbies, time spent (hours)"
hours_W      "Religious activities, time spent (hours)"
hours_X      "Other, time spent (hours)"

```

**Step 7.20.** Create a variable that combines the time spent on domestic work excluding fetching water and the time spent fetching water as a stand-alone activity to enable comparison to baseline when the

two activities were combined (**hours\_OP**). The variable is coded as missing if the respondent is excluded from the workload indicator calculation, based on Step 7.15.

```
Set          hours_OP=0
Replace      hours_OP=hours_O+hours_P
Label values 0 "No"
              1 "Yes"
Label variable "Domestic work incl. fetching water, previous
              day"
```

**Step 7.21.** Create a variable that combines the time spent on all agriculture work activities (i.e., food crop farming, cash crop farming, livestock raising, and fishing or fishpond aquaculture) to align with how the estimates are included in the results tables (**hours\_GHIJ**). The variable is coded as missing if the respondent is excluded from the workload indicator calculation, based on Step 7.15.

```
Set          hours_GHIJ=0 if work_any#missing
Replace      hours_GHIJ=hours_G+hours_H+hours_I+hours_J
Label variable "Farming/livestock/fish work, time spent (hours)"
```

**Step 7.22.** Create a variable that combines the time spent on caring for children and caring for the elderly or sick to align with how the estimates are included in the results tables (**hours\_QR**). The variable is coded as missing if the respondent is excluded from the workload indicator calculation, based on Step 7.15.

```
Set          hours_QR=0 if work_any#missing
Replace      hours_QR=hours_Q+hours_R
Label variable "Caring for children/elderly/sick, time spent
              (hours)"
```

## **Part 2: Computing weighted adequacy scores, empowerment, and the 5DE**

The following steps in Part 2 describe how to create the variables needed to calculate the percentage of women empowered, the average weighted adequacy score, and the 5DE and follow the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_2 5DEcalc.do*.

### **Step 1. Weighted adequacy scores**

**Step 1.1.** Define a variable list (**varlist\_all**) that includes the six A-WEAI indicators created in Part 1.

```
Set          varlist_all=feelinputdecagr, jown_count,
                      credjanydec_any, incdec_count, groupmember_any,
                      npoor_z105
```

**Step 1.2.** Create variables for the A-WEAI indicator weights (**w\_[A-WEAI indicator name]**). Assign a weight of 1/5 for domains 1, 3, 4, and 5 indicators. Assign a weight of 2/15 for the asset ownership indicator and 1/15 for credit access indicator in domain 2. (Weights sum to 1, not to the number of indicators.)

```
For each      variable (var) of variable list varlist_all:
```

```

Set          w_`var'=(1÷5)

Set          w_jown_count=(2÷15)
Set          w_credjanydec_any=(1÷15)

```

**Step 1.3.** Apply the weights created in the previous step to each A-WEAI indicator to create weighted A-WEAI indicator variables (`wg0_[A-WEAI indicator name]`).

```

For each      variable (var) of varlist_all:
Set          wg0_`var'=(`var')*(w_`var')

```

**Step 1.4.** Define a variable list (`wg0_varlist_all`) that includes the six weighted A-WEAI indicators.

```

Set          wg0_varlist_all=wg0_feelinputdecagr,
          wg0_jown_count, wg0_credjanydec_any,
          wg0_incdec_count, wg0_groupmember_any,
          wg0_npoor_z105

```

**Step 1.5.** Create a variable for each A-WEAI indicator indicating whether the respondent is missing a value for the indicator (`[A-WEAI indicator]_miss`).

```

For each      variable (var) of varlist_all:
Set          `var'_miss=1 if `var'=missing
Replace      `var'_miss=0 if `var'≠missing

```

**Step 1.6.** Calculate the weighted adequacy score (`ca`), which is the sum of the six weighted A-WEAI indicators.

```

Set          ca=wg0_feelinputdecagr+wg0_jown_count+
          wg0_credjanydec_any+wg0_incdec_count+
          wg0_groupmember_any+wg0_npoor_z105
Replace      ca=round(ca,0.0001)
Label variable "Weighted adequacy score"

```

**Step 1.7.** Create a variable (`n_missing`) to capture the number of A-WEAI indicators that each respondent is missing.

```

Set          n_missing=0
Replace      n_missing=1 if feelinputdecagr=missing
Replace      n_missing=n_missing+1 if jown_count=missing
Replace      n_missing=n_missing+1 if credjanydec_any=missing
Replace      n_missing=n_missing+1 if incdec_count=missing
Replace      n_missing=n_missing+1 if groupmember_any=missing
Replace      n_missing=n_missing+1 if npoor_z105=missing
Label variable "Number of missing variables"

```

**Step 1.8.** Create a variable (`miss_any`) to indicate whether the respondent is missing any A-WEAI indicators.

```

Set          miss_any=0

```



```

Replace      miss_any=1 if n_missing>0 and n_missing#missing
Label values 0 "No"
              1 "Yes"
Label variable "Individual is missing 1+ A-WEAI indicators"

```

**Step 1.9.** Count the number of women with complete 5DE data—that is, with a value for each of the six A-WEAI indicators; these are the women who will be included in the 5DE indicator calculation. Check for records with missing values for any of the six A-WEAI indicators to see what variables women are missing; women are most often missing a value for A-WEAI Indicator 1 because they did not participate in agricultural activities in the 12 months preceding the survey.

```

Count if miss_any=0
List all variables in variable list varlist_all if miss_any=1

```

**Step 1.10.** Save the data file.

```

Save "FTF ZOI Survey [COUNTRY] [YEAR] aweai depr indicators"

```

## Step 2. Empowerment status and adequacy scores of the disempowered

**Step 2.1.** Create a binary variable that indicates whether the individual is empowered using a weighted adequacy score threshold of 0.8 (**emp\_80p**). Women with a weighted adequacy score (**ca**, created in Step 1.6) equal to or higher than the threshold (i.e.,  $ca \geq 0.8$ ) are considered empowered.

```

Set          emp_80p=0 if ca<float(80/100)
Replace      emp_80p=ca<float(80/100)
Replace      emp_80p=missing if miss_any=1
Label values 0 "No"
              1 "Yes"
Label variable "Woman is empowered, adequacy threshold<math>\geq 0.8</math>"

```

**Step 2.2.** Create a binary variable that indicates whether the individual is disempowered using a threshold of 0.8 (**dis\_80p**). Women with weighted adequacy in the six A-WEAI indicators less than the threshold (i.e.,  $ca < 0.8$ ) are considered disempowered.

```

Set          dis_80p=(ca<float(80/100))
Replace      dis_80p=missing if miss_any=1
Label values 0 "No"
              1 "Yes"
Label variable "Woman is disempowered, adequacy threshold<math>\geq 0.8</math>"

```

**Step 2.3.** Create a variable equal to the weighted adequacy score of those who are disempowered (**ad\_80p**).

```

Set          ad_80p=ca if dis_80p=1
Label variable "Weighted adequacy score (ca) for disempowered,
               adequacy threshold<math>\geq 0.8</math>"

```

### Step 3. 5DE calculation

**Step 3.1.** Define the complex survey design to account for the primary sampling units, the primary adult female decision-maker sampling weight that will be used, and the strata. (Sample code uses Stata syntax.)

```
svyset hhea [pw=wgt_fpdm], strata(strata)
```

**Step 3.2.** Create a variable (**swp\_emp\_80p**) equal to the sample-weighted percentage of women achieving empowerment in the ZOI. The sample code uses Stata syntax; the sample-weighted percentage can be obtained by calculating the mean of the 0/1 binary **emp\_80p** variable, and then the value can be retrieved from position [1,1] of the results matrix e(b).

**Note:** This is the percentage of women empowered indicator reported in the midline indicator assessment template Tables 6.2.1, ESI, and A1.1.

```
svy: mean emp_80p
Set      swp_emp_80p=e(b) [1,1]
Label variable  "Sample-weighted percent of women empowered,
adequacy threshold≥0.8, constant"
```

**Step 3.3.** Create a variable (**swp\_dis\_80p**) equal to the sample-weighted percentage of women who are disempowered in the ZOI. The sample code uses Stata syntax; the sample-weighted percentage can be obtained by calculating the mean of a 0/1 binary **dis\_80p** variable, and then the value can be retrieved from position [1,1] of the results matrix e(b).

```
svy: mean dis_80p
Set      swp_dis_80p=e(b) [1,1]
Label variable  "Sample-weighted percent of women disempowered,
adequacy threshold≥0.8, constant"
```

**Step 3.4.** Create a variable (**swm\_ad\_80p**) equal to the sample-weighted average adequacy score among disempowered women. The sample code uses Stata syntax; the sample-weighted mean can be obtained by calculating the mean of a continuous **ad\_80p** variable, and then the value can be retrieved from position [1,1] of the results matrix e(b).

**Note:** This is the mean weight adequacy count for disempowered women reported in the midline indicator assessment template Tables 6.2.1, ESI, and A1.1.

```
svy: mean ad_80p
Set      swm_ad_80p=e(b) [1,1]
Label variable  "Sample-weighted mean weighted adequacy count
(ca) for disempowered, adequacy threshold≥0.8,
constant"
```

**Step 3.5.** Create a variable (**EA\_80p**) equal to the empowerment index using the equation explained in Section 12.2, Part 2, Step 4:  $EA = He + Hd(Ad)$ .

```
Set      EA_80p=swp_emp_80p+(swp_dis_80p*swm_ad_80p)
Label variable  "5DE (5 domains of empowerment), constant"
```

**Step 3.6.** Save the individual-level data.

```
Save          "FTF ZOI Survey [COUNTRY] [YEAR]
              aweai_individual_indices"
```

### **Part 3: Calculating the 5DE standard (jackknife approach) and performing a test of difference**

Part 3, Step 1 describes how to calculate the 5DE standard error using a jackknife approach and follows the Stata syntax in the corresponding section of *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_2 5DEcalc.do*. Part 3, Step 2 describes how to perform a test of difference using 5DE midline and baseline estimates and follows the Stata syntax in the corresponding section of *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_3 BL\_ML\_compare.do*.

#### **Step 1. Calculating the 5DE standard error (jackknife approach)**

**Step 1.1.** Create a variable that indicates whether the woman is 18-29 years of age or 30 years of age or older using the age variable (see Section 4.3.1).

```
Set          fdm_agecat=missing
Replace      fdm_agecat=1 if age≥18 and age<30
Replace      fdm_agecat=2 if age≥30 and age<98
Label values 1 "18-29 years"
              2 "30+ years"
Label variable "Female PDM's age (18-29, 30+)"
```

**Step 1.2.** Sort the records by cluster and then create a sequential cluster ID variable (*id*) that starts from 1.

```
Sort          hhea
Set           id=group(hhea)
```

**Step 1.3.** Create a variable (*alloc*) that sums the values of the sampling weight by strata if the respondent has complete 5DE data—that is, they are not missing values for any A-WEAI indicators.

```
Set          alloc=total(wgt_pfdm) by strata if miss_any=0
```

**Step 1.4.** Create a variable (*alloc\_AGE*) that sums the values of the sampling weight by strata and age if the respondent has complete 5DE data.

```
By strata fdm_agecat: Set alloc_AGE=total(wgt_pfdm) if miss_any=0
```

**Step 1.5.** Save the data to a data file, AWEAI\_DEPR, in the Temp sub-folder.

```
save Temp\AWEAI_DEPR, replace
```

**Step 1.6.** Create sample-weighted mean and DEFF variables for *emp\_80p*, *dis\_80p*, and *ad\_80p* if the woman is not missing any A-WEAI indicators: (0) overall, (1) women 18 to 29 years of age, and (2) women 30 years of age or older (*mean\_0\_[variable]*, *mean\_1\_[variable]*, *mean\_2\_[variable]*, *deff\_0\_[variable]*, *deff\_1\_[variable]*, *deff\_2\_[variable]*). The sample code uses Stata syntax; the sample-weighted percentages for

**emp\_80p** and **dis\_80p** can be obtained by calculating the sample-weighted mean of the 0/1 binary variables, and the sample-weighted mean of **ad\_80p** can be obtained because it is a continuous variable. The values are saved to matrix **a** for retrieval. The DEFF values are generated when the **svy: mean** commands are run. The DEFF values are saved to matrix **d** for retrieval.

```
foreach emp_80p dis_80p ad_80p:
    svy: mean `var' if miss_any=0
    gen a=e(b)
    gen d=e(deff)
    gen mean_0_`var'=a[1,1]
    gen deff_0_`var'=d[1,1]

    svy: mean `var' if miss_any=0, over(fdm_agecat)
    gen a=e(b)
    gen d=e(deff)
    gen mean_1_`var'=a[1,1]
    gen mean_2_`var'=a[1,2]
    gen deff_1_`var'=d[1,1]
    gen deff_2_`var'=d[1,2]
```

**Step 1.7.** Obtain the values for the variables created in the previous step for each of the three groups, where [X] is a placeholder for 0, 1, and 2 in a single record. The means will be saved in a working data file with one row. In Stata, this can be done using the collapse (mean) command.

```
Save in row 1:  mean_[X]_emp_80p mean_[X]_dis_80p mean_[X]_ad_80p
               deff_[X]_emp_80p deff_[X]_dis_80p deff_[X]_ad_80p
```

**Step 1.8.** Generate the 5DE score variable for each group (**EA\_80p0**, **EA\_80p1**, **EA\_80p2**).

```
Set EA_80p0=mean_0_emp_80p+(mean_0_dis_80p*mean_0_ad_80p)
Set EA_80p1=mean_1_emp_80p+(mean_1_dis_80p*mean_1_ad_80p)
Set EA_80p2=mean_2_emp_80p+(mean_2_dis_80p*mean_2_ad_80p)
```

**Step 1.9.** Generate the DEFF variable for each group by averaging the **emp\_80p** DEFF variable and the **ad\_80p** DEFF variable for each group (**deff0**, **deff1**, **deff2**).

```
Set deff0=(deff_0_emp_80p+deff_0_ad_80p)÷2
Set deff1=(deff_1_emp_80p+deff_1_ad_80p)÷2
Set deff2=(deff_2_emp_80p+deff_2_ad_80p)÷2
```

**Step 1.10.** Keep only the **EA\_80p** and **deff** variables.

```
Keep variables EA_80p0 EA_80p1 EA_80p2 deff0 deff1 deff2
```

**Step 1.11.** Create a variable (ID) equal to one and reshape the data from wide to long format using the ID variable and recreating the age category variable (**fdm\_agecat**), with an additional category for all women. The working data file will be converted from a data file with one row to a data file with three rows—one for each **fdm\_agecat** category (0=all women, 1= women 18-29 years, and 2=women 30+ years). Afterward, drop ID from the data file.

```

Set ID=1
Reshape wide to long
variable stubs EA_80p deff
index variable ID
new variable fdm_agecat
Label variable EA_80p "5DE score"
Label variable deff "5DE design effect"
Label variable fdm_agecat "Woman's age category"
Label values fdm_agecat 0 "All women"
                        1 "Women 18-29 years"
                        2 "Women 20+ years"

Drop variable ID

```

**Step 1.12.** Initialize a counter variable (**replicate**) equal to 0.

```
Set replicate=0
```

**Step 1.13.** Save the data to a data file, 5DE\_JK, in the Temp sub-folder.

```
save Temp\5DE_JK, replace
```

**Step 1.14.** Create a scalar (**P2**) equal to the number of sample clusters (**hhea**) in the analysis. Be sure to update the syntax so that **[XX]** is equal to the number of sample clusters in the survey.

```
Set PS2=[XX]
```

**Step 1.15.** Perform the jackknife computation for the 5DE variables using the following loop. For each cluster included in the analysis, a 5DE score (**EA\_80p**) is generated for all women, women 18-29 years of age, and women 30 years of age or older using a jackknife approach, which involves dropping the cases from one cluster at a time, adjusting the weights (using the following variables: **allocj**, **allocj\_AGE**, **wtadj**, **wtadj\_AGE**), calculating the 5DE scores across the remaining clusters, and appending the 5DE scores to a data file, 5DE\_JK, in the Temp sub-folder that will later be used to calculate the 5DE standard errors (overall and by age group). Note that in this step the **fdm\_agecat** variable in the AWEAI\_DEPR data file has only two values, as created in Part 3, Step 1, but the records for the third **fdm\_agecat** category (0=all women) are created in this step so that the variable will have three categories in subsequent steps.

```

Set local i=1
While `i'<=PS2:
  use Temp\AWEAI_DEPR, replace
  drop record if id=`i'

  By strata: set allocj=total(wgt_pfdm) if
              miss_any=0
  By strata and fdm_agecat: set allocj_AGE=total(wgt_pfdm) if
                           miss_any=0

  Set wtdj=wgt_pfdm*(alloc÷allocj) if miss_any=0
  Set wtadj_AGE=wgt_pfdm*(alloc_AGE÷allocj_AGE) if
    miss_any=0

```

```

Save          Temp\AWEAI_AGE_`i`, replace

Save in row 1: (mean) emp_80p dis_80p ad_80p [aw=wtadj]
Set          fdm_agecat=0
Save          Temp\AWEAI_OVERALL_`i`, replace

Use          Temp\AWEAI_AGE_`I`, clear
By fdm_agecat: Save in row 1: (mean) emp_80p dis_80p ad_80p
               [aw=wtadj_AGE]
Append        using Temp\AWEAI_OVERALL_`i`, force

Set          EA_80p=emp_80p+(dis_80p*ad_80p)
Keep variables fdm_agecat EA_80p
Set          replicate=`i`
Append        using Temp\5DE_JK, force

Save          Temp\5DE_JK, replace
Erase        Temp\AWEAI_AGE_`i`.dta
Erase        Temp\AWEAI_OVERALL_`i`.dta
Set          local i=`i`+1

```

**Step 1.16.** Erase the temporary AWEAI\_DEPR data file.

```

Erase          Temp\AWEAI_DEPR

```

**Step 1.17.** Reshape the data from long to wide format so that there are three **EA20p** variables (**EA\_80p0**, **EA\_80p1**, **EA\_80p2**) and three **DEFF** variables (**deff0**, **deff1**, **deff2**)—one for each value of **fdm\_agecat**—for each cluster included in the analysis.

In the data file in long format, there will be three records for each cluster that were generated using the jackknife approach in Step 1.15, plus three additional records generated in Steps 1.11 and 1.12 that contain the 5DE and DEFF values calculated for the entire sample (i.e., the number of rows in the long format data file=[3\*PS2]+3). When the data file is transformed into wide format, there will be one row for each cluster, with each row containing 5DE and DEFF values for the three **fdm\_agecat** categories (0=all women, 1= women 18-29 years , and 2= women 30+ years).

```

Reshape          long to wide
variable stubs   EA_80p deff
index variable   replicate
suffix variable  fdm_agecat

```

**Step 1.18.** Save the data to the temporary file, 5DE\_JK.

```

Save          Temp\5DE_JK, replace

```

**Step 1.19.** Create a variable list (**varlist\_0**) that includes the three **EA\_80p** variables created in the Jackknife loop for each cluster.

```

Set          varlist_0=EA80p0 EA_80p1 EA_80p2

```

**Step 1.20.** Create variables needed to calculate the standard errors of the three **EA\_80p** variables (**A\_EA80p0**, **A\_EA80p1**, **A\_EA80p2**, **B\_EA80p0**, **B\_EA80p1**, **B\_EA80p2**).

```
For each      variable `var' of varlist_0:
  Set         A_`var'=(PS2*`var'[1]-(PS2-1)*`var')
  Set         B_`var'=(A_`var'-`var'[1])^2
```

**Step 1.21.** Drop the first row in the data file that has the DEFFs and initially calculated 5DE variables.

```
Drop records  if replicate=0
```

**Step 1.22.** Save the sum of the three '**B\_**' variables across all rows (clusters) in row 1 and rename them (**JKSE\_EA\_80p0**, **JKSE\_EA\_80p1**, **JKSE\_EA\_80p2**), dropping all other variables from the working data file. Summing the variables across all rows can be done in Stata using the collapse (sum) command.

```
Save in row 1:  (sum) JKSE_EA_80p0=B_EA_80p0
                  JKSE_EA_80p1=B_EA_80p1
                  JKSE_EA_80p2=B_EA_80p2
Keep variables  JKSE_EA_80p0 JKSE_EA_80p1 JKSE_EA_80p2
```

**Step 1.23.** Create a variable list (**varlist\_1**) that includes the three variables created in Step 1.22.

```
Set           varlist_1=JKSE_EA_80p0 JKSE_EA_80p1 JKSE_EA_80p2
```

**Step 1.24.** Create variables equal to the standard error for each of the three **EA\_80p** variables (**SE\_EA80p0**, **SE\_EA\_80p1**, **SE\_EA\_80p2**) using the variable list created in Step 1.23.

```
For each      variable `var' of varlist_1:
  Set         SE_`var' = sqrt(`var'/(PS2*(PS2-1)))
  Drop variables `var'
```

**Step 1.25.** Create an ID variable (**ID**) equal to 1 and save the data file containing the standard errors.

```
Set           ID=1
Save          Temp\5DE_SE, replace
```

**Step 1.26.** Load the 5DE data file saved in Step 1.18 and keep only the row with the DEFFs.

```
Use           Temp\5DE_JK, replace
Drop records  if replicate≠0
```

**Step 1.27.** Create an ID variable (**ID**) equal to 1 and merge the data file with the 5DE standard error data file with the working data file.

```
Set           ID=1
Merge         1:1
              key variable ID
              data file    Temp\5DE_SE
Drop variables _merge replicate
```

**Step 1.28.** Rename the standard error variables (**SE0**, **SE1**, **SE2**).

```

Rename      SE_JKSE_EA_80p0 to be SE0
Rename      SE_JKSE_EA_80p1 to be SE1
Rename      SE_JKSE_EA_80p2 to be SE2

```

**Step 1.29.** Transform the data file from wide to long format and drop the ID and \_j variables. The working data file will be converted from a data file with one row to a data file with three rows—one for each `fdm_agecat` category (0=all women, 1= women 18-29 years, and 2=women 30+ years). Afterward, drop ID and \_j from the data file.

```

Reshape      wide to long
  variable stubs    EA_80p SE deff
  index variable    ID
  new variable      _j
Drop variables    ID _j

```

**Step 1.30.** Create row labels for the output and an age category variable to match.

```

Set          Label "All women" "Young women" "Older women"
Set          age_cat=1 if Label="All women"
Replace      age_cat=2 if Label="Young women"
Replace      age_cat=3 if Label="Older women"
Label values 1 "All women"
              2 "Women 18-29 years"
              3 "Women 30+ years"
Label variable "Women's age category (A-WEAI)"

```

**Step 1.31.** Create variables equal to the upper bounds (UCI) and lower bounds (LCI) of the 95 percent CIs for the 5DE.

```

Set          LCI=EA_80p- (1.96*SE)
Set          UCI=EA_80p+ (1.96*SE)

```

**Step 1.32.** Save the results to a data file, 5DE\_FINAL, after ordering the variables so that the results for all women are in the first row, the results for women 18-29 years of age are in the second row, and the results for women 30 years or older are in the third row.

```

Order        Label EA_80p SE LCI UCI, first
Save         Results\"FTF ZOI Survey [COUNTRY] [YEAR]
              5DE_FINAL", replace

```

**Step 1.33.** Export the 5DE results (5DE, SE, LCI, and UCI) for the three groups to Excel.

```

Export to Excel file "FTF ZOI Survey [COUNTRY] [YEAR] 5DE
CI.xlsx"

```

## Step 2. Performing a test of difference for 5DE estimates

**Step 2.1.** Ensure that the data file created in Part 3, Step 1.32 with the 5DE results generated for the Midline Survey has an “ML” included in the name of the file and ensure that data file with the same 5DE results generated for the Baseline Survey has a “BL” included in the name of the file. Then load the “ML” data file and append the “BL” data file.



```
Load "\\Results\FTF ZOI Survey [COUNTRY] [YEAR] 5DE_FINAL_ML"
Append "\\Results\FTF ZOI Survey [COUNTRY] [YEAR] 5DE_FINAL_BL"
```

**Step 2.2.** Transform the data file from long to wide format so that there are three records, each with baseline and midline 5DE data (i.e., the estimate, standard error, lower bound of the 95 percent CI, upper bound of the 95 percent CI, and DEFF). The first row is for all women, the second for women 18-29 years of age, and the third for women 30 years of age or older.

```
Reshape          long to wide
variable stubs  EA_80p SE LCI UCI deff
index variable  fdm_agecat
new variable    survey
```

**Step 2.3.** Label the variables in data file.

```
Label variable EA_80p1 "5DE, baseline"
Label variable SE1     "5DE SE, baseline"
Label variable LCI1    "5DE lower CI bound, baseline"
Label variable UCI1    "5DE upper CI bound, baseline"
Label variable deff1   "5DE design effect, baseline"
Label variable EA_80p2 "5DE, midline"
Label variable SE2     "5DE SE, midline"
Label variable LCI2    "5DE lower CI bound, midline"
Label variable UCI2    "5DE upper CI bound, midline"
Label variable deff2   "5DE design effect, midline"
```

**Step 2.4.** Create a variable that is equal to the Z-score (Z), following the guidance in Section 3.2.3, "Analyzing indicators of proportions or means if the data are not available."

```
Set          Z=abs(EA_80p2-EA_80p1)/sqrt(SE2^2 + SE1^2)
Label variable "5DE ML/BL comparison Z-score"
```

**Step 2.5.** Create a variable that is equal to the p-value (P) by subtracting the cumulative standard normal distribution of the absolute value of Z from 1 and multiplying the result by 2.

```
Set          P=2*(1-normal(abs(Z)))
Label variable "5DE ML/BL comparison P-value"
```

#### **Part 4: Calculating adequate achievement in each A-WEAI indicator**

Part 4 describes how to calculate adequate achievement in each A-WEAI indicator among women who have complete 5DE data, disempowered women, and all women, and follows the Stata syntax in the corresponding section of *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_2 5DE\_calc.do*. Stata syntax to compare these estimates over time is included in the Table 6.3 section of *FTF ZOI Survey [COUNTRY] [YEAR] syntax aweai\_3 BL\_ML\_compare.do*.

## Step 1. Calculating among women who have complete 5DE data

### Definitions, women with complete 5DE data

Numerator	Percentage of women who have complete 5DE data who are adequate in each A-WEAI indicator
Denominator	Number of women who have complete 5DE data
Unit of measure	Percentage
Level of data	Individual-level
Sampling weight	Primary adult female decision-maker
Disaggregation levels	Age category (18-29 years, 30+ years)
Treatment of missing data	Primary adult female decision-makers who do not have complete 5DE data (i.e., data for all six A-WEAI indicators) are excluded from the calculations.
Survey variables used	hhea, hhnum, wgt_fpdm
Analytic variables used	feelinputdecagr, jown_count, credjanydec_any, incdec_count, groupmember_any, npoor_z105, miss_any, youth_fdm_dj
Analytic variables created	miss_none

**Step 1.1.** Calculate the sample-weighted percentage of women who have complete 5DE data (i.e., `miss_any=0`) who are adequate in each A-WEAI indicator overall and by age group using the complex sampling design specified in Part 2, Step 3.1 and the variable list created in Part 2, Step 1.1. (Example code uses Stata syntax.)

**Note:** These results reported in the midline indicator assessment template Table 6.3.1.

```
foreach var in $varlist_all:
    svy, subpop(miss_none): tab `var'
    svy, subpop(miss_none): tab `var' youth_fdm_dj, col
```

## Step 2. Calculating among disempowered women

### Definitions, disempowered women

Numerator	Number of disempowered women who are adequate in each A-WEAI indicator
Denominator	Number of women who are disempowered
Unit of measure	Percentage
Level of data	Individual-level
Sampling weight	Primary adult female decision-maker
Disaggregation levels	Age category (18-29 years, 30+ years)
Treatment of missing data	Primary adult female decision-makers who are empowered and those who do not have complete 5DE data are excluded from the calculations.
Survey variables used	hhea, hhnum, wgt_fpdm
Analytic variables used	feelinputdecagr, jown_count, credjanydec_any, incdec_count, groupmember_any, npoor_z105, youth_fdm_dj, dis_80p
Analytic variables created	miss_none

**Step 2.1.** Calculate the sample-weighted percentage of disempowered women (i.e., `dis_80p=1`) who are adequate in each A-WEAI indicator overall and by age group using the complex sampling design specified in Part 2, Step 3.1 and the variable list created in Part 2, Step 1.1. (Example code uses Stata syntax.)

**Note:** These results are reported in the midline indicator assessment template Tables 6.3.2, ESI, and AI.I.

```
foreach var in $varlist_all:
    svy, subpop(dis_80p): tab `var'
    svy, subpop(dis_80p): tab `var' youth_fdm_dj, col
```

### Step 3. Calculating among all women

#### Definitions, all women

Numerator	Number of women who are adequate in each A-WEAI indicator
Denominator	Number of women
Unit of measure	Percentage
Level of data	Individual-level
Sampling weight	Primary adult female decision-maker
Disaggregation levels	Age category (18-29 years, 30+ years)
Treatment of missing data	Primary adult female decision-makers who do not have data for an A-WEAI indicator are excluded from the calculation for that indicator.
Survey variables used	hhea, hhnum, wgt_fpdm
Analytic variables used	feelinputdecagr, jown_count, credjanydec_any, incdec_count, groupmember_any, npoor_z105, youth_fdm_dj
Analytic variables created	n/a

**Step 3.1.** Calculate the sample-weighted percentage of all women who adequate in each A-WEAI indicator overall and by age group using the complex sampling design specified in Part 2, Step 3.1 and the variable list created in Part 2, Step 1.1. (Example code uses Stata syntax.)

**Note:** These results are reported in the midline indicator assessment template Table 6.3.3.

```
foreach var in $varlist_all:
    svy: tab `var'
    svy: tab `var' youth_fdm_dj, col
```

## 13.3 References

Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A., Seymour, G., & Vaz, A. (2012). *The women's empowerment in agriculture Index*. IFPRI discussion paper 01240. Available at:

<http://dlc.dlib.indiana.edu/dlc/bitstream/handle/10535/8724/ifpridp01240.pdf?sequence=1&isAllowed=y>

Heeringa, S., West, B., & Berglund, P. (2010). *Applied survey data analysis*. New York: Chapman and Hall/CRC. <https://doi.org/10.1201/9781420080674>

Hillesland, M. (2016). *Causal mapping of the gender integration framework*. Available at:

[https://agrilinks.org/sites/default/files/resource/files/CausalPathwaysGender\\_fullpaper\\_29Jan2016.pdf](https://agrilinks.org/sites/default/files/resource/files/CausalPathwaysGender_fullpaper_29Jan2016.pdf)

International Food Policy Research Institute. (2018). A-WEAI. Available at:

<https://weai.ifpri.info/versions/a-weai/>

Malapit, H., Kovarik, C., Sproule, K., Meinzen-Dick, R., & Quisumbing, A. (2015). *Instructional guide on the abbreviated women's empowerment in agriculture index (A-WEAI)*. Available at:

<http://www.ifpri.org/publication/instructional-guide-abbreviated-womens-empowerment-agriculture-index-weai>

Malapit, H., Pinkstaff, C., Sproule, K., Kovarik, C., Quisumbing, A., & Meinzen-Dick, R. (2017). *The Abbreviated Women's Empowerment in Agriculture Index (A-WEAI)*. IFPRI discussion paper 01647. Available at: <http://ebrary.ifpri.org/utis/getfile/collection/p15738coll2/id/131231/filename/131442.pdf>

Wolter, K. (2007). The jackknife method. In Wolter, K. (Ed.), *Introduction to variance estimation* (2nd ed., pp. 153-191). New York, NY: Springer. [https://doi.org/10.1007/978-0-387-35099-8\\_4](https://doi.org/10.1007/978-0-387-35099-8_4)

## 14. AGRICULTURAL PRODUCTIVITY INDICATOR

This chapter describes the percent of producers in the targeted area who have applied targeted improved management practices or technologies indicator. The indicator is calculated for selected prioritized agricultural value chain commodities (VCCs) identified by in-country Feed the Future teams. Questions about these prioritized VCCs are included in the agricultural technologies modules of the survey-customized ZOI Survey questionnaire.

This chapter has two sections; the first section describes the guidelines to construct the indicator, and the second section outlines the step-by-step procedures to calculate the indicator. Because the targeted VCCs on which data are collected vary from country to country, the step-by-step section provides examples of how to compute the indicator for specific VCCs. The improved management practices and technologies indicator step-by-step procedures include maize, fishponds, dairy cows, and sheep.

### 14.1 Guidelines to construct the improved management practices and technologies indicator

This section provides the guidelines to construct the improved management practices and technologies indicator. This indicator measures the percentage of producers, such as farmers, ranchers, pastoralists, and other primary sector producers of food and non-food crops, livestock products, fish and other fisheries and aquaculture products, agro-forestry products, and natural resource-based products, who applied targeted agriculture-related improved management practices or technologies in the Feed the Future ZOI in the year preceding the survey.

The management practices and technologies are grouped into 17 management practice and technology type categories. These are listed as follows, with examples of the types of practices and technologies included in each:

- Crop genetics, such as improved or certified seed that could be higher yielding; higher in nutritional content through bio-fortification, such as vitamin A-rich sweet potatoes or rice or high-protein maize; more resilient to climate impacts, such as drought-tolerant maize or stress -tolerant rice; or improved germplasm
- Cultural practices, such as seedling production and transplantation; cultivation practices such as planting density, crop rotation, and mounding; and mulching
- Livestock management, such as improved livestock breeds; livestock health services and products, such as vaccines; improved livestock handling practices and housing; improved feeding practices; and improved grazing practices, improved waste management practices, improved fodder crop, and cultivation of dual-purpose crops
- Wild-caught fisheries management, such as sustainable fishing practices; improved nets, hooks, lines, traps, dredges, and trawls; and improved hand-gathering, netting, angling, spearfishing, and trapping practices
- Aquaculture management, such as improved fingerlings, improved feed and feeding practices, fish health and disease control, improved cage culture, improved pond culture, pond preparation, sampling and harvesting, and management of carrying capacity

- Natural resource or ecosystem management, such as terracing, rock lines, and fire breaks; biodiversity conservation; strengthening of ecosystem services, including stream bank management or restoration or re/afforestation; and woodlot management
- Pest and disease management, such as integrated pest management; improved and environmentally sustainable use of insecticides and pesticides, and improved fungicides; appropriate application of fungicides, improved and environmentally sustainable use of cultural, physical, biological, and chemical insecticides and pesticides; crop rotation; and aflatoxin prevention and control
- Soil-related fertility and conservation, such as integrated soil fertility management; soil management practices that increase biotic activity and soil organic matter levels, such as soil amendments that increase fertilizer-use efficiency, such as soil organic matter and mulching; improved fertilizer; improved fertilizer use practices; inoculant; and erosion control
- Irrigation, such as drip, surface, and sprinkler irrigation and irrigation schemes
- Agriculture water management, non-irrigation-based, such as water harvesting, sustainable water use practices, and practices that improve water quality
- Climate mitigation includes technologies that minimize emission intensities relative to other alternatives (while preventing leakage of emissions elsewhere); examples include low- or no-till practices; restoration of organic soils and degraded lands; efficient nitrogen fertilizer use; practices that promote methane reduction; agroforestry; introduction/expansion of perennials; and practices that promote greater resource use efficiency (e.g., drip irrigation and upgrades of agriculture infrastructure and supply chains)
- Climate adaptation and climate risk management that includes technologies targeted with the explicit objective of reducing risk and minimizing the severity of the impacts of climate change; examples include drought- and flood-resistant varieties, short-duration varieties, adjustment of sowing time, agricultural and climate forecasting, early warning systems, diversification in the use of perennial varieties, agroforestry, and risk insurance
- Marketing and distribution, such as contract farming technologies and practices, improved input purchase technologies and practices, improved commodity sale technologies and practices, and improved market information system technologies and practices
- Post-harvest handling and storage, such as improved transportation, decay and insect control, temperature and humidity control, improved quality control technologies and practices, sorting and grading, and sanitary handling practices
- Value-added processing, such as improved packaging practices and materials, including biodegradable packaging; food and chemical safety technologies and practices; and improved preservation technologies and practices
- Food safety, such as technologies and practices targeted with the explicit objective of preventing and controlling biological, chemical and physical food safety hazards from production, processing and handling, to distribution, storage, and retail; examples include use of natural biocontrol agents (e.g., Aflasafe®) and Good Agricultural Practices; pasteurization, cold chain, and food preservation techniques (e.g., canning); proper handling practices (e.g., use of personal protective equipment, raw meat separation); moisture meters and hermetic storage; application of Hazard Analysis and Critical Control Points principles and other risk assessments, including digital traceability; and sanitary and phytosanitary certification.

- Other, such as improved mechanical and physical land preparation, non-market- and non-climate-related information technology, improved recordkeeping, and improved budgeting and financial management

The management practices and technologies counted under the indicator are only those being targeted by the Feed the Future program in the survey country. This means that the indicator can only be calculated if the Feed the Future team has defined the set of practices and technologies it targeted in the ZOI in the year preceding the survey and has ensured that questions related to those specific practices and technologies were included in the ZOI Survey questionnaire during the country customization process. If the Feed the Future team has not defined the set of targeted practices and technologies, do not compute this indicator during ZOI Survey data analysis.

Because it is common for Feed the Future programs to promote more than one improved management practice or technology to producers, the Feed the Future ZOI indicator allows tracking the percentage of producers who apply any targeted improved practice or technology in the ZOI. It also allows tracking the percentage of producers that apply targeted improved management practices or technologies in specific categories.

This indicator is designed to capture the application of targeted improved management practices and technologies by producers on their individual plots. This indicator is not intended to capture producers who are part of a group or members of an organization that apply targeted improved management practices or technologies on a demonstration plot or other common plot.

When calculating the indicator:

- Count each producer in the sample who applied a targeted improved practice or technology only once in the overall indicator and the applicable sex and age disaggregate categories, regardless of the number of targeted improved practices or technologies applied, to track the percentage of producers who applied Feed the Future-targeted practices or technologies. If more than one producer in a household applied targeted improved practices or technologies, count each producer in the household who did so.
- For the commodity disaggregate, count each producer once under each commodity for which they applied one or more Feed the Future-targeted practices or technologies. For example, if a producer used targeted improved seed for maize and beans, count that producer once under maize and once under beans.
- For the practice or technology type category disaggregate, count each producer once under each practice or technology type category for which they applied one or more Feed the Future-targeted practices or technologies. Producers can be counted under more than one practice and technology type category. For example:
  - If a producer applied more than one targeted improved management practice or technology in the year preceding the survey that are categorized under different categories, count the producer under each relevant category, for example, once under pest and diseases management and once under irrigation.
  - If Feed the Future is promoting a practice or technology for multiple benefits, the producers applying the practice or technology should be reported under each category

for which the practice or technology is being targeted. For example, a producer practicing mulching should be reported under cultural practices (weed control) and soil-related fertility and conservation (organic content), if the technology is being targeted for both purposes. If the practice is being targeted for only one purpose, the producer should be reported under only the relevant category. If Feed the Future is promoting a practice or technology for a single benefit even though it could be targeted for multiple benefits, be sure that producers applying the practice or technology are reported under only the one category for which the practice or technology is being targeted.

- Count a producer once, regardless of how many times they applied a targeted improved management practice or technology during the year preceding the survey. For example, Feed the Future is promoting irrigation and use of improved seed. A producer has access to irrigation and can now cultivate a second crop during the dry season, in addition to during the rainy season. Whether the producer applies Feed the Future-targeted improved seed during one season and not the other, or in both seasons, the producer should be counted only once in the crop genetics category and once under the irrigation category.
- Count a producer once per category, regardless of how many targeted improved management practices or technologies under that category the producer applied. For example, Feed the Future is promoting improved plant spacing and planting on ridges for a specific crop—two practices that fall under the cultural practices category. A producer who applies both practices should be counted only once under the cultural practices category.

## 14.2 Step-by-step procedure to calculate the targeted improved management practices and technologies indicator

This section describes the detailed step-by-step procedures to calculate the targeted improved management practices and technologies indicator. Because it is essential that the Percent of producers who have applied targeted improved management practices or technologies indicator only counts the application of improved management practices and technologies that are being targeted by the Feed the Future program in the specific country, this guide cannot provide standard instructions for computing the indicator that can be applied across ZOI Surveys. In this section, we provide an example of how to compute the indicator for a ZOI Survey with maize, fishpond aquaculture, dairy cows, and sheep as the priority value chains using illustrative sets of targeted improved management practices and technologies in **Table 17** (maize), **Table 18** (fishpond aquaculture), **Table 19** (dairy cows), and **Table 20** (sheep). These examples must be adapted to each ZOI Survey's specific context; they should never be applied as is unless the Feed the Future program is supporting the same value chains and targeting the same set of improved management practices and technologies for the same reasons (i.e., under the same categories). Throughout the step-by-step procedures, “category” refers to improved management practice and technology type category, and “targeted practice” refers to targeted improved management practice or technology.



### 14.2.1 Maize calculations

In this example, the steps to calculate the percentage of maize producers who applied targeted practices during the year preceding the ZOI Survey are presented. The step-by-step procedures follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax agtech\_maize.do*. Variables from Module 7.1 are used for the calculations.

#### Definitions

Numerator	Number of maize producers who applied any targeted improved management practices or technologies
Denominator	Number of individuals responsible for cultivating maize (i.e., maize producers) in the 12 months preceding the survey
Unit of measure	Percentage
Level of data	Individual (producer)
Sampling weight	Maize producer
Disaggregation levels	Sex* Age (15-29, 30+)* Improved management practice or technology type category*
Treatment of missing data	All missing and “don’t know” responses to targeted practice questions are considered to be “no.” They are included in the indicator denominator but not in the numerator.
Survey variables used	<code>hhea, hhnum, m1_line, strata, wgt_maize</code> , all variables from Module 7.1
Analytic variables used	<code>sex, age15_29y, hhmem_dj, vcc_maize</code>
Analytic variables created	<code>imp_maize_impseed, imp_maize_plantrow, imp_maize_orgfert, imp_maize_mulchsoil, imp_maize_terrace, imp_maize_mulchweed, imp_maize_pestche, imp_maize_irrigdrip, imp_maize_irrigpump, imp_maize_zerotill, imp_maize_agdealerseed, imp_maize_soldhusks, imp_maize_mechdry, imp_maize_hermetic, imp_maize_warehouse, imp_maize_mechshuck, imp_maize_mechtill, imp_maize_mechharvest, imp_maize_genetics, imp_maize_culture, imp_maize_ecosys, imp_maize_pest, imp_maize_fert, imp_maize_irrig, imp_maize_water, imp_maize_cmitigate, imp_maize_cadapt, imp_maize_markdist, imp_maize_pharvest, imp_maize_valadd, imp_maize_fsafety, imp_maize_other, imp_tot_maize, imp_cat_maize, imp_any_maize</code>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Review Module 7.1 of the country-customized ZOI Survey questionnaire and confirm which questions and response options capture the practices that Feed the Future is targeting in the ZOI. Also confirm under which category or categories the individual practices are being targeted.<sup>92</sup> A practice might fit under multiple categories but may only be targeted under one category in the particular ZOI context. If this information is not readily available (e.g., in a table or list provided to the Data Analyst by the Country Lead), the information must be confirmed by the Data Analyst before they can begin data analysis. Consultations with the Bureau for Resilience and Food Security (RFS) or in-country Feed the Future may be required.

**Table 17** shows the relevant variables and corresponding values used to categorize targeted practices for maize in the step-by-step procedures and the *FTF ZOI Survey [COUNTRY] [YEAR] syntax agtech\_maize* and *FTF ZOI Survey COUNTRY YEAR syntax agtech\_final* Stata syntax files. Note that these are only examples and **MUST** be adapted for the country context.

If Feed the Future is targeting a practice for multiple benefits, the producers applying the practice should be reported under **each category** for which the practice is being targeted. If the practice is being targeted for only one purpose, the producer should be reported under **only the relevant category**. If Feed the Future is targeting a practice for a single benefit even though it could be targeted for multiple benefits, be sure that producers applying the practice are reported under **only the one category** for which the practice is being targeted.

**Table 17: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Maize Cultivation, by Category**

Management practice and technology type category (maize)		Feed the Future-targeted practices and technologies for maize (illustrative)
1	Crop genetics	<b>Practice 1: Used improved maize seeds</b> ( <i>imp_maize_impseed</i> ): v7107 includes “B” (Improved open pollinated) or v7107 includes “C” (Hybrid)
2	Cultural practices	<b>Practice 2: Planted maize in rows</b> ( <i>imp_maize_plantrow</i> ): v7109=1 (Plant in rows) or v7109=3 (Some in rows and some randomly broadcast)
3	Natural resource or ecosystem management	<b>Practice 3: Used organic fertilizer</b> ( <i>imp_maize_orgfert</i> ): v7111b includes “A” (Soil based organic) or v7111b includes “C” (Foliar feeds organic) <b>Practice 4: Practiced mulching for soil health</b> ( <i>imp_maize_mulchsoil</i> ): v7121b=1 (Mulching for soil health) <b>Practice 5: Used terracing soil bands or trenches</b> ( <i>imp_maize_terrace</i> ): v7121a=1 (Terracing) or v7121c=1 (Soil bands/trenches)

<sup>92</sup> The targeted improved management practices and technologies targeted by the Feed the Future program should have been established during the questionnaire customization stage to ensure that the practices and technologies are adequately captured in the survey.

Management practice and technology type category (maize)		Feed the Future-targeted practices and technologies for maize (illustrative)
4	Pest and disease management	<p><b>Practice 6: Practiced mulching for weed control</b> (imp_maize_mulchweed): v7119 includes “C” (Weed control mulching)</p> <p><b>Practice 7: Used chemical pest control</b> (imp_maize_pestchem): v7115=1 (Chemical pest control) or v7119 includes “B” (Herbicide)</p>
5	Soil-related fertility and conservation	<p><b>Practice 4: Practiced mulching for soil health</b> (imp_maize_mulchsoil): v7121_b=1 (Mulching for soil health)</p> <p><b>Practice 5: Used terracing soil bands or trenches</b> (imp_maize_terrace): v7121a=1 (Terracing) or v7121c=1 (Soil bands/trenches)</p>
6	Irrigation	<p><b>Practice 8: Used drip irrigation</b> (imp_maize_irrigdrip): v7123 includes “A” (Drip irrigation)</p> <p><b>Practice 9: Used pump irrigation</b> (imp_maize_irrigpump): v7123 includes “D” (Pump system)</p>
7	Agriculture water management (non-irrigation)	<p><b>Practice 4: Practiced mulching for soil health</b> (imp_maize_mulchsoil): v7121_b=1 (Mulching for soil health)</p> <p><b>Practice 5: Used terracing soil bands or trenches</b> (imp_maize_terrace): v7121 includes “A” (Terracing) or v7121 includes “C” (Soil bands/trenches)</p>
8	Climate mitigation	<p><b>Practice 8: Used drip irrigation</b> (imp_maize_irrigdrip): v7123 includes “A” (Drip irrigation)</p> <p><b>Practice 10: Used zero-tillage land preparation</b> (imp_maize_zerotill): v7101 includes “B” (Zero tillage)</p>
9	Climate adaptation/climate risk management	<p><b>Practice 1: Used improved maize seeds</b> (imp_maize_impseed): v7107 includes “B” (Improved open pollinated) or v7107 includes “C” (Hybrid)</p> <p><b>Practice 9: Used pump irrigation</b> (imp_maize_irrigpump): v7123 includes “D” (Pump system)</p>
10	Marketing and distribution	<p><b>Practice 11: Bought seeds from an ag dealer</b> (imp_maize_agdealerseed): v7106=4 (Bought from ag dealer with cash) or v7106=5 (Bought from ag dealer with voucher)</p> <p><b>Practice 12: Sold maize stalks or husks</b> (imp_maize_soldhusks): v7124a includes “G” (Stalks harvested and sold to others) or v7124c includes “E” (Husks sold/traded with others as animal feed)</p>
11	Post-harvest handling and storage	<p><b>Practice 13: Used solar or mechanized dryers</b> (imp_maize_mechdry): v7126 includes “H” (Solar dryers) or v7126 includes “I” (Mechanized dryers)</p> <p><b>Practice 14: Used improved storage containers</b> (imp_maize_hermetic): v7129 includes “C” (Hermetic bag)</p>

Management practice and technology type category (maize)		Feed the Future-targeted practices and technologies for maize (illustrative)
		<b>Practice 15: Used improved storage location</b> (imp_maize_warehouse): v7130e=1 (Warehouses)
12	Value-added processing	<b>Practice 16: Shelled maize by machine</b> (imp_maize_mechshuck): v7127 includes "C" (Shelling by machine)
13	Food safety	<b>Practice 14: Used improved storage containers</b> (imp_maize_hermetic): v7129 includes "C" (Hermetic bag)
14	Other	<b>Practice 17: Prepared land with a motorized tiller or tractor</b> (imp_maize_mechtill): v7105 includes "C" (Land prep with motorized tiller) or v7105 includes "D" (Tractor)  <b>Practice 18: Used a machine to harvest maize</b> (imp_maize_mechharvest): v7124=2 (Harvested with a machine only) or v7124=3 (Some by hand, some with a machine)

**Step 2.** Prepare the data—that is, load the individual-level data file, drop records for producers who did not cultivate maize in the year preceding the survey, and drop variables not required to calculate the targeted practices indicator for maize. Also review the `vcc_maize` variable that should have already been created and included in the individual-level data file.

```
Load          "FTF ZOI Survey [Country] [Year] persons data
               analytic"
Keep record   if vcc_maize=1
Keep variables hhea hhnum m1_line strata sex age15_29y hhmem_dj
               wgt_maize vcc_maize and all variables that begin
               with 'v71'
```

**Step 3.** Create 18 binary variables—one for each targeted practice to flag producers who applied the practice to cultivate their maize crop during the 12 months preceding the survey (yes=1; no=0).

**Step 3.1.** Create a binary variable to flag maize producers who used improved maize seeds (`imp_maize_impseed`).

```
Set           imp_maize_impseed=0
Replace       imp_maize_impseed=1 if v7107 includes "B" or "C"
Label values  0 "No"
              1 "Yes"
Label variable "Used improved or hybrid maize seeds"
```

**Step 3.2.** Create a binary variable to flag maize producers who planted their maize in rows (`imp_maize_plantrow`).

```
Set           imp_maize_plantrow=0
Replace       imp_maize_plantrow=1 if v7109=1 or v7109=3
Label values  0 "No"
              1 "Yes"
Label variable "Planted maize in rows"
```

**Step 3.3.** Create a binary variable to flag maize producers who used organic fertilizer to cultivate their maize (`imp_maize_orgfert`).

```
Set                imp_maize_orgfert=0
Replace            imp_maize_orgfert=1 if v7111b includes "A" or "C"
Label values      0 "No"
                  1 "Yes"
Label variable     "Used organic fertilizer on maize"
```

**Step 3.4.** Create a binary variable to flag maize producers who used mulching to manage soil and water for their maize crop (`imp_maize_mulchsoil`).

```
Set                imp_maize_mulchsoil=0
Replace            imp_maize_mulchsoil=1 if v7121_b=1
Label values      0 "No"
                  1 "Yes"
Label variable     "Used mulching for soil and water management for
maize"
```

**Step 3.5.** Create a binary variable to flag maize producers who used terracing, soil bands, or trenches to cultivate their maize crop (`imp_maize_terrace`).

```
Set                imp_maize_terrace=0
Replace            imp_maize_terrace=1 if v7121a=1 or v7121c=1
Label values      0 "No"
                  1 "Yes"
Label variable     "Used terracing, soil bands, or trenches on maize
plots"
```

**Step 3.6.** Create a binary variable to flag maize producers who practiced mulching to control weeds on their maize plots (`imp_maize_mulchweed`).

```
Set                imp_maize_mulchweed=0
Replace            imp_maize_mulchweed=1 if v7119 includes "C"
Label values      0 "No"
                  1 "Yes"
Label variable     "Practiced mulching to control weeds on maize
plots"
```

**Step 3.7.** Create a binary variable to flag maize producers who used chemical pest control, including herbicides, to cultivate their maize crop (`imp_maize_pestchem`).

```
Set                imp_maize_pestchem=0
Replace            imp_maize_pestchem=1 if v7115=1 or v7119 includes
                  "B"
Label values      0 "No" 1 "Yes"
Label variable     "Used chemical pest control on maize"
```

**Step 3.8.** Create a binary variable to flag maize producers who used drip irrigation on their maize plots (`imp_maize_irrigdrip`).

```

Set          imp_maize_irrigdrip=0
Replace      imp_maize_irrigdrip=1 if v7123 includes "A"
Label values 0 "No"
              1 "Yes"
Label variable "Used drip irrigation on maize plots"

```

**Step 3.9.** Create a binary variable to flag maize producers who used pump irrigation on their maize plots (**imp\_maize\_irrigpump**).

```

Set          imp_maize_irrigpump=0
Replace      imp_maize_irrigpump=1 if v7123 includes "D"
Label values 0 "No"
              1 "Yes"
Label variable "Used pump irrigation on maize plots"

```

**Step 3.10.** Create a binary variable to flag maize producers who used zero tillage land preparation (**imp\_maize\_zerotill**).

```

Set          imp_maize_zerotill=0
Replace      imp_maize_zerotill=1 if v7101 includes "B"
Label values 0 "No"
              1 "Yes"
Label variable "Used zero tillage land preparation for maize plots"

```

**Step 3.11.** Create a binary variable to flag maize producers who bought seeds from an agrodealer with either cash or a voucher (**imp\_maize\_agdealerseed**).

```

Set          imp_maize_agdealerseed=0
Replace      imp_maize_agdealerseed=1 if v7106=4 or v7106=5
Label values 0 "No"
              1 "Yes"
Label variable "Bought maize seeds from an ag dealer"

```

**Step 3.12.** Create a binary variable to flag maize producers who sold maize stalks or sold or traded maize husks (**imp\_maize\_soldhusks**).

```

Set          imp_maize_soldhusks=0
Replace      imp_maize_soldhusks=1 if v7124a includes "G" or
              v7124c includes "E"
Label values 0 "No"
              1 "Yes"
Label variable "Sold maize stalks or sold/traded maize husks"

```

**Step 3.13.** Create a binary variable to flag maize producers who used solar or mechanized dryers to dry their maize (**imp\_maize\_mechdry**).

```

Set          imp_maize_mechdry=0
Replace      imp_maize_mechdry=1 if v7126 includes "H" or "I"
Label values 0 "No"
              1 "Yes"

```

Label variable "Used solar or mechanized dryers for maize"

**Step 3.14.** Create a binary variable to flag producers who used improved storage containers—hermetic bags (`imp_maize_hermetic`).

```
Set          imp_maize_hermetic=0
Replace      imp_maize_hermetic=1 if v7129 includes "C"
Label values 0 "No"
              1 "Yes"
Label variable "Used hermetic bags for maize"
```

**Step 3.15.** Create a binary variable to flag maize producers who used improved storage location—stored in a warehouse (`imp_maize_warehouse`).

```
Set          imp_maize_warehouse=0
Replace      imp_maize_warehouse=1 if v7130e=1
Label values 0 "No"
              1 "Yes"
Label variable "Stored maize in a warehouse"
```

**Step 3.16.** Create a binary variable to flag maize producers who shelled their maize crop by machine (`imp_maize_mechshuck`).

```
Set          imp_maize_mechshuck=0
Replace      imp_maize_mechshuck=1 if v7127 includes "C"
Label values 0 "No"
              1 "Yes"
Label variable "Shelled maize by machine"
```

**Step 3.17.** Create a binary variable to flag maize producers who prepared their maize plots using a motorized tiller or tractor (`imp_maize_mechtill`).

```
Set          imp_maize_mechtill=0
Replace      imp_maize_mechtill=1 if v7105 includes "C" or "D"
Label values 0 "No"
              1 "Yes"
Label variable "Prepared maize plots with motorized tiller or tractor"
```

**Step 3.18.** Create a binary variable to flag maize producers who used a machine to harvest their maize crop (`imp_maize_mechharvest`).

```
Set          imp_maize_mechharvest=0
Replace      imp_maize_mechharvest=1 if v7124=2 or v7124=3
Label values 0 "No"
              1 "Yes"
Label variable "Used a machine to harvest maize"
```

**Step 4.** Create a binary variable for each category to flag producers who applied any practices targeted by Feed the Future to cultivate maize under the relevant category (yes=1, no=0).

**Step 4.1.** Create a binary variable to flag maize producers who applied a targeted practice related to crop genetics (`imp_maize_genetics`).

```
Set                imp_maize_genetics=0
Replace            imp_maize_genetics=1 if imp_maize_impseed=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved crop genetics
practices"
```

**Step 4.2.** Create a binary variable to flag maize producers who applied a targeted practice related to cultural practices (`imp_maize_culture`).

```
Set                imp_maize_culture=0
Replace            imp_maize_culture=1 if imp_maize_plantrow=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved cultural practices"
```

**Step 4.3.** Create a binary variable to flag maize producers who applied a targeted practice related to natural resource or ecosystem management (`imp_maize_ecosys`).

```
Set                imp_maize_ecosys=0
Replace            imp_maize_ecosys=1 if imp_maize_orgfert=1 or
                  imp_maize_mulchsoil=1 or imp_maize_terrace=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved natural resources and
ecosystem management practices"
```

**Step 4.4.** Create a binary variable to flag maize producers who applied a targeted practice related to improved pest and disease management (`imp_maize_pest`).

```
Set                imp_maize_pest=0
Replace            imp_maize_pest=1 if imp_maize_mulchweed=1 or
                  imp_maize_pestchem=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved pest and disease
management practices"
```

**Step 4.5.** Create a binary variable to flag maize producers who applied a targeted practice related to soil-related fertility and conservation (`imp_maize_fert`).

```
Set                imp_maize_fert=0
Replace            imp_maize_fert=1 if imp_maize_mulchsoil=1 or
                  imp_maize_terrace=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved soil related fertility
and conservation practices"
```



**Step 4.6.** Create a binary variable to flag maize producers who applied a targeted practice related to irrigation (**imp\_maize\_irrig**).

```
Set          imp_maize_irrig=0
Replace      imp_maize_irrig=1 if imp_maize_irrigdrip=1 or
             imp_maize_irrigpump=1
Label values 0 "No"
             1 "Yes"
Label variable "Applied targeted improved irrigation practices"
```

**Step 4.7.** Create a binary variable to flag maize producers who applied a targeted practice related to water management (non-irrigation) (**imp\_maize\_water**).

```
Set          imp_maize_water=0
Replace      imp_maize_water=1 if imp_maize_mulchsoil=1 or
             imp_maize_terrace=1
Label values 0 "No"
             1 "Yes"
Label variable "Applied targeted improved water management (non-
irrigation) practices"
```

**Step 4.8.** Create a binary variable to flag maize producers who applied a targeted practice related to climate mitigation (**imp\_maize\_cmitigate**).

```
Set          imp_maize_cmitigate=0
Replace      imp_maize_cmitigate=1 if imp_maize_irrigdrip=1 or
             imp_maize_zerotill=1
Label values 0 "No"
             1 "Yes"
Label variable "Applied targeted improved climate mitigation
practices"
```

**Step 4.9.** Create a binary variable to flag maize producers who applied a targeted practice related to climate adaption (**imp\_maize\_cadapt**).

```
Set          imp_maize_cadapt=0
Replace      imp_maize_cadapt=1 if imp_maize_impseed=1 or
             imp_maize_irrigpump=1
Label values 0 "No"
             1 "Yes"
Label variable "Applied targeted improved climate adaption
practices"
```

**Step 4.10.** Create a binary variable to flag maize producers who applied a targeted practice related to marketing and distribution (**imp\_markdist**).

```
Set          imp_maize_markdist=0
Replace      imp_maize_markdist=1 if imp_maize_agdealerseed=1
             or imp_maize_soldhusks=1
Label values 0 "No"
             1 "Yes"
```

```
Label variable    "Applied targeted improved marketing and
                  distribution practices"
```

**Step 4.11.** Create a binary variable to flag maize producers who applied a targeted practice related to post-harvest handling and storage (`imp_maize_pharvest`).

```
Set              imp_maize_pharvest=0
Replace          imp_maize_pharvest=1 if imp_maize_mechdry=1 or
Label values     0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved post-harvest handling
                  and storage practices"
```

**Step 4.12.** Create a binary variable to flag maize producers who applied a targeted practice related to value-added processing (`imp_maize_valadd`).

```
Set              imp_maize_valadd=0
Replace          imp_maize_valadd=1 if imp_maize_mechshuck=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved value-added processing
                  practices"
```

**Step 4.13.** Create a binary variable to flag maize producers who applied a targeted practice related to food safety (`imp_maize_fsafety`).

```
Set              imp_maize_fsafety=0
Replace          imp_maize_fsafety=1 if imp_maize_hermetic=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Applied targeted improved food safety practices"
```

**Step 4.14.** Create a binary variable to flag maize producers who applied a targeted practice related to other targeted practices (`imp_maize_other`).

```
Set              imp_maize_other=0
Replace          imp_maize_other=1 if imp_maize_mechtill=1 or
Label values     0 "No"
                  1 "Yes"
Label variable    "Applied other targeted improved management
                  practices"
```

**Step 5.** Create a variable list (`IMP_maize`) that includes all targeted practices for maize. Adapt this step as needed to reflect the practices targeted in the specific ZOI country context.

```
Set variable list IMP_maize=imp_maize_impseed
                  imp_maize_plantrow imp_maize_orgfert
                  imp_maize_mulchsoil imp_maize_terrace
                  imp_maize_mulchweed imp_maize_pestchem
```

```

imp_maize_irrigdrip imp_maize_irrigpump
imp_maize_zerotill imp_maize_agdealerseed
imp_maize_soldhusk imp_maize_mechdry
imp_maize_hermetic imp_maize_warehouse
imp_maize_mechshuck imp_maize_mechtill
imp_maize_mechharvest

```

**Step 6.** Create a count variable to capture the total number of targeted practices each maize producer applied to cultivate maize (**imp\_tot\_maize**)—that is, create a variable that sums all the variables in the (**IMP\_maize**) variable list created in Step 5.

```

Set          imp_tot_maize=count of all IMP_maize variables=1
Label variable "Number of targeted improved practices and
               technologies applied (maize)"

```

**Step 7.** Create a categorical variable to categorize maize producers by the number of targeted practices they applied (**imp\_cat\_maize**) for Table 7.2.4. The categories used should be adapted based on the survey data; they do not have to be 0, 1-3, 4-6, 7-9, and 10 or more targeted practices.

```

Set          imp_cat_maize=missing
Replace      imp_cat_maize=0 if imp_tot_maize=0
Replace      imp_cat_maize=1 if imp_tot_maize≥1 and
               imp_tot_maize≤3
Replace      imp_cat_maize=2 if imp_tot_maize≥4 and
               imp_tot_maize≤6
Replace      imp_cat_maize_3 if imp_tot_maize≥7 and
               imp_tot_maize≤9
Replace      imp_cat_maize_4 if imp_tot_maize≥10 and
               imp_tot_maize does not equal missing
Label value  0 "None"
              1 "1-3 practices"
              2 "4-6 practices"
              3 "7-9 practices"
              4 "10+ practices"
Label variable "Number of targeted improved practices and
               technologies applied (maize, categorical)"

```

**Step 8.** Create a binary variable to indicate whether maize producers applied any targeted practices to cultivate maize (**imp\_any\_maize**).

```

Set          imp_any_maize=0
Replace      imp_any_maize=1 if imp_tot_maize>0 and
               imp_tot_maize≠missing
Label values 0 "No"
              1 "Yes"
Label variable "Applied any targeted improved practice or
               technology (maize)"

```

**Step 9.** After applying the maize producer sampling weight, calculate the percentage of maize producers who are de jure household members who applied at least one targeted practice to cultivate maize

during the 12 months preceding the survey using `imp_any_maize`. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

Also calculate the percentage of maize producers who are de jure household members who applied each targeted practice (i.e., for all variables created in Step 3) and the percentage of maize producers who are de jure household members who applied targeted practices by category (i.e., for all variables created in Step 4). (Sample code uses Stata syntax.)

```
svyset hhea [pw=wt_maize], strata(strata)
svy, subpop(hhmem_dj): tab imp_any_maize
svy, subpop(hhmem_dj): tab imp_any_maize age15_29y, col
svy, subpop(hhmem_dj): tab imp_any_maize sex, col
```

```
For each variable that starts with 'imp_maize' (var):
    svy, subpop(hhmem_dj): tab 'var'
```

**Step 10.** Also calculate the percentage distribution of maize producers who are de jure household members by the number of targeted practices they applied to cultivate maize during the 12 months preceding the survey using `imp_cat_maize`. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

```
svy, subpop(hhmem_dj): tab imp_cat_maize
svy, subpop(hhmem_dj): tab imp_cat_maize age15_29y, col
svy, subpop(hhmem_dj): tab imp_cat_maize sex, col
```

**Step 11.** Keep only the variables that are necessary to calculate the final overall indicator across all VCCs and save the data.

```
Keep          hhea hhnum m1_line strata sex age15_29y hhmem_dj
               wgt_maize vcc_maize imp_tot_maize imp_cat_maize
               imp_any_maize and all variables that begin with
               'imp_maize_'
Save          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_maize"
```

### 14.2.2 Fishpond calculations

In this example, the steps to calculate the percentage of fishpond producers who applied targeted practices during the year preceding the ZOI Survey are presented. The step-by-step procedures follow the Stata syntax `FTF ZOI Survey [COUNTRY] [YEAR] syntax agtech_fishpond.do`. Variables from Module 7.80 are used for the calculations.

#### Definitions

Numerator	Number of fishpond producers who applied any targeted improved management practices or technologies
Denominator	Number of individuals responsible for raising fish (i.e., fishpond producers) in the 12 months preceding the survey
Unit of measure	Percentage
Level of data	Individual (producer)
Sampling weight	Fishpond producer

Disaggregation levels	Sex* Age (15-29, 30+)* Improved management practice or technology type category*
Treatment of missing data	All missing and “don’t know” responses to targeted practice questions are considered to be “no.” They are included in the indicator denominator but not in the numerator.
Survey variables used	hhea, hhnum, m1_line, strata, all variables from Module 7.80
Analytic variables used	sex, age15_29y, hhmem_dj, wgt_fish, vcc_fish
Analytic variables created	imp_fish_ponddrain, imp_fish_pondmanure, imp_fish_certhatch, imp_fish_species, imp_fish_fedsupp, imp_fish_diseasecontrol, imp_fish_pondmonqual, imp_fish_pondmainqual, imp_fish_separate, imp_fish_partharv, imp_fish_records, imp_fish_useguts, imp_fish_aquam, imp_fish_other, imp_tot_fish, imp_cat_fish, imp_any_fish

\*Standard Feed the Future disaggregate

## Calculations

**Step 1. 1.** Review the Module 7.80 of the country-customized ZOI Survey questionnaire and confirm which questions and response options capture the practices that Feed the Future is targeting in the ZOI. Also confirm which category or categories the practices are being targeted under.<sup>93</sup> A practice might fit under multiple categories but may only be targeted under one category in the particular ZOI context. If this information is not readily available (e.g., in a table or list provided to the Data Analyst by the Country Lead), the information must be confirmed by the Data Analyst before they can begin data analysis. Consultations with RFS or in-country Feed the Future may be required.

**Table 18** shows the relevant variables and corresponding values used to categorize targeted practices for fishponds in the step-by-step procedures and the *FTF ZOI Survey [COUNTRY] [YEAR] syntax agtech\_fishpond.do* and *FTF ZOI Survey COUNTRY YEAR syntax agtech\_final.do* Stata syntax files. Note that these are only examples and **MUST** be adapted for the country context.

For fishpond aquaculture, most targeted practices will be included under one category—aquaculture management—but some may fall under other categories (e.g., the “other” category, which is meant to capture targeted practices that do not fit under any of the other category). Please see the *Feed the Future Indicator Handbook* or the *Guide to Feed the Future Statistics for P2-ZOI Midline Surveys* for the list of all categories, including examples.

If Feed the Future is targeting a practice for multiple benefits, the fishpond producers applying the practice should be reported under **each category** for which the practice is being targeted. If the practice is being targeted for only one purpose, the fishpond producer should be reported under **only the relevant category**. If Feed the Future is targeting a practice for a single benefit even though it could be targeted for multiple benefits, be sure that producers applying the practice are reported under **only the one category** for which the practice is being targeted.

<sup>93</sup> The practices targeted by the Feed the Future program should have been established during the questionnaire customization stage to ensure that the practices and technologies are adequately captured in the survey.

**Table 18: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Fishpond Aquaculture, by Category**

Management practice and technology type category		Feed the Future-targeted practices and technologies for fishponds (illustrative)
1	Aquaculture management	<p><b>Practice 1: Drained ponds at least once</b> (imp_fish_ponddrain): v78014&gt;0 (Number of times pond drained)</p> <p><b>Practice 2: Added animal manure to ponds</b> (imp_fish_pondmanure): v78017=1 (Added manure)</p> <p><b>Practice 3: Obtained fish from a registered/certified hatchery</b> (imp_fish_certhatch): v78002c=1 (Certified hatchery)</p> <p><b>Practice 4: Raised carp, tilapia, or catfish</b> (imp_fish_species): v78009 includes "A" (Carp) or "B" (Tilapia) or "C" (Catfish)</p> <p><b>Practice 5: Fed fish supplemental food to promote faster growth</b> (imp_fish_fedsupp): v78004=1 (Fish supplement)</p> <p><b>Practice 6: Controlled disease or parasites with salt or formalin</b> (imp_fish_diseasecontrol): v78010b includes "A" (Salt for disease) or "B" (Formalin for disease) or v78011b includes "C" (Formalin for parasites) or "B" (Salt for parasite)</p> <p><b>Practice 7: Monitored pond water quality</b> (imp_fish_pondmonqual): v78012=1 (Monitored water quality)</p> <p><b>Practice 8: Took steps to maintain good pond water quality</b> (imp_fish_pondmainqual): v78013=1 (Maintained water quality)</p> <p><b>Practice 9: Improved fish produced using sex or age separation</b> (imp_fish_separate): v78016a=1 (Sex separation) or v78016b=1 (Age separation)</p> <p><b>Practice 10: Harvested fish using partial harvests</b> (imp_fish_partharv): v78018a=2 (Conducted partial harvests)</p>
2	Other	<p><b>Practice 11: Kept regular records on fish</b> (imp_fish_records): v78024=1 (Kept records)</p> <p><b>Practice 12: Used or sold fish guts, skin, or scales after harvest</b> (imp_fish_useguts): v78023a includes "A" (Used guts as compost) or "B" (Used guts as animal feed) or v78023b includes "A" (Used scales/skin as compost) or "B" (Used scales/skin as animal feed) or "C" (Sold scales/skin)</p>

**Step 2.** Prepare the data—that is, load the individual-level data file, drop records for producers who did not raise fish in fishponds in the year preceding the survey, and drop variables not required to calculate the targeted practices indicator for fishponds. Also review the `vcc_fish` variable that should have already been created and included in the individual-level data file.

```

Load          "FTF ZOI Survey [Country] [Year] persons data
               analytic"
Keep record   if vcc_fish=1

```

```
Keep variables   hhea hhnum m1_line strata sex age15_29y hhmem_dj
                  wgt_fish vcc_fish, and all variables that begin
                  with 'v780'
```

**Step 3.** Create 12 binary variables—one for each targeted practice to flag producers who applied the practice to raise fish in ponds during the 12 months preceding the survey (yes=1, no=0).

**Step 3.1.** Create a binary variable to flag fishpond producers who drained their fishponds at least once (*imp\_fish\_ponddrain*).

```
Set              imp_fish_ponddrain=0
Replace          imp_fish_ponddrain=1 if v78014>0 and
                  v78014≠missing
Label value      0 "No"
                  1 "Yes"
Label variable    "Drained fishponds 1+ times"
```

**Step 3.2.** Create a binary variable to flag fishpond producers who added manure to their fishponds (*imp\_fish\_pondmanure*).

```
Set              imp_fish_pondmanure=0
Replace          imp_fish_pondmanure=1 if v78017=1
Label value      0 "No"
                  1 "Yes"
Label variable    "Added manure to fishponds"
```

**Step 3.3.** Create a binary variable to flag fishpond producers who obtained their fish from a registered/certified hatchery (*imp\_fish\_certhatch*).

```
Set              imp_fish_certhatch=0
Replace          imp_fish_certhatch=1 if v78002c=1
Label value      0 "No"
                  1 "Yes"
Label variable    "Obtained fish from a registered/certified
                  hatchery"
```

**Step 3.4.** Create a binary variable to flag fishpond producers who raised carp, tilapia, or catfish (*imp\_fish\_species*).

```
Set              imp_fish_species=0
Replace          imp_fish_species=1 if v78009 includes "A" or "B"
                  or "C"
Label value      0 "No"
                  1 "Yes"
Label variable    "Raised carp, tilapia, or catfish"
```

**Step 3.5.** Create a binary variable to flag fishpond producers who fed their fish supplemental food to promote faster growth (*imp\_fish\_fedsupp*).

```
Set              imp_fish_fedsupp=0
Replace          imp_fish_fedsupp=1 if v78004=1
```

```

Label value      0 "No"
                  1 "Yes"
Label variable   "Fed fish supplemental food"

```

**Step 3.6.** Create a binary variable to flag fishpond producers who controlled disease or parasites with salt or formalin (`imp_fish_diseasecontrol`).

```

Set              imp_fish_diseasecontrol=0
Replace          imp_fish_diseasecontrol=1 if (v78010b includes
Label value      0 "No"
                  1 "Yes"
Label variable   "Controlled disease or parasites with salt or
                  formalin"

```

**Step 3.7.** Create a binary variable to flag fishpond producers who monitored water quality in their fishponds (`imp_fish_pondmonqual`).

```

Set              imp_fish_pondmonqual=0
Replace          imp_fish_pondmonqual=1 if v78012=1
Label value      0 "No"
                  1 "Yes"
Label variable   "Monitored pond water quality"

```

**Step 3.8.** Create a binary variable to flag fishpond producers who took steps to maintain good water quality in their fishponds (`imp_fish_pondmainqual`).

```

Set              imp_fish_pondmainqual=0
Replace          imp_fish_pondmainqual=1 if v78013=1
Label value      0 "No"
                  1 "Yes"
Label variable   "Took steps to maintain good pond water quality"

```

**Step 3.9.** Create a binary variable to flag fishpond producers who improved their fish produced using sex or age separation (`imp_fish_separate`).

```

Set              imp_fish_separate=0
Replace          imp_fish_separate=1 if v78016a=1 or v78016b=1
Label value      0 "No"
                  1 "Yes"
Label variable   "Improved fish produced using sex or age
                  separation"

```

**Step 3.10.** Create a binary variable to flag fishpond producers who harvested their fish using partial harvests (`imp_fish_partharv`).

```

Set              imp_fish_partharv=0
Replace          imp_fish_partharv=1 if v78018a=2
Label value      0 "No"
                  1 "Yes"
Label variable   "Harvested fish using partial harvests"

```



**Step 3.1.1.** Create a binary variable to flag fishpond producers who kept regular records on their fish (`imp_fish_records`).

```
Set          imp_fish_records=0
Replace      imp_fish_records=1 if v78024=1
Label value  0 "No"
             1 "Yes"
Label variable "Kept regular records on fish"
```

**Step 3.1.2.** Create a binary variable to flag fishpond producers who used or sold fish guts, skin, or scales after harvest (`imp_fish_useguts`).

```
Set          imp_fish_useguts=0
Replace      imp_fish_useguts=1 if (v78023a includes "A" or
Label value  0 "No"
             1 "Yes"
Label variable "Used or sold fish guts, skin, or scales after
harvest"
```

**Step 4.** Create a binary variable for each category to flag producers who applied any practices targeted by Feed the Future to cultivate fishponds under the relevant category (yes=1, no=0). Almost all targeted aquaculture practices are categorized under a single category: aquaculture management.<sup>94</sup>

**Step 4.1.** Create a binary variable to flag fishpond producers who applied a targeted practice related to pond aquaculture management (`imp_fish_aquam`). **Note:** If all practices were targeted under aquaculture management, this variable will be duplicative of the one created in Step 6 (`imp_any_fish`).

```
Set          imp_fish_aquam=0
Replace      imp_fish_aquam=1 if imp_fish_ponddrain=1 or
Label values 0 "No"
             1 "Yes"
Label variable "Applied improved aquaculture management
practices to raise fish"
```

**Step 4.2.** Create a binary variable to flag fishpond producers who applied a targeted practice that is not captured under aquaculture management or another category (`imp_fish_other`).

---

<sup>94</sup> Aquaculture producers could also apply improved management practices or technologies that are targeted under other management practice and technology type categories; for example, using a Feed-the-Future-supported mobile phone application to get on-demand production and marketing information and advice might be categorized under the "Other" category if it is determined not to be targeted under another category.

```

Set          imp_fish_other=0
Replace      imp_fish_other=1 if imp_fish_records=1 or
             imp_fish_useguts=1
Label values 0 "No"
             1 "Yes"
Label variable "Applied other improved practices or technologies
              to raise fish"

```

**Step 5.** Create a variable list (**IMP\_fish**) that includes all applicable targeted practices for fishponds. Adapt this step as needed to reflect the practices targeted in the specific ZOI country context.

```

Set variable list IMP_fish=imp_fish_ponddrain
                  imp_fish_pondmanure imp_fish_certhatch
                  imp_fish_species imp_fish_fedsupp
                  imp_fish_diseasecontrol imp_fish_pondmonqual
                  imp_fish_pondmainqual imp_fish_separate
                  imp_fish_partharv imp_fish_records
                  imp_fish_useguts

```

**Step 6.** Create a count variable to capture the total number of targeted practices each fishpond producer applied to raise fish in fishponds (**imp\_tot\_fish**)—that is, create a variable that sums all the variables in the (**IMP\_fish**) variable list created in Step 5.

```

Set          imp_tot_fish=count of all IMP_fish variables=1
Label variable "Number of targeted improved practices and
              technologies applied (fishponds)"

```

**Step 7.** Create a categorical variable to categorize fishpond producers by the number of targeted practices they applied (**imp\_cat\_fish**) for Table 7.3.4. The categories used should be adapted based on the survey data; they do not have to be 0, 1-3, 4-6, 7-9, and 10 or more targeted practices.

```

Set          imp_cat_fish=missing
Replace      imp_cat_fish=0 if imp_tot_fish=0
Replace      imp_cat_fish=1 if imp_tot_fish≥1 and
             imp_tot_fish≤3
Replace      imp_cat_fish=2 if imp_tot_fish≥4 and
             imp_tot_fish≤6
Replace      imp_cat_fish_3 if imp_tot_fish≥7 and
             imp_tot_fish≤9
Replace      imp_cat_fish_4 if imp_tot_fish≥10 and
             imp_tot_fish≠missing
Label value  0 "None"
             1 "1-3 practices"
             2 "4-6 practices"
             3 "7-9 practices"
             4 "10+ practices"
Label variable "Number of targeted improved practices and
              technologies applied (fish, categorical)"

```

**Step 8.** Create a binary variable to indicate whether each fishpond producer applied any targeted practices to raise fish in ponds (**imp\_any\_fish**).

```

Set          imp_any_fish=0
Replace      imp_any_fish=1 if imp_tot_fish>0 and
            imp_tot_fish#missing
Label values 0 "No"
            1 "Yes"
Label variable "Applied any targeted improved practice or
            technology (fishponds)"

```

**Step 9.** After applying the fishpond producer sampling weight, calculate the percentage of fishpond producers who are de jure household members who applied at least one targeted practice to raise fish in ponds during the year preceding the survey using the `imp_any_fish` analytic variable. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

Also calculate the percentage of fishpond producers who are de jure household members who applied each targeted practice (i.e., for all variables created in Step 3) and the percentage of fishpond producers who are de jure household members who applied targeted practices by category (i.e., for all variable created in Step 4). (Example code uses Stata syntax.)

```

svyset hhea [pw=wgt_fish], strata(strata)
svy, subpop(hhmem_dj): tab imp_any_fish
svy, subpop(hhmem_dj): tab imp_any_fish age15_29y, col
svy, subpop(hhmem_dj): tab imp_any_fish sex, col

For each variable that starts with 'imp_fish_' (var):
    svy, subpop(hhmem_dj): tab 'var'

```

**Step 10.** Calculate the percent distribution of fishpond producers who are de jure household members by the number of targeted practices they applied to raise fish in ponds during the 12 months preceding the survey using `imp_cat_fish`. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

```

svy, subpop(hhmem_dj): tab imp_cat_fish
svy, subpop(hhmem_dj): tab imp_cat_fish age15_29y, col
svy, subpop(hhmem_dj): tab imp_cat_fish sex, col

```

**Step 11.** Keep only the variables that are necessary to calculate the final overall indicator across all VCCs and save the data.

```

Keep          hhea hhnum m1_line wgt_fish strata vcc_fish
            age15_29y sex imp_tot_fish imp_cat_fish
            imp_any_fish and all variables that begin with
            'imp_fish_'
Save          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_fishpond"

```

### 14.2.3 Dairy cow calculations

In this example, the steps to calculate the percentage of dairy cow producers who applied targeted practices during the year preceding the ZOI Survey are presented. The step-by-step procedures follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax agtech\_dairycow.do*. Variables from Module 7.50 are used for the calculations.

## Definitions

Numerator	Number of dairy cow producers who applied any targeted improved management practices or technologies
Denominator	Number of individuals responsible for raising dairy cows (i.e., dairy cow producers) during the 12 months preceding the survey
Unit of measure	Percentage
Level of data	Individual (producer)
Sampling weight	Dairy cow producer
Disaggregation levels	Sex* Age (15-29, 30+)* Improved management practice or technology type category*
Treatment of missing data	All missing and “don’t know” responses to targeted practice questions are considered to be “no.” They are included in the indicator denominator but not in the numerator.
Survey variables used	<code>hhea, hhnum, m1_line, strata</code> , all variables from Module 7.50
Analytic variables used	<code>sex, age15_29y, hhmem_dj, wgt_dairy, vcc_dairy</code>
Analytic variables created	<code>imp_dairy_artinsem, imp_dairy_selectbreed, imp_dairy_healthserv, imp_dairy_medicine, imp_dairy_vaccinated, imp_dairy_prevmast, imp_dairy_roof, imp_dairy_fedbyprod, imp_dairy_fedconcentrate, imp_dairy_fedvitmin, imp_dairy_collmanure, imp_dairy_fedfodder, imp_dairy_pasteurized, imp_dairy_soldmilk, imp_dairy_soldmanure, imp_dairy_records, imp_dairy_livestm, imp_dairy_markdist, imp_dairy_valadd, imp_dairy_fsafety, imp_dairy_other, imp_tot_dairy, imp_cat_dairy, imp_any_dairy</code>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Review Module 7.50 of the country-customized ZOI Survey questionnaire and confirm which questions and response options capture the practices that Feed the Future is targeting in the ZOI. Also confirm under which category or categories the individual practices are being targeted.<sup>95</sup> A practice might fit under multiple categories but may only be targeted under one category in the particular ZOI context. If this information is not readily available (e.g., in a table or list provided to the Data Analyst by the Country Lead), the information must be confirmed by the Data Analyst before they can begin data analysis. Consultations with RFS or in-country Feed the Future may be required.

**Table 19** shows the relevant variables and corresponding values used to categorize targeted practices for dairy cows in the step-by-step procedures and the *FTF ZOI Survey [COUNTRY] [YEAR] syntax agtech\_dairycow.do* and *FTF ZOI Survey COUNTRY YEAR syntax agtech\_final.do* Stata syntax files. Note that these are only examples and **MUST** be adapted for the country context.

<sup>95</sup> The targeted improved management practices and technologies targeted by the Feed the Future program should have been established during the questionnaire customization stage to ensure that the practices and technologies are adequately captured in the survey.

For dairy cows, most targeted practices will be included under one category—livestock management—but some may fall under other categories (e.g., food safety or “other,” which is meant to capture targeted practices that do not fit under any of the other categories). Please see the *Feed the Future Indicator Handbook* or the *Guide to Feed the Future Statistics for P2-ZOI Midline Surveys* for the list of all categories, including examples.

If Feed the Future is promoting a practice for multiple benefits, the dairy producers applying the targeted practice should be reported under **each category** for which the practice is being targeted. If Feed the Future is targeting a practice for multiple benefits, the dairy producers applying the practice should be reported under **each category** for which the practice is being targeted. If the practice is being targeted for only one purpose, the dairy producer should be reported under **only the relevant category**. If Feed the Future is promoting a practice for a single benefit even though it could be targeted for multiple benefits, be sure that producers applying the practice are reported under **only the one category** for which the practice is being targeted.

**Table 19: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Raising Dairy Cows, by Category**

Management practice or technology type category (livestock)		Feed the Future-targeted practices and technologies for dairy cows (illustrative)
I	Livestock management	<p><b>Practice 1: Bred cows with artificial insemination</b> (imp_dairy_artinsem): v75008 includes “B” (Artificial insemination)</p> <p><b>Practice 2: Selectively chose bulls for breeding</b> (imp_dairy_selectbreed): v75011=3 (Bull has good body size, composition), 4 (Bull is son of a high-producing cow), or 5 (Bull is known to have good fertility)</p> <p><b>Practice 3: Used trained health service providers</b> (imp_dairy_healthserv): v75028=1 (Yes, obtained health services from trained provider)</p> <p><b>Practice 4: Gave dairy cows medicine</b> (imp_dairy_medicine): v75030=1 (Yes, gave medicines)</p> <p><b>Practice 5: Vaccinated some or all cattle</b> (imp_dairy_vaccinated): v75033=2 (Some cattle vaccinated) or 3 (All cattle vaccinated)</p> <p><b>Practice 6: Prevented mastitis using udder wash, teat dip, somatic cell counts</b> (imp_dairy_prevmast): v75034b includes “A” (Wash udder to prevent mastitis), “B” (Teat dip to prevent mastitis), or “C” (Somatic cell counts to prevent mastitis)</p> <p><b>Practice 7: Used housing with a roof for dairy cows</b> (imp_dairy_roof): v75016=4 (Roof only, no sides), 5 (Roof and sides, dirt floor), or 6 (Roof and sides, concrete floor)</p> <p><b>Practice 8: Fed dairy cows crop by-products daily/weekly</b> (imp_dairy_fedbyprod): v75021a=1 (Fed crop by-products daily) or 2 (Fed crop by-products weekly)</p> <p><b>Practice 9: Fed dairy cows mixed concentrates</b></p>

Management practice or technology type category (livestock)		Feed the Future-targeted practices and technologies for dairy cows (illustrative)
		<p><b>daily/weekly (imp_dairy_fedconcentrate): v75024a=1</b> (Fed mixed concentrates daily) or 2 (Fed mixed concentrates weekly)</p> <p><b>Practice 10: Fed dairy cows vitamins or minerals daily/weekly (imp_dairy_fedvitmin): v75026a=1</b> (Fed vitamins or minerals daily) or 2 (Fed vitamins or minerals weekly)</p> <p><b>Practice 11: Collected dairy cows' manure and put in a designated, covered area (imp_dairy_collmanure): v75027b=2</b> (Put in heap in covered area), 3 (Put in pit/lagoon), 4 (Put into a tank), or 5 (Put into a biogas-producing digester)</p> <p><b>Practice 12: Fed dairy cows improved fodder crops<sup>a</sup> (imp_dairy_fedfodder): v75019</b> includes "B" (Conserved rice straw), "C" (Conserved maize stover), "D" (Legume haulms or stovers), "E" (Forage legumes), "F" (Napier grass), "G" (Guinea grass), "H" (Cut fresh grass), or "I" (Tree fodder)</p>
2	Marketing and distribution	<p><b>Practice 13: Sold dairy cows' milk (imp_dairy_soldmilk): v75036=1</b> (Yes, sold milk)</p> <p><b>Practice 14: Sold dairy cows' manure (imp_dairy_soldmanure): v75027c=05</b> (Sell to friends/neighbors) or 06 (Sell at market)</p>
3	Value-added processing	<b>Practice 15: Pasteurized dairy cow's milk (imp_dairy_pasteurized): v75035a</b> includes "B" (Yes, pasteurize)
4	Food safety	<b>Practice 15: Pasteurized dairy cow's milk (imp_dairy_pasteurized): v75035a</b> includes "B" (Yes, pasteurize)
5	Other	<b>Practice 16: Kept written records on dairy cows (imp_dairy_records): v75045=1</b> (Yes, kept written records)

<sup>a</sup> Also manually check response option X (Other) to determine whether any manually entered forages should be considered targeted improved fodder crops and included in the calculations.

**Step 2.** Prepare the data—that is, load the individual-level data file, drop records for producers who did not raise dairy cows in the year preceding the survey, and drop variables not required to calculate the targeted practices indicator for dairy cows. Also review the `vcc_dairy` variable that should have already been created and included in the individual-level data file.

```

Load          "FTF ZOI Survey [Country] [Year] persons data
              analytic"
Keep record   if vcc_dairy=1
Keep          hhea hnum m1_line strata sex age15-19y hhmem_dj
              wgt_dairy vcc_dairy and all variables that begin
              with `v750'
```

**Step 3.** Create 16 binary variables—one for each targeted practice to flag producers who applied the practice to raise dairy cows during the 12 months preceding the survey (yes=1, no=0).

**Step 3.1.** Create a binary variable to flag dairy cow producers who bred their dairy cows with artificial insemination (`imp_dairy_artinsem`).

```
Set          imp_dairy_artinsem=0
Replace      imp_dairy_artinsem=1 if v75008 includes "B"
Label values 0 "No"
              1 "Yes"
Label variable "Used artificial insemination to breed dairy cows"
```

**Step 3.2.** Create a binary variable to flag dairy cow producers who selectively chose bulls for breeding—that is, chose a bull because he had good body size or composition, was the son of a high-producing cow, or was known to have good fertility (`imp_dairy_selectbreed`).

```
Set          imp_dairy_selectbreed=0
Replace      imp_dairy_selectbreed=1 if v75011=3 or v75011=4
              or v75011=5
Label values 0 "No"
              1 "Yes"
Label variable "Selectively chose bulls for breeding dairy cows"
```

**Step 3.3.** Create a binary variable to flag dairy cow producers who used trained health service providers for their dairy cows (`imp_dairy_healthserv`).

```
Set          imp_dairy_healthserv=0
Replace      imp_dairy_healthserv=1 if v75028=1
Label values 0 "No"
              1 "Yes"
Label variable "Obtained health services for dairy cows from a trained provider"
```

**Step 3.4.** Create a binary variable to flag dairy cow producers who gave dairy cows medicine (`imp_dairy_medicine`).

```
Set          imp_dairy_medicine=0
Replace      imp_dairy_medicine=1 if v75030=1
Label values 0 "No"
              1 "Yes"
Label variable "Gave dairy cows medicine"
```

**Step 3.5.** Create a binary variable to flag dairy cow producers who vaccinated some or all of their cattle (`imp_dairy_vaccinated`).

```
Set          imp_dairy_vaccinated=0
Replace      imp_dairy_vaccinated=1 if v75033=2 or v75033=3
Label values 0 "No"
              1 "Yes"
Label variable "Vaccinated some or all cattle"
```

**Step 3.6.** Create a binary variable to flag dairy cow producers who prevented mastitis in their dairy cows using udder wash, teat dip, or somatic cell counts (`imp_dairy_prevmast`).

```

Set          imp_dairy_prevmast=0
Replace      imp_dairy_prevmast=1 if v75034b includes "A,"
              "B," or "C"
Label values 0 "No"
              1 "Yes"
Label variable "Prevented mastitis in dairy cows (udder wash,
              teat dip, somatic cell counts)"

```

**Step 3.7.** Create a binary variable to flag dairy cow producers who used housing with a roof for their dairy cows (**imp\_dairy\_roof**).

```

Set          imp_dairy_roof=0
Replace      imp_dairy_roof=1 if v75016=4 or v75016=5 or
              v75016=6
Label values 0 "No"
              1 "Yes"
Label variable "Used housing with a roof for dairy cows"

```

**Step 3.8.** Create a binary variable to flag dairy cow producers who fed their dairy cows crop by-products daily or weekly (**imp\_dairy\_fedbyprod**).

```

Set          imp_dairy_fedbyprod=0
Relace       imp_dairy_fedbyprod if v75021a=1 or v75021a=2
Label values 0 "No"
              1 "Yes"
Label variable "Fed dairy cows crop by-products daily or weekly"

```

**Step 3.9.** Create a binary variable to flag dairy cow producers who fed their dairy cows mixed concentrates daily or weekly (**imp\_dairy\_fedconcentrate**).

```

Set          imp_dairy_fedconcentrate=0
Replace      imp_dairy_fedconcentrate=1 if v75024a=1 or
              v75024a=2
Label values 0 "No"
              1 "Yes"
Label variable "Fed dairy cows mixed concentrates daily or
              weekly"

```

**Step 3.10.** Create a binary variable to flag dairy cow producers who fed their dairy cows vitamins or minerals daily or weekly (**imp\_dairy\_fedvitmin**).

```

Set          imp_dairy_fedvitmin=0
Replace      imp_dairy_fedvitmin=1 if v75026a=1 or v75026a=2
Label values 0 "No"
              1 "Yes"
Label variable "Fed dairy cows vitamins or minerals daily or
              weekly"

```

**Step 3.11.** Create a binary variable to flag dairy cow producers who collected their dairy cows' manure and put it in a designated covered area, pit or lagoon, tank, or biogas-producing digester (**imp\_dairy\_collmanure**).



```

Set          imp_dairy_collmanure=0
Replace      imp_dairy_collmanure=1 if v75027b=2 or v75027b=3
              or v75027b=4 or v75027b=5
Label values 0 "No"
              1 "Yes"
Label variable "Collected dairy cows' manure and put in a
               designated covered area"

```

**Step 3.12.** Create a binary variable to flag dairy cow producers who fed dairy cows improved fodder crops—that is, conserved rice straw; conserved maize stover; legume haulms or stovers; forage legumes; napier, guinea, or fresh cut grass; or tree fodder (**imp\_dairy\_fedfodder**).

```

Set          imp_dairy_fedfodder=0
Replace      imp_dairy_fedfodder=1 if v75019 includes "B,"
              "C," "D," "E," "F," "G," "H," or "I"
Label values 0 "No"
              1 "Yes"
Label variable "Fed dairy cows improved fodder"

```

**Step 3.13.** Create a binary variable to flag dairy cow producers who pasteurized their dairy cows' milk (**imp\_dairy\_pasteurized**).

```

Set          imp_dairy_pasteurized=0
Replace      imp_dairy_pasteurized=1 if v75035a includes B
Label values 0 "No"
              1 "Yes"
Label variable "Pasteurized milk from dairy cows"

```

**Step 3.14.** Create a binary variable to flag dairy cow producers who sold their dairy cows' milk (**imp\_dairy\_soldmilk**).

```

Set          imp_dairy_soldmilk=0
Replace      imp_dairy_soldmilk=1 if v75036=1
Label values 0 "No"
              1 "Yes"
Label variable "Sold milk from dairy cows"

```

**Step 3.15.** Create a binary variable to flag dairy cow producers who sold their dairy cows' manure to their friends or neighbors or at a market (**imp\_dairy\_soldmanure**).

```

Set          imp_dairy_soldmanure=0
Replace      imp_dairy_soldmanure=1 if v75027c=05 or
              v75027c=06
Label values 0 "No"
              1 "Yes"
Label variable "Sold dairy cows' manure"

```

**Step 3.16.** Create a binary variable to flag dairy cow producers who kept written records on their dairy cows (**imp\_dairy\_records**).

```

Set          imp_dairy_records=0

```

```

Replace      imp_dairy_records=1 if v75045=1
Label values 0 "No"
              1 "Yes"
Label variable "Kept written records on dairy cows"

```

**Step 4.** Create a binary variable for each category to flag producers who applied any practices targeted by Feed the Future to raise dairy cows under the relevant category (yes=1, no=0). Almost all targeted practices for raising dairy cows are categorized under a single category: livestock management.<sup>96</sup>

**Step 4.1.** Create a binary variable to flag dairy cow producers who applied a targeted practice related to livestock management (**imp\_dairy\_livestm**). **Note:** If all practices were targeted under livestock management, this variable will be duplicative of the one created in Step 7 (**imp\_any\_dairy**).

```

Set          imp_dairy_livestm=0
Replace      imp_dairy_livestm=1 if imp_dairy_artinsem=1 or
              imp_dairy_selectbreed=1 or imp_dairy_healthserv=1
              or imp_dairy_medicine=1 or imp_dairy_vaccinated=1
              or imp_dairy_prevmast=1 or imp_dairy_roof=1 or
              imp_dairy_fedbyprod=1 or
              imp_dairy_fedconcentrate=1 or
              imp_dairy_fedvitmin=1 or imp_dairy_collmanure=1
              or imp_dairy_fedfodder=1
Label values 0 "No"
              1 "Yes"
Label variable "Applied improved livestock management"

```

**Step 4.2.** Create a binary variable to flag dairy cow producers who applied a targeted practice related to marketing and distribution (**imp\_dairy\_markdist**).

```

Set          imp_dairy_markdist =0
Replace      imp_dairy_markdist=1 if imp_dairy_soldmilk=1 or
              imp_dairy_soldmanure=1
Label values 0 "No"
              1 "Yes"
Label variable "Applied improved marketing and distribution"

```

**Step 4.3.** Create a binary variable to flag dairy cow producers who applied a targeted practice related to value-added processing (**imp\_dairy\_valadd**).

```

Set          imp_dairy_valadd=0
Replace      imp_dairy_valadd=1 if imp_dairy_pasteurized=1
Label values 0 "No"
              1 "Yes"
Label variable "Applied improved marketing and distribution"

```

---

<sup>96</sup> Dairy cow producers could also apply practices or technologies in the “Other” management practice and technology type category; for example, by using a Feed-the-Future-supported mobile phone app to get on-demand production and marketing information and advice.

**Step 4.4.** Create a binary variable to flag dairy cow producers who applied a targeted practice related to food safety (`imp_dairy_fsafety`).

```
Set                imp_dairy_fsafety=0
Replace            imp_dairy_fsafety=1 if imp_dairy_pasteurized=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Applied improved food safety practices"
```

**Step 4.5.** Create a binary variable to flag dairy cow producers who applied a targeted practice not captured under any other category (`imp_dairy_other`).

```
Set                imp_dairy_other=0
Replace            imp_dairy_other=1 if imp_dairy_records=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Applied other improved practice or technology"
```

**Step 5.** Create a variable list (`IMP_dairy`) that includes all targeted practices for dairy cows. Adapt this step as needed to reflect the practices targeted in the specific ZOI country context.

```
Set variable list IMP_dairy = imp_dairy_artinsem
                        imp_dairy_selectbreed imp_dairy_healthserv
                        imp_dairy_medicine imp_dairy_vaccinated
                        imp_dairy_prevmast imp_dairy_roof
                        imp_dairy_fedbyprod imp_dairy_fedconcentrate
                        imp_dairy_fedvitmin imp_dairy_collmanure
                        imp_dairy_fedfodder imp_dairy_pasteurized
                        imp_dairy_soldmilk imp_dairy_soldmanure
                        imp_dairy_records
```

**Step 6.** Create a count variable to capture the total number of targeted practices each dairy cow producer applied to raise dairy cows (`imp_tot_dairy`)—that is, create a variable that sums all the variables in the (`IMP_dairy`) variable list created in Step 5.

```
Set                imp_tot_dairy=count of all IMP_dairy variables
Label variable    "Number of targeted improved practices and
                  technologies applied (dairy cows)"
```

**Step 7.** Create a categorical variable to categorize dairy cow producers by the number of targeted practices they applied (`imp_cat_dairy`) for Table 7.4.4. The categories used should be adapted based on the survey data; they do not have to be 0, 1-3, 4-6, 7-9, and 10 or more targeted practices.

```
Set                imp_cat_dairy=missing
Replace            imp_cat_dairy=0 if imp_tot_dairy=0
Replace            imp_cat_dairy=1 if imp_tot_dairy≥1 and
                  imp_tot_dairy≤3
Replace            imp_cat_dairy=2 if imp_tot_dairy≥4 and
                  imp_tot_dairy≤6
Replace            imp_cat_dairy_3 if imp_tot_dairy≥7 and
                  imp_tot_dairy≤9
```

```

Replace      imp_cat_dairy_4 if imp_tot_dairy≥10 and
              imp_tot_dairy≠missing
Label value  0 "None"
              1 "1-3 practices"
              2 "4-6 practices"
              3 "7-9 practices"
              4 "10+ practices"
Label variable "Number of targeted improved practices and
               technologies applied (dairy cows, categorical)"

```

**Step 8.** Create a binary variable to indicate whether each dairy cow producer applied any targeted practices to raise dairy cows (`imp_any_dairy`).

```

Set          imp_any_dairy=0
Replace      imp_any_dairy=1 if imp_tot_dairy>0 and
              imp_tot_dairy≠missing
Label values 0 "No"
              1 "Yes"
Label variable "Applied any targeted management practice or
               technology (dairy cows)"

```

**Step 9.** After applying the dairy cow producer sampling weight, calculate the percentage of dairy cow producers who are de jure household members who applied at least one targeted practice to raise dairy cows during the year preceding the survey using `imp_any_dairy`. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

Also calculate the percentage of dairy cow producers who are de jure household members who applied each targeted practice (i.e., for all variables created in Step 3) and the percentage of dairy cow producers who are de jure household members who applied targeted practices by category (i.e., for all variables created in Step 4). (Example code uses Stata syntax.)

```

svyset hhea [pw=wgt_cow], strata(strata)
svy, subpop(hhmem_dj): tab anyimp_dairy
svy, subpop(hhmem_dj): tab anyimp_dairy age15_29y, col
svy, subpop(hhmem_dj): tab anyimp_dairy sex, col

```

```

For each variable that starts with 'imp_dairy_' (var):
    svy, subpop(hhmem_dj): tab 'var'

```

**Step 10.** Calculate the percentage distribution of dairy cow producers who are de jure household members by the number of targeted practices they applied to cultivate dairy cows during the 12 months preceding the survey using `imp_cat_dairy`. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

```

svy, subpop(hhmem_dj): tab imp_cat_dairy
svy, subpop(hhmem_dj): tab imp_cat_dairy age15_29y, col
svy, subpop(hhmem_dj): tab imp_cat_dairy sex, col

```

**Step 11.** Keep only the variables that are necessary to calculate the final overall indicator across all VCCs and save the data.

```

Keep      hhea hhnum m1_line strata sex age15_29y hhmem dj
          wgt_dairy vcc_dairy imp_tot_dairy imp_cat_dairy
          imp_any_dairy and all variables that begin with
          'imp_dairy_'
Save      "FTF ZOI Survey [COUNTRY] [YEAR] agtech_dairycow"

```

## 14.2.4 Sheep calculations

### Definitions

Numerator	Number of sheep producers who applied any targeted improved management practices or technologies
Denominator	Number of individuals responsible for raising sheep (i.e., sheep producers) during the 12 months preceding the survey
Unit of measure	Percentage
Level of data	Individual (producer)
Sampling weight	Sheep producer
Disaggregation levels	Sex* Age (15-29, 30+)* Management practice or technology type category*
Treatment of missing data	All missing and “don’t know” responses to management practice or technology type questions are considered to be “no.” They are included in the indicator denominator but not in the numerator.
Survey variables used	<b>hhea, hhnum, m1_line, strata</b> , all variables from Module 7.52
Analytic variables used	<b>sex, age15_29y, hhmem_dj, wgt_sheep, vcc_sheep</b>
Analytic variables created	<b>imp_sheep_artinsem, imp_sheep_selectbreed, imp_sheep_healthserv, imp_sheep_vaccinated, imp_sheep_roof, imp_sheep_fedvitmin, imp_sheep_pipedwater, imp_sheep_pasture, imp_sheep_fodder, imp_sheep_pasteurized, imp_sheep_soldmilk, imp_sheep_soldmanure, imp_sheep_livestm, imp_sheep_markdist, imp_sheep_valadd, imp_sheep_fsafety, imp_sheep_other, imp_cat_sheep, imp_tot_sheep, imp_any_sheep</b>

\*Standard Feed the Future disaggregate

### Calculations

In this example, the steps to calculate the percentage of sheep producers who applied targeted improved management practices or technologies during the year preceding the ZOI Survey are presented. The step-by-step procedures follow the Stata syntax *FTF ZOI Survey [COUNTRY][YEAR] syntax agtech\_sheep.do*. Variables from Module 7.52 are used for the calculations.

**Step 1.** Review Module 7.52 of the country-customized ZOI Survey questionnaire and confirm which questions and response options capture the practices that Feed the Future is targeting in the ZOI. Also confirm under which category or categories the individual practices are being targeted.<sup>97</sup> A practice

<sup>97</sup> The targeted improved management practices and technologies targeted by the Feed the Future program should have been established during the questionnaire customization stage to ensure that the practices and technologies are adequately captured in the survey.

might fit under multiple categories but may only be targeted under one category in the particular ZOI context. If this information is not readily available (e.g., in a table or list provided to the Data Analyst by the Country Lead), the information must be confirmed by the Data Analyst before they can begin data analysis. Consultations with RFS or in-country Feed the Future may be required.

**Table 20** shows the relevant variables and corresponding values used to categorize targeted practices for sheep in the step-by-step procedures and the *FTF ZOI Survey [COUNTRY][YEAR] syntax agtech\_sheep.do* and *FTF ZOI Survey COUNTRY YEAR syntax agtech\_final.do* Stata syntax files. Note that these are only examples and **MUST** be adapted for the country context.

For sheep, most targeted practices will be included under one category—livestock management—but some may fall under other categories (e.g., food safety or “other,” which is meant to capture targeted practices that do not fit under any of the other types). Please see the *Feed the Future Indicator Handbook* or the *Guide to Feed the Future Statistics for P2-ZOI Midline Surveys* for the list of all categories, including examples.

If Feed the Future is targeting a practice for multiple benefits, the sheep producers applying the technology should be reported under **each category** for which the practice is being targeted. If the practice is being targeted for only one purpose, the sheep producer should be reported under **only the relevant category**. If Feed the Future is targeting a practice for a single benefit even though it could be targeted for multiple benefits, be sure that producers applying the practice are reported under **only the one category** for which the practice is being targeted.

**Table 20: ZOI Survey Variables and Values to Identify Illustrative Targeted Practices for Raising Sheep, by Category**

Management practice or technology type category		Feed the Future-targeted practices and technologies for sheep (illustrative)
I	Livestock management	<p><b>Practice 1: Bred sheep with artificial insemination</b> (<i>imp_sheep_artinsem</i>): v75206 includes “D” (Artificial insemination)</p> <p><b>Practice 2: Selectively chose rams for breeding</b> (<i>imp_sheep_selectbreed</i>): v75207 includes “C” (Ram has good body size, composition), “D” (Ram is son of high-producing sheep), or “E” (Ram is known to have good fertility)</p> <p><b>Practice 3: Used trained health service providers</b> (<i>imp_sheep_healthserv</i>): v75225=1 (Yes, obtained health services from trained provider)</p> <p><b>Practice 4: Vaccinated some or all sheep</b> (<i>imp_sheep_vaccinated</i>): v75230=2 (Some sheep vaccinated) or v75230=3 (All sheep vaccinated)</p> <p><b>Practice 5: Used housing with a roof for sheep</b> (<i>imp_sheep_roof</i>): v75210=5 (Roof and sides, dirt floor) or v75210=6 (Roof and sides, wood floor) or v75210=7 (Roof and sides, concrete floor)</p> <p><b>Practice 6: Fed sheep vitamins or minerals</b> (<i>imp_sheep_fedvitmin</i>): v75222=1 (Fed vitamins or</p>

Management practice or technology type category		Feed the Future-targeted practices and technologies for sheep (illustrative)
		minerals in the past one year) <b>Practice 7: Piped drinking water to sheep</b> (imp_sheep_pipedwater): v75211=4 (piped drinking water to sheep) <b>Practice 8: Improved grazing pasture quality</b> (imp_sheep_pasture): v75212b=1 (Yes, improved the quality of the pasture) <b>Practice 9: Fed sheep improved fodder crops<sup>a</sup></b> (imp_sheep_fodder): v75215=1 (Fed crop fodder to sheep) or v75213 includes "B" (Fed wheat bran forages to sheep)
2	Marketing and distribution	<b>Practice 10: Sold sheep's milk</b> (imp_sheep_soldmilk): v75238=1 (Yes, sold milk) <b>Practice 11: Sold sheep's manure</b> (imp_sheep_soldmanure): v75224=05 (Sell to friends/neighbors) or v75224=06 (Sell at market)
3	Value-added processing	<b>Practice 12: Pasteurized sheep's milk</b> (imp_sheep_pasteurized): v75240 includes "B" (pasteurized sheep milk)
9	Food safety	<b>Practice 12: Pasteurized sheep's milk</b> (imp_sheep_pasteurized): v75240 includes "B" (pasteurized sheep milk)
10	Other	<b>Practice 13: Kept written records on sheep</b> (imp_sheep_records): v75241=1 (Yes, kept written records)

<sup>a</sup> Also manually check response option X (Other) to determine if any manually entered forages should be considered targeted improved fodder crops and included in the calculations.

**Step 2.** Prepare the data—that is, load the individual-level data file, drop records for producers who did not raise sheep in the year preceding the survey, and drop variables not required to calculate the improved management practices or technologies indicator for sheep. Also review the **vcc\_sheep** variable that should have already been created and included in the individual-level data file.

```

Load          "FTF ZOI Survey [Country] [Year] persons data
               analytic"
Keep record   if vcc_sheep=1
Keep          hhea hhnum m1_line strata sex age15-29y hhmem_dj
               vcc_sheep wgt_sheep and all variables that begin
               with 'v752'
```

**Step 3.** Create 13 binary variables—one for each targeted practice to flag producers who used the practice to raise sheep during the 12 months preceding the survey (yes=1, no=0).

**Step 3.1.** Create a binary variable to flag sheep producers who bred their sheep with artificial insemination (**imp\_sheep\_artinsem**).

```

Set           imp_sheep_artinsem=0
```

```

Replace      imp_sheep_artinsem=1 if v75206 includes "D"
Label values 0 "No"
              1 "Yes"
Label variable "Used artificial insemination to breed sheep"

```

**Step 3.2.** Create a binary variable to flag sheep producers who selectively chose rams for breeding their sheep (`imp_sheep_selectbreed`).

```

Set          imp_sheep_selectbreed=0
Replace      imp_sheep_selectbreed=1 if v75207 includes "C,"
              "D," or "E"
Label values 0 "No"
              1 "Yes"
Label variable "Selectively chose rams for breeding sheep"

```

**Step 3.3.** Create a binary variable to flag sheep producers who used trained health service providers for their sheep (`imp_sheep_healthserv`).

```

Set          imp_sheep_healthserv=0
Replace      imp_sheep_healthserv=1 if v75225=1
Label values 0 "No"
              1 "Yes"
Label variable "Obtained health services for sheep from trained
              provider"

```

**Step 3.4.** Create a binary variable to flag sheep producers who vaccinated some or all of their sheep (`imp_sheep_vaccinated`).

```

Set          imp_sheep_vaccinated=0
Replace      imp_sheep_vaccinated=1 if v75230=2 or v75230=3
Label values 0 "No"
              1 "Yes"
Label variable "Vaccinated some or all sheep"

```

**Step 3.5.** Create a binary variable to flag sheep producers who used housing with a roof for their sheep (`imp_sheep_roof`).

```

Set          imp_sheep_roof=0
Replace      imp_sheep_roof=1 if v75210=5 or v75210=6 or
              v75210=7
Label values 0 "No"
              1 "Yes"
Label variable "Used housing with a roof for sheep"

```

**Step 3.6.** Create a binary variable to flag sheep producers who fed sheep vitamins or minerals to their sheep (`imp_sheep_fedvitmin`).

```

Set          imp_sheep_fedvitmin=0
Replace      imp_sheep_fedvitmin=1 if v75222=1
Label values 0 "No"
              1 "Yes"

```



Label variable    "Fed sheep vitamins or minerals in the past one year"

**Step 3.7.** Create a binary variable to flag sheep producers who piped drinking water to sheep (**imp\_sheep\_pipedwater**).

Set                    imp\_sheep\_pipedwater=0  
Replace              imp\_sheep\_pipedwater=1 if v75211=4  
Label values        0 "No"  
                         1 "Yes"  
Label variable      "Piped drinking water to sheep"

**Step 3.8.** Create a binary variable to sheep producers who improved the pasture quality where their sheep graze (**imp\_sheep\_pasture**).

Set                    imp\_sheep\_pasture=0  
Replace              imp\_sheep\_pasture=1 if v75212b=1  
Label values        0 "No"  
                         1 "Yes"  
Label variable      "Improved pasture quality for sheep grazing"

**Step 3.9.** Create a binary variable to flag sheep producers who fed sheep improved fodder crops or wheat bran forages (**imp\_sheep\_fodder**).

Set                    imp\_sheep\_fodder=0  
Replace              imp\_sheep\_fodder=1 if v75215=1 or v75213 includes  
                         "B"  
Label values        0 "No"  
                         1 "Yes"  
Label variable      "Fed crop fodder or wheat bran forages to sheep"

**Step 3.10.** Create a binary variable to sheep producers who sold their sheep's milk (**imp\_sheep\_soldmilk**).

Set                    imp\_sheep\_soldmilk=0  
Replace              imp\_sheep\_soldmilk=1 if v75238=1  
Label values        0 "No"  
                         1 "Yes"  
Label variable      "Sold sheep's milk"

**Step 3.11.** Create a binary variable to sheep producers who sold their sheep's manure (**imp\_sheep\_soldmanure**).

Set                    imp\_sheep\_soldmanure=0  
Replace              imp\_sheep\_soldmanure=1 if v75224=05 or v75224=06  
Label values        0 "No"  
                         1 "Yes"  
Label variable      "Sold sheep's manure to friends or neighbors or at market"

**Step 3.12.** Create a binary variable to sheep producers who pasteurized their sheep's milk (`imp_sheep_pasteurized`).

```
Set          imp_sheep_pasteurized=0
Replace      imp_sheep_pasteurized=1 if v75240 includes "B"
Label values 0 "No"
              1 "Yes"
Label variable "Pasteurized sheep's milk"
```

**Step 3.13.** Create a binary variable to flag sheep producers who kept written records on their sheep (`imp_sheep_records`).

```
Set          imp_sheep_records=0
Replace      imp_sheep_records=1 if v75241=1
Label values 0 "No"
              1 "Yes"
Label variable "Kept written records on sheep"
```

**Step 4.** Create a binary variable for each category to flag producers who used any practices targeted by Feed the Future to raise sheep under the relevant category (yes=1, no=0). Almost all targeted practices for raising sheep are categorized under a single category: livestock management.<sup>98</sup>

**Step 4.1.** Create a binary variable to indicate whether each sheep producer used any targeted practices under livestock management (`imp_sheep_livestm`). **Note:** If all practices were targeted under livestock management, this variable will be duplicative of the one created in Step 6 (`imp_any_sheep`).

```
Set          imp_sheep_livestm=0
Replace      imp_sheep_livestm=1 if imp_sheep_artinsem=1 or
              imp_sheep_selectbreed=1 or imp_sheep_healthserv=1
              or imp_sheep_vaccinated=1 or imp_sheep_roof=1 or
              imp_sheep_fedvitmin=1 or imp_sheep_pipedwater=1
              or imp_sheep_pasture=1 or imp_sheep_fodder=1
Label values 0 "No"
              1 "Yes"
Label variable "Used improved livestock management"
```

**Step 4.2.** Create a binary variable to flag sheep producers who applied a targeted practice related to marketing and distribution (`imp_sheep_markdist`).

```
Set          imp_sheep_markdist=0
Replace      imp_sheep_markdist=1 if imp_sheep_soldmilk=1 or
              imp_sheep_soldmanure=1
Label values 0 "No"
              1 "Yes"
Label variable "Used improved marketing and distribution"
```

---

<sup>98</sup> Sheep producers could also apply targeted practices that are targeted under other categories; for example, using a Feed the Future-supported mobile phone application to get on-demand production and marketing information and advice might be categorized under the "Other" type if it is determined not to be targeted under another category.

**Step 4.3.** Create a binary variable to flag sheep producers who applied a targeted practice related to value added processing (`imp_sheep_valadd`).

```
Set                imp_sheep_valadd=0
Replace            imp_sheep_valadd=1 if imp_sheep_pasteurized=1
Label values       0 "No"
                   1 "Yes"
Label variable     "Used improved value-added processing"
```

**Step 4.4.** Create a binary variable to flag sheep producers who applied a targeted practice related to food safety (`imp_sheep_fsafety`).

```
Set                imp_sheep_fsafety=0
Replace            imp_sheep_fsafety=1 if imp_sheep_pasteurized=1
Label values       0 "No"
                   1 "Yes"
Label variable     "Used improved food safety practices"
```

**Step 4.5.** Create a binary variable to flag sheep producers who applied a targeted practice not captured under any other category (`imp_sheep_other`).

```
Set                imp_sheep_other=0
Replace            imp_sheep_other=1 if imp_sheep_records=1
Label values       0 "No"
                   1 "Yes"
Label variable     "Used other improved practice or technology"
```

**Step 5.** Create a variable list (`IMP_sheep`) that includes all applicable targeted practices for sheep. Adapt this step as needed to reflect different livestock value chains and targeted practices technologies targeted in the ZOI country context.

```
Set variable list IMP_sheep = imp_sheep_artinsem
                        imp_sheep_selectbreed imp_sheep_healthserv
                        imp_sheep_vaccinated imp_sheep_roof
                        imp_sheep_fedvitmin imp_sheep_pipedwater
                        imp_sheep_pasture imp_sheep_fodder
                        imp_sheep_pasteurized imp_sheep_soldmilk
                        imp_sheep_soldmanure imp_sheep_records
```

**Step 6.** Create a count variable to capture the total number of targeted practices each producer used to raise sheep (`imp_tot_sheep`). That is, create a variable that sums all the variables in the (`IMP_sheep`) variable list created in Step 5.

```
Set                imp_tot_sheep=count of all IMP_sheep variables=1
Label values       0 "No"
                   1 "Yes"
Label variable     "Number of targeted improved practices and
                    technologies applied (sheep)"
```

**Step 7.** Create a categorical variable to categorize sheep producers by the number of targeted practices they used (`imp_cat_sheep`) for Table 7.4.4 adapted for sheep. The categories used should

be adapted based on the survey data; they do not have to be 0, 1-3, 4-6, 7-9, and 10 or more targeted practices.

```

Set          imp_cat_sheep=missing
Replace      imp_cat_sheep=0 if imp_tot_sheep=0
Replace      imp_cat_sheep=1 if imp_tot_sheep≥1 and
             imp_tot_sheep≤3
Replace      imp_cat_sheep=2 if imp_tot_sheep≥4 and
             imp_tot_sheep≤6
Replace      imp_cat_sheep=3 if imp_tot_sheep≥7 and
             imp_tot_sheep≤9
Replace      imp_cat_sheep=4 if imp_tot_sheep≥10 and
             imp_tot_sheep≠missing
Label value  0 "None"
             1 "1-3 practices"
             2 "4-6 practices"
             3 "7-9 practices"
             4 "10+ practices"
Label variable "Number of targeted improved practices and
              technologies used (sheep, categorical)"

```

**Step 8.** Create a binary variable to indicate whether each sheep producer applied any targeted practices to raise sheep (**imp\_any\_sheep**).

```

Set          imp_any_sheep=0
Replace      imp_any_sheep=1 if imp_tot_sheep>0 and
             imp_tot_sheep≠missing
Label values 0 "No"
             1 "Yes"
Label variable "Used any targeted management practice or
              technology (sheep)"

```

**Step 9.** After applying the sheep producer sampling weight, calculate the percentage of sheep producers who are de jure household members who applied at least one targeted practice to raise sheep during the year preceding the survey using **imp\_any\_sheep**. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

Also calculate the percentage of sheep producers who are de jure household members who applied each targeted practice (i.e., for all variables created in Step 3) and the percentage of sheep producers who are de jure household members who applied targeted practices by category (i.e., for all variables created in Step 4). (Example code uses Stata syntax.)

```

svyset hhea [pw=wgt_sheep], strata(strata)
svy, subpop(hhmem_dj): tab imp_any_sheep
svy, subpop(hhmem_dj): tab imp_any_sheep age15_29y, col
svy, subpop(hhmem_dj): tab imp_any_sheep sex, col

For each      variable that starts with 'imp_sheep_' (var):
    svy, subpop(hhmem_dj): tab 'var'

```

**Step 10.** Calculate the percentage distribution of sheep producers who are de jure household members by the number of targeted practices they used to cultivate sheep during the 12 months preceding the survey using `imp_cat_sheep`. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

```
svy, subpop(hhmem_dj): tab imp_cat_sheep
svy, subpop(hhmem_dj): tab imp_cat_sheep age15_29y, col
svy, subpop(hhmem_dj): tab imp_cat_sheep sex, col
```

**Step 11.** Keep only the variables that are necessary to calculate the final overall indicator across all VCCs and save the data.

```
Keep variables  hhea hhnum m1_line strata sex age15_29y hhmem_dj
                wgt_sheep vcc_sheep imp_tot_sheep imp_cat_sheep
                imp_any_sheep and all variables that begin with
                'imp_sheep_'
Save "FTF ZOI Survey [COUNTRY] [YEAR] agtech_sheep"
```

#### 14.2.5 Overall indicator calculations<sup>99</sup>

In this example, the steps to calculate the percentage of producers who applied targeted improved management practices or technologies across all targeted VCCs during the year preceding the ZOI Survey are presented. The step-by-step procedures follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax agtech\_final.do*.

#### Definitions

Numerator	<p>Number of producers who applied targeted improved management practices or technologies</p> <p>For step-by-step instructions: Number of maize, fishpond, dairy cow, and sheep producers who applied any targeted improved management practices or technologies</p>
Denominator	<p>Number of individuals responsible for cultivating any targeted VCC (i.e., VCC producers) in the 12 months preceding the survey</p> <p>For step-by-step instructions: Number of individuals responsible for cultivating maize or raising fish in fishponds, dairy cows, or sheep (i.e., maize, fishpond, dairy cow, or sheep producers) during the 12 months preceding the survey</p>
Unit of measure	Percentage
Level of data	Individual (producer)
Sampling weight	Producer of any selected agricultural VCC
Disaggregation levels	<p>Sex*</p> <p>Age (15-29, 30+)*</p> <p>Management practice or technology type category*</p> <p>Commodity*</p>

<sup>99</sup> Before moving to this step, perform the calculations for each value chain included in the ZOI Survey following the guidance in Sections 14.2.1-14.2.4, adapting as needed to reflect different crop, livestock, or fishpond value chains and improved management practices and technologies targeted in the ZOI.

Treatment of missing data	All missing and “don’t know” responses to management practice or technology type questions are considered to be “no.” They are included in the indicator denominator but not in the numerator.
Survey variables used**	<b>hhea, hhnum, ml_line, strata, wgt_vcc</b>
Analytic variables used**	<b>sex, age15_29y, hhmem_dj, vcc_maize, vcc_fish, vcc_dairy, vcc_sheep</b> , all variables that begin with ‘imp_’ created in Sections 14.2.1-14.2.4
Analytic variables created**	<p><b>Practices:</b></p> <p><b>imp_crop_impseed, imp_crop_plantrow, imp_crop_orgfert, imp_crop_mulchsoil, imp_crop_terrace, imp_crop_mulchweed, imp_crop_pestchem, imp_crop_irrigdrip, imp_crop_irrigpump, imp_crop_zerotill, imp_crop_agdealerseed, imp_crop_soldhusks, imp_crop_mechdry, imp_crop_hermetic, imp_crop_warehouse, imp_crop_mechshuck, imp_crop_mechtill, imp_crop_mechharvest, imp_ls_artinsem, imp_ls_selectbreed, imp_ls_healthserv, imp_ls_medicine, imp_ls_vaccinated, imp_ls_prevmast, imp_ls_roof, imp_ls_fedbyprod, imp_ls_fedconcentrate, imp_ls_fedvitmin, imp_ls_pipedwater, imp_ls_pasture, imp_ls_collmanure, imp_ls_fodder, imp_ls_pasteurized, imp_ls_soldmilk, imp_ls_soldmanure, imp_ls_records</b></p> <p><b>Management practice and technology type categories:</b></p> <p><b>imp_genetics, imp_culture, imp_ecosys, imp_pest, imp_fert, imp_irrig, imp_water, imp_cmitigate, imp_cadapt, imp_markdist, imp_pharvest, imp_valadd, imp_fsafety, imp_other, imp_aquam, imp_livestm</b></p> <p><b>Other:</b></p> <p><b>imp_tot_vcc, imp_any_vcc, imp_cat_vcc</b></p>

\* Standard Feed the Future disaggregate

\*\*Depends on the targeted VCCs included in the ZOI Survey

## Calculations

**Step 1.** Create a data file, *agtech\_all*, that includes one entry for each crop producer with variables for all targeted practices and categories for each crop and across crops.

**Step 1.1.** If there is only one crop VCC, skip to Step 1.5. If there is more than one crop VCC, create a data file that includes the data for all crop VCCs. Append all crop targeted practice data files (e.g., as created in Step 1.1 of Section 14.2.1). Maize and millet are used here as an example. Millet is not otherwise included in this Guide, but it is used as an example throughout Step 1.

```
Load          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_maize"
Append using  "FTF ZOI Survey [COUNTRY] [YEAR] agtech_millet"
```

**Step 1.2.** Aggregate the data so that there is only one record per crop producer. That is, if a crop producer cultivated multiple crop VCCs and therefore has multiple records in the data file, collapse the data summing each variable that begins with ‘imp.’ The summed variables will have a value of either 0 or 1, and the resulting data file will have one record per crop producer.

```
By hhea hhnum m1_line: sum each variable that begins with 'imp_'
Save "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"
```

**Step 1.3.** Create a binary variable for each targeted practice that has “crop” in the variable name instead of the name of the crop (e.g. “maize”) and captures whether the crop producer used the practice for any of the crops for which it is targeted.

The first variable created below in this step, `imp_crop_impseed`, applies to both maize and millet. The second variable created below in this step, `imp_crop_maize_mulchweed`, applies only to maize, but a variable is still created so that “crop” is consistently included in the name of all variables that will be used for the overall indicator calculation.

Be sure to expand the template syntax below as applicable to create a variable for each targeted practice. This should include the variables created in Step 3 of Section 14.2.1: `imp_crop_impseed`, `imp_crop_plantrow`, `imp_crop_orgfert`, `imp_crop_mulchsoil`, `imp_crop_terrace`, `imp_crop_mulchweed`, `imp_crop_pestchem`, `imp_crop_irrigdrip`, `imp_crop_irrigpump`, `imp_crop_zerotill`, `imp_crop_agdealerseed`, `imp_crop_soldhusks`, `imp_crop_mechdry`, `imp_crop_hermetic`, `imp_crop_warehouse`, `imp_crop_mechshuck`, `imp_crop_mechtill`, and `imp_crop_mechharvest`.

```
Set          imp_crop_impseed=0
Replace      imp_crop_practice1=1 if imp_maize_impseed=1 or
              imp_millet_impseed=1
Label values 0 "No"
              1 "Yes"
Label variable "Used improved seeds (maize, millet)"

Set          imp_crop_mulchweed=imp_maize_mulchweed
Label values 0 "No"
              1 "Yes"
Label variable "Used mulching to control weeds (maize only)"
```

**Step 1.4.** Create a binary variable for each category without the name of the crop in the variable name and save the data file. Be sure to expand the template syntax below to create a variable for each category applicable to the crop VCCs included in the survey. This should include the variables created in Step 4 of Section 14.2.1: `imp_genetics`, `imp_culture`, `imp_ecosys`, `imp_pest`, `imp_fert`, `imp_irrig`, `imp_water`, `imp_cmitigate`, `imp_cadapt`, `imp_markdist`, `imp_pharvest`, `imp_valadd`, `imp_fsafety`, and `imp_other`.

```
Set          imp_genetics=0
Replace      imp_genetics=1 if imp_maize_genetics=1 or
              imp_millet_genetics=1
Label values 0 "No"
              1 "Yes"
Label variable "Used targeted improved crop genetics practices
              (maize, millet)"
```

```

Set          imp_pest=imp_maize_pest
Label values 0 "No"
              1 "Yes"
Label variable "Used targeted improved pest management (maize
              only)"

Save          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"

```

**Step 1.5.** If there is more than one crop VCC, skip this step. If there is only one crop VCC, clone the variable for each category and save the data file. Be sure to expand the template syntax below to create a variable for each category applicable to the crop VCCs. This should include the variables created in Step 4 of Section 14.2.1: *imp\_genetics*, *imp\_culture*, *imp\_ecosys*, *imp\_pest*, *imp\_fert*, *imp\_irrig*, *imp\_water*, *imp\_cmitigate*, *imp\_cadapt*, *imp\_markdist*, *imp\_pharvest*, *imp\_valadd*, *imp\_fsafety*, and *imp\_other*.

```

Set          imp_genetics=imp_maize_genetics
Label values 0 "No"
              1 "Yes"
Label variable "Used targeted improved crop genetics practices
              (maize only)"

Set          imp_pest=imp_maize_pest
Label values 0 "No"
              1 "Yes"
Label variable "Used targeted improved pest management (maize
              only)"

Save          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"

```

**Step 2.** Skip to Step 3 if fishponds are not a value chain included in the survey. If fishponds are a value chain, create the variables needed for the overall indicator calculation and then add an observation for each fishpond producer to the *agtech\_all* data file created in Step 1.5.

**Step 2.1.** Load the fishpond data file created in Step 11 of Section 14.2.2 and create a binary variable for each category without 'fish' in the variable name and save the data file. Be sure to expand the template syntax below to create a variable for each category applicable to fishponds. Most targeted improved practices will be targeted under aquaculture management (*imp\_aquam*), but some may be targeted under other categories, such as "other" (*imp\_other*).

```

Load          "FTF ZOI Survey [COUNTRY] [YEAR]
              agtech_fishponds"
Set          imp_aquam=imp_fish_aquam
Label values 0 "No"
              1 "Yes"
Label variable "Used improved aquaculture management practices"

Set          imp_other=imp_fish_other
Label values 0 "No"
              1 "Yes"
Label variable "Used other improved practice or technology"

```



```
Save "FTF ZOI Survey [COUNTRY] [YEAR] fishpond_all"
```

**Step 2.2.** Append the fishpond data file to the *agtech\_all* data file created in Step 1.5 and save the data file.

```
Load "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"
Append using "FTF ZOI Survey [COUNTRY] [YEAR]
agtech_fishpond_all"
Save "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"
```

**Step 3.** Skip to Step 4 if there are no livestock value chains included in the survey. If there are one or more livestock value chains, create a data file, *agtech\_livestock\_all*, that includes one entry for each livestock producer with variables for all targeted practices and categories for each livestock and across livestock. Then add one observation for each livestock producer to the *agtech\_all* data file created in Step 1.5 (and modified in Step 2.2, if applicable).

**Step 3.1.** If there is only one livestock VCC, skip to Step 3.5. Create a data file that includes the data for all livestock VCCs included in the survey. Append all livestock targeted practice data files (e.g., as created in Step 11 of Sections 14.2.3 and 14.2.4). Dairy cows and sheep are used here as an example.

```
Load "FTF ZOI Survey [Country] [Year] agtech_dairycow"
Append using "FTF ZOI Survey [Country] [Year] agtech_sheep"
```

**Step 3.2.** Aggregate the data so that there is only one record per livestock producer—that is, if a livestock producer raised multiple livestock VCCs and therefore has multiple records in the data file, collapse the data summing each variable that begins with ``imp_``. The summed variables will have a value of either 0 or 1, and the resulting data file will have one record per livestock producer.

```
By hhea hhnum m1_line: sum each variable that begins with `imp_`
Save "FTF ZOI Survey [COUNTRY] [YEAR] livestock_all"
```

**Step 3.3.** Create a binary variable for each targeted practice that has “1s” in the variable name instead of the name of the livestock (e.g., “**sheep**”) and that captures whether the livestock producer used the practice for any of the livestock for which it is targeted.

The first variable created below in this step, `imp_1s_artinsem`, applies to both dairy cows and sheep. The second variable created below in this step, `imp_1s_waste`, applies only to dairy cows, but a variable is still created so that “1s” is consistently included in the name of all variables that will be used for the overall indicator calculation.

Be sure to expand the template syntax that follows as applicable to create a variable for each targeted practice. This should include the variables created in Step 3 of Sections 14.2.3 and 14.2.4:

```
imp_1s_artinsem, imp_dairy_selectbreed, imp_1s_healthserv,
imp_dairy_medicine, imp_1s_vaccinated, imp_1s_prevmast, imp_1s_roof,
imp_1s_fedbyprod, imp_1s_fedconcentrate, imp_1s_fedvitmin,
imp_1s_collmanure, imp_1s_fodder, imp_1s_pastuerized, imp_1s_soldmilk,
imp_1s_soldmanure, and imp_1s_records.
```

```

Set          imp_ls_artinsem=0
Replace      imp_ls_artinsem=1 if imp_dairy_artinsem=1 or
             imp_sheep_artinsem=1
Label values 0 "No"
             1 "Yes"
Label variable "Used artificial insemination to breed (sheep,
              dairy cows)"

Set          imp_lswaste=imp_dairy_lswaste
Label values 0 "No"
             1 "Yes"
Label variable "Used improved waste management (dairy cows
              only)"

```

**Step 3.4.** Create a binary variable for each category without the name of the VCC in the variable name and save the data file. Be sure to expand the template syntax below to create a variable for each category applicable to the livestock value chains included in the survey. This should include the variables created in Step 4 of Sections 14.2.3 and 14.2.4. Most targeted practices will be targeted under livestock management (`imp_livestm`), but some may be targeted under other categories, such as `imp_fsafety` or `imp_other`.

```

Set          imp_livestm=0
Replace      imp_livestm=1 if imp_dairy_livestm=1 or
             imp_sheep_livestm=1
Label values 0 "No"
             1 "Yes"
Label variable "Used improved livestock management practices
              (sheep, dairy cows)"

Set          imp_fsafety=0
Replace      imp_fsafety=1 if imp_dairy_fsafety=1 or
             imp_sheep_fsafety=1
Label values 0 "No"
             1 "Yes"
Label variable "Used improved food safety practices (sheep,
              dairy cows)"

Set          imp_other=0
Replace      imp_other=1 if imp_dairy_other=1 or
             imp_sheep_other=1
Label values 0 "No"
             1 "Yes"
Label variable "Used other improved practice or technology
              (sheep, dairy cows)"

Save          "FTF ZOI Survey [COUNTRY] [YEAR] livestock_all"

```

**Step 3.5.** If there is more than one livestock VCC included, skip this step. If there is only one livestock VCC included in the ZOI Survey, clone the variable for each category and save the data file. Be sure to expand the template syntax below to create a variable for each category applicable to livestock value chains. This should include the variables created in Step 4 of Sections 14.2.3 and 14.2.4. Most targeted practices and technologies will be targeted under livestock management (`imp_livestm`), but some may be targeted under other categories, such as `imp_fsafety` or `imp_other`.

```
Set          imp_livestm=imp_dairy_livestm
Label values 0 "No"
              1 "Yes"
Label variable "Used improved livestock management practices
              (sheep, dairy cows)"

Set          imp_fsafety=imp_dairy_fsafety
Label values 0 "No"
              1 "Yes"
Label variable "Used improved food safety practices (sheep,
              dairy cows)"

Set          imp_other=imp_dairy_other
Label values 0 "No"
              1 "Yes"
Label variable "Used other improved practice or technology
              (sheep, dairy cows)"

Save          "FTF ZOI Survey [COUNTRY] [YEAR] livestock_all"
```

**Step 3.6.** Append the livestock data file to the `agtech_all` data file and save the `agtech_all` data file.

```
Load          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"
Append using  "FTF ZOI Survey [COUNTRY] [YEAR] livestock_all"
Save          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"
```

**Step 4.** Aggregate the `agtech_all` data so that there is only one record per producer across all value chains. That is, collapse the data summing the values for each variable that begins with `'imp.'` This will result in a data file with one record per producer with the summed total of each targeted practice that producers adopted across all VCCs. The `agtech_all` data file should now include variables for each targeted practice and each category for each value chain and across all value chains.

```
By hhea hnum m1_line: sum each variable that begins with 'imp_'
Save          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"
```

**Step 5.** Set the denominator for each targeted practice variable and category variable to the total number of VCC farmers who completed the survey—that is, for each targeted practice and category, set all missing answers to 0 (“No, did not use practice”). Adapt this step as needed to reflect the crop, fishpond, or livestock value chains and practices targeted in the ZOI country context.

```

For each variable `var` of variables that begin with `imp_crop_`
`imp_ls_` and `imp_fish_` and also the variables
imp_genetics imp_culture imp_ecosys imp_pest
imp_fert imp_irrig imp_water imp_cmitigate
imp_cadapt imp_markdist imp_pharvest imp_valadd
imp_fsafety imp_other imp_fish_aquam imp_livestm
imp_markdist:
Replace `var`=0 if `var`=missing

```

**Step 6.** Create a variable list (IMP\_PRACTICE) that includes all targeted practice variables. Adapt this step as needed to reflect the crop, fishpond, or livestock value chains and practices targeted in the ZOI country context.

```

Set variable list IMP_PRACTICE=imp_crop_impseed imp_crop_plantrow
imp_crop_orgfert imp_crop_mulchsoil
imp_crop_terrace imp_crop_mulchweed
imp_crop_pestchem imp_crop_irrigdrip
imp_crop_irrigpump imp_crop_zerotill
imp_crop_agdealerseed imp_crop_soldhusks
imp_crop_mechdry imp_crop_hermetic
imp_crop_warehouse imp_crop_mechshuck
imp_crop_mechtill imp_crop_mechharvest
imp_ls_artinsem imp_dairy_selectbreed
imp_ls_healthserv imp_dairy_medicine
imp_ls_vaccinated imp_ls_prevmast imp_ls_roof
imp_ls_fedbyprod imp_ls_fedconcentrate
imp_ls_fedvitmin imp_ls_pipedwater imp_ls_pasture
imp_ls_collmanure imp_ls_fodder
imp_ls_pasteurized imp_ls_soldmilk
imp_ls_soldmanure imp_ls_records
imp_fish_pondrain imp_fish_pondmanure
imp_fish_certhatch imp_fish_species
imp_fish_fedsupp imp_fish_diseasecontrol
imp_fish_pondmonqual imp_fish_pondmainqual
imp_fish_separate imp_fish_partharv
imp_fish_records imp_fish_useguts

```

**Step 7.** Create a count variable to capture the total number of targeted improved practices used (imp\_tot\_vcc).

```

Set imp_tot_vcc=count of all variables in
IMP_PRACTICE variable list
Label variable "Number of targeted improved practices and
technologies used (total)"

```

**Step 8.** Create a categorical variable to categorize producers by the number of targeted practices they used (imp\_cat\_vcc) for Table 7.1.4. The categories used can be adapted based on the survey data; they do not have to be 0, 1-3, 4-6, 7-9, and 10 or more practices.

```

Set imp_cat_vcc=missing
Replace imp_cat_vcc=0 if imp_tot_vcc=0

```

```

Replace      imp_cat_vcc=1 if imp_tot_vcc≥1 and imp_tot_vcc≤3
Replace      imp_cat_vcc=2 if imp_tot_vcc≥4 and imp_tot_vcc≤6
Replace      imp_cat_vcc_3 if imp_tot_vcc≥7 and imp_tot_vcc≤9
Replace      imp_cat_vcc_4 if imp_tot_vcc≥10 and
              imp_tot_vcc≠missing
Label value  0 "None"
              1 "1-3 practices"
              2 "4-6 practices"
              3 "7-9 practices"
              4 "10+ practices"
Label variable "Number of targeted improved practices and
              Technologies used (total, categorical)"

```

**Step 9.** Create a binary variable to indicate if the producers used any targeted practices (`imp_any_vcc`).

```

Set          imp_any_vcc=1 if imp_tot_vcc>0 and imp_tot_vcc≠
              missing
Label variable "Used any targeted improved management practice
              and technology (total)"

```

**Step 10.** Create a variable list (`IMP_TYPE`) that includes all applicable category variables. Adapt this step as needed to reflect the crop, livestock, or fishpond value chains and practices targeted in the ZOI country context.

```

Set variable list IMP_TYPE=imp_genetics imp_culture imp_ecosys
                  imp_pest imp_fert imp_irrig imp_water
                  imp_cmitigate imp_cadapt imp_markdist
                  imp_pharvest imp_valadd imp_fsafety imp_other
                  imp_fish_aquam imp_livestm

```

**Step 11.** If not already in the data file, add the de jure household member status, age, sex, producer weight, and sample stratum variables and as well as all variables that include 'vcc' in the name from the individual-level analytic data file needed to calculate the overall indicator and its disaggregates and save the data file.

```

Merge        1 to 1
              key variables hhea, hhnum, m1_line
              data file     "FTF ZOI Survey [Country] [Year] persons
                           analytic"
              keep variables hhmem_dj sex age15_29 wgt_vcc strata and all
                           variables with `vcc'
Save          "FTF ZOI Survey [COUNTRY] [YEAR] agtech_all"

```

**Step 12.** After applying the VCC producer sampling weight to adjust for the survey design, calculate the percentage of producers who are de jure household members who applied at least one targeted practice to produce a VCC during the year preceding the survey using the `imp_any_vcc` analytic variable. Repeat the calculation using the sex, age (under 30 years of age, 30 years of age or older), and

category disaggregates. Note that the commodity disaggregate was already calculated in the individual VCC syntax files. (Example code uses Stata syntax.)

```
svyset hhea [pw=wtg_vcc], strata(strata)
svy, subpop(hhmem_dj): tab imp_any_vcc
svy, subpop(hhmem_dj): tab imp_any_vcc sex, col
svy, subpop(hhmem_dj): tab imp_any_vcc age15_29, col
```

```
For each      variable (x) in the variable list IMP_TYPE
               IMP_PRACTICE:
               svy, subpop(hhmem_dj): tab `x', col
```

**Step 13.** Calculate the percentage distribution of VCC producers who are de jure household members by the number of targeted practices they used during the 12 months preceding the survey using `imp_cat_vcc`. Repeat using producers' age (under 30 years of age, 30 years of age or older) and sex as disaggregates.

```
svy, subpop(hhmem_dj): tab imp_cat_vcc
svy, subpop(hhmem_dj): tab imp_cat_vcc age15_29y, col
svy, subpop(hhmem_dj): tab imp_cat_vcc sex, col
```

### 14.3 References

United Nations Food and Agriculture Organization. (2019). *Animal production*. Available at: <http://www.fao.org/animal-production/en/>

United Nations Food and Agriculture Organization. (2019). *Aquaculture production*. Available at: <https://www.fao.org/3/cc0461en/online/sofia/2022/aquaculture-production.html>

United Nations Food and Agriculture Organization. (2018). *Harnessing the power of innovation in agriculture*. Available at: <https://www.fao.org/publications/highlights-detail/en/c/1171251/>

United Nations Food and Agriculture Organization. (2018). *Shaping the future of livestock sustainably, responsibly, efficiently*. Berlin: The 10<sup>th</sup> Global Forum for Food and Agriculture. Available at: <http://www.fao.org/3/i8384en/i8384EN.pdf>

United Nations Food and Agriculture Organization. (2017). *The future of food and agriculture—Trends and challenges*. Available at: <http://www.fao.org/3/a-i6583e.pdf>

United Nations Food and Agriculture Organization. (2009). *The technology challenge. High Level Expert Forum—How to Feed the World in 2050*. Available at: [http://www.fao.org/fileadmin/templates/wsfs/docs/Issues\\_papers/HLEF2050\\_Technology.pdf](http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Technology.pdf)

## 15. HOUSEHOLD FOOD INSECURITY INDICATOR

This chapter describes the prevalence of households with moderate and severe food insecurity using the Food Insecurity Experience Scale (FIES), a Feed the Future phase two indicator. This chapter has two sections; the first section describes the guidelines to construct the indicator, and the second section outlines the step-by-step procedures to calculate the indicator as well as the prevalence of moderate food insecurity and severe food insecurity.

### 15.1 Guidelines to construct the FIES indicator

This section provides the guidelines to construct the Feed the Future food insecurity indicator: the prevalence of moderate and severe food insecurity in the population, based on the FIES.

The prevalence of moderate and severe insecurity is calculated from the respondent's raw score—that is, the number of affirmative responses given to the eight FIES questions. The raw score is simply an integer with a value between 0 and 8, and hence there will always be up to 9 distinct values of respondent parameters, one for each possible raw score (0–8). Although the raw scores and associated respondent parameters depend only on the number of affirmative responses, certain response patterns are expected to conform to the Rasch model's assumptions—that is, when arranged in the order of increasing severity, responses start with “yes” and are followed by “no” (without alternating). This guide provides steps for statistical validation that must precede the calculation of prevalence estimates.

The Food and Agriculture Organization of the United Nations (FAO) produces two FIES indicators for global monitoring: moderate and severe food insecurity and severe food insecurity. In addition to (and the reason for) being a Feed the Future indicator, the moderate and severe food insecurity indicator is a monitoring indicator for Sustainable Development Goal 2 (Target 2.1). For further guidance, the analyst should review the guidance available at <http://www.fao.org/in-action/voices-of-the-hungry/using-fies/en/>.

There are 12 main steps in the step-by-step procedures (see Section 15.2 for details) that cover the key components in working with the FIES, including how to run the Rasch model and assess model fit, calibration of item severity scores to ensure a common scale between Rasch model outputs from multiple surveys, calculation of the prevalence of moderate and severe food insecurity at the aggregate and disaggregate levels, and suggestions for exporting data.

- Step 1: Set up the R environment and read the data into R
- Step 2: Prepare the data and determine how to handle missing data
- Step 3: Run the Rasch model and assess the model fit
- Step 4: Calibrate the item severity scores between Midline and Baseline Surveys<sup>100</sup>
- Step 5: Calibrate the FAO global scale to reference the Baseline Survey scale to obtain moderate and severe food insecurity thresholds<sup>101</sup>
- Step 6: Calculate the prevalence of moderate, severe, and moderate and severe food insecurity
- Step 7: Create binary variables that indicate moderate, severe, and moderate and severe food insecurity

---

<sup>100</sup> If not performing a baseline-midline comparative analysis, this step is not relevant.

<sup>101</sup> If not performing a baseline-midline comparative analysis, the Midline Survey is the reference scale.

- Step 8: Calculate the margins of error (MOE) and conduct tests of significance to assess change over time
- Step 9: Store the aggregate results in a common data frame
- Step 10: Calculate food insecurity prevalence estimates for disaggregate variables
- Step 11: Export dataset or results in CSV format
- Step 12: Export dataset or results in Stata format

The main statistical model used for FIES data assessment and scale construction is an Item Response Function known as the Rasch model. This model is used in Step 3. The Rasch model was developed in educational testing to score a latent trait of ability based on a student responding correctly to a series of questions. In applying the model to food insecurity, the scale is defined as a scale of “severity,” or the degree of negative impact on the household’s or individual’s welfare due to the inability to freely access the food one needs. This is a probabilistic model that scores households based on the number ‘yes’ responses to a series of questions related to the experience of food insecurity. According to this model, the probability of a respondent affirming the  $j$ -th item is modeled as a logistic function of the distance between two parameters, one representing the item’s severity ( $b_j$ ) and one representing the respondent’s severity ( $\theta$ ) (Cafiero, Viviani & Nord, 2017; Onori, Viviani, & Brutti, 2021):

$$P_j(\theta) = P(X_j = 1 | \theta; b_j) = \frac{\exp(\theta - b_j)}{1 + \exp(\theta - b_j)}$$

In Steps 4 and 5, scale calibration is done to ensure that a common scale for the item and raw severity scores is used for different surveys. This is a necessary step to ensure that alike comparisons are being made between different contexts and is particularly important if working with Midline Survey data in comparison to Baseline Survey data. The item and raw severity scores of the Baseline Survey are typically used as the reference scale, and the Midline Survey is adjusted to fit the baseline scale because the baseline was done first.

When performing the calibration, the item severity scores for the two surveys are first compared to identify common items—that is items with similar severity scores for the two surveys. Then the adjusted midline item severity scores ( $z_i$ ) are obtained by subtracting the mean of the midline common item severity scores ( $\bar{x}$ ) from each midline item severity score ( $x_i$ ) and dividing by the product of the population standard deviations (SD) of the common item severity scores for baseline and midline ( $\sigma_y \times \sigma_x$ ) plus the mean of the baseline common item severity scores ( $\bar{y}$ ).

The mean of the midline common item severity scores ( $\bar{x}$ ) is calculated as follows:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Where:

$x_i$  = the severity score of item  $i$

$N$  = the number of common items

The mean of the baseline common item severity scores ( $\bar{y}$ ) is calculated as follows:



$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

Where:

$y_i$  = the severity score of item i

The population SD of the midline common item severity scores ( $\sigma_x$ ) is calculated as follows:

$$\sigma_x = \sqrt{\frac{(x_i - \bar{x})^2}{N}}$$

The population SD of the baseline common item severity scores ( $\sigma_y$ ) is calculated as follows:

$$\sigma_y = \sqrt{\frac{(y_i - \bar{y})^2}{N}}$$

The adjusted midline item severity scores ( $z_i$ ) are calculated as follows, using the values generated in the previous equations:

$$z_i = \frac{x_i - \bar{x}}{\bar{y} + (\sigma_y \times \sigma_x)}$$

The same calculation is used for the adjusted raw score severity scores ( $s_i$ ), in which we substitute the individual raw score severity scores ( $r_i$ ) for the item severity scores ( $x_i$ ).

$$s_i = \frac{r_i - \bar{x}}{\bar{y} + (\sigma_y \times \sigma_x)}$$

The adjusted SD of the raw score severity scores ( $\sigma_s$ ) is calculated by dividing each one by the product of the SD of the common item severity scores for baseline and midline:

$$\sigma_s = \frac{\sigma_r}{(\sigma_y \times \sigma_x)}$$

The same set of equations are used for the calibration of the global scale onto the local scale in Step 5.<sup>102</sup> The FAO global scale comprises the item severity scores obtained from a series of surveys done in multiple countries for the Gallup World Poll from 2014 to 2016. The fifth item (**ATELESS**) and eighth item (**WHLDAY**) of the global scale serve as the thresholds for moderate and severe food insecurity and must be mapped onto the local reference scale, whether baseline or midline. For this reason, the global scale is adjusted to fit onto the local reference scale, as was done with adjusting the midline to the baseline.

<sup>102</sup> The local scale refers to the scale used as the anchor for analysis. The baseline serves as the local scale if calibration between the baseline and midline for comparative analysis is done. If no baseline is used in analysis, then the midline serves as the local scale.

Obtaining the adjusted global scale item severity scores ( $z_i$ ) is done by subtracting the mean of the global scale common item severity scores ( $\bar{g}$ ) from each global scale item severity score ( $g_i$ ) divided by the product of the population SD of the common item severity scores for the global scale onto the local scale ( $\sigma_g \times \sigma_l$ ) plus the mean of the local common item severity scores ( $\bar{l}$ ).

The mean of the common global scores ( $g_i$ ) is calculated as follows:

$$\bar{g} = \frac{1}{N} \sum_{i=1}^N g_i$$

Where:

$g_i$  = the severity score of item i

The mean of the common local scores ( $l_i$ ) is calculated as follows:

$$\bar{l} = \frac{1}{N} \sum_{i=1}^N l_i$$

Where:

$l_i$  = the severity score of item i

The SD of the common global scores ( $\sigma_g$ ) is calculated as follows:

$$\sigma_g = \sqrt{\frac{(g_i - \bar{g})^2}{N}}$$

The SD of the common local scores ( $\sigma_l$ ) is calculated as follows:

$$\sigma_l = \sqrt{\frac{(l_i - \bar{l})^2}{N}}$$

The adjusted global reference scale item severity scores ( $w_i$ ) are calculated as follows, using the values generated in the previous equations:

$$w_i = \frac{g_i - \bar{g}}{\bar{l} + (\sigma_l \times \sigma_g)}$$

## 15.2 Step-by-step procedures to calculate the FIES indicator

This section describes the detailed step-by-step procedures to calculate the prevalence of moderate or severe food insecurity in the population, based on the FIES. Several suggestions in the steps are made in the context of working with Midline Survey data alongside Baseline Survey data or multiple surveys done sequentially in which one can serve as a reference point. If this context does not fit your analysis needs, then it is recommended either to skip particular sections or consult the materials provided on the FAO website in the link provided in Section 15.1.

## Definitions

Numerator	Number of households with a probability of exceeding the moderate food insecurity severity level
Denominator	Total number of surveyed households
Unit of measure	Percentage
Level of data	Household
Sampling weight	Household
Disaggregation levels	Gendered household type* Residency (urban/rural) Wealth quintile Shock exposure severity
Treatment of missing data	Households with missing or refused values for all FIES items are excluded from the indicator calculations. Instructions are provided on handling missing values on the FIES items in Step 2d.
Survey variables used	v301, v302, v303, v304, v305, v306, v307, v308, hhea, hhnum, hhsize_dj, wgt_hh, ahtype, strata
Analytic variables used	genhhtype_dj, awiquint, shock_sev
Analytic variables created	wt, WORRIED, HEALTHY, FEWFOOD, SKIPPED, ATELESS, RUNOUT, HUNGRY, WHLDAY, prob_mod, prod_mod_sev, prob_sev, fi_mod_sev_bin, fi_mod_bin, fi_sev_bin

\*Standard Feed the Future disaggregate

## Calculations

FAO's Voices of the Hungry project provides a free analytical tool for FIES analysis using R software.

Following the FAO technical resources, this guide describes the steps required to compute the prevalence of moderate food insecurity, severe food insecurity, and moderate or severe food insecurity based on the FIES scale. It also describes how to compare FIES estimates over time for overall prevalence and disaggregates.

**Step 1.** Download R, install the required packages, set a working directory, and read the data into R.

**Step 1a.** Go to <https://cran.r-project.org/> to download R. After it is downloaded, R can be used directly in the console or through a user-friendly compiler, RStudio, which needs to be downloaded separately.

**Step 1b.** Go to <https://www.rstudio.com/products/rstudio/download/> to download RStudio, which is an integrated development environment for R. It includes a console, a syntax-highlighting editor that supports direct code execution, and tools for plotting, tracking history, debugging, and managing the workspace.

**Step 1c.** Install and upload the required packages.

Install the “**RM.weights**,” “**survey**,” “**haven**,” and “**tidyverse**” packages if not already installed.

```
install.packages("RM.weights") # for Rasch modeling
install.packages("survey") # for analysis of complex survey
samples
```

```
install.packages("haven") # to read data stored in Stata, SPSS,
or other software
install.packages("tidyverse") # for data manipulation
```

After the packages are installed, upload the packages in the working library.

```
library(RM.weights)
library(survey)
library(haven)
library(tidyverse)
```

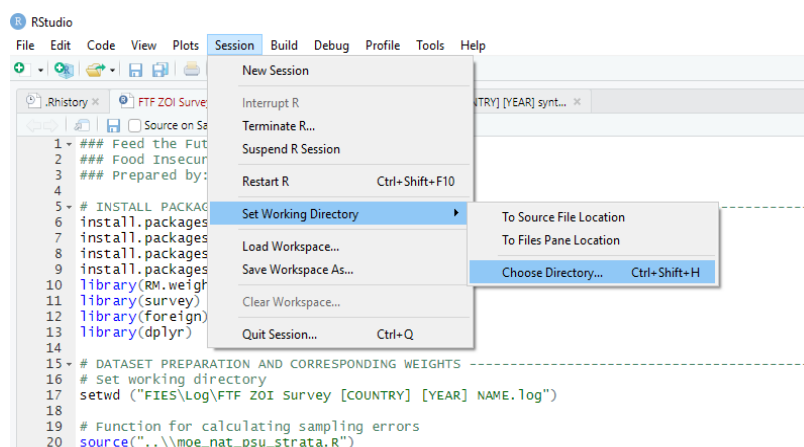
Note that you only need to install the packages above one time. However, each time before you use the FIES code, you need to upload the packages using the `library` function.

**Step 1d.** Specify the working directory in which the data are stored. This will serve as the root directory.

```
setwd("FIES\\FTF ZOI Survey [COUNTRY] [YEAR] NAME")
```

The working directory should be changed to reflect the directory path and folder names. Alternatively, go to RStudio console to set the working directory. Click on Session, and then Set Working Directory and Choose Directory (Figure 1).

**Figure 1: Setting up a Working Directory in R Studio**



**Step 1e.** Load in script to calculate the MOE.

Read the source code of the `moe` function to calculate the CI and DEFF. Before running the code, ensure that the “`moe_nat_psu_strata_v2.R`” file is located in the working directory, and call the `moe` function into the code.<sup>103</sup>

```
source("moe_nat_psu_strata_v2.R")
```

<sup>103</sup> This script is available in the Midline Toolkit.

The **moe** function facilitates the calculations of the CI and DEFF using the survey package of R, which accounts for the complex survey design, such as clustering and unequal probabilities of selection, by specifying the primary sampling unit, strata, and sampling weights.

**Step 1f.** Create functions to compute the population-level SD and variance.

```
#create a new function to compute population standard deviation
var.p <- function(x) var(x) * (length(x)-1) / length(x)
sd.p <- function(x) sqrt(var.p(x))
```

The population SD is used in later steps to calibrate different scales created from the Rasch model and takes into account the entire population.

**Step 1g.** Use the **haven** package to read in the ZOI Survey household-level analytic data file with the Stata labels preserved.

```
ftfzoi_midline_import <- read_dta("FTF ZOI Survey [COUNTRY]
[YEAR] household data analytic.dta")
```

**Step 2.** Prepare the data.

**Note:** For simplicity, Steps 2 and 3 show only preparation and analysis with the Midline Survey data. If this analysis is done alongside Baseline Survey data for comparative analysis, be sure to follow these steps for the baseline data as well. Variable names in the baseline data may differ.

**Step 2a.** Create a data frame (**ftfzoi\_midline**) in which the FIES survey variables are renamed to the corresponding items that will be used in the Rasch model using the **rename** function.

```
ftfzoi_midline <-
  ftfzoi_midline_import %>%
  rename(
    "WORRIED" = "v301",
    "HEALTHY" = "v302",
    "FEWFOOD" = "v303",
    "SKIPPED" = "v304",
    "ATELESS" = "v305",
    "RUNOUT" = "v306",
    "HUNGRY" = "v307",
    "WHLDAY" = "v308")
```

**Step 2b.** Recode all variables in the **ftfzoi\_midline** data frame created in Step 2a. Keep 'YES' responses coded as '1,' and recode 'NO' responses as '0.' 'REFUSED' responses ('7') will be set to missing (NA) and excluded from the analysis.

```
ml_hh_midline <- ml_hh_midline %>% mutate_at(
  dplyr::funs(case_when(
```

```

      x == 1 ~ 1,
      x == 2 ~ 0
    )
  )

```

**Step 2c.** Store the complex survey design variables and disaggregate variables that will be used later for analysis. Remember to edit the variable names, values, and labels as needed.

```

# variables used for complex survey design - weights, household
# size, cluster ID, HHID, strata
ftfzoi_midline$wt <- ftfzoi_midline$wgt_hh
ftfzoi_midline$hhsz_dj <- ftfzoi_midline$hhsz_dj
ftfzoi_midline$hhnum <- ftfzoi_midline$hhnum
ftfzoi_midline$strata <- ftfzoi_midline$strata

# Disaggregates - gendered HH type, residency, shock severity,
# wealth index
ftfzoi_midline$ahtype <- factor(ftfzoi_midline$ahtype,
  levels=c(1,2), labels=c("urban","rural"))
ftfzoi_midline$genhhtype_dj <-
  factor(ftfzoi_midline$genhhtype_dj,
  levels=c(1,2,3,4), labels=c("Male and Female adults","Female
  adults only","Male adults only", "Children only"))
ftfzoi_midline$shock_sev <- factor(ftfzoi_midline$shock_sev,
  levels = c(1,2,3,4), labels = c("None", "Low", "Medium",
  "High"))
ftfzoi_midline$awiquint <- factor(ftfzoi_midline$awiquint, levels
  = c(1,2,3,4,5), labels = c("Poorest", "Second", "Middle",
  "Fourth", "Wealthiest"))

# select variables
ftfzoi_midline <- ftfzoi_midline %>%
  select(wt, mem, hhea, hhnum, strata, ahtype, genhhtype_dj,
    shock_sev, awiquint, WORRIED:WHLDAY)

```

**Step 2d.** Treat missing values on the FIES items.

There are two types of missing values (NA) that may be encountered: (1) all FIES items are missing values, and (2) individual FIES items are missing values but other FIES items have values.

(1) If values are missing for all FIES items, remove the household from the analysis entirely.

```

# remove HHs where all FIES items are missing
ftfzoi_midline <- ftfzoi_midline %>%
  filter_at(dplyr::vars(WORRIED:WHLDAY),
    dplyr::all_vars(!is.na(.)))

```

(2) More caution is required handling individual item non-response. Imputation of values is recommended only in cases in which a pattern is clear. For example, if a household has values of “NO” (0) for all items, except for **ATELESS**, which has a missing value, **ATELESS** can be safely imputed as 0.

To explore the missing values in the data, use the following code:

```
ftfzoi_midline %>%  
  filter_at(dplyr::vars(WORRIED:WHLDAY) ,  
dplyr::any_vars(is.na(.)))
```

To impute a value if **ATELESS** is missing a value and all other FIES items have a value of “NO” (0), the following code may be used:

```
ftfzoi_midline <- ftfzoi_midline %>%  
dplyr::mutate(ATELESS = case_when(WORRIED == 0 & HEALTHY == 0 &  
  FEWFOOD == 0 & SKIPPED == 0 & RUNOUT == 0 & HUNGRY == 0 &  
  WHLDAY == 0 & is.na(ATELESS) ~ 0, TRUE ~  
  as.numeric(ATELESS)))
```

This code identifies households that follow the exact item value pattern specified and assigns 0 to **ATELESS**. For other households that do not have this item value pattern, it keeps **ATELESS** at the original value. Similar code can be used for other items with missing values in cases in which all other items are all “NO” responses.

Another example of acceptable imputation of individual item response is when a household has values of “YES” (1) for less and more severe items surrounding the item with a missing value; this item can be safely imputed as 1.

Assigning values in less absolute cases is not recommended. However, if analysts find themselves working with a high number of missing values, they may want to. When assigning values, it is important to pay close attention to the hierarchy of the items, and where evident patterns emerge, such as a missing value between two “YES” values for the next most severe and next least severe items. To assess item severity patterns without running the Rasch model, calculate the proportion of households with a value of “YES” for each FIES item and order the items by increasing proportion. These proportions will largely follow the item severity scores generated in Step 3 and can provide a sufficient idea of the severity ranking of the FIES items and can be used to justify imputation of individual item missing values in similar cases.

Because the FIES items were converted to binary variables in Step 2b, the **colMeans** function can be used to calculate column proportions (i.e., the proportion of households with a value of “YES” for each FIES item):

```
ftfzoi_midline %>%  
  select(WORRIED:WHLDAY) %>%  
  colMeans()
```

The output will look like this:

WORRIED	HEALTHY	FEWFOOD	SKIPPED	ATELESS
0.37320770	0.22408849	0.18189267	0.04547317	0.10651372
RUNOUT	HUNGRY	WHLDAY		
0.08152397	0.05489553	0.02458009		

The percentages will usually decrease with each subsequent item, although each survey has its own exceptions due to differences in linguistic interpretation and behavior patterns within the surveyed population. In the example above, SKIPPED has the second lowest percentage of “YES” responses among households.

**Step 2e.** Create a data frame (**FIES\_midline**) that includes only the binary FIES items after all imputation has been completed using the **select** function:

```
FIES_midline <- ftfzoi_midline %>%
  select(WORRIED:WHLDAY)
```

**Step 3.** Run the Rasch model to obtain severity scores used for calculating food insecurity prevalence and diagnostic test results to assess model fit.

The Rasch model (1) assesses the suitability of the eight FIES items for scale construction, (2) computes parameter estimates and assessment statistics for each item, (3) generates a scale from the items, and (4) assesses the location of a household along the continuum of the scale that uniquely reflects the food insecurity situation of that household. The model is critical for Step 4, in which item severity scores and the FAO global scale are used to produce the moderate and severe food insecurity prevalence rates.

**Step 3a.** Run the Rasch model on the selected FIES items and households after data preparation and handling missing data, using the data frame created in Step 2e (**FIES\_midline**). Use the **RM.w** function of the **RMweights** package to run the model.

```
FIES_midline_rasch <- RM.w(FIES_midline, write.file = TRUE,
  country = "FTF ZOI SURVEY COUNTRY")
FIES_midline_RS = rowSums(FIES_midline)
```

The **RM.w** function has a **weights** argument that is optional. Given that Rasch models are invariant to population-based weights, they are not used for ZOI Survey analyses.<sup>104</sup> The function automatically exports the Rasch output to a CSV file saved in the working directory. These results can be used to test the quality of data collected.

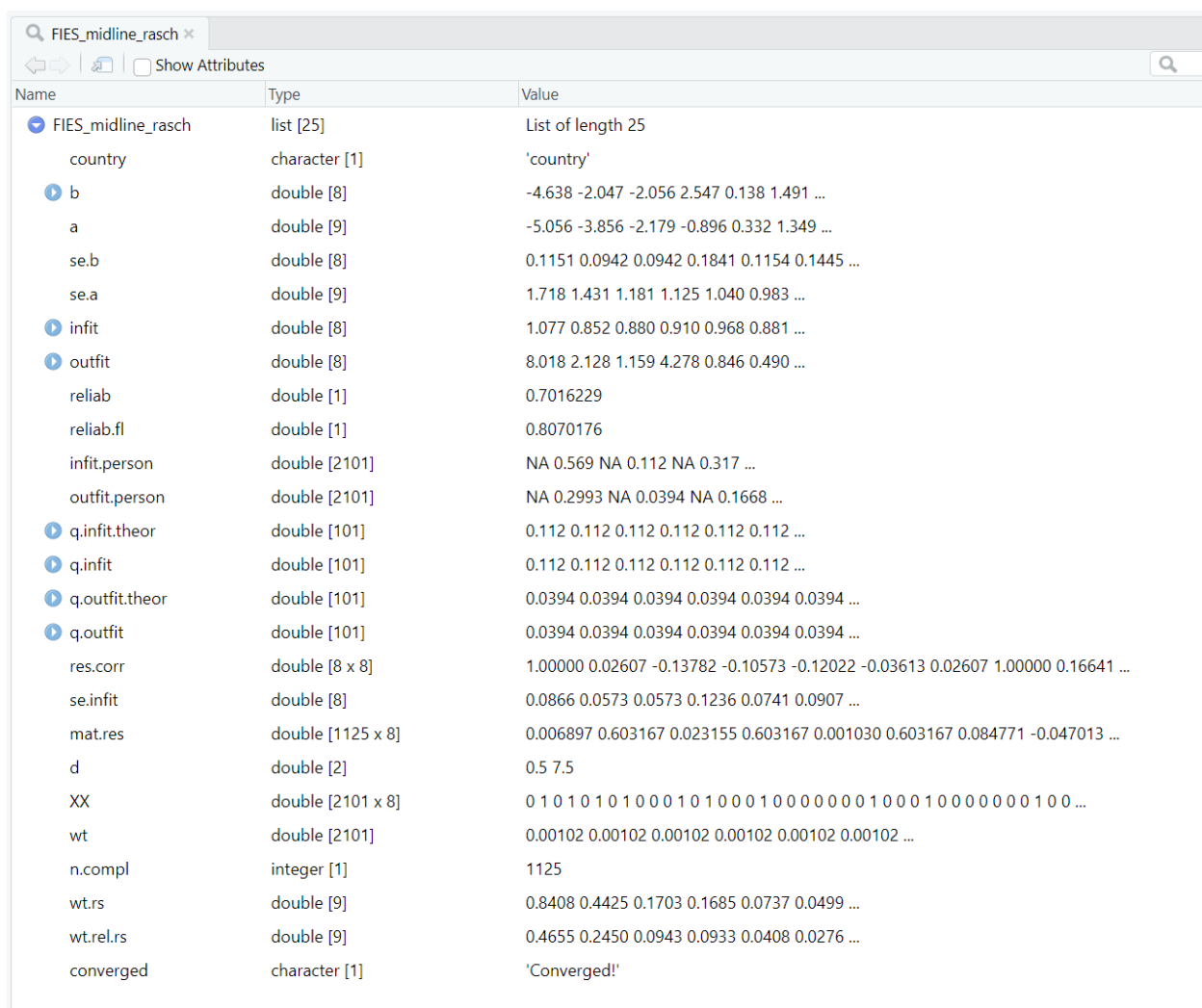
The Rasch model will also be saved in a new object in the R environment (**FIES\_midline\_rasch**) that contains several important outputs (see Figure 2), including the raw score severity scores (**FIES\_midline\_rasch\$a**), item severity scores (**FIES\_midline\_rasch\$b**), the correlation of residuals among items (**FIES\_midline\_rasch\$res.corr**), a residual matrix (**FIES\_midline\_rasch\$mat.res**), the distribution of valid responses, missing data by item, and a detailed output that shows the observed response proportion and the predicted response

<sup>104</sup> If required to include, add the **.wt = x** argument to the function above.



proportion for each raw score and items. In addition, the output includes the results of several important diagnostic tests that are useful to assess the measurement reliability and quality of data.

**Figure 2: Example Rasch Module Output**



Name	Type	Value
FIES_midline_rasch	list [25]	List of length 25
country	character [1]	'country'
b	double [8]	-4.638 -2.047 -2.056 2.547 0.138 1.491 ...
a	double [9]	-5.056 -3.856 -2.179 -0.896 0.332 1.349 ...
se.b	double [8]	0.1151 0.0942 0.0942 0.1841 0.1154 0.1445 ...
se.a	double [9]	1.718 1.431 1.181 1.125 1.040 0.983 ...
infit	double [8]	1.077 0.852 0.880 0.910 0.968 0.881 ...
outfit	double [8]	8.018 2.128 1.159 4.278 0.846 0.490 ...
reliab	double [1]	0.7016229
reliab.fl	double [1]	0.8070176
infit.person	double [2101]	NA 0.569 NA 0.112 NA 0.317 ...
outfit.person	double [2101]	NA 0.2993 NA 0.0394 NA 0.1668 ...
q.infit.theor	double [101]	0.112 0.112 0.112 0.112 0.112 0.112 ...
q.infit	double [101]	0.112 0.112 0.112 0.112 0.112 0.112 ...
q.outfit.theor	double [101]	0.0394 0.0394 0.0394 0.0394 0.0394 0.0394 ...
q.outfit	double [101]	0.0394 0.0394 0.0394 0.0394 0.0394 0.0394 ...
res.corr	double [8 x 8]	1.00000 0.02607 -0.13782 -0.10573 -0.12022 -0.03613 0.02607 1.00000 0.16641 ...
se.infit	double [8]	0.0866 0.0573 0.0573 0.1236 0.0741 0.0907 ...
mat.res	double [1125 x 8]	0.006897 0.603167 0.023155 0.603167 0.001030 0.603167 0.084771 -0.047013 ...
d	double [2]	0.5 7.5
XX	double [2101 x 8]	0 1 0 1 0 1 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 ...
wt	double [2101]	0.00102 0.00102 0.00102 0.00102 0.00102 0.00102 ...
n.compl	integer [1]	1125
wt.rs	double [9]	0.8408 0.4425 0.1703 0.1685 0.0737 0.0499 ...
wt.rel.rs	double [9]	0.4655 0.2450 0.0943 0.0933 0.0408 0.0276 ...
converged	character [1]	'Converged!'

### Step 3b. Assess fit with diagnostic results.

There are four main diagnostics that must be examined thoroughly to assess model fit: infit, outfit, Rasch reliability, and principal component analysis (PCA) on the residual matrix. This step provides a brief description and guidance along with code examples on how to assess each test result. The information provided for each test is not extensive, and analysts may wish to consult other materials given the complexities of working with Rasch models. More thorough background can be found in the following article by FAO outlining statistical validation of the FIES for the Sustainable Development Goals: [https://elearning.fao.org/pluginfile.php/491591/mod\\_scorm/content/5/story\\_content/external\\_files/SDG2.1.2\\_lesson3.pdf](https://elearning.fao.org/pluginfile.php/491591/mod_scorm/content/5/story_content/external_files/SDG2.1.2_lesson3.pdf)

### Infit (and corresponding standard errors):

**Description:** The infit statistic is based on the chi-square statistic with each observation weighted by the overall model variance. As a result, infit statistics are more sensitive to unexpected patterns of observations by households on items close to their cutoff.

**Fit:** If the infit statistic is 1.0 for each item, all items discriminate equally—that is, items are strongly or consistently related to a food insecurity condition—which is one of the main assumptions of the Rasch model. An adequate fit to the Rasch model is indicated by infit and outfit statistics of 0.7-1.3 for each item. Items with infit values of less than 0.7 or greater than 1.3 should not be used for scoring.

**Guidance:** If the item infit is less than 0.7 or greater than 1.3 for any single item, a useful first step is assessing household-level (in R called `person.infit`) statistics to remove households with very low or very high infit patterns and unusual patterns on the items that have high infit scores, or where values do not follow the hierarchical logic of the item severity. If item infit scores remain low or high, consider removing the item from the model and rerunning the Rasch model without it.

Every time a household (row) or item (column) is removed, the Rasch model should be re-run. If there is an item infit value less than 0.7 or greater than 1.3, the item should be dropped from the model and re-run. If there is more than one item with an infit value of greater than 1.3 or less than 0.7, the item with the highest/lowest value should be removed and the model should be re-run. Please note that an analyst should not remove more than three items from the Rasch model.

#### R code:

```
round(FIES_midline_rasch$infit) # to see infit scores for each  
item/question
```

```
round(FIES_midline_rasch$infit.person) # to see individual infit  
scores
```

### Outfit (and corresponding standard errors):

**Description:** Outfit is an outlier-sensitive fit statistic, based on the conventional chi-square statistic. Compared to infit, outfit is more sensitive to unexpected observations for questions that are very easy or very hard to answer (i.e., a “NO” response to item 1 [WORRIED] and “YES” to all other items).

**Fit:** An item outfit of greater than 2 is considered high and may be useful for identifying where unusual patterns are located.

**Guidance:** Unlike item infit, a high item outfit score is typically not a sufficient reason to remove an item. Nevertheless, item infit and outfit should be examined together. As with infit, assessing household-level outfit statistics (in R called `person.outfit`) and potentially removing problematic response patterns can help improve measurement quality. If time permits, examining household-level outfit scores by enumerator can determine if any patterns emerge based on who collected the data.

**R code:**

```
round(FIES_rasch$outfit, digits = 2)

round(FIES_midline_rasch$outfit.person, digits = 2) # to see
individual outfit scores
```

**Rasch reliability (and SD):**

**Description:** Rasch reliability is the proportion of total variance in the ZOI population that is explained by the Rasch model. The flat Rasch reliability statistic can be thought of similarly to how R-squared is used to assess the fit of a linear regression model. The flat Rasch reliability statistic is used because it assumes that each raw score has an equal number of cases and thus provides a comparable measure of model fit, whereas the standard Rasch reliability statistic can be influenced by the total number of cases across the raw scores because it is weighted by the frequency of each raw score.

**Fit:** Like R-squared, the closer that the reliability statistic is to 1, the better the model fits the data. As a rule of thumb, the reliability statistic should be greater than or equal to 0.7 and the SD should be greater than 1.5.

**Guidance:** Usually, a low reliability score occurs because the sample size is too small to establish a reproducible item severity hierarchy. If reliability is lower than the suggested threshold (0.7), a first step is to examine missing (NA) values because they can adversely impact precision and reliability. If one individual item has 10 percent or more missing values, the item can be dropped entirely. If the threshold still is not met, it is possible that removing household observations with missing values on multiple FIES items can improve the reliability. If additional improvements are needed, removing households with high household-level outfit values (i.e., respondents who respond “YES” to items out of line with the severity of the items) can strengthen a hierarchical pattern among the items and increase the reliability score.

**R code:**

```
round(FIES_midline_rasch$reliab.fl, digits = 2) # reliability
stat

round(sd.p(FIES_midline_rasch$b), digits = 2) # calculate pop.
Standard deviation of item scores.
```

**PCA on residual matrix:**

**Description:** Examining residual variance using PCA provides information on the structure of the variance not explained by the model.

**Fit:** The residual variance should not be explained by only one or two principal components. Ideally, there should be a gradual drop-off in the residual variance. The scree plot should not be in an ‘L’ or hockey stick shape; this would suggest that the model is capturing factors that are not explained by the FIES items.

**Guidance:** If the PCA results show a gradual drop-off in the residual variance, analysts can be confident in their results. If the PCA results show a sharp drop-off in residual variance after one or two principal components, the analyst cannot do anything to make the data “better,” but the results should be documented as part of the data quality assessment process.

**R code:**

```
plot(prcomp(FIES_midline_rasch$mat.res)$sd^2)
```

**Final considerations:**

- Although it is not one of the four main diagnostics included in this Guide, reviewing the residual correlation between items can be useful to see if the items successfully capture different aspects of food insecurity. A residual correlation between a pair of items is considered high if it is greater than 0.3. If two items have high residual correlation, it may be necessary to drop one of them and re-run the model.
- Any item or household removed from the Rasch model should be re-run one by one. For example, if there are two items with an infit scores greater than 1.3, the item with the highest score should be removed and the other item with a score greater than 1.3 left in. This is because removing the highest item can lead to a changes in the scores of other items. If the other items are still greater than 1.3, then it is suggested that the other item with a score greater than 1.3 be dropped as well.
- A minimum of five items are required to proceed with analysis; therefore, do not remove more than three items.
- Removal of items or households can be done by calling the column or row number in brackets with the data frame. In the brackets, rows are placed first (before the comma) and columns second (after the comma). Place a minus sign in front of the column or row to be removed.

```
# item (column) removal - WORRIED
FIES_midline <- FIES_midline[, -1]
FIES_midline_rasch <- RM.w(as.matrix(FIES_midline), write.file =
  TRUE, country = "FTF ZOI SURVEY COUNTRY")
# row (household) removal - HH ar row number 30
FIES_midline <- FIES_midline[-30, ]
FIES_midline_rasch <- RM.w(as.matrix(FIES_midline), write.file =
  TRUE, country = "FTF ZOI SURVEY COUNTRY")
```

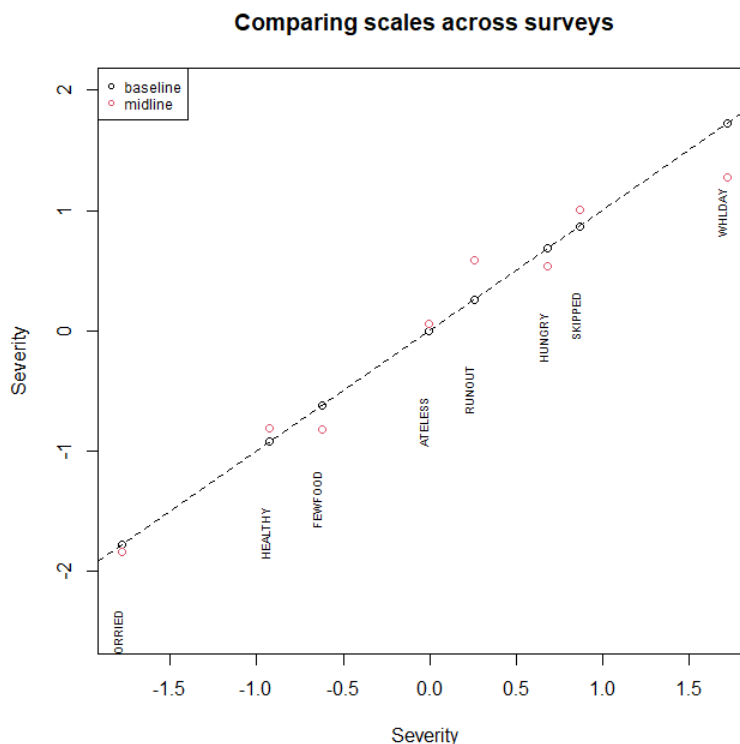
**Step 4.** Use equating processes to adjust the midline data so that they are on the same local scale as the baseline data. The local scale will be the baseline because it serves as the anchor for the analysis. Creating a common scale is crucial to ensuring that prevalence estimates are accurate. A meteorologist would not use two different temperature measurement scales to compare the weather of two different regions. If analyzing only midline data without comparing to baseline data, skip to Step 5.

Before continuing with this step, be sure to create a baseline Rasch model output stored in a list (**FIES\_baseline\_rasch**), following the same steps used to create the midline Rasch object (**FIES\_midline\_rasch**).

**Step 4a.** Create a graph to compare baseline and midline item severity scores divided by the item SD.

```
{
  x <- FIES_baseline_rasch$b/sd.p(FIES_baseline_rasch$b)
  y_bl <- FIES_baseline_rasch$b/sd.p(FIES_baseline_rasch$b)
  y_ml <- FIES_midline_rasch$b/sd.p(FIES_midline_rasch$b)
  plot(x,y_bl, col = 1, ylim = c(-
    2.5,2),ylab="Severity",xlab="Severity")
  points(x, y_ml, col = 2)
  text(x+0.04,x-0.55,colnames(FIES),cex=0.6,pos=2,srt=90)
  abline(0,1,lty = 2)
  legend("topleft",c("baseline","midline"),pch = 1, col = c(1,2),
    cex = 0.75)
  title(main = "Comparing scales across surveys")
}
```

This code produces a graph similar to this:



Look at the difference between midline and baseline item severity scores for each item.

```
abs(round(FIES_baseline_rasch$b/sd.p(FIES_baseline_rasch$b) -
  FIES_midline_rasch$b/sd.p(FIES_midline_rasch$b), 2))
```

Example output:

WORRIED	HEALTHY	FEWFOOD	SKIPPED	ATELESS	RUNOUT	HUNGRY	WHLDAY
0.06	0.11	0.19	0.14	0.06	0.36	0.15	0.44

To equate the baseline and midline data, identify the items that will be included in the common scale. To determine common items, a threshold of 0.35 is frequently used. That is, items with a midline-baseline item severity score difference of 0.35 or less will be considered common items and included in the common scale. A minimum of five common items are required. In the example above, item 6 (**RUNOUT**) and item 8 (**WHLDAY**) have item severity score differences greater than 0.35, and in the graph, the midline values (in red) are farther away from the dotted line, which runs through the baseline items. Therefore, items 6 and 8 will be excluded from the common scale.

**Step 4b.** Create vectors with common items to be used to calculate the mean and SD of the common item severity scores for both baseline and midline.

```
# identify items in common - one where scores are similar
not_common <- c("WHLDAY","RUNOUT")
common <- setdiff(colnames(FIES_midline),not_common)
# mean of common item scores
m.bl <- mean(res_bl$b[common])
m.ml <- mean(res_ml$b[common])
# sd of common item scores
s.bl <- sd.p(res_bl$b[common])
s.ml <- sd.p(res_ml$b[common])
```

**Step 4c.** Adjust the midline item severity scores (**FIES\_midline\_rasch\$b**) and raw score severity scores (**FIES\_midline\_rasch\$a**) so they are on the same scale as the baseline scores. To obtain the adjusted midline scores, the common item severity scores of the midline are subtracted by the mean of the common item severity scores. This is divided by the product of the population SD of the common item severity scores for the baseline and midline plus the mean of the baseline common item severity scores. This calibration process is outlined in greater detail in Section 15.1.

```
# adjust midline scale to the baseline metric
# assign Midline rasch object to Adjusted Midline rasch object
FIES_midline_rasch_adj <- FIES_midline_rasch
# adjusted mean of item severity scores
FIES_midline_rasch_adj$b <- (FIES_midline_rasch$b -
  m.ml)/s.ml*s.bl + m.bl
# adjusted mean of raw score severity scores
FIES_midline_rasch_adj$a <- (FIES_midline_rasch$a -
  m.ml)/s.ml*s.bl + m.bl
# adjusted standard deviation of raw scores
FIES_midline_rasch_adj$se.a <-
  FIES_midline_rasch_adj$se.a/s.ml*s.bl
```

**Step 4d.** Create a matrix with the weighted distributions of baseline and midline raw scores to use for the prevalence calculations.

```
RS_table <- t(cbind(
  "Baseline" = aggregate(ftfzoi_baseline$wt,
    list(FIES_baseline_RS), FUN=sum, na.rm=TRUE)$x
    /sum(ftfzoi_baseline$wt [!is.na(FIES_baseline_RS)]),
  "Midline" = aggregate(ftfzoi_midline$wt, list(FIES_midline_RS),
    FUN=sum, na.rm=TRUE)$x /sum(ftfzoi_midline$wt
    [!is.na(FIES_midline_RS)]))
))
```

**Step 5.** Map the global scale onto the local baseline scale to obtain thresholds for moderate and severe food insecurity, ensuring that measurement is consistent across datasets.

To calculate moderate and severe food insecurity as defined by the international standards of FAO, households must be assigned to a level of food insecurity as defined by set thresholds on the item severity scale. FAO has created a global standard of item parameter values based on survey results from more than 140 countries covered by the Gallup World Poll in 2014-2016, whereby the threshold for moderate food insecurity is set at the severity of the fifth item (**ATELESS**) and the threshold for severe food insecurity is set at the severity of the eighth item (**WHLDAY**) (FAO, 2015).

The calibration process is similar to the process used to equate baseline and midline item severity scores.

**Note:** In this step, the baseline scale is the local scale used. As such, the midline was adjusted to the baseline scale in Step 4. If a baseline is not included in the analysis, the midline can be used as the local scale.

**Step 5a.** Define the local (baseline) and the global scales.<sup>105</sup>

```
loc_st <- FIES_baseline_rasch$b
glob_st <- c("WORRIED"= -1.2230564, "HEALTHY"= -0.847121,
  "FEWFOODS"= -1.1056616, "SKIPPED" = 0.3509848, "ATELESS" =
  -0.3117999, "RUNOUT" = 0.5065051, "HUNGRY" = 0.7546138,
  "WHLDAY" = 1.8755353)
```

**Step 5b.** Obtain the absolute value of the difference of the standardized versions of the local and global scales.

```
# standardized version of both scales - item severity score
  divided by the population SD
```

<sup>105</sup> The values of the global scale and additional details on their construction can be found at: <https://doi.org/10.1016/j.measurement.2017.10.065>

```
abs(round(loc_st/sd.p(loc_st) - glob_st/sd.p(glob_st), 2))
```

Example output:

WORRIED	HEALTHY	FEWFOOD	SKIPPED	ATELESS	RUNOUT	HUNGRY	WHLDAY
0.55	0.08	0.28	0.31	0.52	0.25	0.07	0.34

To calibrate the local and global scales, identify the items that will be included in the common scale. To determine common items, use a threshold of 0.35 as was done for equating the baseline and midline data—that is, items with a local-global item severity score difference of 0.35 or less will be considered common items and included in the common scale. In the example above, items 1 (**WORRIED**), and 5 (**ATELESS**) have item severity score differences greater than 0.35 and would, therefore, be excluded from the common scale.

**Step 5c.** Calculate the mean and SD of the common item severity scores for both the global and local scales.

```
# produce mean and sd for each scale among common items used in
  equating
# columns 2,3,4,6,7,8
glob_st.m <- mean(glob_st[c(2:4, 6:8)])
glob_st.s <- sd.p(glob_st[c(2:4, 6:8)])
m.bl <- mean(loc_st[c(2:3, 6:8)])
s.bl <- sd.p(loc_st[c(2:3, 6:8)])
```

**Step 5d.** Adjust the global scale to fit the local scale. The thresholds are defined as items 5 and 8 of the adjusted scale based on FAO's thresholds for moderate and severe food insecurity.

```
# mapping the thresholds from the global scale onto the local
  (baseline) scale
glob_st_adj <- (glob_st - glob_st.m)/(glob_st.s * s.bl + m.bl)
newthres <- glob_st_adj[c(5,8)]
```

**Step 6.** Calculate the prevalence of moderate and severe food insecurity, severe food insecurity, and moderate food insecurity.

For this, the **pnorm** function uses the normal distribution to assess the probability of a household being over the moderate and severe thresholds given the threshold level, mean raw score severity score, and SD of the raw score severity scores.

**Step 6a.** Assign a probability to each household that it is beyond the moderate food insecurity threshold and then calculate the prevalence of moderate and severe food insecurity using matrix multiplication of each household's score probabilities against the frequency of raw scores in the sample:

```
# moderate+severe FI
# midline
```



```

glo_probs_ml_mod_sev <- 1-pnorm(newthres[1], mean =
  FIES_midline_rasch_adj$a, sd = FIES_midline_rasch_adj$se.a)

glo_prev_ml_adj_mod_sev <- glo_probs_ml_mod_sev %*%
  FIES_midline_RS[2,]

#baseline (if needed)
glo_probs_bl_mod_sev <- 1-pnorm(newthres[1], mean =
  FIES_baseline_rasch$a, sd = FIES_baseline_rasch$se.a)

glo_prev_bl_mod_sev <- glo_probs_bl_mod_sev %*%
  FIES_baseline_RS[1,]

```

**Step 6b.** Follow a similar process to calculate the prevalence of severe food insecurity.

```

# severe FI
# midline
glo_probs_ml_sev <- 1-pnorm(newthres[2], mean =
  FIES_midline_rasch_adj$a, sd = FIES_midline_rasch_adj$se.a)

glo_prev_ml_adj_sev <- glo_probs_ml_sev %*% FIES_midline_RS[2,]

#baseline (if needed)
glo_probs_bl_sev <- 1-pnorm(newthres[2], mean =
  FIES_baseline_rasch$a, sd = FIES_baseline_rasch$se.a)

glo_prev_bl_sev <- glo_probs_bl_sev %*% FIES_baseline_RS[1,]

```

**Step 6c.** Calculate the prevalence of moderate food insecurity by obtaining the difference between the moderate and severe food insecurity prevalence and severe food insecurity prevalence.

```

# create moderate only <- mod+sev - sev
# prevalence
glo_prev_bl_mod <- glo_prev_bl_mod_sev - glo_prev_bl_sev
glo_prev_ml_adj_mod <- glo_prev_ml_adj_mod_sev -
  glo_prev_ml_adj_sev

# probability of moderate food insecurity at each raw score
glo_probs_bl_mod <- glo_probs_bl_mod_sev - glo_probs_bl_sev
glo_probs_ml_mod <- glo_probs_ml_mod_sev - glo_probs_ml_sev

```

**Step 6d.** Add the prevalence results to a common object and assign probabilities to each household in the original data frame read in Step 1g.

```

# put into common object

```

```

glo_prev_bl <- c(glo_prev_bl_mod_sev, glo_prev_bl_mod,
  glo_prev_bl_sev)
glo_prev_ml <- c(glo_prev_ml_adj_mod_sev, glo_prev_ml_adj_mod,
  glo_prev_ml_adj_sev)

#Attaching probabilities to each case/HH
ftfzoi_baseline$prob_mod_sev <- NULL
ftfzoi_midline$prob_mod_sev <- NULL

ftfzoi_baseline$prob_mod <- NULL
ftfzoi_midline$prob_mod <- NULL

ftfzoi_baseline$prob_sev <- NULL
ftfzoi_midline$prob_sev <- NULL

for (rs in 0:8) {
  ftfzoi_baseline$prob_mod[ftfzoi_baseline$RS == rs] =
    glo_probs_bl_mod[rs+1]
  ftfzoi_midline$prob_mod[ftfzoi_midline$RS == rs] =
    glo_probs_ml_mod[rs+1]
  ftfzoi_baseline$prob_mod_sev[ftfzoi_baseline$RS == rs] =
    glo_probs_bl_mod_sev[rs+1]
  ftfzoi_midline$prob_mod_sev[ftfzoi_midline$RS == rs] =
    glo_probs_ml_mod_sev[rs+1]
  ftfzoi_baseline$prob_sev[ftfzoi_baseline$RS == rs] =
    glo_probs_bl_sev[rs+1]
  ftfzoi_midline$prob_sev[ftfzoi_midline$RS == rs] =
    glo_probs_ml_sev[rs+1]
}
table(ftfzoi_baseline$prob_mod, RS, useNA = "ifany")
table(ftfzoi_baseline$prob_mod_sev, RS, useNA = "ifany")
table(ftfzoi_baseline$prob_sev, RS, useNA = "ifany")

```

**Step 7.** Now that each household has a probability assigned for each level of severity, an optional step is to assign binary variables to the different food insecurity levels. This can be helpful when trying to use food insecurity as a disaggregate in the analysis of other variables. To do this, use an **ifelse** statement that assigns 1 if the unit is .5 or over and 0 if the unit is below.

```

Ftfzoi_midline$fi_mod_sev_bin <-
  ifelse(ftfzoi_midline$prob_mod_sev <= .5, 1, 0)
ftfzoi_midline$fi_mod_bin <- ifelse(ftfzoi_midline$prob_mod <=
  .5, 1, 0)
ftfzoi_midline$fi_sev_bin <- ifelse(ftfzoi_midline$prob_sev <=
  .5, 1, 0)

```

```

ftfzoi_baseline$fi_mod_sev_bin <-
  ifelse(ftfzoi_baseline$prob_mod_sev <= .5, 1, 0)
ftfzoi_baseline$fi_mod_bin <- ifelse(ftfzoi_baseline$prob_mod <=
  .5, 1, 0)
ftfzoi_baseline$fi_sev_bin <- ifelse(ftfzoi_baseline$prob_sev <=
  .5, 1, 0)

```

**Step 8.** Calculate the baseline and midline MOE and use them to detect statistically significant changes in the food insecurity estimates over time.

The MOE is calculated using the `moe` function, which was read into R in Step 1f. When calling the function, specify the confidence level (the default is 90 percent; in the example that follows, 95 percent is specified), and define the primary sampling unit and strata variables.

```

Modsev_bl_moe_95 <-
  moe(ftfzoi_baseline$prob_mod_sev, ftfzoi_baseline$RS, ftfzoi_
    baseline$wt * 10^6, conf.level = .95, psu =
    ftfzoi_baseline$psu, strata = ftfzoi_baseline$strata)$moe *
    100

modsev_ml_moe_95 <-
  moe(ftfzoi_midline$prob_mod_sev, ftfzoi_midline$RS, ftfzoi_mi
    dline$wt * 10^6, conf.level = .95, psu =
    ftfzoi_midline$psu, strata = ftfzoi_midline$strata)$moe *
    100

```

To assess statistically significant changes in food insecurity prevalence over time, from baseline to midline, subtract the baseline estimate from the midline estimate and compare the difference to the midline MOE. If the difference is greater than the midline MOE, the change is considered statistically significant at the  $p < .05$  level.

To assess statistically significant changes in food insecurity prevalence over time, from baseline to midline, subtract the baseline estimate from the midline estimate and compare the difference to the midline MOE. If the difference is greater than the midline MOE, the change is considered statistically significant at the  $p < .05$  level.

```

Fies_diff <- glo_prev_bl[1] - glo_prev_ml[1] # calculate
  difference
ifelse(fies_diff > modsev_ml_moe_95, TRUE, FALSE) # output will
  indicate whether statement is TRUE or FALSE

```

The MOE for each food insecurity level and survey can be calculated and the value can be stored into an object as follows.

```

# Baseline
modsev_moe_bl_95 <-
  moe(ftfzoi_baseline$prob_mod_sev, ftfzoi_baseline$RS, ftfzoi_
    baseline$wt * 10^6, conf.level = .95,
      psu = ftfzoi_baseline$psu, strata =
        ftfzoi_baseline$strata)$moe * 100
mod_moe_bl_95 <-
  moe(ftfzoi_baseline$prob_mod, ftfzoi_baseline$RS, ftfzoi_base
    line$wt * 10^6, conf.level = .95,
      psu = ftfzoi_baseline$psu, strata =
        ftfzoi_baseline$strata)$moe * 100
sev_moe_bl_95 <-
  moe(ftfzoi_baseline$prob_sev, ftfzoi_baseline$RS, ftfzoi_base
    line$wt * 10^6, conf.level = .95,
      psu = ftfzoi_baseline$psu, strata =
        ftfzoi_baseline$strata)$moe * 100
moe_bl_95 <- c(modsev_moe_bl_95, mod_moe_bl_95, sev_moe_bl_95)

# Midline
modsev_moe_ml_95 <-
  moe(ftfzoi_midline$prob_mod_sev, ftfzoi_midline$RS, ftfzoi_mi
    dline$wt * 10^6, conf.level = .95,
      psu = ftfzoi_midline$psu, strata =
        ftfzoi_midline$strata)$moe * 100
mod_moe_ml_95 <-
  moe(ftfzoi_midline$prob_mod, ftfzoi_midline$RS, ftfzoi_midlin
    e$wt * 10^6, conf.level = .95,
      psu = ftfzoi_midline$psu, strata =
        ftfzoi_midline$strata)$moe * 100
sev_moe_ml_95 <-
  moe(ftfzoi_midline$prob_sev, ftfzoi_midline$RS, ftfzoi_midlin
    e$wt * 10^6, conf.level = .95,
      psu = ftfzoi_midline$psu, strata =
        ftfzoi_midline$strata)$moe * 100
moe_ml_95 <- c(modsev_moe_ml_95, mod_moe_ml_95, sev_moe_ml_95)

```

**Step 9.** Store all results (i.e., the baseline and midline weighted and unweighted number of observations, the estimates for moderate and severe food insecurity, moderate food insecurity, and severe food insecurity, and the MOE and 95 percent CI for each estimate) in a data frame, `AGG_df`.

```

#### Calculate aggregate prevalence levels and put in df ##### #

# create empty DF
AGG_df <- matrix(NA, nrow = 2, ncol = 17)
# add column and rownames

```

```

colnames(AGG_df) = c("Moderate+Severe_Food_Insecurity",
  "Moderate_Food_Insecurity", "Severe_Food_Insecurity",
  "N","WN",
  "MSFI_MoE", "MFI_MoE","SFI_MoE",
  "MSFI_CI_Low", "MFI_CI_Low", "SFI_CI_Low",
  "MSFI_CI_High", "MFI_CI_High",
  "SFI_CI_High",
  "MSFI_Sig", "MFI_Sig", "SFI_Sig")
rownames(AGG_df) = c("Baseline","Midline")
# add in values
AGG_df[1, c(1,2, 3)] <- glo_prev_bl * 100
AGG_df[1, c(4)] <- nrow(FIES_baseline_rasch$XX)
AGG_df[1, c(5)] <- round(sum(ftfzoi_baseline$wt * 10^6), 2)
AGG_df[1, c(6, 7, 8)] <- moe_bl_95
AGG_df[1, c(9, 10, 11)] <- (glo_prev_bl * 100) - moe_bl_95
AGG_df[1, c(12, 13, 14)] <- (glo_prev_bl * 100) + moe_bl_95
AGG_df[1, c(15, 16, 17)] <- c(NA, NA, NA)

AGG_df[2, c(1,2, 3)] <- glo_prev_ml * 100
AGG_df[2, c(4)] <- nrow(res_ml$XX)
AGG_df[2, c(5)] <- round(sum(FIES_midline_rasch$wt * 10^6), 2)
AGG_df[2, c(6, 7, 8)] <- moe_ml_95
AGG_df[2, c(9, 10, 11)] <- (glo_prev_ml * 100) - moe_ml_95
AGG_df[2, c(12, 13, 14)] <- (glo_prev_ml * 100) + moe_ml_95
AGG_df[2, c(15, 16, 17)] <- ifelse(((glo_prev_ml * 100) -
  (glo_prev_bl * 100)) > moe_ml_95, "T", "F")

# Convert matrix to dataframe and convert row ID to column.
AGG_df <- AGG_df %>%
  as.data.frame() %>%
  rownames_to_column("Survey_Round")

# view results

head(AGG_df)

```

#### Step 10. Calculate midline estimates for disaggregates.

Assessing food insecurity prevalence when considering key characteristics of households is an important component of analysis. In this step, the disaggregate variables read into R in Step 2 (i.e., residence, gendered household type, shock severity, and wealth quintile) are used to disaggregate moderate and severe food insecurity at midline. A similar process can be repeated for the baseline data, different classifications of food insecurity, or additional disaggregates.

**Step 10a.** Define the disaggregates to be included in this analysis.

If more disaggregates are needed, create additional groups (**group7**, **group8**, etc.), as done with **group1** to **group6**.

```
# Computing prevalence and MoEs by groups
# survey round
# urban rural
group1 = ftfzoi_midline$urbanrural

group2 = ftfzoi_midline$genhhtype_dj
group2 <- factor(group4, levels = c(1,2,3,4), labels = c("De jure
male and female adults", "De jure female, no male", "De jure
male, no female", "De jure children only"))

group3 <-
  ftfzoi_midline$shock_sev[!is.na(ftfzoi_midline$shock_sev)]
group3 <- factor(group5, levels = c(1,2,3,4), labels = c("None",
"Low", "Medium", "High"))

group4 <-
  ftfzoi_midline$awiquint[!is.na(ftfzoi_midline$awiquint)]
group4 <- factor(group6, levels = c(1,2,3,4,5), labels =
  c("Poorest", "Second", "Middle", "Fourth", "Wealthiest"))
```

**Step 10b.** Add the disaggregate variables to the **group\_list** and **groups** objects so that they are included in the loop function.

```
Group_list <- list(group1, group2, group3, group4)

groups <- c(unique(as.character(group1)),
  unique(as.character(group2)),
  unique(as.character(group3)), unique(as.character(group4)))
```

**Step 10c.** Create an empty data frame (**mod\_sev-fi\_ml**) and use a loop function to calculate prevalence, unweighted and weighted number of observations, and the MOE for each disaggregate at midline and store the results in the data frame.

```
# create empty data frame for results
mod_sev-fi_ml <- data.frame()
# loop function
for (I in 1:length(group_list)) {
  for (dis in unique(groups)) {
    if(!(dis %in% group_list[[i]])) next # skip if disaggregate
    not in group
    disag = dis # store disaggregate name
```

```

    fltr = which(group_list[[i]] == dis) #define rows/HHs to
    calc. values
    prob_mods = ftfzoi_midline$prob_mod_sev[fltr] # define
    probabilities
    wt = ftfzoi_midline$wt[fltr]*10^6 # define weights
    rs = ftfzoi_midline$RS[fltr] # define raw scores
    psu = ftfzoi_midline$psu[fltr] # define primary sampling
    units for MOE
    strata = ftfzoi_midline$strata[fltr] # define strata for
    MOE
# results
    output_1 = disag
    output_2 = length(fltr)
    output_3 = sum(wt)
    output_4 = wtd.mean(prob_mods,wt) * 100
    output_5 = moe(prob_mods,rs,wt,psu=psu,strata=strata,
    conf.level = .95)$moe * 100
    tot_output = c(output_1, output_2, output_3,
    round(output_4, 1), round(output_5, 1)) # assign results to
    vector
    mod_sev-fi_ml <- rbind(mod_sev-fi_ml, tot_output) # assign
    vector to data frame row
  }
  colnames(mod_sev-fi_ml) =
  c("Disaggregat","Midline_""Midline_W""Midline_MSF""
  Midline_Mo") # assign column names
}

# view results
head(mod_sev-fi_ml)

```

**Step 10d.** If also analyzing baseline data, customize the code in Steps 10a-c to run with the baseline data. Then ensure that the column names have the survey round in each name so that they are unique as shown above (“Midline\_”, “Baseline\_”) and merge the two data frames using either the `left_join` or `bind_rows` function so that the midline and baseline estimates can be compared and changes over time assessed.

```

# create empty data frame for results
mod_sev-fi-bl <- data.frame()
# loop function
for (i in 1:length(group_list)) {
  for (dis in unique(groups)) {
    if(!(dis %in% group_list[[i]])) next # skip if
    disaggregate not in group
    disag = dis # store disaggregate name
  }
}

```

```

        fltr = which(group_list[[i]] == dis) #define
rows/HHs to calc. values
        prob_mods = ftfzoi_baseline$prob_mod_sev[fltr] #
define probabilities
        wt = ftfzoi_baseline$wt[fltr]*10^6 # define
weights
        rs = ftfzoi_baseline$RS[fltr] # define raw scores
        psu = ftfzoi_baseline$psu[fltr] # define primary
sampling units for MOE
        strata = ftfzoi_baseline$strata[fltr] # define
strata for MOE
        # results
        output_1 = disag
        output_2 = length(fltr)
        output_3 = sum(wt)
        output_4 = wtd.mean(prob_mods,wt) * 100
        output_5 =
moe(prob_mods,rs,wt,psu=psu,strata=strata, conf.level =
.95)$moe * 100
        tot_output = c(output_1, output_2, output_3,
round(output_4, 1), round(output_5, 1)) # assign results to
vector
        mod_sev-fi-bl <- rbind(mod_sev-fi-bl, tot_output)
# assign vector to data frame row
    }
colnames(mod_sev-fi-bl) =
    c("Disaggregat","Baseline_W""Baseline_MS
F""Baseline_Mo") # assign column names
}

# view results
head(mod_sev-fi-bl)

# create variable identifying survey
mod_sev-fi-ml$Survey <- "Midline"
mod_sev-fi-bl$Survey <- "Baseline"
# join into combined disaggregate data frame
mod_sev-fi-combined <- left_join(mod_sev-fi-bl, mod_sev-fi-ml, by
= c("Disaggregate", "Survey"))

```

**Step 10e.** After combining the baseline and midline data frames, determine whether the change over time is statistically significant for each disaggregate category by calculating the difference between the midline and baseline estimates and then running an `ifelse` statement as done for the aggregate level in Step 8.



```
# calculate difference- midline-- baseline
mod_sev-fi-combined$Diff <- mod_sev-fi-combined$ Midline--
mod_sev-fi-combined$Baseline_MSFI
# significance test
mod_sev-fi-combined$Sig_test <- ifelse(mod_sev-fi-combined$Diff >
mod_sev-fi-combined$Midline_MoE, "T", "F")
```

**Step 11.** Export the midline results to two CSV files—one with the aggregate results and one with the disaggregate results—using the `write_csv` function. The data can then be used for further analysis in another statistical software program or for generating tables. Adjust the code and repeat the step if the baseline results also must be exported to CSV files.

```
# define working directory to save results in a separate folder
setwd("FIES\\FTF ZOI Survey [COUNTRY] [YEAR] NAME\\Result")

write_csv(as.data.frame(AGG_df), "prevalence_food_insecurity_aggr
egate.csv")
write_csv(as.data.frame(mod_sev-fi), "prevalence_mod_sev_food_ins
ecurity_disaggregates.csv")
```

**Step 12.** Export the results to two Stata data files—one with household-level results and one with ZOI-level results—using the `write_dta` function.

**Step 12a.** Ensure that all variables are labeled. This can be done with the `var_label` function of the ‘labelled’ library, which is a subpackage of the ‘haven’ library. For example:

```
var_label(ftfzoi_midline$prob_mod_sev) <- "Probability of HH
being moderate or severe food insecure"
var_label(ftfzoi_midline$fi_mod_sev_bin) <- "Binary variable for
moderate & severe food insecurity"
```

**Step 12b.** Ensure that all values are labeled using the `val_label` function. For example:

```
val_label(ftfzoi_midline$fi_mod_sev_bin, 1) <- "Yes"
val_label(ftfzoi_midline$fi_mod_sev_bin, 0) <- "No" # set yes/no
labels for the binary variable of moderate/severe food
insecurity
```

**Step 12c.** Include the following variables in the `ftfzoi_midline` data frame and export to a Stata data file, *FTF ZOI Survey [COUNTRY] [YEAR] FIES*, to be added to the larger household-level data file.

```
Core variables:
survey - "Survey round"
hhnum - "Household ID number"
psu - "Primary sampling unit"
wt - "HH sampling weight"
hysize_dj - "Number of members in household"
strata - "Stratum"
```

fcluster - "Cluster ID number"

**FIES items:**

WORRIED - "Past 12 months: ever worried not enough food"  
HEALTHY - "Past 12 months: ever unable to eat healthy foods"  
FEWFOOD - "Past 12 months: ever limited variety of food"  
SKIPPED - "Past 12 months: ever skipped a meal"  
ATELESS - "Past 12 months: ever ate less than should"  
RUNOUT - "Past 12 months: ever did' 't have food"  
HUNGRY - "Past 12 months: ever hungry but did not eat"  
WHLDAY - "Past 12 months: ever did' 't eat for a whole day"  
RS - "Raw score: sum of "yes" responses to FIES Items  
prob\_mod - "Probability of being moderately food secure"  
prob\_mod\_sev - "Probability of HH being moderately or severely food insecure"  
prob\_sev - "Probability of HH being severely food insecure  
fi\_mod\_sev\_bin - "Moderately or severely food insecure - disaggregate"  
fi\_mod\_bin - "Moderately food insecure - disaggregate"  
fi\_sev\_bin - "Severely food insecure - disaggregate"

**Disaggregates (if calculated):**

ahtype - "Residency type (urban/rural)"  
genhhtype\_dj - "Gender household type - de jure household members"  
awiquint - "Wealth quintile disaggregate"  
shock\_sev - "Shock severity exposure disaggregate"

write\_dta(ftfzoi\_midline, "FTF ZOI Survey [COUNTRY] [YEAR] FIES.dt")

**Step 12d.** Include the following aggregate ZOI-level variables in the AGG\_df data frame and export to a Stata data file, *FTF ZOI Survey [COUNTRY] [YEAR] NAME Food Insecurity Aggregate Results*.

Survey\_Round - "Survey Round: Baseline/Midline"  
Moderate+Severe\_Food\_Insecurity - "Estimate of Moderate & Severe Food Insecurity Rate"  
Moderate\_Food\_Insecurity - "Estimate of Moderate Food Insecurity Rate"  
Severe\_Food\_Insecurity - "Estimate of Severe Food Insecurity Rate"  
N - "Unweighted N"  
WN - "Weighted N"  
MSFI\_MoE - "Margin of Error - Moderate & Severe Food Insecurity"  
MFI\_MoE-- "Margin of Error - Moderate Food Insecurity"  
SFI\_MoE-- "Margin of Error - Severe Food Insecurity"  
MSFI\_CI\_Low - "Lower Confidence Interval - Moderate & Severe Food Insecurity"  
MFI\_CI\_Low-- "Lower Confidence Interval - Moderate Food Insecurity"  
SFI\_CI\_Low-- "Lower Confidence Interval - Severe Food Insecurity"

```

MSFI_CI_High - "Upper Confidence Interval - Moderate & Severe
Food Insecurity"
MFI_CI_High-- "Upper Confidence Interval - Moderate Food
Insecurity"
SFI_CI_High - "Upper Confidence Interval - Severe Food
Insecurity"
MSFI_Sig-- "Significance Test - Moderate & Severe Food
Insecurity: TRUE/FALSE"
MFI_Sig - "Significance Test - Moderate Food Insecurity:
TRUE/FALSE"
SFI_Sig - "Significance Test - Severe Food Insecurity: TRUE/FALSE"

write_dta(AGG_df,"FTF ZOI Survey [COUNTRY] [YEAR] NAME Food Insecurity
Aggregate Results.dt")

```

**Step 12e.** Include the following aggregate ZOI-level variables in the `mod_sev-fi_combined` data frame and export to a Stata data file, *FTF ZOI Survey [COUNTRY] [YEAR] NAME Food Insecurity Disaggregate Results..* **Note:** The variables shown are for only moderate and severe food insecurity; be sure to include the same variables for the other two levels of food insecurity (moderate, and severe) for each disaggregate.

```

Disaggregate - "Name of disaggregate"
FI_Baseline - "Moderate and Severe Food Insecurity Rate,
Baseline"
CI_Baseline - "Confidence Interval: Moderate and Severe Food
Insecurity Rate, Baseline"
N_Baseline - "Unweighted N, Baseline"
FI_Midline-- "Moderate and Severe Food Insecurity Rate, Midline"
CI_Midline-- "Confidence Interval: Moderate and Severe Food
Insecurity Rate, Midline"
N_Midline-- "Unweighted N, Midline"
Diff - "Difference between Midline and Baseline in Moderate and
Severe Food Insecurity Rate"
Sig - "Significance Test from Midline to Baseline - Moderate and
Severe Food Insecurity Rate"
MoE_Baseline - "Margin of Error-- Moderate and Severe Food
Insecurity Rate, Baseline"
MoE_Midline - "Margin of Error-- Moderate and Severe Food
Insecurity Rate, Midline"
write_dta(mod_sev-fi_combined,"FTF ZOI Survey [COUNTRY] [YEAR]
NAME Moderate & Severe Food Insecurity Disaggregate
Results.dta")

```

### 15.3 References

Ballard, T., Kepple, A.W., & Cafiero, C. (2013). *The food insecurity experience scale: Development of a global standard for monitoring hunger worldwide*. Technical Paper. Rome: Food and Agriculture Organization of the United Nations. Available at:

[http://www.fao.org/fileadmin/templates/ess/voh/FIES\\_Technical\\_Paper\\_v1.1.pdf](http://www.fao.org/fileadmin/templates/ess/voh/FIES_Technical_Paper_v1.1.pdf)

Cafiero, C., Viviani, S., & Nord, M. (2017). *Food security measurement in a global context: The Food Insecurity Experience Scale, Measurement*. Available at: doi: <https://doi.org/10.1016/j.measurement.2017.10.065>

Cafiero, C., et al. (2016). *Methods for estimating comparable prevalence rates of food insecurity experienced by adults throughout the world*. Voices of the Hungry Technical Report No 1/2016. Rome: Food and Agriculture Organization of the United Nations. Available at: <http://www.fao.org/3/a-i4830e.pdf>

Food and Agriculture Organization of the United Nations (FAO). (2019). *Voices of the hungry: Applying the FIES*. Rome: FAO. Available at: <http://www.fao.org/in-action/voices-of-the-hungry/using-fies/en/>

Food and Agriculture Organization of the United Nations (FAO). (2015). *Modeling food insecurity in bivariate and regression analyses*. Rome: FAO. Available at: <https://www.fao.org/3/bp091e/bp091e.pdf>

INDDEx Project. (2018). *Food insecurity experience scale (FIES)*. Data4Diets: Building Blocks for Diet-related Food Security Analysis. Tufts University, Boston, MA. Available at: <https://inddex.nutrition.tufts.edu/data4diets/indicator/food-insecurity-experience-scale-fies>

Nord, M. (2014). *Introduction to item response theory applied to food security measurement: Basic concepts, parameters and statistics*. Rome: Food and Agriculture Organization of the United Nations. Available at: <http://www.fao.org/3/a-i3946e.pdf>

Onori, F., Viviani S., & Brutti P. (2021). *Towards global monitoring: equating the Food Insecurity Experience Scale (FIES) and food insecurity scales in Latin America*. Available at: <https://arxiv.org/pdf/2102.10005.pdf>

Viviani, S. (2016). *Manual for the implementation of the FAO Voices of the Hungry methods to estimate food insecurity: RM.weights package in R. Version 2*. Rome: Food and Agriculture Organization of the United Nations. Available at: [http://www.sesric.org/imgs/news/1752\\_Manual\\_on\\_RM\\_Weights\\_Package\\_EN.pdf](http://www.sesric.org/imgs/news/1752_Manual_on_RM_Weights_Package_EN.pdf)

## 16. CHILDREN'S AND WOMEN'S DIETARY INTAKE INDICATORS

This chapter describes the children's and women's nutrition indicators related to diet. This chapter has two sections; the first section describes the guidelines to construct the indicators, and the second section outlines the step-by-step procedures to calculate the indicators.

### 16.1 Guidelines to construct the indicators

This section provides the guidelines to construct the Feed the Future children's and women's nutrition indicators related to diet.

#### 16.1.1 Prevalence of exclusive breastfeeding of children under 6 months of age

Exclusive breastfeeding for children's first 6 months of life provides them with significant health and nutritional benefits, including protection from gastrointestinal infections and reduced risk of mortality due to infectious disease. This indicator estimates the percentage of children under 6 months of age (0-5 months) who were exclusively breastfed during the day and night preceding the survey. Exclusive breastfeeding means that the infant received breast milk, including milk expressed or from a wet nurse, and may have received oral rehydration solution, vitamins, minerals, and medicines, but did not receive any other food or liquid, including water, during the day and night preceding the survey.

#### 16.1.2 Percent of children 6-23 months of age receiving a minimum acceptable diet

Appropriate feeding of children 6-23 months of age is multidimensional. The minimum acceptable diet indicator combines standards of dietary diversity, which is a proxy for nutrient density; meal frequency, which is a proxy for energy density, by breastfeeding status and age; and milk feeds among non-breastfed children, which is to ensure that they are receiving adequate amounts of calcium and other important nutrients, and thus provides a useful way to track progress in improving the key quality and quantity dimensions of children's diets.

If children meet the minimum dietary diversity requirements and the minimum meal frequency requirements for their breastfeeding status and age, plus the minimum milk feeding frequency requirements if they are not breastfed, they are considered to have received a minimum acceptable diet. Tabulation of the indicator requires that data be collected on breastfeeding, dietary diversity, number of semi-solid and solid feeds, and number of milk feeds for children 6-23 months of age during the day and night preceding the survey. The approach to calculating the indicator has been updated from the previous version of the *Guide to Feed the Future Statistics* to reflect updates to the World Health Organization (WHO) and UNICEF's infant and young child feeding practice indicator definitions and measurement methods (WHO & UNICEF, 2021). Previously, minimum dietary diversity was defined as having received four out of seven food groups. Currently, minimum dietary diversity is defined as having received five out of eight food groups, with breastfeeding included as the eighth group (WHO, 2017) to remove differences in the indicator for breastfeeding children compared with non-breastfeeding

children. In addition, the definition for minimum meal frequency for non-breastfed children changed to require that at least one of the feedings is solid, semi-solid, or soft foods.

### **Minimum dietary diversity**

Minimum dietary diversity is defined as the percentage of children 6-23 months of age who consumed foods and beverages from at least five of eight defined food groups during the day or night preceding the survey, and is based on the following eight defined food groups:

- Grains, roots, tubers, and plantains
- Pulses (peas, beans, and lentils), nuts, and seeds
- Dairy products (milk, infant formula, yogurt, cheese)
- Flesh foods (meat, fish, poultry, and organ meats)
- Eggs
- Vitamin A-rich fruits and vegetables
- Other (non-vitamin A-rich) fruits and vegetables
- Breast milk

### **Minimum meal frequency**

Minimum meal frequency is defined as the percentage of children 6-23 months of age who consumed solid, semi-solid, or soft foods (but also including milk feeds for non-breastfed children) at least the minimum number of times for their breastfeeding status and age during the day or night preceding the survey.

The minimum number of times is defined as:

- Breastfed infants 6-8 months of age: two feedings of solid, semi-solid, or soft foods
- Breastfed children 9-23 months of age: three feedings of solid, semi-solid, or soft foods
- Non-breastfed children 6-23 months of age: four feedings of solid, semi-solid, soft foods, or milk feeds, whereby at least one feed must be a solid, semi-solid or soft feed

### **Minimum milk feeding frequency**

Minimum milk feeding frequency is defined as the percentage of non-breastfed children 6-23 months of age who consumed at least two milk feeds during the day preceding the survey. Infant formula, animal milk, yogurt drinks, and semi-solid yogurt all count as milk feeds.

### **Minimum acceptable diet**

Minimum acceptable diet is defined as the percentage of children 6-23 months of age who consumed a minimum acceptable diet during the day preceding the survey, whereby a minimum acceptable diet depends on the child's breastfeeding status:

- Breastfed children: Received at least the minimum dietary diversity and minimum meal frequency for their age during the day preceding the survey

- Non-breastfed children: Received at least the minimum dietary diversity and minimum meal frequency for their age during the day preceding the survey, as well as the minimum milk feeding frequency

### **16.1.3 Percent of women of reproductive age consuming a diet of minimum diversity**

Dietary diversity is a key characteristic of a high-quality diet with adequate micronutrient content and is thus important to ensuring the health and nutrition of both women and their children. This indicator captures the percentage of women of reproductive age (15-49 years) in the population who are consuming a diet of minimum diversity. A woman of reproductive age is considered to consume a diet of minimum diversity if she consumed at least 5 of 10 specific food groups during the day and night preceding the survey. The 10 food groups included in the indicator are as follows:

- Grains, white roots and tubers, and plantains
- Pulses (beans, peas, and lentils)
- Nuts and seeds (including groundnuts)
- Dairy
- Meat, poultry, and fish
- Eggs
- Dark green leafy vegetables
- Other vitamin A-rich fruits and vegetables (besides dark green leafy vegetables)
- Other (non-vitamin A-rich) vegetables
- Other (non-vitamin A-rich) fruits

Note that although Feed the Future usually considers groundnuts as part of a legume value chain, for this indicator, it is classified in the nuts and seeds group.

## **16.2 Step-by-step procedures to calculate nutrition indicators**

This section describes the detailed step-by-step procedures to calculate the nutrition indicators.

### **16.2.1 Prevalence of exclusive breastfeeding of children under 6 months of age**

This indicator estimates the percentage of children under 6 months of age (0-5 months) who are exclusively breastfed in the ZOI population. Exclusive breastfeeding is defined according to the UNICEF and WHO guidelines as receiving only breastmilk and no other liquid, solid food, or plain water during the first 6 months of life. It is based on information collected in the core ZOI Survey questionnaire Module 5, *Children's nutrition*. The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_CHN DIET.do* file.

#### **Definitions**

Numerator	Number of children 0-5 months of age who are de facto household members who received only breastmilk and no other liquid, solid food, or plain water in the day and night preceding the survey
-----------	--

Denominator	Number of children 0-5 months of age who are de facto household members
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Children 0-5 months of age
Disaggregation levels	Child's sex* Wealth quintile Shock exposure severity
Treatment of missing data	Children with missing information about breastfeeding or with a "don't know" response are assumed to be not breastfeeding. Children with missing information on a specific food given or "don't know" response are assumed to be not given that food item. Children missing all data for all food and drink variables are omitted from the numerator and denominator.
Survey variables used	<b>hhea, hhnum, strata, wgt_c0_5m, v521, v522, v526, v527, v529, v531-v560</b>
Analytic variables used	<b>c0_5m, hhmem_df, sex, awiquint, shock_sev</b>
Analytic variables created	<b>v526x, v527x, v529x, v531x, v531ax, v532x, v533x, v534ax, v535x, v536x, v537x, v539x-v560x, chn_bf, chn_water, chn_thermilk, chn_nonmilk, chn_food, chn_num_ebf_miss, chn_ebf_miss, chn_ebf, hhmem_c05m_df</b>

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Create a binary variable that flags children 0-5 months of age who were breastfed during the day and night preceding the survey (**chn\_bf**). This includes being breastfed by the mother or by another woman (**v521**) or receiving breastmilk in a spoon, cup, or bottle (**v522**).

```

Set          chn_bf=missing
Replace      chn_bf=0 if c0_5m=1
Replace      chn_bf=1 if (v521=1 or v522=1) and c0_5m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months consumed breastmilk"

```

**Step 2.** Create a binary variable that flags children 0-5 months of age who received any milk other than breastmilk (**chn\_thermilk**). This includes formula (**v527**); milk, such as tinned, powdered, or fresh animal milk (**v529**); or yogurt, including yogurt drinks (**v533**) and other yogurt (**v534a**).

```

Set          chn_thermilk=missing
Replace      chn_thermilk=0 if c0_5m=1
Replace      chn_thermilk=1 if (v527=1 or v529=1 or v533=1 or v534a=1) and c0_5m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months consumed other milk products (non-breastmilk)"

```

**Step 3.** Create a binary variable that flags children 0-5 months of age who received plain water (**chn\_water**).



```

Set          chn_water=missing
Replace      chn_water=0 if c0_5m=1
Replace      chn_water=1 v526=1 and c0_5m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months consumed plain water"

```

**Step 4.** Create a binary variable that flags children 0-5 months of age who received non-milk liquid (chn\_nonmilk). Non-milk liquids include juice or juice drinks (v531); sugary drinks such as soda pop, sports drinks, or malt drinks (v531a); clear broth (v532); thin porridge (v535); and other water-based liquids, such as glucose water or sugar water (v536, v537).

```

Set          chn_nonmilk=missing
Replace      chn_nonmilk=0 if c0_5m=1
Replace      chn_nonmilk=1 if (v531=1 or v531a or v532=1 or
v535=1 or v536=1 or v537=1) and c0_5m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months consumed non-milk liquids"

```

**Step 5.** Create a binary variable that flags children 0-5 months of age who received any food (chn\_food), which includes any food reported to be eaten in variables v539 through v560.

```

Set          chn_food=missing
Replace      chn_food=0 if c0_5m=1
Replace      chn_food=1 if (v539=1 or v540=1 or v541=1 or
v541a=1 or v542=1 or v543=1 or v544=1 or v545=1
or v546=1 or v547=1 or v548=1 or v549=1 or v550=1
or v551=1 or v552=1 or v553=1 or v554=1 or v555=1
or v556=1 or v557=1 or v558=1 or v559=1 or
v560=1) and c0_5m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months consumed any food (non-liquid)"

```

**Step 6.** Create a binary variable that flags children 0-5 months of age who are missing data for all food and drink variables, excluding breastmilk (chn\_ebf\_miss).

**Step 6.1.** Create variables that recode responses that have a value of 'no' (2) or 'don't know' (8) to be 'no' (0) and missing (9) responses to be blank (missing).

```

For each      variable (var) of variable list v526 v527 v529
              v531 v531a v532 v533 v534a v535 v536 v537 v539-
              v560:
Set          `var`x=`var`
Replace      `var`x=0 if `var`=2 or `var`=8
Replace      `var`x=missing if `var`=9
Label values 0 "No"
              1 "Yes"
Label variable "Recode of `var` ([`var` label])"

```

**Step 6.2.** Create a variable that counts the number of recoded food and drink variables created in Step 6.1 that are missing for children 0-5 months of age (`chn_num_ebf_miss`).

```
Set          chn_num_ebf_miss=missing
Replace      chn_num_ebf_miss=number of variables in
              parentheses=missing(v526x v527x v529x v531x
              v531ax v532x v533x v534ax v535x v536x v537x
              v539x-v560x) and c0_5m=1
Label variable "Children 0-5 months, number of food/drink
              variables missing values"
```

**Step 6.3.** Create a variable that flags whether a child 0-5 months of age is missing all food variables (`chn_ebf_miss`)—that is, if the value of `chn_num_ebf_miss` is less than the total number of variables being summed (i.e., 34 in the pseudo code and template analysis Stata do file).

```
Set          chn_ebf_miss=missing
Replace      chn_ebf_miss=0 if chn_num_ebf_miss<34 and c0_5m=1
Replace      chn_ebf_miss=1 if chn_num_ebf_miss=34 and c0_5m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months is missing all food data"
```

**Step 7.** Create a binary variable that flags children 0-5 months of age who were exclusively breastfed (`chn_ebf`).

```
Set          chn_ebf=missing
Replace      chn_ebf=0 if c0_5m=1
Replace      chn_ebf=1 if chn_bf=1 and chn_water=0 and
              chn_thermilk=0 and chn_nonmilk=0 and chn_food=0
              and c0_5m=1
Replace      chn_ebf=missing if chn_ebf_miss=1 and c0_5m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months was exclusively breastfed"
```

**Step 8.** Create a variable that captures the sub-population being examined for the breastfeeding indicator calculation—that is, children 0-5 months of age who are de facto household members among all children 0-5 months of age surveyed (`hhmem_c05m_df`).

```
Set          hhmem_c05m_df=missing
Replace      hhmem_c05m_df=0 if c0_5m=1
Replace      hhmem_c05m_df=1 if c0_5m=1 and hhmem_df=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 0-5 months old is a de facto household
              member"
```

**Step 9.** After applying the children 0-5 month weight (`wgt_c0_5m`), calculate the percentage of children 0-5 months of age who are de facto household members who were exclusively breastfed using

the `chn_ebf` analytic variable. Repeat using the child's sex, wealth quintile, and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svyset hhea [pw=wgt_c0_5m], strata(strata)
svy, subpop(hhmem_c05m_df) : tab chn_ebf
svy, subpop(hhmem_c05m_df) : tab chn_ebf sex, col
svy, subpop(hhmem_c05m_df) : tab chn_ebf awiquint, col
svy, subpop(hhmem_c05m_df) : tab chn_ebf shock_sev, col
```

## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

UNICEF. Report: *Breastfeeding*. Available at: <https://www.unicef.org/reports/breastfeeding>

World Health Organization. (2017). *Global nutrition monitoring framework: Operational guidance for tracking progress in meeting targets for 2025*. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/nutrition/publications/operational-guidance-GNMF-indicators/en/>

World Health Organization. (2015). *Nutrition: Breastfeeding*. Available at: [http://www.who.int/nutrition/topics/exclusive\\_breastfeeding/en/](http://www.who.int/nutrition/topics/exclusive_breastfeeding/en/)

World Health Organization & UNICEF. (2021). Indicators for assessing infant and young child feeding practices: definitions and measurement methods. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789240018389>

### 16.2.2 Percentage of children 6-23 months of age receiving a minimum acceptable diet

This indicator estimates the percentage of children 6-23 months of age who meet the requirements of a minimum acceptable diet in the ZOI population based on their breastfeeding status and age. It is based on information collected in the core ZOI Survey questionnaire Module 5, *Children's nutrition*. The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_CHN DIET.do* file.

## Definitions

Numerator	Number of children 6-23 months of age who are de facto household members who achieve minimum dietary diversity and minimum meal frequency for their breastfeeding status and age, as well as minimum milk feeding frequency if they are not breastfed
Denominator	All children 6-23 months of age who are de facto household members
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Children 6-23 months of age
Disaggregation levels	Child's sex* Child's age category (6-11, 12-17, 18-23 months) Child's breastfeeding status

	Wealth quintile Shock exposure severity
Treatment of missing data	Children with missing breastfeeding information or “don’t know” response are assumed to be not breastfeeding. Children with missing information on a specific food given or with a “don’t know” response are assumed to be not given that food item. Children missing <b>all</b> data food and drink data are omitted from the numerator and denominator. Children with missing information or with a “don’t know” response for the number of times that food was given are assumed to have had zero servings.
Survey variables used	hhea, hhnum, strata, wgt_c6_23m, v521, v522, v527, v529, v533, v534a, v535, v539-v563
Analytic variables used	c6_23m, hhmem_df, sex, cage_mad, awiquint, shock_sev
Analytic variables created	v526x-v537x, v539x-v560x, v563x, bf_stat, chn_foodgrp1-chn_foodgrp8, chn_mmff, chn_nbf_milk, chn_mfreq_milkplus, chn_foodsum, chn_num, mad_miss, chn_mad_miss, c6_8m, c9_23m, chn_bf_grp1, chn_bf_grp2, chn_bf_grp3, chn_mmff, chn_mdd, chn_mad, hhmem_c623m_df

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Create a breastfeeding status variable that indicates whether children 6-23 months of age were breastfed during the day and night preceding the survey (**bf\_stat**).

```

Set          bf_stat=missing
Replace      bf_stat=0 if c6_23m=1
Replace      bf_stat=1 if (v521=1 or v522=1) and c6_23m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child is currently breastfeeding"

```

**Step 2.** Identify the variables associated with the eight food groups used to calculate children’s dietary diversity score in the data file. Pay particular attention if the questionnaire was adapted to include local foods to ensure that they are assigned to the appropriate food group and that those variables are included in the analysis. **Table 21** lists the food groups and their corresponding variables according to the core ZOI Midline Survey questionnaire. Questionnaire customization may also have resulted in changes in the numbering of the questions in the questionnaire, so the variable numbering in **Table 21** should also be reviewed for any needed adjustment.

**Table 21: Eight Food Groups Used to Calculate Minimum Dietary Diversity for Children**

Food group		Variables
1	Grains, roots, and tubers <sup>106</sup>	v535=porridge v539=grains v541=roots or tubers v541a=plantain or green banana
2	Legumes and nuts	v552=legumes

<sup>106</sup> Includes plantains and green bananas, which can be starchy staple foods in some areas

Food group		Variables
		v553=nuts
3	Dairy products	v527=formula v529=milk v533=yogurt drinks v534a=yogurt, non-drinks v554=dairy products (cheese or other milk products, other than yogurt)
4	Flesh foods	v546=organ meats from domesticated animals v547=any meat from domesticated animals v548=organ meats from wild animals v549=any flesh meat from wild animals v551=fresh or dried fish, shellfish, or seafood
5	Eggs	v550=eggs
6	Vitamin A-rich fruits and vegetables	v540=pumpkin, carrots, squash, or sweet potatoes that are yellow or orange inside v542=any dark green leafy vegetables v544=ripe mangoes or papayas v559=foods made with red palm oil, red palm nuts, or red palm nut pulp sauce <sup>107</sup>
7	Other fruits and vegetables	v543=any other vegetables v545=any other fruits
8	Breast milk	v521=breastfed yesterday v522=consumed breastmilk yesterday

**Step 3.** Create the variables for the eight food groups (chn\_foodgrp1-chn\_foodgrp8).

**Step 3.1.** Create a variable to flag children 6-23 months of age who ate grains, roots, or tubers (chn\_foodgrp1).

```

Set          chn_foodgrp1=0 if c6_23m=1
Replace      chn_foodgrp1=1 if (v535=1 or v539=1 or v541=1 or
              v541a=1) and c6_23m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 6-23 months ate grains, roots, tubers"

```

**Step 3.2.** Create a variable to flag children 6-23 months of age who ate legumes or nuts (chn\_foodgrp2).

```

Set          chn_foodgrp2=0 if c6_23m=1
Replace      chn_foodgrp2=1 if (v552=1 or v553=1) and c6_23m=1

```

<sup>107</sup> Variable v559, which captures the consumption of foods made with red palm oil, red palm nuts, or red palm nut pulp sauce, is included in food group 6, vitamin A-rich fruits and vegetables. This deviates from minimum acceptable diet indicator calculation guidance (WHO & UNICEF, 2021), which states that red palm fruit and red palm pulp should be included but red palm oil should not be included. In the core ZOI Survey questionnaires, red palm oil, red palm nuts, and red palm nut pulp sauce are asked about in the same question, so red palm oil cannot be excluded unless red palm nuts and red palm nut pulp sauce are also excluded.

```

Label values      0 "No"
                  1 "Yes"
Label variable    "Child 6-23 months ate legumes, nuts"

```

**Step 3.3.** Create a variable to flag children 6-23 months of age who ate dairy products (chn\_foodgrp3).

```

Set              chn_foodgrp3=0 if c6_23m=1
Replace          chn_foodgrp3=1 if (v527=1 or v529=1 or v533=1 or
Label values      0 "No"
                  1 "Yes"
Label variable    "Child 6-23 months ate dairy products"

```

**Step 3.4.** Create a variable to flag children 6-23 months of age who ate flesh foods (chn\_foodgrp4).

```

Set              chn_foodgrp4=0 if c6_23m=1
Replace          chn_foodgrp4=1 if (v546=1 or v547=1 or v548=1 or
Label values      0 "No"
                  1 "Yes"
Label variable    "Child 6-23 months ate flesh foods"

```

**Step 3.5.** Create a variable to flag children 6-23 months of age who ate eggs (chn\_foodgrp5).

```

Set              chn_foodgrp5=0 if c6_23m=1
Replace          chn_foodgrp5=1 if v550=1 and c6_23m=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Child 6-23 months ate eggs"

```

**Step 3.6.** Create a variable to flag children 6-23 months of age who ate vitamin A-rich fruits or vegetables (chn\_foodgrp6).

```

Set              chn_foodgrp6=0 if c6_23m=1
Replace          chn_foodgrp6=1 if (v540=1 or v542=1 or v544=1 or
Label values      0 "No"
                  1 "Yes"
Label variable    "Child 6-23 months ate vitamin A-rich fruits and
                  vegetables"

```

**Step 3.7.** Create a variable to flag children 6-23 months of age who ate other non-vitamin A-rich fruits or vegetables (chn\_foodgrp7).

```

Set              chn_foodgrp7=0 if c6_23m=1
Replace          chn_foodgrp7=1 if (v543=1 or v545=1) and c6_23m=1
Label values      0 "No"
                  1 "Yes"
Label variable    "Child 6-23 months ate other fruits and
                  vegetables"

```

**Step 3.8.** Create a variable to flag children 6-23 months of age who consumed breastmilk (`chn_foodgrp8`).

```
Set                chn_foodgrp8=bf_stat
Label values       0 "No"
                   1 "Yes"
Label variable     "Child 6-23 months consumed breastmilk"
```

**Step 4.** Create a binary variable that flags children 6-23 months of age who are missing data for all food and drink variables, excluding breastmilk (`chn_mad_miss`).

**Step 4.1.** Create variables that recode responses that have a value of 'no' (2) or 'don't know' (8) to be 'no' (0) and missing (9) responses to be blank (missing) if they do not already exist in the data file.

```
For each           variable (var) of variable list v526 v527 v529
                  v531 v531a v532 v533 v534a v535 v536 v537 v539-
                  v560:
Set                `var'x=`var'
Replace            `var'x=0 if `var'=2
Replace            `var'x=missing if `var'=8 or `var'=9
Label values       0 "No"
                   1 "Yes"
Label variable     "Recode of `var' ([`var' label])"
```

**Step 4.2.** Create a variable that counts the number of recoded food and drink variables created in Step 4.1 that are missing for children 6-23 months of age (`chn_num_mad_miss`).

```
Set                chn_num_mad_miss=missing
Replace            chn_num_mad_miss=number of variables in
                  parentheses= missing(v526x v527x v529x v531x
                  v531ax v532x v533x v534ax v535x v536x v537x
                  v539x-v560x) and c6_23m=1
Label variable     "Children 6-23 months, number of food/drink
                  variables missing values"
```

**Step 4.3.** Create a variable that flags whether a child 6-23 months of age is missing all food variables (`chn_mad_miss`)—that is, if the value of `chn_num_mad_miss` is less than the total number of variables being summed (i.e., 34 in the pseudo code and template analysis Stata do file).

```
Set                chn_mad_miss=missing
Replace            chn_mad_miss=0 if chn_num_mad_miss<34 and
                  c6_23m=1
Replace            chn_mad_miss=1 if chn_num_mad_miss=34 and
                  c6_23m=1
Label values       0 "No"
                   1 "Yes"
Label variable     "Child 6-23 months is missing all food data"
```

**Step 5.** Create a variable that sums the number of food groups that children 6-23 months of age consumed (`chn_foodsum`).

```

Set          chn_foodsum=0 if c6_23m=1
Replace      chn_foodsum=total(chn_foodgrp1-chn_foodgrp8)
Label variable "Child 6-23 months, number of food groups
consumed"

```

**Step 6.** Create a variable that flags children 6-23 months of age who meet the minimum dietary diversity threshold (**chn\_mdd**)—that is, they consumed at least five of the eight specified food groups.

```

Set          chn_mdd=missing
Replace      chn_mdd=0 if chn_foodsum>0 and chn_foodsum<5
Replace      chn_mdd=1 if chn_foodsum≥5 and chn_foodsum≤8
Replace      chn_mdd=missing if chn_mad_miss=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 6-23 months ate 5+ food groups (achieved
MDD)"

```

**Step 7.** Create a variable that flags children 6-23 months of age who achieve minimum feeding frequency (**chn\_mmff**)—that is: 2 feedings of solid, semi-solid, or soft foods for breastfed infants 6–8 months of age; 3 feedings of solid, semi-solid, or soft foods for breastfed children 9–23 months of age; and 4 feedings of solid, semi-solid, soft foods, or milk feeds for non-breastfed children 6–23 months of age, 1 of which must be a solid, semi-solid or soft feed.

**Step 7.1.** Create analytic variables that can be used to create **chn\_mmff** (**v528x**, **v530x**, **v534x**, **v534bx**, and **v563x**). Because values will be summed across variables to get a total number of feedings, reassign the numeric values that have missing (i.e., 9 or 99) or “don’t know” (i.e., 8 or 98) values so that they are not included in the total. For **v528**, **v530**, **v534**, and **v534b**, recode “don’t know” responses (98) and missing responses (99) to zero (0) under the assumption that children had no milk feeds. For **v563**, if **v562** is “no” (2), “don’t know” (8), or “missing” (9), or if **v563** is “don’t know” (98) or “missing” (99), recode **v563** to zero (0) under the assumption that children did not eat any solid, semi-solid, or soft foods.

```

Set          v528x=0 if v528=98 or v528=99
Set          v530x=0 if v530=98 or v530=99
Set          v534x=0 if v534=98 or v534=99
Set          v534bx=0 if v534b=98 or v534b=99
Set          v563x=0 if v562=2 or v562=8 or v562=9 or v563=98
              or v563=99

```

**Step 7.2.** Create a variable that counts the total number of feeds, including milk feeds plus soft, solid, or semi-solid food feeds that children 6-23 months of age received (**chn\_mfreq\_milkplus**).

```

Set          chn_mfreq_milkplus=0 if c6_23m=1
Replace      chn_mfreq_milkplus=(v528x+v530x+v534x+v563x) if
              c6_23m=1
Label variable "Child 6-23 months, number of feeds (milk+food)"

```

**Step 7.3.** Create variables that categorize children 6-23 months of age by age and breastfeeding status.



**Step 7.3.1.** First create a variable that flags children 6-8 months of age (**c6\_8m**) and then create a variable that flags breastfed children 6-8 months of age (**chn\_bf\_grp1**).

```

Set          c6_8m=0
Replace      c6_8m=1 if cage_months_int≥6 and
              cage_months_int≤8
Label values 0 "No"
              1 "Yes"
Label variable "Child is 6-8 months"

Set          chn_bf_grp1=0 if c6_8m=1
Replace      chn_bf_grp1=1 if bf_stat=1 and c6_8m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 6-8 months, breastfeeds"

```

**Step 7.3.2.** First create a variable that flags children 9-23 months of age (**c9\_23m**) and then create a variable that flags breastfed children 9-23 months of age (**chn\_bf\_grp2**).

```

Set          c9_23m=0
Replace      c9_23m=1 if cage_months_int≥9 and
              cage_months_int≤23
Replace      c9_23m=1 if ((v104a≥9 and v104a≤12) or v104=1)
              and c9_23m=0
Label values 0 "No"
              1 "Yes"
Label variable "Child is 9-23 months"

Set          chn_bf_grp2=0 if c9_23m=1
Replace      chn_bf_grp2=1 if bf_stat=1 and c9_23m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 9-23 months, breastfeeds"

```

**Step 7.3.3.** Create a variable that flags non-breastfed children 6-23 months of age (**chn\_bf\_grp3**).

```

Set          chn_bf_grp3=0 if c6_23=1
Replace      chn_bf_grp3=1 if bf_stat=0 and c6_23m=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 6-23 months, does not breastfeed"

```

**Step 7.4.** Create a binary variable that flags children 6-23 months of age who received the minimum meal frequency based on their age and breastfeeding status (**chn\_mmf**).

```

Set          chn_mmf=0 if c6_23m=1
Replace      chn_mmf=1 if chn_bf_grp1=1 and (v563x≥2 and
              v563≠missing)
Replace      chn_mmf=1 if chn_bf_grp2=1 and (v563x≥3 and
              v563≠missing)

```

```

Replace      chn_mmf=1 if chn_bf_grp3=1 and
              (chn_mfreq_milkplus≥4 and mfreq_milkplus≠missing)
              and (v563x≥1 and v563x≠missing)
Replace      chn_mmf=missing if chn_mad_miss=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 6-23 months meets minimum meal frequency
              (MMF) criteria"

```

**Step 7.5.** Create a variable that counts the total number of milk feeds non-breastfed children 6-23 months of age received (chn\_nbf\_milk).

```

Set          chn_nbf_milk=0 if chn_bf_grp3=1
Replace      chn_nbf_milk=(v528x+v530x+v534x+v534bx) if
              chn_bf_grp3=1
Label variable "Non-breastfed child 6-23 months, number of milk
              feeds"

```

**Step 7.6.** Create a variable (chn\_mmff) that flags non-breastfed children 6-23 months of age who received minimum milk feeding frequency—that is, they received at least two milk feeds.

```

Set          chn_mmff=0 if chn_bf_grp3=1
Replace      chn_mmff=1 if chn_nbf_milk≥2 and chn_nbf≠missing
              and chn_bf_grp3=1
Replace      chn_mmff=missing if chn_mad_miss=1
Label values 0 "No"
              1 "Yes"
Label variable "Non-breastfed child 6-23 months meets minimum
              milk feeding frequency (MMFF)"

```

**Step 8.** Create a binary variable (chn\_mad) that flags children 6-23 months of age who received a minimum acceptable diet—that is, they achieved both minimum dietary diversity and minimum meal frequency given their age and breastfeeding status.

```

Set          chn_mad=0 if c6_23=1
Replace      chn_mad=1 if chn_mdd=1 and chn_mmf=1 and
              (chn_bf_grp1=1 or chn_bf_grp2=1)
Replace      chn_mad=1 if chn_mdd=1 and chn_mmf=1 and
              chn_mmff=1 and chn_bf_grp3=1
Replace      chn_mad=missing if chn_mad_miss=1
Label values 0 "No"
              1 "Yes"
Label variable "Child 6-23 months meets minimum acceptable diet
              (MAD) criteria"

```

**Step 9.** Create a variable that captures the sub-population being examined for the minimum acceptable diet indicator calculation—that is, children 6-23 months of age who are de facto household members among all children 6-23 months of age surveyed (hhmem\_c623m\_df).

```

Set          hhmem_c623m_df=missing
Replace      hhmem_c623m_df=0 if c6_23m=1

```

```

Replace          hhmem_c623m_df=1 if c6_23m=1 and hhmem_df=1
Label values     0 "No"
                  1 "Yes"
Label variable   "Child 6-23 months is a de facto household
                  member"

```

**Step 10.** After applying the children 6-23 months weight (`wgt_c6_23m`), calculate the percentage of children 6-23 months of age who are de facto household members who received a minimum acceptable diet using the `chn_mad` analytic variable. Repeat using the child's sex, child's age, gendered household type, wealth index and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```

svyset hhea [pw=wgt_c6_23m], strata(strata)
svy, subpop(hhmem_c623m_df): tab chn_mad
svy, subpop(hhmem_c623m_df): tab chn_mad sex, col
svy, subpop(hhmem_c623m_df): tab chn_mad cage_mad, col
svy, subpop(hhmem_c623m_df): tab chn_mad awiquint, col
svy, subpop(hhmem_c623m_df): tab chn_mad shock_sev, col

```

## References

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: <https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm>

INDDEx Project. (2018). *Minimum acceptable diet (MAD) for children 6-23 months old*. Data4Diets: Building blocks for diet-related food security analysis. Tufts University, Boston, MA. Available at: <https://inddex.nutrition.tufts.edu/data4diets/indicator/minimum-acceptable-diet-mad>

World Health Organization (WHO). (2017). *Global nutrition monitoring framework: Operational guidance for tracking progress in meeting targets for 2025*. Geneva: WHO. License: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/nutrition/publications/operational-guidance-GNMF-indicators/en/>

World Health Organization. (2010). *Indicators for assessing infant and young child feeding practices, Part 2, Measurement*. Available at: [http://whqlibdoc.who.int/publications/2010/9789241599290\\_eng.pdf](http://whqlibdoc.who.int/publications/2010/9789241599290_eng.pdf)

World Health Organization & UNICEF. (2021). *Indicators for assessing infant and young child feeding practices: definitions and measurement methods*. License: CC BYNC-SA 3.0 IGO. Available at: <https://www.who.int/publications/i/item/9789240018389>

### 16.2.3 Percent of women of reproductive age consuming a diet of minimum dietary diversity

This indicator estimates the percentage of women of reproductive age in the ZOI population who consume a diet of minimum dietary diversity. A woman of reproductive age is considered to achieve a diet of minimum diversity if she consumed at least 5 of 10 specified food groups during the day and night preceding the survey. It is based on information collected in the core ZOI Survey Module 4, *Women's nutrition*. The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_WHN DIET.do* file.

## Definitions

Numerator	Number of women 15-49 years of age who are de facto household members who consumed at least 5 of 10 specified food groups during the day and night preceding the survey
Denominator	Number of women 15-49 years of age who are de facto household members
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Women of reproductive age
Disaggregation levels	Age category (15-19, 20-49)* 5-year age category Wealth quintile Shock exposure severity
Treatment of missing data	Women missing information about what they consumed the day and night preceding the survey are excluded from the numerator and denominator. All “don’t know” responses are considered to be “no” responses when calculating this indicator. Women missing <b>all</b> data regarding food consumed during the day and night preceding the survey will be omitted from the numerator and denominator.
Survey variables used	hhea, hhnum, strata, wgt_wra, v409, v410, v411, v411a, v412, v413, v414, v415, v416, v417, v418, v419, v420, v421, v422, v423, v424, v429
Analytic variables used	hhmem_df, wra, wra_cage, agegrp_wra, genhhtype_dj, awiquint, shock_sev
Analytic variables created	v409x, v410x, v411x, v411ax, v412x, v413x, v414x, v415x, v416x, v417x, v418x, v419x, v420x, v421x, v422x, v423x, v424x, v429x, whn_foodgrp1-whn_foodgrp10, whn_num_fmiss, whn_fmiss, whn_fscore, whn_mddw, hhmem_wra_df

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Identify the variables associated with the 10 food groups used to generate women’s food score in the persons data file. Pay particular attention if the questionnaire was adapted to include local foods.

**Table 22** lists the food groups and their corresponding variables according to the core ZOI Midline Survey questionnaire.

**Table 22: Ten Food Groups Used to Generate the Women’s Food Score**

Food group		Variables
1	Grains, white roots, tubers <sup>108</sup>	v409=grains v411=white roots or tubers v411a=plantains or bananas
2	Pulses	v422=beans, peas, lentils, or legumes
3	Nuts and seeds (including groundnuts)	v423=nuts or seeds

<sup>108</sup> Also includes plantains and green bananas, which can be starchy staple foods in some areas

Food group		Variables
4	Dairy products	v424=dairy products (milk, cheese, yogurt)
5	Meat, poultry, and fish	v416=organ meats from domesticated animals v417=any meat from domesticated animals v418=organ meats from wild animals v419=any flesh meat from wild animals v421=fresh or dried fish, shellfish, or seafood
6	Eggs	v420=eggs
7	Dark green leafy vegetables	v412=any dark green leafy vegetables
8	Other vitamin A-rich fruits and vegetables	v410=pumpkin, carrots, squash, or sweet potatoes that are yellow or orange inside v414=ripe mangoes, ripe papayas v429=foods made with red palm oil, red palm nuts, or red palm nut pulp sauce <sup>109</sup>
9	Other vegetables	v413=any other vegetables
10	Other fruits	v415=any other fruits

**Step 2.** Create 10 binary variables, 1 for each food group variable (**whn\_foodgrp1-whn\_foodgrp10**).

**Step 2.1.** Create a variable to flag women who ate grains, white roots, or tubers (**whn\_foodgrp1**).

```
Set          whn_foodgrp1=0
Replace      whn_foodgrp1=1 if v409=1 or v411=1 or v411a=1
Label values 0 "No"
              1 "Yes"
Label variable "Woman ate grains, white roots, or tubers"
```

**Step 2.2.** Create a variable to flag women who ate pulses (**whn\_foodgrp2**).

```
Set          whn_foodgrp2=0
Replace      whn_foodgrp2=1 if v422=1
Label values 0 "No"
              1 "Yes"
Label variable "Woman ate pulses"
```

**Step 2.3.** Create a variable to flag women who ate nuts or seeds (**whn\_foodgrp3**).

```
Set          whn_foodgrp3=0
Replace      whn_foodgrp3=1 if v423=1
```

<sup>109</sup> Variable v429, which captures the consumption of foods made with red palm oil, red palm nuts, or red palm nut pulp sauce, is included in food group 8, other vitamin A-rich fruits and vegetables. This deviates from women's minimum dietary diversity indicator calculation guidance (Food and Agriculture Organization of the United Nations, 2021), which states that red palm fruit and red palm pulp should be included but red palm oil should not be included. In the core ZOI Survey questionnaires, red palm oil, red palm nuts, and red palm nut pulp sauce are asked about in the same question, so red palm oil cannot be excluded unless red palm nuts and red palm nut pulp sauce are also excluded.

```

Label values      0 "No"
                  1 "Yes"
Label variable    "Woman ate nuts or seeds"

```

**Step 2.4.** Create a variable to flag women who ate dairy products (**whn\_foodgrp4**).

```

Set              whn_foodgrp4=0
Replace          whn_foodgrp4=1 if v424=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Woman ate dairy products"

```

**Step 2.5.** Create a variable to flag women who ate meat, poultry, or fish (**whn\_foodgrp5**).

```

Set              whn_foodgrp5=0
Replace          whn_foodgrp5=1 if v416=1 or v417=1 or v418=1 or
                  v419=1 or v421=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Woman ate meat, poultry, or fish"

```

**Step 2.6.** Create a variable to flag women who ate eggs (**whn\_foodgrp6**).

```

Set              whn_foodgrp6=0
Replace          whn_foodgrp6=1 if v420=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Woman ate eggs"

```

**Step 2.7.** Create a variable to flag women who ate dark leafy green vegetables (**whn\_foodgrp7**).

```

Set              whn_foodgrp7=0
Replace          whn_foodgrp7=1 if v412=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Woman ate dark green leafy vegetables"

```

**Step 2.8.** Create a variable to flag women who ate vitamin A-rich fruits or vegetables other than dark leafy green vegetables (**whn\_foodgrp8**).

```

Set              whn_foodgrp8=0
Replace          whn_foodgrp8=1 if v410=1 or v414=1 or v429=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Woman ate other vitamin A-rich
                  fruits/vegetables"

```

**Step 2.9.** Create a variable to flag women who ate other vegetables—that is, those not rich in vitamin A (**whn\_foodgrp9**).

```

Set              whn_foodgrp9=0
Replace          whn_foodgrp9=1 if v413=1

```

```

Label values      0 "No"
                  1 "Yes"
Label variable    "Woman ate other vegetables"

```

**Step 2.10.** Create a variable to flag women who ate other fruits—that is, those not rich in vitamin A (`whn_foodgrp10`).

```

Set              whn_foodgrp10=0
Replace          whn_foodgrp10=1 if v415=1
Label values     0 "No"
                  1 "Yes"
Label variable    "Woman ate other fruits"

```

**Step 3.** Create a binary variable that flags women who are missing data for all food variables (`whn_fmiss`).

**Step 3.1.** Create variables that recode responses that have a value of 'no' (2) or 'don't know' (8) to be '0' and missing (9) responses to be blank (missing) if they do not already exist in the data file.

```

For each          variable (var) of variable list: v409 v410 v411
                  v411a v412 v413 v414 v415 v416 v417 v418 v419
                  v420 v421 v422 v423 v424 v429
Set               `var'x=`var'
Replace           `var'x=0 if `var'=2 or `var'=8
Replace           `var'x=missing if `var'=9
Label values     0 "No"
                  1 "Yes"
Label variable    "Recode of `var' ([`var' label])"

```

**Step 3.2.** Create a variable that counts the number of recoded food variables created in Step 3.1 that are missing (`whn_num_fmiss`).

```

Set              whn_num_fmiss=missing
Replace          whn_num_fmiss=number of variables in parentheses=
                  missing(v409x v410x v411x v411ax v412x v413x
                  v414x v415x v416x v417x v418x v419x v420x v421x
                  v422x v423x v424x v429x)
Label variable    "Number of woman's food variables missing values"

```

**Step 3.3.** Create a variable that flags whether a woman is missing all food variables (`whn_fmiss`)—that is, if the value of `whn_num_fmiss` is less than the total number of variables being summed (i.e., 18 in the pseudo code and template analysis Stata do file).

```

Set              whn_fmiss=missing
Replace          whn_fmiss=0 if whn_num_fmiss<18
Replace          whn_fmiss=1 if whn_num_fmiss=18
Label values     0 "No"
                  1 "Yes"
Label variable    "Woman is missing all food data"

```

**Step 4.** Calculate each woman's minimum dietary diversity food score (**whn\_fscore**) by summing the number of food groups consumed during the day and night preceding the survey.

```
Set          whn_fscore=(whn_foodgrp1+whn_foodgrp2+
                        whn_foodgrp3+ whn_foodgrp4+whn_foodgrp5+
                        whn_foodgrp6+whn_foodgrp7+
                        whn_foodgrp8+whn_foodgrp9+whn_foodgrp10)
Replace      whn_fscore=missing if whn_fmiss==1
Label variable "Woman's minimum dietary diversity food score"
```

**Step 5.** Create a binary variable that flags a woman who achieved a minimum dietary diversity (**whn\_mddw**).

```
Set          whn_mddw=missing
Replace      whn_mddw=0 if whn_fscore<5
Replace      whn_mddw=1 if (whn_fscore≥5 and
                        whn_fscore≠missing)
Replace      whn_mddw=missing if whn_fmiss=1
Label values 0 "No"
              1 "Yes"
Label variable "Woman achieved minimum dietary diversity"
```

**Step 6.** Create a variable that captures the sub-population being examined for the indicator calculation—that is, women of reproductive age who are de facto household members among all women of reproductive age surveyed (**hhmem\_wra\_df**).

```
Set          hhmem_wra_df=missing
Replace      hhmem_wra_df=0 if wra=1
Replace      hhmem_wra_df=1 if wra=1 and hhmem_df=1
Label values 0 "No"
              1 "Yes"
Label variable "Woman 15-49 years is a de facto HH member"
```

**Step 7.** After applying the women of reproductive age weight (**wgt\_wra**), calculate the percentage of de facto women of reproductive age who achieved the minimum diet diversity score using the **whn\_mddw** analytic variable. Repeat using the two age category disaggregates as well as the wealth quintile and shock exposure severity disaggregates. (Sample code uses Stata syntax.)

```
svyset hhea [pweight=wgt_w], strata(strata)
svy, subpop(hhmem_wra_df): tab whn_mddw
svy, subpop(hhmem_wra_df): tab whn_mddw wra_cage, col
svy, subpop(hhmem_wra_df): tab whn_mddw agegrp_wra, col
svy, subpop(hhmem_wra_df): tab whn_mddw awiquint, col
svy, subpop(hhmem_wra_df): tab whn_mddw shock_sev, col
```

## References

Arimond, M., et al. (2010). Developing simple measures of women's diet quality in developing countries: Methods and findings. *J Nutr*, 140, 2059S-69 S. Available at: <https://doi.org/10.3945/jn.110.123414>



Food and Agriculture Organization of the United Nations (FAO). (2021.) *Minimum dietary diversity for women*. Rome: FAO. Available at: <https://doi.org/10.4060/cb3434en>

Food and Nutrition Technical Assistance III Project. (2014). *Introducing the minimum dietary diversity–women (MDD-W) global dietary diversity indicator for women*. Available at: [https://www.fantaproject.org/sites/default/files/resources/Introduce-MDD-W-indicator-brief-Sep2014\\_0.pdf](https://www.fantaproject.org/sites/default/files/resources/Introduce-MDD-W-indicator-brief-Sep2014_0.pdf)

INDDEx Project. (2018). *Minimum Dietary Diversity for Women (MDD-W)*. Data4Diets: Building blocks for diet-related food security analysis. Tufts University, Boston, MA. Available at: <https://inddex.nutrition.tufts.edu/data4diets/indicator/minimum-dietary-diversity-women-mdd-w?back=/data4diets/indicators>

Kennedy, G., Ballard, T., & Dop, M.C. (2011). *Guidelines for measuring household and individual dietary diversity*. Rome: Food and Agriculture Organization of the United Nations. Available at: <http://www.fao.org/docrep/014/i1983e/i1983e00.pdf>

Ruel, M. From the WDDS to the MDD-W: Get to know the new indicator for measuring women's dietary diversity [Blog post]. Available at: <https://a4nh.cgiar.org/2015/09/22/get-to-know-the-new-indicator-for-measuring-womens-dietary-diversity/>

## 17. CHILDREN'S AND WOMEN'S NUTRITIONAL STATUS INDICATORS

This chapter describes the four children's and women's nutrition indicators related to anthropometry:

- Prevalence of stunted ( $HAZ < -2$ ) children under 5 (0-59 months)
- Prevalence of wasted ( $WHZ < -2$ ) children under 5 (0-59 months)
- Prevalence of healthy weight ( $WHZ \leq 2$  and  $\geq -2$ ) among children under 5 (0-59 months)
- Prevalence of underweight ( $BMI < 18.5$ ) women of reproductive age (15-49 years of age)

This chapter has two sections; the first section describes the guidelines to construct the indicators using secondary data from The Demographic and Health Surveys (DHS) Program, and the second section outlines the step-by-step procedures to calculate the indicators.

### 17.1 Guidelines to construct the indicators using secondary data

This section provides the guidelines to construct the Feed the Future children's and women's nutritional status indicators using secondary data—specifically from surveys implemented by The DHS Program. In addition to calculating midline ZOI-level estimates, and to enable baseline-midline comparisons, data analysts should also use DHS data to compute baseline ZOI-level estimates for the four anthropometry indicators. If, however, DHS surveys were not conducted at appropriate points in time to enable the calculation of baseline and midline ZOI-level estimates, the indicators may not be calculated.<sup>110</sup> Consult the specific survey protocol to determine whether appropriate DHS data have been identified to produce anthropometry indicator estimates for the indicator assessment and to confirm the data sources that will be used. If data from a recently implemented DHS survey are not available at the time of analysis, the ZOI anthropometry results may be added as an addendum to the midline indicator assessment report at a future date.

#### Potential analysis and sampling issues

DHS Program surveys are designed to provide reliable estimates for most of the key DHS indicators at the national level and at the country's first administrative level (i.e., region, state, or province). The total sample size for a DHS survey depends on the country's demographic profile and the number of regions, states, or provinces.

The ZOI is not a study domain considered during the design of a DHS survey, so caution must be exercised when using DHS data to produce estimates for the ZOI. The reliability of estimates at the ZOI level cannot be guaranteed because the survey was not explicitly designed to produce estimates at this level.

---

<sup>110</sup> The general guidance is that DHS data should be used to calculate the nutritional status indicators for the midline indicator assessment if fieldwork for a DHS survey is conducted within 1 year of the fieldwork for the Midline Survey and fieldwork for the previous DHS survey was conducted within 5 years of fieldwork for the Midline Survey.

The sample-weighted indicator estimates will be generated using Stata's `svy, subpop()` : option to reflect additional variability due to the analysis being performed at ZOI level, which was not controlled for during the sample design. The estimates produced using the `survey, subpop()` option will be the same as those produced using `svy:` command with an "if" statement to define the sub-population. However, the standard errors produced using the `svy, subpop()` : option will be larger because they reflect the additional variability in the ZOI-level estimates.

### 17.1.1 Point-to-polygon geospatial overlay method to identify DHS clusters in the Feed the Future ZOI

To be able to calculate ZOI-level estimates using DHS data, data analysts must first identify the sampled DHS clusters located in the ZOI to include in their analysis. The DHS Program, however, considers the names of clusters to be personally identifiable information and cannot share them with third parties. In its datasets, the DHS Program does make available cluster-level location data, which are the approximate geographic cluster center points based on locations of sampled households that are then offset 0-2 km for urban clusters and 0-5 km for rural clusters (with 1 percent of rural households offset 0-10 km).<sup>111</sup> In all cases, the offset is restricted to remain within the country's second administrative level (Admin 2) area boundary. If random offsets result in a geographic data point falling outside the relevant Admin 2 area's boundary, its placement is adjusted to remain within the Admin 2 boundary.

To determine which DHS clusters are located in the ZOI, data analysts can use shape files for the administrative units that comprise the ZOI and execute an intersection overlay of the DHS data geospatial cluster points onto the administrative units. To accomplish the geospatial overlay, the R programming language spatial package *Simple Feature* (the 'sf') can be used. See Section 17.2.1 for more information.

### 17.1.2 Children's nutritional status indicators

The prevalence of stunted, wasted, and healthy weight children under 5 years of age (0-59 months) are three key Feed the Future children's nutritional status indicators.

**Stunting** is an indicator of suboptimal linear growth that results from cumulative exposure to adverse conditions. These include a range of underlying and basic determinants (e.g., suboptimal hygiene and sanitation, poor access to health services, and food insecurity) that lead to poor health and inadequate diet. However, knowledge about specific determinants of stunting and their causal pathways remains limited. Stunting is sensitive to these factors, but not specific to any one cause. Reducing the prevalence of stunting among children, particularly in the first 1,000 days, is important because height deficits accrued early in life are generally irreversible.

**Wasting** is an indicator of acute undernutrition among children under 5 years of age, resulting from inadequate dietary intake or disease. Exposure to repeated episodes of wasting in childhood can contribute to stunting in the long term. It is a robust predictor of under-5 mortality.

---

<sup>111</sup> [The DHS Program - GPS Data Collection](#) guidance

**Healthy weight** is an indicator of well-nourished children under 5 years of age—those who are neither wasted nor overweight. A population that is well-nourished is essential to enhance human potential, health, and productivity.

To obtain nutritional status indicators from height or length,<sup>112</sup> weight, and age data, the World Health Organization (WHO) growth reference standards (WHO Multicentre Growth Reference Study Group, 2006) are used to compute three nutritional scores, or z-scores. These are the height-for-age z-score (HAZ), relating to stunting, and the weight-for-height z-score (WHZ), relating to wasting and healthy weight. The methodology for calculating z-scores is described as follows.

Each z-score is calculated by comparing the child's height, length, or weight with the median value of the WHO 2006 reference population. The difference is divided by the standard deviation (SD) of the reference population, as shown in the following formula:

$$\text{z-score} = (\text{Individual value of the child} - \text{Median value of children in the reference population}) \div (\text{SD of the reference population})$$

Using this formula, the z-scores are then calculated as follows:

$$\text{HAZ} = (\text{Height of child in the sample} - \text{Median value of height of children in the reference population having the same age and sex}) \div (\text{SD of height of children in the reference population having the same age and sex})$$

$$\text{WHZ} = (\text{Weight of child in the sample} - \text{Median value of weight of children in the reference population having the same height and sex}) \div (\text{SD of weight of children in the reference population having the same height and sex})$$

After calculating the z-scores, outlier z-scores are flagged and excluded from the computation of the nutritional status indicators. The purpose of excluding outliers is to eliminate extreme values that are likely due to measurement or data entry errors. The cutoffs for z-score outliers are shown in **Table 23** (WHO, 2006). Cases less than the minimum z-score cutoff or greater than the maximum z-score cutoff are excluded from the calculation for that nutritional status indicator. Z-scores that are outliers will be set to missing during variable generation in the post-processing stage. The number of cases excluded and the reason should be tracked and reported in the survey report.

**Table 23: Z-score Cutoffs for Children's Nutritional Status Indicators**

Nutritional status indicator	Z-score	Minimum cutoff	Maximum cutoff
Stunting	HAZ	-6 SD	+6 SD
Wasting and healthy weight	WHZ	-5 SD	+5 SD

Source: WHO. (2006). *The WHO child growth standards*.

<sup>112</sup> Length is obtained for children who are measured lying down. Height is obtained for children who are measured standing up. Children 0-23 months of age are measured lying down, and children 24 months of age or older are measured standing up unless they are unable to stand still for the measurement; then they are also measured lying down.

### 17.1.3 Prevalence of underweight women of reproductive age

This indicator provides information about the extent to which women's diets meet their caloric requirements. Adequate energy in the diet is necessary to support the continuing growth of adolescent girls and women's ability to provide optimal care for their children and participate fully in income-generating activities. Undernutrition among women of reproductive age is associated with increased morbidity and poor food security, and undernutrition can result in adverse birth outcomes in future pregnancies. Improvements in women's nutritional status are expected to improve women's work productivity, which should also have benefits for agricultural production.

This indicator measures the percentage of non-pregnant women of reproductive age (15-49 years) who are underweight, as defined by a body mass index (BMI) < 18.5. To calculate an individual's BMI, weight and height data are needed:  $BMI = \frac{\text{weight (in kg)}}{\text{height (in meters)}^2}$ . Criteria for categorizing women's nutritional status using BMI are shown in **Table 24**. Women who have a BMI less than 12 or greater than 60 are excluded from the indicator calculation (Croft, et al., 2018).

**Table 24: Women's Nutritional Status Category by BMI**

Women's nutritional status	BMI
Moderately and severely underweight	12.0 to <17.0
Mildly underweight	17.0 to <18.5
Normal weight	18.5 to <25.0
Overweight	25.0 to <30.0
Obese	≥30.0 to ≤60.0

## 17.2 Step-by-step procedures to calculate nutritional status indicators

This section describes the detailed step-by-step procedures to perform the geospatial overlay to identify the DHS clusters in the ZOI to include in the analysis (Section 17.2.1) and to then calculate the key Feed the Future ZOI nutritional status indicators (Sections 17.2.2-17.2.5), as well as additional anthropometric indicators that are included in the midline indicator assessment reports (Sections 17.2.6 [children] and 17.2.7 [women of reproductive age]). These indicators include the following for children under 5 years of age: the proportion who are severely stunted, the proportion who are severely wasted, the proportion who are overweight for their height, mean HAZ, and mean WHZ; and the following for women of reproductive age: mean BMI, the proportion who are of normal weight, the proportion who are overweight, and the proportion who are obese.

### 17.2.1 Geospatial overlay to identify DHS clusters in the ZOI

The step-by-step instructions in this section follow those included in the R syntax file *Feed the Future ZOI Surveys [COUNTRY] [YEAR] syntax geospatial overlay anthropometry.R*.

**Step 1.** Install and load needed libraries, or packages, in R. The principal packages used in R for this process are the 'sf' package for the geospatial overlay and the 'haven' package for reading in Stata datasets (i.e., .dta files). The 'tidyverse' package contains key packages used for data manipulation and cleaning.

```
install.packages("tidyverse")
```

```
install.packages("sf")
install.packages("haven")

library(tidyverse)
library(sf)
library(haven)
```

**Step 2.** Read the required DHS dataset files and subnational administrative shapefile into R.

The DHS datasets required for this process are the ASCII flat files, which are designated by “FL” and are named as follows, where [CC] is the two-letter country code and [VV] is the dataset version.

- Women 15-49 years of age: [CC]IR[VV]FL.dta
- Household members (used for children): [CC]PR[VV]FL.dta.
- Cluster geospatial shapefile: [CC]GE[VV]FL.shp

A subnational administrative shapefile is also needed so that the administrative units that comprise the ZOI can be defined in Step 3. Two common sources of this file are the government website for the country or the [Humanitarian Data Exchange](#), which many international organizations and governments use to upload datasets and shapefiles.

```
# clear environment
rm(list = ls())

# read data
setwd("[DIRECTORY PATH]")
[COUNTRY]_dhs_female_[YEAR] <- read_dta(paste0(getwd(),
"/[CC]IR[VV]FL.DTA"))

[COUNTRY]_dhs_pr_[YEAR] <- read_dta(paste0(getwd(),
"/[CC]PR[VV]FL.DTA"))

[COUNTRY]_dhs_geospatial_[YEAR] <- read_sf(paste0(getwd(),
"/[CC]GE[VV]FL/[CC]GE[VV]FL.shp"))

[COUNTRY]_adm2_shp <- read_sf("[COUNTRY]_
adm2_1m_gov_[DATE].shp")
```

**Step 3.** In the subnational administrative shapefile, create a binary variable **FTF\_ZOI** that assigns 1 (Yes) to the administrative units in the ZOI and 0 (No) to those outside the ZOI. The administrative units that comprise the ZOI can be found in the sample design document. The variable for the administrative unit used to identify the ZOI (**ADM2** in the template syntax) must be made into a character type variable, and administrative units in the ZOI must be spelled exactly as they are in the shapefile. In the template syntax, **ADM2** is the variable used to define the ZOI, and 10 administrative units in the ZOI are identified as a starting point. Be sure to update the variable name and the number and names of the administrative units as relevant for the country.

```

# convert to character variable
[COUNTRY]_adm2_shp <-
  [COUNTRY]_adm2_shp %>%
  mutate(ADM2 = as.character(ADM2))

# create binary variable FTF_ZOI to identify admin units inside
  and outside the ZOI
[COUNTRY]_adm2_shp <-
  [COUNTRY]_adm2_shp %>%
  mutate(FTF_ZOI = case_when(
    ADM2=="[Admin Unit 1]"~1, ADM2=="[Admin Unit 2]"~1,
    ADM2=="[Admin Unit 3]"~1, ADM2=="[Admin Unit 4]"~1,
    ADM2=="[Admin Unit 5]"~1, ADM2=="[Admin Unit 6]"~1,
    ADM2=="[Admin Unit 7]"~1, ADM2=="[Admin Unit 8]"~1,
    ADM2=="[Admin Unit 9]"~1, ADM2=="[Admin Unit 10]"~1,
    TRUE ~ 0))

# check results
[COUNTRY]_adm3_shp %>%
  select(ADM2, ADM1, FTF_ZOI) %>%
  st_drop_geometry() %>%
  filter(FTF_ZOI == 1) %>%
  arrange(ADM1, ADM2) %>%
  as.data.frame()

sum([COUNTRY]_adm2_shp$FTF_ZOI)

```

**Step 4.** Conduct a point-to-polygon geospatial intersection that identifies the administrative unit for each DHS cluster. This process will also join the relevant data, including the `ftf_zoi` variable, from the subnational shapefile object created in Step 3 to the DHS cluster geospatial shapefile object. To ensure that personally identifiable information is protected, remove any variables (columns) for administrative levels below the Admin 1 level after the overlay.

```

# do join in temporary shp/sf
temp <- st_join(st_as_sf([COUNTRY]_dhs_geospatial_[YEAR]),
  [COUNTRY]_adm2_shp, join = st_intersects)

# view column names
names([COUNTRY]_dhs_geospatial_[YEAR])
# check results - see frequency table on FTF ZOI variable
table(temp$FTF_ZOI)

# name new shapefile/sf object original name
[COUNTRY]_dhs_geospatial_[YEAR] <- temp
# delete temporary shp/sf

```

```
rm(temp)

# drop administrative units levels below Admin 1 to protect PII
[COUNTRY]_dhs_geospatial_[YEAR] <-
  [COUNTRY]_dhs_geospatial_[YEAR] %>%
  select(-contains("ADM2"))
```

**Step 5.** Merge the new DHS cluster geospatial shapefile object created in Step 4 to the DHS household members and women’s datasets read into R in Step 2. The merge with the women’s dataset is done using the `v001` variable in the women’s dataset and the `DHSClust` variable in the DHS geospatial shapefile object. The merge with the DHS household members dataset is done using the `hv001` variable in the household members dataset and the `DHSClust` variable in the DHS cluster geospatial shapefile object.

```
# join geospatial dhs cluster info onto female dhs df
# check names
names([COUNTRY]_dhs_female_[YEAR])

# check to make sure all clusters have match
[COUNTRY]_dhs_female_[YEAR] %>%
  anti_join([COUNTRY]_dhs_geospatial_[YEAR] %>%
    st_drop_geometry(), by = c("v001" = "DHSClust"))

# returns zero - all of it joins - now use left_join
[COUNTRY]_dhs_female_[YEAR] <-
[COUNTRY]_dhs_female_[YEAR] %>%
  left_join([COUNTRY]_dhs_geospatial_[YEAR], by = c("v001" =
    "DHSClust"))

# join geospatial data onto persons df
# check to make sure all clusters have match
[COUNTRY]_dhs_pr_[YEAR] %>%
  anti_join([COUNTRY]_dhs_geospatial_[YEAR], by = c("hv001" =
    "DHSClust"))

# all clusters matched - now merge
[COUNTRY]_dhs_pr_[YEAR] <-
  [COUNTRY]_dhs_pr_[YEAR] %>%
  left_join([COUNTRY]_dhs_geospatial_[YEAR], by = c("hv001" =
    "DHSClust"))
```

**Step 6.** Using R’s `write_dta()` function, create Stata versions of the women’s and household member datasets created in Step 5 to enable the calculation of the anthropometric measures in Stata.

```
# remove geometry to make STATA compatible
[COUNTRY]_dhs_female_[YEAR] <- [COUNTRY]_dhs_female_[YEAR] %>%
```



```

st_drop_geometry() %>%
as_tibble()

[COUNTRY]_dhs_pr_[YEAR] <- [COUNTRY]_dhs_pr_[YEAR] %>%
  st_drop_geometry() %>%
  as_tibble()

haven::write_dta([COUNTRY]_dhs_female_[YEAR],
  path = paste0( getwd(), "[CC]IR[VV]FL_FTF_ZOI.DTA"))
haven::write_dta([COUNTRY]_dhs_pr_[YEAR],
  path = paste0( getwd(), "[CC]PR[VV]FL_FTF_ZOI.DTA"))

```

### 17.2.2 Prevalence of stunted children under 5 years of age

This indicator estimates the percentage of children under 5 years of age (0-59 months) who are moderately to severely stunted. The indicator is calculated using DHS survey data. The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax DHS\_CHN.do* file.

#### Definitions

Numerator	Number of children 0-59 months of age whose HAZ is more than 2 SD below the 2006 WHO Child Growth Standards population median
Denominator	Number of children 0-59 months of age in surveyed households who are de facto household members
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Children under 5 years of age
Disaggregation levels	Child sex* Child age category (0-23 months, 24-59 months)* Child age category (0-11, 12-23, 24-35, 36-47, 48-59 months) Wealth quintile
Treatment of missing data	Children whose height was not measured or responses with missing height information are excluded from the numerator and denominator. Children who are missing an age in months are excluded from the numerator and denominator. Children with out-of-range or invalid z-scores are excluded from the numerator and denominator.
Survey variables used	hv103, hv001, hv005, hv022, hv270, hc1, hc27, hc70
Analytic variables used	ftf_zoi
Analytic variables created	chn_stunted, agegrp_cu5_2grp, agegrp_cu5_5grp

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Load the household member data file that was created in Section 17.2.1.

```
Load "[CC]PR[VV]FL_FTF_ZOI.DTA"
```

**Step 2.** Create a variable to indicate whether a child is stunted (`chn_stunted`). Children with a HAZ score less than -2 (i.e., `hc70<-200`) are categorized as stunted. Children with a HAZ score determined to be invalid (HAZ score `<-6` or HAZ score `>+6`) or missing (`hc70≥9996`) are set to missing.

```
Set          chn_stunted=0 if hv103=1
Replace      chn_stunted=1 if hc70<-200 and hv103=1
Replace      chn_stunted=missing if hc70≥9996
Label values 0 "No"
              1 "Yes"
Label variable "Stunted child under 5 years"
```

**Step 3.** Create a variable that categorizes children under 5 years of age into two age categories (0-23 months and 24-59 months) (`agegrp_cu5_2grp`).

```
Set          agegrp_cu5_2grp=missing
Replace      agegrp_cu5_2grp=0 if hc1<24
Replace      agegrp_cu5_2grp=1 if hc1≥24 and hc1<60
Label values 0 "0-23 months"
              1 "24-59 months"
Label variable "Child under 5 age category (0-23, 24-59 months)"
```

**Step 4.** Create a variable that categorizes children under 5 years of age into 12-month age categories (`agegrp_cu5_5grp`).

```
Set          agegrp_cu5_5grp=missing
Replace      agegrp_cu5_5grp=1 if hc1<12
Replace      agegrp_cu5_5grp=2 if hc1≥12 and hc1<24
Replace      agegrp_cu5_5grp=3 if hc1≥24 and hc1<36
Replace      agegrp_cu5_5grp=4 if hc1≥36 and hc1<48
Replace      agegrp_cu5_5grp=5 if hc1≥48 and hc1<60
Label values 1 "0-11 months"
              2 "12-23 months"
              3 "24-35 months"
              4 "36-47 months"
              5 "48-59 months"
Label variable "Child under 5 age category (12-month groups)"
```

**Step 5.** After applying the household sampling weight, calculate the percentage of children under 5 years of age who are stunted (de facto household members only). Repeat using the child sex (`hc27`) and child age category disaggregates (`agegrp_cu5_2grp` and `agegrp_cu5_5grp`), as well as the wealth quintile disaggregate (`hv270`). Note that in DHS datasets, `hv005` is the household sampling weight without any decimal places, `hv001` is the cluster variable, `hv022` is the strata variable, and `hv103` is the de facto household member variable. (Sample code uses Stata syntax.)

```
Set          wgt_hh=hv005/1000000
svyset hv001 [pw=wgt_hh], strata(hv022) singleunit(scaled)
```

```

svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_stunted
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_stunted hc27, col
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_stunted
    agegrp_cu5_2grp, col
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_stunted
    agegrp_cu5_5grp, col
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_stunted hv270, col

```

### 17.2.3 Prevalence of wasted children under 5 years of age

This indicator estimates the percentage of children under 5 years of age (0-59 months) who are moderately to severely wasted. The indicator is calculated using DHS survey data. The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR]* syntax *DHS\_CHN.do* file.

#### Definitions

Numerator	Number of children 0-59 months of age whose WHZ is more than 2 SD below the 2006 WHO Child Growth Standards population median
Denominator	Number of children 0-59 months of age in surveyed households who are de facto household members
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Children under 5 years of age
Disaggregation levels	Child sex* Child age category (0-23 months, 24-59 months)* Child age category (0-11, 12-23, 24-35, 36-47, 48-59 months) Wealth quintile
Treatment of missing data	Children whose weight was not measured or the response is missing weight information are excluded from the numerator and denominator. Children whose height was not measured or the response is missing height information are excluded from the numerator and denominator. Children who are missing an age in months are excluded from the numerator and denominator. Children with out-of-range or invalid z-scores are excluded from the numerator and denominator.
Survey variables used	hv103, hv001, hv005, hv022, hv270, hc1, hc27, hc72
Analytic variables used	ftf_zoi
Analytic variables created	chn_wasted, agegrp_cu5_2grp, agegrp_cu5_5grp

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Using the same data file used to calculate the stunting indicator in Section 17.2.2, create a variable to indicate whether a child is wasted (**chn\_wasted**). Children with a WHZ score less than -2 (i.e., **hc72**<-200) are categorized as wasted. Children with a WHZ score determined to be invalid (WHZ score <-5 or HAZ score >+5) or missing (**hc72**≥9996) are set to missing.

```
Set          chn_wasted=0 if hv103=1
```

```

Replace      chn_wasted=1 if hc72<-200 and hv103=1
Replace      chn_wasted=missing if hc72>9996
Label values 0 "No"
              1 "Yes"
Label variable "Wasted child under 5 years"

```

**Step 2.** After applying the household sampling weight, calculate the percentage of children under 5 years of age who are wasted (de facto household members only). Repeat using the child sex (**hc27**) and child age category disaggregates (**agegrp\_cu5\_2grp** and **agegrp\_cu5\_5grp**), as well as the wealth quintile disaggregate (**hv270**). Note that in DHS datasets, **hv005** is the household sampling weight without any decimal places, **hv001** is the cluster variable, **hv022** is the strata variable, and **hv103** is the de facto household member variable. (Sample code uses Stata syntax.)

```

Set          wgt_hh=hv005/1000000
svyset hv001 [pw=wgt_hh], strata(hv022) singleunit(scaled)
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_wasted
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_wasted hc27, col
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_wasted
                                agegrp_cu5_2grp, col
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_wasted
                                agegrp_cu5_5grp, col
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_wasted hv270, col

```

#### 17.2.4 Prevalence of healthy weight children under 5 years of age

This indicator estimates the percentage of children under 5 years of age (0-59 months) who have a healthy weight in the ZOI population. The indicator is calculated using DHS survey data. The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR]* syntax *DHS\_CHN.do* file.

#### Definitions

Numerator	Number of children 0-59 months of age whose WHZ is less than or equal to 2 SD below or above the 2006 WHO Child Growth Standards population median
Denominator	Number of children 0-59 months of age in surveyed households who are de facto household members
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Children under 5 years of age
Disaggregation levels	Child sex* Child age category (0-23 months, 24-59 months)* Child age category (0-11, 12-23, 24-35, 36-47, 48-59 months) Wealth quintile
Treatment of missing data	Children not weighed or with missing weight information are excluded from the numerator and denominator. Children whose height was not measured or the response is missing height information are excluded from the numerator and

	denominator. Children who are missing an age in months are excluded from the numerator and denominator. Children with out-of-range or invalid z-scores are excluded from the numerator and denominator.
Survey variables used	hv103, hv001, hv005, hv022, hv270, hc1, hc27, hc72
Analytic variables used	ftf_zoi
Analytic variables created	chn_hw, agegrp_cu5_2grp, agegrp_cu5_5grp

\*Standard Feed the Future disaggregate

## Calculations

**Step 1.** Using the same data file used to calculate the stunting indicator in Section 17.2.2 and wasting indicator in Section 17.2.3, create a new variable to indicate whether a child is of healthy weight (**chn\_hw**). Children with a WHZ score greater or equal to -2 and less than or equal to +2 (i.e.,  $hc72 \geq -200$  and  $hc72 \leq +200$ ) are categorized as being of healthy weight. Children with a WHZ score determined to be invalid (WHZ score  $< -5$  or WHZ score  $> +5$ ) or missing ( $hc72 \geq 9996$ ) are set to missing.

```

Set          chn_hw=0 if hv103=1
Replace      chn_hw=1 if hc72≥-200 and hc72≤+200
Replace      chn_hw=missing if hc72≥9996
Label values 0 "No"
              1 "Yes"
Label variable "Healthy weight child under 5 years"

```

**Step 2.** After applying the household sampling weight, calculate the percentage of children under 5 years of age who are a healthy weight (de facto household members only). Repeat using the child sex (**hc27**) and child age category disaggregates (**agegrp\_cu5\_2grp** and **agegrp\_cu5\_5grp**), as well as the wealth quintile disaggregate (**hv270**). Note that in DHS datasets, **hv005** is the household sampling weight without any decimal places, **hv001** is the cluster variable, **hv022** is the strata variable, **hv103** is the de facto household member variable, **hc27** is the child sex variable, and **hv270** is the wealth quintile variable. (Sample code uses Stata syntax.)

```

Set          wgt_hh=hv005/1000000
svyset hv001 [pw=wgt_hh], strata(hv022) singleunit(scaled)
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_hw
svy, subpop(if ftf_zoi==1 & hv103 ==1): tab chn_hw hc27, col
svy, subpop(if ftf_zoi==1 & hv103 ==1): tab chn_hw
                                agegrp_cu5_2grp, col
svy, subpop(if ftf_zoi==1 & hv103==1): tab chn_hw
                                agegrp_cu5_5grp, col
svy, subpop(if ftf_zoi==1 & hv103 ==1): tab chn_hw hv270, col

```

**Step 3.** Save the data file with the children's anthropometric variables created in Sections 17.2.2-17.2.4.

```

Save "[CC]PR[VV]FL_FTF_ZOI_CHN.DTA"

```

### 17.2.5 Prevalence of underweight women of reproductive age

This indicator estimates the percentage of non-pregnant women of reproductive age (15-49 years) who are underweight, according to their BMI, in the ZOI population. Women's BMI is calculated by dividing women's weight in kilograms by the square of their height in meters. The indicator is calculated using DHS survey data. The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax DHS\_WHN.do* file.

#### Definitions

Numerator	Number of non-pregnant women of reproductive age (15-49 years) who have a BMI less than 18.5
Denominator	Number of non-pregnant women of reproductive in surveyed households who are de facto household members
Unit of measure	Percentage
Level of data	Individual
Sampling weight	Women of reproductive age 15-49 years of age
Disaggregation levels	Age category (15-19, 20-49)* 5-year age categories Wealth quintile
Treatment of missing data	Women who are missing height or weight information are excluded from the numerator and denominator. Women who are missing information about whether they are pregnant or who respond that they are pregnant are excluded from the calculation.
Survey variables used	b3_01, b19_01, v005, v008, v022, v012, v190, v213, v445
Analytic variables used	ftf_zoi
Analytic variables created*	cage_youngest, whn_uw, agegrp_wra_2grp, agegrp_wra_7grp

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Load the woman's data file that was created in Section 17.2.1.

```
Load " [CC] IR[VV] FL_FTF_ZOI.DTA"
```

**Step 2.** Create a variable that indicates the age of each woman's youngest child (**cage\_youngest**).

The variable will be used in the next step to exclude women who gave birth to a child during the 2 months preceding the survey. To create **cage\_youngest**, use **b19\_01** (current age in months of the woman's youngest child) if it exists in the data file. If it does not exist, subtract the date of birth of the woman's youngest child (**b3\_01**) from the date of the interview (**v008**). The variable **cage\_youngest** will have a value of missing if a woman does not have any children.

```
Set cage_youngest=b19_01 if b19_01 is in data file
Set cage_youngest=v008-b3_01 if b19_01 is not in data file
Label variable "Age in months of woman's youngest child"
```

**Step 3.** Create a variable that indicates whether a woman 15-49 years of age is underweight according to their BMI ( $\text{BMI} < 18.5 \text{ kg/m}^2$ ) (`whn_uw`). That is, use `v445` (women's BMI without decimals) to identify women who are underweight. Women with a `v455` value between 1200 and 1849 are underweight; women with a `v455` value between 1850 and 6000 are not underweight. Women with a `v455` value less than 1200 or greater than 6000 are excluded from the calculation. Be sure to exclude women who reported that they were pregnant at the time of interview (`v213=1`) or gave birth to a child during the 2 months preceding the survey (`cage_youngest<2`).

```
Set          whn_uw=missing
Replace      whn_uw=0 if v445≥1200 and v445≤6000
Replace      whn_uw=1 if v445≥1200 and v445≤1849
Replace      whn_uw=missing if v213=1 or cage_youngest<2
Label values 0 "No"
              1 "Yes"
Label variable "Underweight woman, BMI<18.5kg/m2"
```

**Step 4.** Create a variable that categorizes women 15-49 years of age into two age categories (15-19 and 20-49 years of age) (`agegrp_wra_2grp`) using survey variable `v012`, women's age in years.

```
Set          agegrp_wra_2grp=missing
Replace      agegrp_wra_2grp=1 if v012≥15 and v012≤19
Replace      agegrp_wra_2grp=2 if v012≥20 and v012≤49
Label values 1 "15-19"
              2 "20-49"
Label variable "Women 15-49 age category (15-19, 20-49 years)"
```

**Step 5.** Create a variable that categorizes women 15-49 years of age into 5-year age categories (`agegrp_wra_7grp`).

```
Set          agegrp_wra_7grp=missing
Replace      agegrp_wra_7grp=1 if v012≥15 and v012≤19
Replace      agegrp_wra_7grp=2 if v012≥20 and v012≤24
Replace      agegrp_wra_7grp=3 if v012≥25 and v012≤29
Replace      agegrp_wra_7grp=4 if v012≥30 and v012≤34
Replace      agegrp_wra_7grp=5 if v012≥35 and v012≤39
Replace      agegrp_wra_7grp=6 if v012≥40 and v012≤44
Replace      agegrp_wra_7grp=7 if v012≥45 and v012≤49
Label values 1 "15-19"
              2 "20-24"
              3 "25-29"
              4 "30-34"
              5 "35-39"
              6 "40-44"
              7 "45-49"
Label variable "Women 15-49 age category (5-year groups)"
```

**Step 6.** After applying the women's individual sampling weight, calculate the percentage of women 15-49 years of age who are underweight according to their BMI. Repeat using the women's age category disaggregates (`agegrp_wra_2grp` and `agegrp_wra_7grp`), as well as the wealth quintile disaggregate (`v190`). Note that in DHS women's data files, `v005` is the women's individual sampling weight without any decimal places, `v001` is the cluster variable, and `v022` is the strata variable. Because the women's file contains records for only de facto household members, there is no need to specify de facto household members in the sub-population definition. (Sample code uses Stata syntax.)

```
Set                wgt_wra=v005/1000000
svyset v001 [pw=wgt_wra], strata(v022) singleunit(scaled)
svy, subpop(fff_zoi): tab whn_uw
svy, subpop(fff_zoi): tab whn_uw agegrp_wra_2grp, col
svy, subpop(fff_zoi): tab whn_uw agegrp_wra_7grp, col
svy, subpop(fff_zoi): tab whn_uw v190, col
```

**Step 7.** Save the data file with the women's anthropometry variables created in this section.

```
Save "[CC]IR[VV]FL_FTF_ZOI_WHN.DTA"
```

## 17.2.6 Additional children's anthropometric indicators

This section describes how to create additional children's anthropometric indicators that are presented in midline indicator assessment reports. They can be added to the data file, `[CC]PR[VV]FL_FTF_ZOI_CHN.DTA`, created in Section 17.2.4, Step 4, that also has the other children's anthropometric indicator variables, and the sample-weighted estimates can be generated using the same approach taken for the key Feed the Future children's anthropometric indicators (e.g., see Section 17.2.2, Step 5).

**Severely stunted.** Create a variable to indicate whether a child is severely stunted (`chn_sev_stunted`). Children with a HAZ score less than -3 (i.e., `hc70<-300`) are categorized as severely stunted. Children with a HAZ score determined to be invalid (HAZ score `<-6` or HAZ score `>+6`) or missing (`hc70≥9996`) are set to missing.

```
Set                chn_sev_stunted=0 if hv103=1
Replace            chn_sev_stunted=1 if hc70<-300 & hv103=1
Replace            chn_sev_stunted=missing if hc70≥9996
Label values       0 "No"
                   1 "Yes"
Label variable      "Severely stunted child under 5 years"
```

**Mean HAZ.** Create a variable that can be used to calculate the mean HAZ score for children under 5 years of age (`chn_haz`). Children with a HAZ score determined to be invalid (HAZ score `<-6` or HAZ score `>+6`) or missing (`hc70≥9996`) are set to missing.

```
Set                chn_haz=hc70/100 if hc70<9996
Label variable      "Mean z-score for height-for-age for children
under 5 years"
```



**Severely wasted.** Create a variable to indicate whether a child is severely wasted (**chn\_sev\_wasted**). Children with a WHZ score less than -3 (i.e., **hc72<-300**) are categorized as severely wasted. Children with a WHZ score determined to be invalid (WHZ score <-5 or WHZ score >+5) or missing (**hc72≥9996**) are set to missing.

```
Set          chn_sev_wasted=0 if hv103=1
Replace      chn_sev_wasted=1 if hc72<-300 and hv103=1
Replace      chn_sev_wasted=missing if hc72≥9996
Label values 0 "No"
              1 "Yes"
Label variable "Severely wasted child under 5 years"
```

**Overweight for height.** Create a variable to indicate whether a child is overweight for their height (**chn\_ow**). Children with a WHZ score greater than +2 (i.e., **hc72>+200**) are categorized as overweight for their height. Children with a WHZ score determined to be invalid (WHZ score <-5 or WHZ score >+5) or missing (**hc72≥9996**) are set to missing.

```
Set          chn_ow=0 if hv103=1
Replace      chn_ow=1 if hc72>+200 and hv103=1 and hc72<9996
Replace      chn_ow=missing if hc72≥9996
Label values 0 "No"
              1 "Yes"
Label variable "Overweight for height child under 5 years"
```

**Obese for height.** Create a variable to indicate whether a child is obese for their height (**chn\_obese**). Children with a WHZ score greater than +3 (i.e., **hc72>+300**) are categorized as obese for their height. Children with a WHZ score determined to be invalid (WHZ score <-5 or WHZ score >+5) or missing (**hc72≥9996**) are set to missing.

```
Set          chn_obese=0 if hv103=1
Replace      chn_obese=1 if hc72>+300 and hv103=1 and
              hc72<9996
Replace      chn_obese=missing if hc72≥9996
Label values 0 "No"
              1 "Yes"
Label variable "Obese for height child under 5 years"
```

**Mean WHZ.** Create a variable that can be used to calculate the mean WHZ score for children under 5 years of age (**chn\_whz**). Children with a WHZ score determined to be invalid (WHZ score <-5 or WHZ score >+5) or missing (**hc72≥9996**) are set to missing.

```
Set          chn_whz=hc72/100 if hc72<9996
Label variable "Mean z-score for weight-for-height for children
              under 5 years"
```

### 17.2.7 Additional women's anthropometric indicators

This section describes how to create additional women's anthropometric indicators that are presented in midline indicator assessment reports. They can be added to the data file, [CC]IR[VV]FL\_FTF\_ZOI\_CHN.DTA, created in Section 17.2.4, Step 4, that also has the other women's anthropometric indicator variables, and the sample-weighted estimates can be generated using the same approach taken for the key Feed the Future women's anthropometric indicator (e.g., see Section 17.2.5, Step 6).

**Mean BMI.** Create a variable that can be used to calculate the mean BMI for women 15-49 years of age (**whn\_bmi**).

```
Set          whn_bmi=missing
Replace      whn_bmi=v445 if v445≥1200 and v445≤6000 and v213
              ≠1 and (v208=0 or cage_youngest≥2)
Label variable "Mean BMI, women 15-49 years"
```

**Normal weight.** Create a variable that indicates whether a woman 15-49 years of age is of normal weight according to their BMI ( $BMI \geq 18.5 \text{ kg/m}^2$  and  $BMI < 25.0 \text{ kg/m}^2$ ) (**whn\_hw**). That is, use **v445** (women's BMI without decimals) to identify women who are of normal weight. Women with a **v455** value greater than or equal to 1850 and less than 2500 are of normal weight; women with a **v455** value greater than or equal to 1200 and less than 1850 or between 2500 and 6000 are not of normal weight. Women with a **v455** value less than 1200 or greater than 6000 are excluded from the calculation. Be sure to exclude women who reported that they were pregnant at the time of interview (**v213=1**) or gave birth to a child during the 2 months preceding the survey (**cage\_youngest<2**).

```
Set          whn_hw=missing
Replace      whn_hw=0 if v445≥1200 and v445≤6000
Replace      whn_hw=1 if v445≥1850 and v445<2500
Replace      whn_hw=missing if v213=1 or cage_youngest<2
Label values 0 "No"
              1 "Yes"
Label variable "Normal weight (BMI>=18.5kg/m2, BMI<25.0kg/m2)"
```

**Overweight.** Create a variable that indicates whether a woman 15-49 years of age is overweight according to their BMI ( $BMI \geq 25.0 \text{ kg/m}^2$  and  $BMI < 30.0 \text{ kg/m}^2$ ) (**whn\_ow**). That is, use **v445** (women's BMI without decimals) to identify women who are overweight. Women with a **v455** value greater than or equal to 2500 and less than 3000 are overweight; women with a **v455** value greater than or equal to 1200 and less than 2500 or between 3000 and 6000 are not overweight. Women with a **v455** value less than 1200 or greater than 6000 are excluded from the calculation. Be sure to exclude women who reported that they were pregnant at the time of interview (**v213=1**) or gave birth to a child during the 2 months preceding the survey (**cage\_youngest<2**).

```
Set          whn_ow=missing
Replace      whn_ow=0 if v445≥1200 and v445≤6000
Replace      whn_ow=1 if v445≥2500 and v445<3000
```

```

Replace      whn_ow=missing if v213=1 or cage_youngest<2
Label values 0 "No"
              1 "Yes"
Label variable "Overweight (BMI>=25.0kg/m2, BMI<30.0kg/m2) "

```

**Obese.** Create a variable that indicates whether a woman 15-49 years of age is obese according to their BMI ( $BMI \geq 30.0 \text{ kg/m}^2$ ) (**whn\_obese**). That is, use **v445** (women's BMI without decimals) to identify women who are obese. Women with a **v445** value greater than or equal to 3000 and less than 6000 are obese; women with a **v445** value greater than or equal to 1200 and less than 3000 are not obese. Women with a **v445** value less than 1200 or greater than 6000 are excluded from the calculation. Be sure to exclude women who reported that they were pregnant at the time of interview (**v213=1**) or gave birth to a child during the 2 months preceding the survey (**cage\_youngest<2**).

```

Set          whn_obese=missing
Replace      whn_obese =0 if v445≥1200 and v445≤6000
Replace      whn_obese=1 if v445≥3000 and v445<6000
Replace      whn_obese=missing if v213=1 or cage_youngest<2
Label values 0 "No"
              1 "Yes"
Label variable "Obese (BMI>=30.0kg/m2) "

```

**BMI Category.** Create a categorical variable that captures the percent distribution women by their nutritional status according to their BMI category (**bmi\_cat**). The variable includes four categories: underweight, healthy weight, overweight, and obese.

```

Set          bmi_cat=1 if whn_uw=1
Replace      bmi_cat=2 if whn_hw=1
Replace      bmi_cat=3 if whn_ow=1
Replace      bmi_cat=4 if whn_obese=1
Label values 1 "Underweight"
              2 "Healthy weight"
              3 "Overweight"
              4 "Obese"
Label variable "BMI Category"

```

### 17.3 References

Assaf, S. (2019). NT\_CH\_NUT.do [Source code]. Available at: [https://github.com/DHSProgram/DHS-Indicators-Stata/blob/master/Chap11\\_NT/NT\\_CH\\_NUT.do](https://github.com/DHSProgram/DHS-Indicators-Stata/blob/master/Chap11_NT/NT_CH_NUT.do)

Assaf, S. (2020). NT\_WN\_NUT.do [Source code]. Available at: [https://github.com/DHSProgram/DHS-Indicators-Stata/blob/master/Chap11\\_NT/NT\\_WM\\_NUT.do](https://github.com/DHSProgram/DHS-Indicators-Stata/blob/master/Chap11_NT/NT_WM_NUT.do)

Croft, T.N., Marshall, A.M.J., Allen, C.K., et al. (2018). *Guide to DHS statistics*. Rockville, Maryland, USA: ICF. Available at: [https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.htm#t=Guide\\_to\\_DHS\\_Statistics\\_DHS-7.htm](https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.htm#t=Guide_to_DHS_Statistics_DHS-7.htm)

Pebesma, E., & Bivand, R. (2023). *Spatial data science: With applications in R* (1st ed). Chapman and Hall/CRC. Available at: <https://doi.org/10.1201/9780429459016>

World Health Organization. (2006). *The WHO child growth standards*. Available at: <https://www.who.int/tools/child-growth-standards>

World Health Organization. (2016). *Factsheet: Obesity and overweight*. Available at: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

World Health Organization. (2019). *Global database on child growth and malnutrition*. Available at: <https://platform.who.int/nutrition/malnutrition-database>

World Health Organization. (2023). *The Global Health Observatory: Body mass index (BMI)*. Available at: <https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/body-mass-index>

World Health Organization & UNICEF. (2009). *WHO child growth standards and the identification of severe acute malnutrition in infants and children*. A joint statement by the World Health Organization and UNICEF. Available at: [https://apps.who.int/iris/bitstream/handle/10665/44129/9789241598163\\_eng.pdf?ua=1](https://apps.who.int/iris/bitstream/handle/10665/44129/9789241598163_eng.pdf?ua=1)

World Health Organization & UNICEF. (2021). *Indicators for assessing infant and young child feeding practices: definitions and measurement methods*. Available at: <https://www.who.int/publications/i/item/9789240018389>

## 18. CLIMATE ADAPTATION

### 18.1 Background

The United States Agency for International Development's Climate Strategy guides its approach to confront the climate crisis (United States Agency for International Development, 2022). Included in its standard ZOI-level reporting, Feed the Future reports the proportion of producers of targeted value chains that apply improved climate adaptation management practices and technologies under the climate adaption technology type disaggregate category. As part of the Bureau for Resilience and Food Security's effort to further build an evidence base around climate knowledge and the types of programs being implemented to adapt to climate change in areas that Feed the Future targets, additional climate-related questions were incorporated into ZOI Midline Surveys. The descriptive findings related to the climate-related data collected are presented for the ZOI in Section 5.5 of the midline indicator assessment reports. The findings span the following topics:

- Knowledge, perception, and concern about the impacts of climate change on daily life and income
- Receipt of information to help prepare for and manage climate and weather-related challenges
- Participation in training, receipt of support, and adoption of practices to help manage climate and weather-related challenges related to crop production
- Participation in training, receipt of support, and adoption of practices to help manage climate and weather-related challenges related to raising livestock or other farm animals
- Barriers faced in adopting crop or farm animal practices to help manage the impacts of climate and weather-related problems
- Participation in groups or committees working to strengthen communities' ability to manage climate and weather-related problems
- Insurance obtained to protect against weather-related problems related to cultivating crops or raising livestock or other farm animals
- Receipt of payments for climate and weather-related problems that affected crop or farm animal production
- Participation in groups or committees that monitor or influence land use
- Participation in training or receipt of support to help diversify sources of income

### 18.2 Step-by-step instructions for populating the climate tables

This section presents the step-by-step procedures for calculating the sample-weighted estimates that are presented in climate adaptation results tables in Section 5.5 of the midline indicator assessment report template. The instructions follow the syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax climate.do*. Variables from Module 2A, *Climate adaptation*, are used for the calculations. All template syntax for calculating sample-weighted estimates in this section uses Stata syntax; the template syntax for other analysis commands (e.g., for creating variables, merging files, and reshaping data) uses pseudo-syntax. For all estimates populated in the tables, report the unweighted number of observations included in the calculation in the corresponding number of observations (i.e., the "n") table field.

## Definitions

Level of data	Household
Sampling weight	Household
Treatment of missing data	All missing and refused responses are excluded from calculations presented in this section
Survey variables used	<b>hhea</b> , <b>wgt_hh</b> , <b>strata</b> , <b>m1_line</b> , <b>v233</b> , <b>v225</b> , <b>v225a</b> , all survey variables that begin with `v2a`
Analytic variables used	<b>age15_29y</b> , <b>sex</b>
Analytic variables created	<b>v2a108ax-v2a108gx</b> , <b>v2a108xx</b> , <b>v2a110ax-v2a110fx</b> , <b>v2a110xx</b> , <b>v2a112ax-v2a112gx</b> , <b>v2a112xx</b> , <b>v2a114ax-v2a114fx</b> , <b>v2a114xx</b> , <b>ag_barriers</b>

## Calculations

**Step 1.** Load the household-level analytic data file.

```
Load          "FTF ZOI Survey [COUNTRY] [YEAR] household data  
              analytic", clear
```

**Step 2.** Apply the complex survey design with the household sample weight (**wgt\_hh**), where **hhea** is the primary sampling unit or cluster and **strata** is the strata variable.

```
svyset hhea [pw=wgt_hh], strata(strata)
```

### 18.2.1 Table 5.5.1: Knowledge, Perception, and Concern about Impacts of Climate Change on Daily Life and Income among Households in the ZOI

**Step 3.** Calculate the sample-weighted estimates required for Table 5.5.1.

**Step 3.1.** Add the age and sex variables for the Module 2A respondents to the working data file from the persons-level analytic data file so that the results in Table 5.5.1 can be disaggregated by respondents' sex and age (15-29 years of age and 30 years of age or older). To do so, first create a copy of the Module 2A respondent line number variable (**v2a100rsp**) that is the same as the household member's line number variable in the persons-level analytic data file (**m1\_line**) so that the key variables have the same names in both the persons- and household-level analytic data files. Then add the variables **age15\_29y** and **sex**.

```
Set          m1_line=v2a100rsp  
Merge       1 to 1  
key variables hhea hhnum m1_line  
data file    "FTF ZOI Survey [COUNTRY] [YEAR] persons data  
              analytic"  
keep variables age15_29y sex
```

**Step 3.2.** Calculate the sample-weighted percent distribution of respondents by their self-reported knowledge about climate change using **v2a101**. Populate the table with the percentage of all respondents with no knowledge, which is indicated by a value of 1 (nothing); the percentage of all respondents with little knowledge, which is indicated by a value of 2 (a little); the percentage of all

respondents with some knowledge, which is indicated by a value of 3 (some/moderate); and the percentage of all respondents with a lot of knowledge, which is indicated by a value of 4 (a lot). Also populate the table with the same percentages for male respondents, female respondents, respondents 15-29 years of age, and respondents 30 years of age or older.

```
svy: tab v2a101, perc format(%6.1f)
svy: tab v2a101 sex, col perc format(%6.1f)
svy: tab v2a101 age15_29y, col perc format(%6.1f)
```

**Step 3.2.** Calculate the sample-weighted percentage of respondents who perceive that weather has become more problematic over their lifetime using **v2a102**. Populate the table with the percentage of all respondents who believe that weather has become more problematic, which is indicated by a value of 1 (gotten worse). Also populate the table with the same percentage for male respondents, female respondents, respondents 15-29 years of age, and respondents 30 years of age or older.

```
svy: tab v2a102, perc format(%6.1f)
svy: tab v2a102 sex, col perc format(%6.1f)
svy: tab v2a102 age15_29y, col perc format(%6.1f)
```

**Step 3.3.** Calculate the sample-weighted percent distribution of respondents who are concerned about the impacts of climate change on their income using **v2a106**. Populate the table with the percentages of all respondents who are not concerned, which is indicated by a value of 1 (not concerned); who are somewhat concerned, which is indicated by a value of 2 (somewhat concerned); who are moderately concerned, which is indicated by a value of 3 (moderately concerned); and who are very concerned, which is indicated by a value of 4 (very concerned). Also populate the table with the same percentages for male respondents, female respondents, respondents 15-29 years of age, and respondents 30 years of age or older.

```
svy: tab v2a106, perc format(%6.1f)
svy: tab v2a106 sex, col perc format(%6.1f)
svy: tab v2a106 age15_29y, col perc format(%6.1f)
```

**Step 3.4.** In the “Number of households (n)” row, populate the unweighted numbers of respondents for each population included in the calculations (i.e., all respondents, male respondents, female respondents, respondents 15-29 years of age, and respondents 30 years of age or older) in the corresponding column of the table. The unweighted number of respondents should be the same across results presented in the table for a particular population; if the numbers vary, add a table note that provides information about the missing observations.

### **18.2.2 Table 5.5.2: Percent of Households in the ZOI that Received Information to Help Prepare for or Manage Weather-related Challenges During the 12 months Preceding the Survey**

**Step 4.** Calculate the sample-weighted estimates required for Table 5.5.2.

**Step 4.1.** Calculate the sample-weighted percentage of households that received information to help prepare for or manage weather-related problems during the 12 months preceding the survey using **v2a103**. Populate the table with the percentage of households with a value of 1 (yes) for **v2a103**.

```
svy: tab v2a103, perc format(%6.1f)
```

**Step 4.2.** In the “Number of households (n)” row, populate the unweighted number of households included in the calculation.

**Step 4.3.** Calculate the sample-weighted percentages of households that received information by mechanism using the binary yes/no variables created during data processing for each v2a105 response option (e.g., v2a105a, v2a105b, v2a105c, v2a105d, v2a105e, v2a105f, v2a105g, v2a105h, v2a105i, v2a105j, and v2a105x). For each mechanism, populate the table with the percentage of households with a value of 1 (yes) for the corresponding mechanism variable.

```
svy: tab v2a105a, perc format(%6.1f)
svy: tab v2a105b, perc format(%6.1f)
svy: tab v2a105c, perc format(%6.1f)
svy: tab v2a105d, perc format(%6.1f)
svy: tab v2a105e, perc format(%6.1f)
svy: tab v2a105f, perc format(%6.1f)
svy: tab v2a105g, perc format(%6.1f)
svy: tab v2a105h, perc format(%6.1f)
svy: tab v2a105i, perc format(%6.1f)
svy: tab v2a105j, perc format(%6.1f)
svy: tab v2a105x, perc format(%6.1f)
```

**Step 4.4.** In the “Number of households that received information (n)” row, populate the unweighted number of households included in the calculations.

### 18.2.3 Table 5.5.3: Percent of Households in the ZOI that Received Training or Support or Adopted Practices to Help Manage Weather-related Challenges Related to Crop Production During the 12 Months Preceding the Survey

**Step 5.** Calculate the sample-weighted estimates required for Table 5.5.3.

**Step 5.1.** Calculate the sample-weighted percentage of households that cultivated any crops during the 12 months preceding the survey that received any training or support to help manage weather-related challenges to crop production during this time period using v2a108. Populate the table with the percentage of households with a value of 1 (yes) for v2a108.

```
svy: tab v2a108, perc format(%6.1f)
```

**Step 5.2.** Create variables v2a108ax-v2a108gx and v2a108xx that set all households with a value of 2 (no) or 8 (don’t know) for the corresponding survey variable to have a value of 0 (no).

```
For each      x of variable list v2a108a-v2a108g v2a108x:
  Set          `x' x=0 if `x'=2 or `x'=8
  Replace      `x' x=1 if `x'=1
  Label values 0 "No"
               1 "Yes"
  Label variable ``x' for all HHs that cultivated crops"
```



**Step 5.3.** Calculate the sample-weighted percentages of households that cultivated any crops during the 12 months preceding the survey that received any training or support to help manage weather-related challenges to crop production during the 12 months preceding the survey by topic (i.e., **v2a108ax**, **v2a108bx**, **v2a108cx**, **v2a108dx**, **v2a108ex**, **v2a108fx**, **v2a108gx**, and **v2a108xx**). For each topic, populate the table with the percentage of households with a value of 1 (yes) for the corresponding topic variable created in Step 5.2.

```
svy: tab v2a108ax, perc format(%6.1f)
svy: tab v2a108bx, perc format(%6.1f)
svy: tab v2a108cx, perc format(%6.1f)
svy: tab v2a108dx, perc format(%6.1f)
svy: tab v2a108ex, perc format(%6.1f)
svy: tab v2a108fx, perc format(%6.1f)
svy: tab v2a108gx, perc format(%6.1f)
svy: tab v2a108xx, perc format(%6.1f)
```

**Step 5.4.** Calculate the sample-weighted percentage of households that cultivated any crops during the 12 months preceding the survey that adopted any practices to manage the impacts of weather-related problems on their crop production during this time period using **v2a110**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a110**.

```
svy: tab v2a110, perc format(%6.1f)
```

**Step 5.5.** Create variables **v2a110ax-v2a110fx** and **v2a110xx** that set all households with a value of 2 (no) or 8 (don't know) for the corresponding survey variable to have a value of 0 (no).

```
For each      x of variable list v2a110a-v2a110f v2a110x:
  Set          `x'x=0 if `x'=2 or `x'=8
  Replace      `x'x=1 if `x'=1
  Label values 0 "No"
               1 "Yes"
  Label variable ``x' for all HHs that cultivated crops"
```

**Step 5.6.** Calculate the sample-weighted percentages of households that cultivated any crops during the 12 months preceding the survey that adopted any practices to manage the impacts of weather-related problems on their crop production by topic (i.e., **v2a110ax**, **v2a110bx**, **v2a110cx**, **v2a110dx**, **v2a110ex**, **v2a110fx**, and **v2a110xx**). For each topic, populate the table with the percentage of households with a value of 1 (yes) for the corresponding topic variable created in Step 5.5.

```
svy: tab v2a110ax, perc format(%6.1f)
svy: tab v2a110bx, perc format(%6.1f)
svy: tab v2a110cx, perc format(%6.1f)
svy: tab v2a110dx, perc format(%6.1f)
svy: tab v2a110ex, perc format(%6.1f)
svy: tab v2a110fx, perc format(%6.1f)
svy: tab v2a110xx, perc format(%6.1f)
```

**Step 5.7.** In the “Number of households that grew crops (n)” row, populate the unweighted number of households included in the calculations. The unweighted number of households should be the same for

the two columns; if the numbers differ, add a table note that provides information about the missing observations.

#### **18.2.4 Table 5.5.4: Percent of Households in the ZOI that Received Training or Support or Adopted Practices to Manage Weather-related Challenges Related to Raising Livestock or Other Farm Animals During the 12 Months Preceding the Survey**

**Step 6.** Calculate the sample-weighted estimates required for Table 5.5.4.

**Step 6.1.** Calculate the sample-weighted percentage of households that raised livestock or other farm animals during the 12 months preceding the survey that received any training or support to help manage weather-related challenges to farm animal production during that time period using **v2a112**. Populate the table with the percentage of households with a value of 1 (yes) for **v2a112**.

```
svy: tab v2a112, perc format(%6.1f)
```

**Step 6.2.** Create variables **v2a112ax-v2a112gx** and **v2a112xx** that set all households with a value of 2 (no) or 8 (don't know) for the corresponding survey variable to have a value of 0 (no).

```
For each      x of variable list v2a112a-v2a112g v2a112x:
  Set          `x' x=0 if `x'=2 or `x'=8
  Replace      `x' x=1 if `x'=1
  Label values 0 "No"
               1 "Yes"
  Label variable "`x' for all HHs that raised farm animals"
```

**Step 6.3.** Calculate the sample-weighted percentages of households that raised livestock or other farm animals during the 12 months preceding the survey that received any training or support to help manage weather-related challenges to farm animal production during the 12 months preceding the survey by topic (i.e., **v2a112ax**, **v2a112bx**, **v2a112cx**, **v2a112dx**, **v2a112ex**, **v2a112fx**, **v2a112gx**, and **v2a112xx**). For each topic, populate the table with the percentage of households with a value of 1 (yes) for the corresponding topic variable created in Step 6.2.

```
svy: tab v2a112ax, perc format(%6.1f)
svy: tab v2a112bx, perc format(%6.1f)
svy: tab v2a112cx, perc format(%6.1f)
svy: tab v2a112dx, perc format(%6.1f)
svy: tab v2a112ex, perc format(%6.1f)
svy: tab v2a112fx, perc format(%6.1f)
svy: tab v2a112gx, perc format(%6.1f)
svy: tab v2a112xx, perc format(%6.1f)
```

**Step 6.4.** Calculate the sample-weighted percentage of households that raised livestock or other farm animals during the 12 months preceding the survey that adopted any practices to manage the impacts of weather-related problems on their farm animal production during the 12 months preceding the survey using **v2a114**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a114**.

```
Svy: tab v2a114, perc format(%6.1f)
```

**Step 6.5.** Create variables **v2a114ax-v2a114fx** and **v2a114xx** that set all households with a value of 2 (no) or 8 (don't know) for the corresponding survey variable to have a value of 0 (no).

```
For each      x of variable list v2a114a-v2a114f v2a114x:
  Set          `x' x=0 if `x'=2 or `x'=8
  Replace      `x' x=1 if `x'=1
  Label values 0 "No"
               1 "Yes"
  Label variable "`x' for all HHs that raised farm animals"
```

**Step 6.6.** Calculate the sample-weighted percentages of households that raised livestock or other farm animals during the 12 months preceding the survey that adopted any practices to manage the impacts of weather-related problems on their farm animal production during the 12 months preceding the survey by topic (i.e., **v2a114ax**, **v2a114bx**, **v2a114cx**, **v2a114dx**, **v2a114ex**, **v2a114fx**, and **v2a114xx**). For each topic, populate the table with the percentage of all with a value of 1 (yes) for the corresponding topic variable created in Step 6.5.

```
svy: tab v2a114ax, perc format(%6.1f)
svy: tab v2a114bx, perc format(%6.1f)
svy: tab v2a114cx, perc format(%6.1f)
svy: tab v2a114dx, perc format(%6.1f)
svy: tab v2a114ex, perc format(%6.1f)
svy: tab v2a114fx, perc format(%6.1f)
svy: tab v2a114xx, perc format(%6.1f)
```

**Step 6.7.** In the “Number of households that raised livestock or other farm animals (n)” row, populate the unweighted number of households included in the calculations. The unweighted number of households should be the same for the two columns; if the numbers differ, add a table note that provides information about the missing observations.

### 18.2.5 Table 5.5.5: Percent of Households in the ZOI that Faced Barriers to Adopting Crop or Farm Animal Practices to Help Manage the Impacts of Weather-related Problems During the 12 Months Preceding the Survey

**Step 7.** Calculate the sample-weighted estimates required for Table 5.5.5.

**Step 7.1.** Calculate the sample-weighted percentage of households that cultivated crops or raised farm animals during the 12 months preceding the survey that faced barriers to adopting crop or farm animal practices to help manage the impacts of weather-related problems during that time period. To do so, create an analytic variable, **ag\_barriers**, that excludes households with a **v2a115** response of “not applicable” (i.e., households that did not try to adopt new practices to help manage the impacts of weather-related problems). Populate the table with the percentage of households with a value of 1 (yes) for **ag\_barriers**.

```
Set          ag_barriers=v2a115
Replace      ag_barriers=missing if v2a115=3
Replace      ag_barriers=0 if v2a115=2 or v2a115=8
Label values 0 "No"
               1 "Yes"
```

Label variable "HH faced barriers to adopting climate practices"

```
svy: tab ag_barriers, perc format(%6.1f)
```

**Step 7.2.** In the "Number of households that grew crops or raised farm animals (n)" row, populate the unweighted number of households included in the calculation.

**Step 7.3.** Calculate the sample-weighted percentages of households that faced barriers to adopting new practices to help manage the impacts of weather-related problems during the 12 months preceding the survey by barrier using the binary yes/no variables created during data processing for each **v2a116** response option (i.e., **v2a116a**, **v2a116b**, **v2a116c**, **v2a116d**, **v2a116e**, **v2a116f**, **v2a116g**, **v2a116h**, and **v2a116x**). For each barrier, populate the table with the percentage of all households with a value of 1 (yes) for the corresponding barrier variable.

```
svy: tab v2a116a, perc format(%6.1f)
svy: tab v2a116b, perc format(%6.1f)
svy: tab v2a116c, perc format(%6.1f)
svy: tab v2a116d, perc format(%6.1f)
svy: tab v2a116e, perc format(%6.1f)
svy: tab v2a116f, perc format(%6.1f)
svy: tab v2a116g, perc format(%6.1f)
svy: tab v2a116h, perc format(%6.1f)
svy: tab v2a116x, perc format(%6.1f)
```

**Step 7.4.** In the "Number of households that faced barriers (n)" row, populate the unweighted number of households included in the calculations.

#### **18.2.6 Table 5.5.6: Percent of Households in the ZOI that Participated in a Group or Committee Working to Strengthen the Community's Ability to Manage Weather-related Problems During the 12 Months Preceding the Survey**

**Step 8.** Calculate the sample-weighted estimates required for Table 5.5.6.

**Step 8.1.** Calculate the sample-weighted percentage of households with a household member who participated in a development group or committee working to strengthen the community's ability to manage the impacts of weather-related problems in the 12 months preceding the survey using **v2a117**. Populate the table with the percentage of households with a value of 1 (yes) for **v2a117**.

```
svy: tab v2a117, perc format(%6.1f)
```

**Step 8.2.** In the "Number of households (n)" row, populate the unweighted number of households included in the calculation.

**Step 8.3.** Calculate the sample-weighted percentage of households with a household member who participated in a development group or committee that made a formal plan to manage the impact of weather-related problems on their community among households that had a household member who participated in a development group or committee working to strengthen the community's ability to

manage the impacts of weather-related problems in the 12 months preceding the survey using **v2a118**. Populate the table with the percentage of households with a value of 1 (yes) for **v2a118**.

```
svy: tab v2a118, perc format(%6.1f)
```

**Step 8.4.** Calculate the sample-weighted percentage of households with a household member who participated in a development group or committee that took any action to strengthen their community's ability to manage the impact of weather-related problems among households that had a household member who participated in a development group or committee working to strengthen the community's ability to manage the impacts of weather-related problems in the 12 months preceding the survey using **v2a119**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a119**.

```
svy: tab v2a119, perc format(%6.1f)
```

**Step 8.5.** In the "Number of households with at least one member participating in a group (n)" row, populate the unweighted number of households included in the Step 8.3 and Step 8.4 calculations. The unweighted number of households should be the same for the two calculations; if the numbers differ, add a table note that provides information about the missing observations.

**Step 8.6.** Calculate the sample-weighted percent distribution of household members who were group or committee participants by sex and age using the **v2a117a** variables. To do so:

**Step 8.6.1.** Keep only the required variables.

```
Keep variables      hhea hhnum v2a117a_1-v2a117a_6
```

**Step 8.6.2.** Reshape the data from wide to long format.

```
Reshape             wide to long
variable stub       v2a117a_
index variable      hhea hhnum
```

**Step 8.6.3.** Drop any observations for which the value of **v2a117a\_** is missing.

```
Drop observation    if v2a117a_ =missing
```

**Step 8.6.4.** Rename **v2a117a\_** to be **m1\_line** in preparation for the merge that will take place in the next step.

```
Rename              v2a117a_ to be m1_line
```

**Step 8.6.5.** Add the age (**age15\_29y**) and sex (**sex**) variables for the household members who were group or committee participants from the persons-level analytic data file so that their characteristics can be tabulated.

```
Merge              1 to 1
key variables      hhea hhnum m1_line
data file           "FTF ZOI Survey [COUNTRY] [YEAR] persons data
                    analytic"
```

```
keep variables age15_29y sex
```

**Step 8.6.6.** Drop any observations that may have been merged if they are in only the persons data file and not in the working data file.

```
Drop observation if observation is only in persons data file
```

**Step 8.6.7.** Apply the complex survey design with the household sample weight (**wgt\_hh**), where **hhea** is the primary sampling unit or cluster and **strata** is the strata variable, calculate the sample-weighted percent distribution of household members who were group or committee participants by sex and age, and populate the table with the results.

```
svyset hhea [pw=wgt_hh], strata(strata)
svy: tab sex, perc format(%6.1f)
svy: tab age15_29y, perc format(%6.1f)
```

**Step 8.7.** In the “Number of group participants (n)” row, populate the unweighted number of participants included in the calculations.

### **18.2.7 Table 5.5.7: Percent of Households in the ZOI that Obtained Insurance, Type of Insurance and Whether Support was Received, and Whether Insurance for Crops or Farm Animals Was Obtained to Protect Against Weather-related Problems During the 12 Months Preceding the Survey**

**Step 9.** Calculate the sample-weighted values required for Table 5.5.7.

**Step 9.1.** Calculate the sample-weighted percentage of households that obtained health, life, crop, livestock, or other farm animal insurance during the 12 months preceding the survey using **v2a120**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a120**.

```
svy: tab v2a120, perc format(%6.1f)
```

**Step 9.2.** In the “Number of households (n)” row, populate the unweighted number of households included in the calculation.

**Step 9.3.** Calculate the sample-weighted percentages of households that obtained insurance during the 12 months preceding the survey by type of insurance using the binary yes/no variables created during data processing for each **v2a122** response option (e.g., **v2a122a**, **v2a122b**, **v2a122c**, and **v2a122d**). For each type of insurance, populate the table with the percentages of all households with a value of 1 (yes) for the corresponding type of insurance variable.

```
svy: tab v2a122a, perc format(%6.1f)
svy: tab v2a122b, perc format(%6.1f)
svy: tab v2a122c, perc format(%6.1f)
svy: tab v2a122d, perc format(%6.1f)
```

**Step 9.4.** Calculate the sample-weighted percentage of households that obtained health, life, crop, livestock, or other farm animal insurance in the 12 months preceding the survey that received support

to do so using **v2a121**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a121**.

```
svy: tab v2a121, perc format(%6.1f)
```

**Step 9.5.** In the “Number of households that obtained insurance (n)” row, populate the unweighted number of households included in the Step 9.3 and Step 9.4 calculations. The unweighted number of households should be the same for the two calculations; if the numbers differ, add a table note that provides information about the missing observations.

**Step 9.6.** Calculate the sample-weighted percentage of households that obtained crop or livestock or other farm animal of insurance in the 12 months preceding the survey that obtained the insurance to help manage the impacts of weather-related problems on their crop or farm animal production using **v2a124**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a124**.

```
svy: tab v2a124, perc format(%6.1f)
```

**Step 9.7.** In the “Number of households (n)” row, populate the unweighted number of households included in the calculation.

**Step 9.8.** Calculate the sample-weighted percent distribution of household members who received support to obtain health, life, crop, livestock, or other farm animal insurance during the 12 months preceding the survey by sex and age using the **v2a121a** variables. To do so:

**Step 9.8.1.** Keep only the required variables.

```
Keep variables      hhea hhnum v2a121a_1-2a121a_6
```

**Step 9.8.2.** Reshape the data from wide to long format.

```
Reshape            wide to long
variable stub      v2a121a_
index variable     hhea hhnum
```

**Step 9.8.3.** Drop any observations for which the value of **v2a121a\_** is missing.

```
Drop observation    if v2a121a_ =missing
```

**Step 9.8.4.** Rename **v2a121a\_** to be **m1\_line** in preparation for the merge that will take place in the next step.

```
Rename              v2a121a_ to be m1_line
```

**Step 9.8.5.** Add the age (**age15\_29y**) and sex (**sex**) variables for the household members who received support to obtain health, life, crop, livestock, or other farm animal insurance from the persons-level analytic data file so that their characteristics can be tabulated.

```
Merge              1 to 1
key variables       hhea hhnum m1_line
```

```
data file      "FTF ZOI Survey [COUNTRY] [YEAR] persons data
                analytic"
keep variables age15_29y sex
```

**Step 9.8.6.** Drop any observations that may have been merged if they are in only the persons data file and not in the working data file.

```
Drop observation if observation is only in persons data file
```

**Step 9.8.7.** Apply the complex survey design with the household sample weight (`wgt_hh`), where `hhea` is the primary sampling unit or cluster and `strata` is the strata variable, calculate the sample-weighted percent distribution of household members who received support to obtain health, life, crop, livestock, or other farm animal insurance during the 12 months preceding the survey by sex and age, and populate the table with the results.

```
svyset hhea [pw=wgt_hh], strata(strata)
svy: tab sex, perc format(%6.1f)
svy: tab age15_29y, perc format(%6.1f)
```

**Step 9.9.** In the “Number of individuals who received support to obtain insurance (n)” row, populate the unweighted number of individuals included in the calculations.

#### 18.2.8 Table 5.5.8: Percent of Households in the ZOI that Obtained Payments for Weather-related Problems that Affected Their Crop or Farm Animal Production During the 12 Months Preceding the Survey

**Step 10.** Calculate the sample-weighted estimates required for Table 5.5.8.

**Step 10.1.** Calculate the sample-weighted percentage of households that raised farm animals or cultivated any crops that obtained payments for weather-related problems that affected their crop or farm animal production during the 12 months preceding the survey using `v2a125`. Populate the table with the percentage of all households that raised farm animals or cultivated any crops (i.e., `v225=1` [yes], `v225a=1` [yes], or `v233=1` [yes]) with a value of 1 [yes] for `v2a125`.

```
svy, subpop(if v233==1 | v225==1 | v225a==1): tab v2a125, perc
format(%6.1f)
```

**Step 10.2.** In the “Number of households that grew crops or raised farm animals (n)” row, populate the unweighted number of households included in the calculation.

**Step 10.3.** Calculate the sample-weighted percentages of households that received a transfer or payment for weather-related problems that affected their crop or farm animal production during the 12 months preceding the survey by type of payment (i.e., `v2a126a`, `v2a126b`, `v2a126c`, `v2a126d`, and `v2a126x`). For each type of payment, populate the table with the percentage of all households with a value of 1 (yes) for the corresponding type of payment variable.

```
svy: tab v2a126a, perc format(%6.1f)
svy: tab v2a126b, perc format(%6.1f)
svy: tab v2a126c, perc format(%6.1f)
svy: tab v2a126d, perc format(%6.1f)
```



```
svy: tab v2a126x, perc format(%6.1f)
```

**Step 10.4.** Calculate the sample-weighted percentages of households that received a transfer or payment for weather-related problems that affected their crop or farm animal production during the 12 months preceding the survey by source (i.e., **v2a127a**, **v2a127b**, **v2a127c**, **v2a127d**, **v2a127e**, and **v2a127x**). For each source, populate the table with the percentage of with a value of 1 (yes) for the corresponding source.

```
svy: tab v2a127a, perc format(%6.1f)
svy: tab v2a127b, perc format(%6.1f)
svy: tab v2a127c, perc format(%6.1f)
svy: tab v2a127d, perc format(%6.1f)
svy: tab v2a127e, perc format(%6.1f)
svy: tab v2a127x, perc format(%6.1f)
```

**Step 10.5.** In the “Number of households that received cash payment (n)” row, populate the unweighted number of households included in the Step 10.3 and Step 10.4 calculations. The unweighted number of households should be the same for the two calculations; if the numbers differ, add a table note that provides information about the missing observations.

**Step 10.6.** Calculate the sample-weighted percent distribution of household members who received a payment to help them recover from weather-related problems that limited their crop or farm animal production by sex and age using the **v2a125a** variables.

**Step 10.6.1.** Keep only the required variables.

```
Keep variables      hhea hhnum v2a125a_1-2a125a_6
```

**Step 10.6.2.** Reshape the data from wide to long format.

```
Reshape            wide to long
variable stub      v2a125a_
index variable     hhea hhnum
```

**Step 10.6.3.** Drop any observations for which the value of **v2a125a\_** is missing.

```
Drop observation    if v2a125a_=missing
```

**Step 10.6.4.** Rename **v2a125a\_** to be **m1\_line** in preparation for the merge that will take place in the next step.

```
Rename              v2a125a_ to be m1_line
```

**Step 10.6.5.** Add the age (**age15\_29y**) and sex (**sex**) variables for the household members who received a payment to help them recover from weather-related problems that limited their crop or farm animal production from the persons-level analytic data file so that their characteristics can be tabulated.

```
Merge              1 to 1
key variables       hhea hhnum m1_line
```

```
data file      "FTF ZOI Survey [COUNTRY] [YEAR] persons data
               analytic"
keep variables age15_29y sex
```

**Step 10.6.6.** Drop any observations that may have been merged if they are in only the persons data file and not in the working data file.

```
Drop observation if observation is only in persons data file
```

**Step 10.6.7.** Apply the complex survey design with the household sample weight (`wgt_hh`), where `hhea` is the primary sampling unit or cluster and `strata` is the strata variable, calculate the sample-weighted percent distribution of household members who received a payment to help them recover from weather-related problems that limited their crop or farm animal production during the 12 months preceding the survey by sex and age, and populate the table with the results.

```
svyset hhea [pw=wgt_hh], strata(strata)
svy: tab sex, perc format(%6.1f)
svy: tab age15_29y, perc format(%6.1f)
```

**Step 10.7.** In the “Number of payment recipients (n)” row, populate the unweighted number of individuals included in the calculations.

### **18.2.9 Table 5.5.9: Percent of Households in the ZOI that Participated in a Group or Committee Working to Monitor or Influence Land Use During the 12 Months Preceding the Survey**

**Step 11.** Calculate the sample-weighted estimates required for Table 5.5.9.

**Step 11.1.** Calculate the sample-weighted percentage of households with a household member who participated in a group or committee working to monitor or influence land use in the 12 months preceding the survey using `v2a128`. Populate the table with the percentage of all households with a value of 1 (yes) for `v2a128`.

```
svy: tab v2a128, perc format(%6.1f)
```

**Step 11.2.** In the “Number of households (n)” row, populate the unweighted number of households included in the calculation.

**Step 11.3.** Calculate the sample-weighted percentage of households with a household member who participated in a group or committee working to monitor or influence land use that made a formal plan to change land use in their community among households with a household member who participated in a group or committee working to monitor or influence land use in the 12 months preceding the survey using `v2a129`. Populate the table with the percentage of all households with a value of 1 (yes) for `v2a129`.

```
svy: tab v2a129, perc format(%6.1f)
```

**Step 11.4.** Calculate the sample-weighted percentage of households with a household member who participated in a group or committee to monitor or influence land use that took action to change land use in their area among households with a household member who participated in a group or

committee working to monitor or influence land use in the 12 months preceding the survey using **v2a130**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a130**.

```
Svy: tab v2a130, perc format(%6.1f)
```

**Step 11.5.** In the “Number of households with at least one group participant (n)” row, populate the unweighted number of households included in the Step 11.3 and Step 11.4 calculations. The unweighted number of households should be the same for the two calculations; if the numbers differ, add a table note that provides information about the missing observations.

**Step 11.6.** Calculate the sample-weighted percent distribution of household members who participated in a group or committee working to monitor or influence land use during the 12 months preceding the survey using the **v2a128a** variables. To do so:

**Step 11.6.1.** Keep only the required variables.

```
Keep variables      hhea hhnum v2a128a_1-2a128a_6
```

**Step 11.6.2.** Reshape the data from wide to long format.

```
Reshape             wide to long
variable stub       v2a128a_
index variable      hhea hhnum
```

**Step 11.6.3.** Drop any observations for which the value of **v2a128a\_** is missing.

```
Drop observation    if v2a128a_ =missing
```

**Step 11.6.4.** Rename **v2a128a\_** to be **m1\_line** in preparation for the merge that will take place in the next step.

```
Rename              v2a128a_ to be m1_line
```

**Step 11.6.5.** Add the age (**age15\_29y**) and sex (**sex**) variables for the household members who participated in a group or committee working to monitor or influence land use during the 12 months preceding the survey from the persons-level analytic data file so that their characteristics can be tabulated.

```
Merge               1 to 1
key variables       hhea hhnum m1_line
data file            "FTF ZOI Survey [COUNTRY] [YEAR] persons data
                     analytic"
keep variables       age15_29y sex
```

**Step 11.6.6.** Drop any observations that may have been merged if they are in only the persons data file and not in the working data file.

```
Drop observation    if observation is only in persons data file
```

**Step 11.6.7.** Apply the complex survey design with the household sample weight (**wgt\_hh**), where **hhea** is the primary sampling unit or cluster and **strata** is the strata variable, calculate the sample-weighted percent distribution of household members who participated in a group or committee working to monitor or influence land use during the 12 months preceding the survey by sex and age, and populate the table with the results.

```
Svyset hhea [pw=wgt_hh] , strata(strata)

svy: tab sex, perc format(%6.1f)
svy: tab age15_29y, perc format(%6.1f)
```

**Step 11.7.** In the “Number of group participants (n)” row, populate the unweighted number of group participants included in the calculations.

**18.2.10 Table 5.5.10: Percent of Households in the ZOI that Participated in a Training or Received Support to Help Diversify Their Sources of Income During the 12 Months Preceding the Survey, Feed the Future Phase Two ZOI Midline Survey**

**Step 12.** Calculate the sample-weighted estimates required for Table 5.5.10.

**Step 12.1.** Calculate the sample-weighted percentage of households that accessed new sources of income during the 12 months preceding the survey using **v2a131**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a131**.

```
svy: tab v2a131, perc format(%6.1f)
```

**Step 12.2.** Calculate the sample-weighted percentages of households that accessed new sources of income during the 12 months preceding the survey by source using the binary yes/no variables created during data processing for each **v2a132** response option (i.e., **v2a132a**, **v2a132b**, **v2a132c**, **v2a132d**, **v2a132e**, **v2a132f**, **v2a132g**, **v2a132h**, **v2a132i**, **v2a132j**, **v2a132k**, **v2a132l**, **v2a132m**, and **v2a132x**). For each source, populate the table with the percentages of all households with a value of 1 (yes) for the corresponding source variable.

```
svy: tab v2a132a, perc format(%6.1f)
svy: tab v2a132b, perc format(%6.1f)
svy: tab v2a132c, perc format(%6.1f)
svy: tab v2a132d, perc format(%6.1f)
svy: tab v2a132e, perc format(%6.1f)
svy: tab v2a132f, perc format(%6.1f)
svy: tab v2a132g, perc format(%6.1f)
svy: tab v2a132h, perc format(%6.1f)
svy: tab v2a132i, perc format(%6.1f)
svy: tab v2a132j, perc format(%6.1f)
svy: tab v2a132k, perc format(%6.1f)
svy: tab v2a132l, perc format(%6.1f)
svy: tab v2a132m, perc format(%6.1f)
svy: tab v2a132x, perc format(%6.1f)
```

**Step 12.3.** Calculate the sample-weighted percentage of households that participated in a training or received support to help diversify their sources of income during the 12 months preceding the survey using **v2a133**. Populate the table with the percentage of all households with a value of 1 (yes) for **v2a133**.

```
svy: tab v2a133, perc format(%6.1f)
```

**Step 12.4.** In the “Number of households with at least one group participant (n)” row, populate the unweighted number of households included in Steps 12.1-12.3. The unweighted number of households should be the same for the three calculations; if the numbers differ, add a table note that provides information about the missing observations.

**Step 12.5.** Calculate the sample-weighted percent distribution of household members who participated in training or received support to help diversify their sources of income using the **v2a133a** variables. To do so:

**Step 12.5.1.** Keep only the required variables.

```
Keep variables      hhea hhnum v2a133a_1-v2a133a_6
```

**Step 12.5.2.** Reshape the data from wide to long format.

```
Reshape             wide to long
variable stub       v2a133a_
index variable      hhea hhnum
```

**Step 12.5.3.** Drop any observations for which the value of **v2a128a\_** is missing.

```
Drop observation    if v2a133a_ =missing
```

**Step 12.5.4.** Rename **v2a133a\_** to be **m1\_line** in preparation for the merge that will take place in the next step.

```
Rename              v2a133a_ to be m1_line
```

**Step 12.5.5.** Add the age (**age15\_29y**) and sex (**sex**) variables for the household members who participated in a training or received support to help diversify their sources of income during the 12 months preceding the survey from the persons-level analytic data file so that their characteristics can be tabulated.

```
Merge               1 to 1
key variables       hhea hhnum m1_line
data file            "FTF ZOI Survey [COUNTRY] [YEAR] persons data
                     analytic"
keep variables       age15_29y sex
```

**Step 12.5.6.** Drop any observations that may have been merged if they are in only the persons data file and not in the working data file.

```
Drop observation    if observation is only in persons data file
```

**Step 12.5.7.** Apply the complex survey design with the household sample weight (**wgt\_hh**), where **hhea** is the primary sampling unit or cluster and **strata** is the strata variable, calculate the sample-weighted percent distribution of household members who participated in a training or received support to help diversify their sources of income during the 12 months preceding the survey by sex and age, and populate the table with the results.

```
svyset hhea [pw=wgt_hh], strata(strata)
svy: tab sex, perc format(%6.1f)
svy: tab age15_29y, perc format(%6.1f)
```

**Step 12.6.** In the “Number of individuals who received income diversification training or support (n)” row, populate the unweighted number of individuals included in the calculations.

### 18.3 References

United States Agency for International Development. (2022.) *USAID climate strategy 2022-2030*. Available at: <https://www.usaid.gov/sites/default/files/2022-11/USAID-Climate-Strategy-2022-2030.pdf>

## APPENDIX A. ABBREVIATED WOMEN'S EMPOWERMENT IN AGRICULTURE INDEX SUMMARY INFORMATION

**Table A1: Summary of Abbreviated Women's Empowerment in Agriculture Index Domains, Indicators, Survey Questions, Variables, Definitions, and Weights**

Domain	Indicator name	Survey questions	ZOI Survey questions	ZOI Survey variables	Adequacy criteria	Inadequacy criteria	Weight
Decision-making over production	Indicator 1.1: Input in productive decisions	"When decisions are made regarding food crop farming, cash crop farming, livestock raising, and fishing or fishpond culture, who is it that normally takes the decision?"	V6202 (1, 2, 3, 6)	v6202_1- v6202_3, v6202_6	For at least one activity: decides alone; OR participates and has input into some; or most or all decisions regarding the activity; OR someone else decides but feels could decide to a medium or high extent if wanted to	Participates but does not have input into some; or most or all decisions regarding the activity; OR does not make the decision NOR feels he or she could to a medium or high extent (93 'no decision made' coded as missing)	1/5
		"How much input did you have in making decisions about food crop farming, cash crop farming, livestock raising, and fishing or fishpond culture?"	V6203 (1, 2, 3, 6)	v6202_1- v6202_3, v6202_6			
		"To what extent do you feel you can make your own decisions regarding these aspects of household life if you want(ed) to: food crop farming, cash crop farming, livestock raising, and fishing or fishpond culture if you wanted to?"	V6204 (1, 2, 3, 6)	v6202_1- v6202_3, v6202_6			

Domain	Indicator name	Survey questions	ZOI Survey questions	ZOI Survey variables	Adequacy criteria	Inadequacy criteria	Weight
Access to resources	Indicator 2.1: Ownership of assets	<p>“Does anyone in your household currently have any [ITEM]?: agricultural land, large livestock, small livestock, chickens/ducks/turkeys/ pigeons, fishpond or fishing equipment, hand tools, non-mechanized farm equipment, mechanized farm equipment, non-farm business equipment, house, large consumer durable goods, small consumer durable goods, cell phone, other land or structures, and means of transportation”</p> <p>“Do you own any of the item either by yourself or jointly with someone else?”</p>	<p>V6301 (1-15)</p> <p>V6303 (1-15)</p>	<p>v6301_01–v6301_15</p> <p>v6303_01–v6303_15</p>	Owns—alone or jointly—at least one large asset or two types of small assets (small assets are chickens/ducks/turkeys/pigeons, hand tools, non-mechanized farm equipment, and small consumer durable goods)	Does not own any assets; OR owns only one type of small asset alone or jointly	2/15
	Indicator 2.2: Access to and decisions over credit	<p>“Has anyone in your household taken any loans or borrowed cash/in-kind from [SOURCE] in the past 12 months?: NGO, informal lender, formal lender, friends or relatives, group-based microfinance or lending (savings/credit group), informal credit/savings groups such as merry-go-rounds, tontines, funeral societies, etc.”</p> <p>“Who made the decision to borrow from [SOURCE]?”</p> <p>“Who makes the decision about what to do with the money/item borrowed from [SOURCE]?”</p>	<p>V6308 (1-6)</p> <p>V6309 (1-6)</p> <p>V6310 (1-6)</p>	<p>v6308_1–v6303_6</p> <p>v6309_1–v6309_6</p> <p>v6310_1–v6310_6</p>	Can alone or jointly make at least one decision regarding at least one source of credit	Household has no credit; OR household has credit but respondent did not participate in any decision about it	1/15
Control over income	Indicator 3: Control of use of income	“How much input did you have in decisions on the use of income generated from food crop farming, cash crop farming, livestock raising, non-farm economic activities, wage and salary employment, and fishing or fishpond culture?”	V6205 (1-6)	v6205_1–v6206_6,	Has input into some; or most or all decisions on use of income for at least one productive/	Participates in activity but has no input in decisions about income, OR feels she or he has no	1/5



Domain	Indicator name	Survey questions	ZOI Survey questions	ZOI Survey variables	Adequacy criteria	Inadequacy criteria	Weight
		“To what extent do you feel you can make your own personal decisions regarding these aspects of household life if you want(ed) to?: non-farm activities, own wage and salary employment, major household expenditures, and minor household expenditures”	V6204 (4-5) V6204 (7-8)	v6204_4 , v6204_5 , v6204_7 , v6204_8	economic activity; OR feels can make decisions to medium or high extent if respondent wanted for at least one income or expenditure decision—excludes minor household expenditures	or very little input into the decision regarding income from non-farm activities, wage and salary employment, or decisions regarding major household expenditures even if she or he wanted to	
Group membership and leadership	Indicator 4.1: Membership in economic or social group	“Are you an active member of an agricultural/livestock/fisheries producers’ group, waters users’ group, forest users’ group, credit/microfinance group, mutual help/insurance group, trade and business association, trade and business association, civic groups, local government, religious group, other women’s/men’s group, or any other formal or informal organization?”	V6405 (I-II)	v6405_01– v6405_11	Is an active member of at least one group	Is not an active member of at least one group	1/5
Time allocation	Indicator 5.1: Workload	The survey collected information on respondents’ time allocation for a 24-hour period. Information was collected for primary activities and reported in 15-minute intervals.	V6601	v6601p_15_[hour] , v6601p_30_[hour] , v6601p_45_[hour] , v6601p_60_[hour] where [hour] is a value 1–24	Works less than or equal to 10.5 hours in 24-hour period	Works more than 10.5 hours in 24-hour period	1/5

Source: Adapted from Malapit et al., 2015.

## APPENDIX B. 5DE CONFIDENCE INTERVAL AND DESIGN EFFECT CALCULATIONS

This appendix provides guidelines for calculating the 95 percent confidence interval (CI) and design effect (DEFF) for the five domains of empowerment (5DE). Due to the complicated nature of the non-linear 5DE, non-standard methods must be used to compute the CI and DEFF. Section B1 outlines how the 95 percent CI is estimated, Section B2 outlines how the DEFF is estimated, and Section B3 outlines how differences and tests of significance between two estimates are calculated. The step-by-step Stata instructions are included in Section 12\3.2.3.

### B1. Confidence interval estimation

To compute the CI for the 5DE, standard error must be estimated using a replication method such as the jackknife repeated replication (JRR) method. In the JRR method, the standard error is the square root of the replicate variance among the indicator estimates calculated based on several sample replicates of the full sample (Heeringa, et al 2010; Wolter, 2007). In each replicate, all but one sample cluster are considered in estimation of the indicator. Where  $C$  is the total number of sample clusters,  $C$  sample replicates should be created, and the variance of the estimated indicator  $\hat{\theta}$  can be calculated as follows:

$$var(\hat{\theta}) = \frac{1}{C(C-1)} \sum_{c=1}^C (\hat{\theta}_c - \hat{\theta})^2$$

in which

$$\hat{\theta}_c = C\hat{\theta} - (C-1)\hat{\theta}_{(c)}$$

where  $\hat{\theta}$  is the estimate computed from the full sample of  $C$  clusters and  $\hat{\theta}_{(c)}$  is the estimate computed from the reduced sample of  $(C-1)$  clusters ( $c^{\text{th}}$  cluster excluded)

To estimate the CI for the 5DE, the algorithm for the JRR method follows a sequence of six steps:

1. The indicator of interest, the 5DE, is estimated based on the full sample as  $\hat{\theta}$ .
2. A total of  $C$  sample replicates is defined, where replicate  $c$  is a reduced sample of  $(C-1)$  clusters where  $c^{\text{th}}$  cluster is excluded.
3. For each sample replicate  $c$ , a replicate weight is calculated by multiplying the full sample survey weight by a strata-level adjustment factor  $\frac{W_h}{W_{h(c)}}$ , where  $W_h$  is the summation of the survey weight for all units within stratum  $h$  and  $W_{h(c)}$  is the same summation but after cluster  $c$  is excluded.

4. From each sample replicate  $c$ , the indicator of interest, the 5DE, is calculated as  $\hat{\theta}_{(c)}$  using the replicate weights.
5. The JRR replicated variance estimation formula is used to compute the standard error based on the estimates from the full sample and  $C$  replicates, computed in Steps 1 and 4.
6. 95 percent CIs are estimated based on the estimated indicators and standard error as  $CI(\hat{\theta}) = \hat{\theta} \pm 1.96 \sqrt{var(\hat{\theta})}$ .

## B2. Design effect estimation

Due to the reasons mentioned previously, there is not a constructed method to calculate the DEFF for the 5DE. Because sets of linear variables are used in the computation of both indicators, the average of DEFFs for these variables can be used as approximates for DEFFs as follows:

Calculated as the average of DEFFs for the two variables that are involved in the calculation of the 5DE: **ch\_20p** and **a\_20p**.

## B3. Differences in indicator estimates over time and test of significance

The 5DE estimates can be compared over time, such as, for women at the time of the Baseline Survey and at the time of the Midline Survey. Hypothesis testing can be used to determine whether the difference between the estimated indicators for women at baseline compared to at midline is statistically significant.

Assuming that the 5DE follows the standard normal distribution, let  $\hat{\theta}_1$  denote the estimate computed for baseline, and  $\hat{\theta}_2$  denote the estimate computed for midline. To test the null hypothesis,  $H_0: \hat{\theta}_1 = \hat{\theta}_2$ , versus the alternative hypothesis,  $H_a: \hat{\theta}_1 \neq \hat{\theta}_2$ , the test statistic (z-score) can be calculated as follows:

$$Z = \frac{|\hat{\theta}_1 - \hat{\theta}_2|}{\sqrt{SE1^2 + SE2^2}}$$

where SE1 is the standard error for  $\hat{\theta}_1$

SE2 is the standard error for  $\hat{\theta}_2$

Using the standard normal tables, the associated two-tailed p-value of the z-score can be found as  $2 * P(Z > Z_\alpha)$  where  $Z_\alpha$  is the critical Z value for the chosen  $\alpha$  level significance (e.g.,  $Z_\alpha = 1.96$  for the 95 percent confidence level and  $Z_\alpha = 1.645$  for the 90 percent confidence level). The p-value is then assessed in comparison with the significance level (0.05 for 95 percent confidence level, or 0.10 for 90 percent confidence level), where the null hypothesis is accepted if the p-value is larger than the significance level, and rejected otherwise. If the null hypothesis is accepted, any observed difference is due to sampling or experimental error, but if the null hypothesis is rejected, it is likely that the estimated indicator for the two points in time being compared are truly different.

## B4. References

Heeringa, S., West, B., & Berglund, P. (2010). *Applied survey data analysis*. New York: Chapman and Hall/CRC. <https://doi.org/10.1201/9781420080674>

Wolter, K. (2007). The jackknife method. In Wolter, K. (Ed.), *Introduction to variance estimation* (2nd ed., pp. 153-191). New York, NY: Springer. [https://doi.org/10.1007/978-0-387-35099-8\\_4](https://doi.org/10.1007/978-0-387-35099-8_4)