Feed the Future

Survey Implementation

Document

Guide to

Feed the Future Statistics

Zone of Influence Survey

August 2019

Last revised October 2020

This document was produced for review by the United States Agency for International Development. It was prepared by ME&A and its sub-contractor ICF for Feed the Future Global Program Evaluation for Effectiveness and Learning Task Order (PEEL-TO), Contract Number GS-23F-8144H/AID-OAA-M-12-00006.

**Recommended Citation**

Kirsten Zalisk, Genevieve Dupuis, Madeleine Gauthier, Jasbir Kaur, Nizam Khan, Anne Swindale, and Kiersten B. Johnson. 2019.

*Feed the Future Zone of Influence Surveys: Guide to Feed the Future Statistics.*

Washington, DC: Bureau for Food Security, USAID

**Contact Information**

Feed the Future

1300 Pennsylvania Ave, NW

Washington, DC 20004

[www.feedthefuture,gov](about:blank)

# ACKNOWLEDGEMENTS

The authors are grateful to Andrew Bisson, Trevor Croft, Cristina de la Torre, Navin Hada, Gary Jahn, Tyrell Kahan, Ricardo Lardizabal, Shivaun Leonard, John Peters, Gheda Temsah and Hailu Wordofa for their essential contributions to this work. This document benefitted greatly from their investments of time, expertise, and careful attention to detail.

# TABLE OF CONTENTS

[ACKNOWLEDGEMENTS i](#_heading=h.1fob9te)

[TABLE OF CONTENTS ii](#_heading=h.3znysh7)

[List of figures x](#_heading=h.2et92p0)

[List of tables x](#_heading=h.tyjcwt)

[Abbreviations xi](#_heading=h.3dy6vkm)

[Part I. 1](#_heading=h.4d34og8)

[INTRODUCTION TO ZOI SURVEY DATA AND VARIABLES 1](#_heading=h.2s8eyo1)

[1.](#_heading=h.17dp8vu) Introduction 2

[1.1.](#_heading=h.26in1rg) Feed the Future population-based ZOI-level performance indicators 3

[1.2.](#_heading=h.1ksv4uv) Feed the Future ZOI Survey 2018-2019 core questionnaire 5

[1.3.](#_heading=h.44sinio) Data management and results reporting 5

[2.](#_heading=h.3j2qqm3) Organization of ZOI Survey data 7

[2.1](#_heading=h.1y810tw) Structure of the Feed the Future ZOI Survey data files 7

[2.1.1](#_heading=h.4i7ojhp) Hierarchical files 7

[2.1.2](#_heading=h.2xcytpi) Flat files 8

[2.1.3](#_heading=h.3whwml4) Formats and naming convention of datasets exported from CSPro 9

[2.1.4](#_heading=h.3as4poj) Matching relationships 9

[2.2](#_heading=h.1pxezwc) Variable name conventions 10

[2.2.1](#_heading=h.49x2ik5) General variable naming 10

[2.2.2](#_heading=h.2p2csry) Select-all questions 13

[2.2.3](#_heading=h.147n2zr) Special response options and variables that are not applicable 13

[3.](#_heading=h.23ckvvd) Data analysis information 15

[3.1](#_heading=h.ihv636) Sampling weights 15

[3.2](#_heading=h.32hioqz) Descriptive statistics and statistical tests of difference 16

[3.2.1](#_heading=h.1hmsyys) Analyzing indicators of proportions 17

[3.2.2](#_heading=h.vx1227) Analyzing indicators of means 28

[3.2.3](#_heading=h.3fwokq0) Analyzing indicators of proportions or means if the data are not available 38

[3.3](#_heading=h.1v1yuxt) Data analysis folder structure and file naming conventions 41

[3.3.1](#_heading=h.4f1mdlm) Folder structure 41

[3.3.2](#_heading=h.2u6wntf) File naming conventions 42

[4.](#_heading=h.19c6y18) Key analytic variables 43

[4.1.](#_heading=h.28h4qwu) Person-level variables 43

[4.1.1](#_heading=h.nmf14n) De jure household member 43

[4.1.2](#_heading=h.37m2jsg) De facto household member 43

[4.1.3](#_heading=h.1mrcu09) Sex 44

[4.1.4](#_heading=h.46r0co2) Age in years 44

[4.1.5](#_heading=h.2lwamvv) Age by 5-year categories 45

[4.1.6](#_heading=h.111kx3o) Adults, total and by sex 46

[4.1.7](#_heading=h.3l18frh) Youth, total and by sex 46

[4.1.8](#_heading=h.206ipza) Women of reproductive age 47

[4.1.9](#_heading=h.4k668n3) Children under 2 years of age, total and by sex 47

[4.1.10](#_heading=h.2zbgiuw) Children under 5 years of age, total and by sex 48

[4.1.11](#_heading=h.1egqt2p) Children 5 years of age or older, total and by sex 49

[4.1.12](#_heading=h.3ygebqi) Child age in days 50

[4.1.13](#_heading=h.2dlolyb) Child age in months 52

[4.1.14](#_heading=h.sqyw64) Child age by age group 52

[4.1.15](#_heading=h.3cqmetx) Household member completed primary education 54

[4.1.16](#_heading=h.1rvwp1q) Household member was attending school at the time of the survey 54

[4.1.17](#_heading=h.4bvk7pj) Household member’s highest level of educational attainment 55

[4.1.18](#_heading=h.2r0uhxc) Primary adult decisionmaker, by sex 56

[4.1.19](#_heading=h.1664s55) Primary caregiver of child under 5 years of age 57

[4.1.20](#_heading=h.3q5sasy) Targeted VCC producer, any VCC 58

[4.1.21](#_heading=h.25b2l0r) Targeted VCC producer, specific VCCs 59

[4.1.22](#_heading=h.kgcv8k) Maize farming decisionmakers 60

[4.2](#_heading=h.34g0dwd) Household-level variables 61

[4.2.1](#_heading=h.1jlao46) Number of de jure adults in household, total and by sex 61

[4.2.2](#_heading=h.43ky6rz) Number of de facto adults in household, total and by sex 62

[4.2.3](#_heading=h.2iq8gzs) Number of de jure women of reproductive age in household 63

[4.2.4](#_heading=h.xvir7l) Number of de jure children under 2 years of age in household 63

[4.2.5](#_heading=h.3hv69ve) Number of de jure children under 5 years of age in household 64

[4.2.6](#_heading=h.1x0gk37) Number of de jure children 5-17 years of age in household 64

[4.2.7](#_heading=h.4h042r0) Number of de jure youth in household 65

[4.2.8](#_heading=h.2w5ecyt) Number of de jure producers of targeted VCCs in household 65

[4.2.9](#_heading=h.1baon6m) Household size—de jure 66

[4.2.10](#_heading=h.3vac5uf) Household size—de facto 67

[4.2.11](#_heading=h.2afmg28) Household size category—de jure 67

[4.2.12](#_heading=h.pkwqa1) Household size category—de facto 68

[5.](#_heading=h.39kk8xu) Disaggregate variables 69

[5.1](#_heading=h.1opuj5n) Household 69

[5.1.1](#_heading=h.48pi1tg) Gendered household type—de jure household members 69

[5.1.2](#_heading=h.2nusc19) Gendered household type—de facto household members 69

[5.1.3](#_heading=h.1302m92) Household education—de jure household members 70

[5.1.4](#_heading=h.3mzq4wv) Household education—de facto household members 71

[5.1.5](#_heading=h.2250f4o) Wealth quintile 72

[5.1.6](#_heading=h.haapch) Poverty status 72

[5.1.7](#_heading=h.319y80a) Food insecurity 72

[5.1.8](#_heading=h.1gf8i83) Shock exposure index 72

[5.1.9](#_heading=h.40ew0vw) Farm size 73

[5.1.10](#_heading=h.2fk6b3p) Agricultural land ownership, by category 74

[5.1.11](#_heading=h.upglbi) Livestock ownership 75

[5.1.12](#_heading=h.3ep43zb) VCC production 77

[5.2](#_heading=h.1tuee74) Primary adult decisionmakers 78

[5.2.1](#_heading=h.4du1wux) De jure primary adult decision makers, by sex 78

[5.2.2](#_heading=h.2szc72q) Age category, by sex (de jure only) 78

[5.2.3](#_heading=h.184mhaj) Youth, by sex (de jure only) 79

[5.2.4](#_heading=h.3s49zyc) Education, by sex (de jure only) 80

[5.3](#_heading=h.279ka65) Women of reproductive age 81

[5.3.1](#_heading=h.meukdy) Age category—5-year age categories 82

[5.3.2](#_heading=h.36ei31r) Age category—15-18 years and 19-49 years 82

[5.3.3](#_heading=h.1ljsd9k) Education 83

[5.3.4](#_heading=h.45jfvxd) Pregnancy status 83

[5.4](#_heading=h.2koq656) Children under 5 84

[5.4.1](#_heading=h.zu0gcz) Sex 84

[5.4.2](#_heading=h.3jtnz0s) Age category—12 months 84

[5.4.3](#_heading=h.1yyy98l) Age category—6 months 85

[5.4.4](#_heading=h.4iylrwe) Age category—0-23 months and 24-59 months 85

[5.4.5](#_heading=h.2y3w247) Caregiver’s education 86

[5.5](#_heading=h.1d96cc0) Targeted VCC producers 87

[5.5.1](#_heading=h.3x8tuzt) Sex 87

[5.5.2](#_heading=h.2ce457m) Age category 87

[5.5.3](#_heading=h.rjefff) Youth 87

[5.5.4](#_heading=h.3bj1y38) Education 88

[5.5.5](#_heading=h.1qoc8b1) Commodity 88

[Part II. 91](#_heading=h.4anzqyu)

[INDICATORS 91](#_heading=h.2pta16n)

[6.](#_heading=h.14ykbeg) Demographic indicators 92

[6.1](#_heading=h.j8sehv) Household demographic characteristics 92

[6.1.1](#_heading=h.338fx5o) Mean household size 92

[6.1.2](#_heading=h.1idq7dh) Mean number of adult male household members 93

[6.1.3](#_heading=h.42ddq1a) Mean number of adult female household members 93

[6.1.4](#_heading=h.2hio093) Mean number of women of reproductive age 94

[6.1.5](#_heading=h.wnyagw) Mean number of children under 2 years 94

[6.1.6](#_heading=h.3gnlt4p) Mean number of children under 5 years 95

[6.1.7](#_heading=h.1vsw3ci) Mean number of children 5-17 years of age 96

[6.1.8](#_heading=h.4fsjm0b) Mean number of youth 15-29 years of age 96

[6.1.9](#_heading=h.2uxtw84) Mean number of producers of any targeted commodity 97

[6.1.10](#_heading=h.1a346fx) Percent of adults who are male 98

[6.1.11](#_heading=h.3u2rp3q) Percent of adults who are female 98

[6.1.12](#_heading=h.2981zbj) Percent distribution of households by household size category—de jure 99

[6.1.13](#_heading=h.odc9jc) Percent distribution of households by highest level of educational attainment category 100

[6.2](#_heading=h.38czs75) Characteristics of primary adult male and female decisionmakers 100

[6.2.1](#_heading=h.1nia2ey) Percent distribution of primary adult male decisionmakers by age group 101

[6.2.2](#_heading=h.47hxl2r) Percent distribution of primary adult female decisionmakers by age group 101

[6.2.3](#_heading=h.2mn7vak) Percent distribution of primary adult male decisionmakers by their current marital status 102

[6.2.4](#_heading=h.11si5id) Percent distribution of primary adult female decisionmakers by their current marital status 103

[6.2.5](#_heading=h.3ls5o66) Percent distribution of primary adult male decisionmakers by their highest completed level of education 104

[6.2.6](#_heading=h.20xfydz) Percent distribution of primary adult female decisionmakers by their highest completed level of education 105

[6.2.7](#_heading=h.4kx3h1s) Percent of primary adult male decisionmakers by participation in economic activity 105

[6.2.8](#_heading=h.302dr9l) Percentage of primary adult female decisionmakers by participation in economic activity 107

[6.2.9](#_heading=h.1f7o1he) Percent distribution of primary adult male decisionmakers by type of economic activity 108

[6.2.10](#_heading=h.3z7bk57) Percent distribution of primary adult female decisionmakers by type of economic activity 109

[6.3](#_heading=h.2eclud0) Education of household members 111

[6.3.1](#_heading=h.3dhjn8m) Percent distribution of de facto household members who are currently attending school, by age category 111

[6.3.2](#_heading=h.1smtxgf) Percent distribution of de facto household members 10 years of age or older who completed a primary-level education, by age group 112

[7.](#_heading=h.4cmhg48) Water, sanitation, and hygiene indicators 114

[7.1](#_heading=h.2rrrqc1) Percent of households using an improved water source 114

[7.2](#_heading=h.16x20ju) Percent of households using a correct water treatment practice or technology 116

[7.3](#_heading=h.3qwpj7n) Percent of households with soap and water at a handwashing station commonly used by family members (Feed the Future phase two indicator) 117

[7.4](#_heading=h.261ztfg) Percent of households with access to a basic sanitation service (Feed the Future phase two indicator) 118

[7.5](#_heading=h.l7a3n9) Percent of households using an improved but shared sanitation facility 120

[7.6](#_heading=h.356xmb2) Percent of households using an unimproved sanitation facility 121

[7.7](#_heading=h.1kc7wiv) Percent of households practicing open defecation 123

[8.](#_heading=h.44bvf6o) Dwelling and household characteristic indicators 125

[8.1](#_heading=h.2jh5peh) Percent of households using solid fuels for cooking 125

[8.2](#_heading=h.ymfzma) Mean number of de jure household members per sleeping room 126

[8.3](#_heading=h.3im3ia3) Percent distribution of households by dwelling roof materials 127

[8.4](#_heading=h.1xrdshw) Percent distribution of households by dwelling exterior wall materials 128

[8.5](#_heading=h.4hr1b5p) Percent distribution of households by dwelling floor materials 129

[References 130](#_heading=h.2wwbldi)

[9.](#_heading=h.1c1lvlb) The consumption aggregate and the poverty indicators 131

[9.1](#_heading=h.3w19e94) Guidelines to construct the poverty indicators 131

[9.1.1](#_heading=h.2b6jogx) Creating the consumption aggregate 131

[9.1.2](#_heading=h.1pgrrkc) Poverty indicators 141

[9.2](#_heading=h.3nqndbk) Step-by-step procedure to calculate the poverty indicators 148

[References 148](#_heading=h.22vxnjd)

[10.](#_heading=h.i17xr6) The comparative wealth index 149

[10.1](#_heading=h.320vgez) Guidelines to construct Feed the Future’s CWI indicator 149

[10.2](#_heading=h.2gb3jie) Step-by-step procedures to calculate the CWI indicator 153

[References 173](#_heading=h.1ulbmlt)

[11.](#_heading=h.4ekz59m) Resilience indicators 174

[11.1](#_heading=h.2tq9fhf) Guidelines to construct the resilience indicators 174

[11.1.1](#_heading=h.18vjpp8) Ability to recover from shocks and stresses index 174

[11.1.2](#_heading=h.3sv78d1) Index of social capital at the household level 176

[11.1.3](#_heading=h.n5rssn) Percent of households that believe local government will respond effectively to future shocks and stresses 178

[11.1.4](#_heading=h.375fbgg) Percent of households participating in group-based savings, micro-finance, or lending programs 178

[11.2](#_heading=h.1maplo9) Step-by-step procedures to calculate resilience indicators 179

[11.2.1](#_heading=h.46ad4c2) Ability to recover from shocks and stresses index 179

[11.2.2](#_heading=h.2lfnejv) Index of social capital at the household level 182

[11.2.3](#_heading=h.10kxoro) Percent of households that believe local government will respond effectively to future shocks and stresses 184

[11.2.4](#_heading=h.3kkl7fh) Percent of households participating in group-based savings, micro-finance, or lending programs 185

[References 187](#_heading=h.4jpj0b3)

[12.](#_heading=h.2yutaiw) Abbreviated Women’s Empowerment in Agriculture Index 188

[12.1](#_heading=h.1e03kqp) Background 188

[12.2](#_heading=h.3c9z6hx) Guidelines to construct the A-WEAI indicators 189

[12.2.1](#_heading=h.1rf9gpq) A-WEAI 189

[12.2.2](#_heading=h.393x0lu) Average percentage of women achieving adequacy across the six indicators of the A-WEAI 199

[12.3](#_heading=h.1o97atn) Step-by-step instructions for computing the A-WEAI indicators 199

[12.3.1](#_heading=h.488uthg) A-WEAI 199

[12.3.2](#_heading=h.3rnmrmc) Average percentage of women achieving adequacy across the six indicators of the A-WEAI 218

[References 220](#_heading=h.26sx1u5)

[13.](#_heading=h.ly7c1y) Agricultural productivity indicators 222

[13.1](#_heading=h.35xuupr) Guidelines to construct the indicators 222

[13.1.1](#_heading=h.1l354xk) Percent of producers in the targeted area who have applied targeted improved management practices or technologies 222

[13.1.2](#_heading=h.452snld) Yield of targeted agricultural commodities 225

[13.2](#_heading=h.2k82xt6) Step-by-step procedure to calculate agricultural productivity indicators 227

[13.2.1](#_heading=h.zdd80z) Percent of producers in the targeted area who have applied targeted improved management practices or technologies 227

[13.2.2 Yield of targeted agricultural commodities 243](#_heading=h.r2r73f)

[14.](#_heading=h.14hx32g) Household food insecurity indicators 254

[14.1](#_heading=h.3ohklq9) Guidelines to construct the indicators 254

[14.1.1](#_heading=h.23muvy2) Prevalence of moderate and severe hunger in the population, based on the Household Hunger Scale 254

[14.1.2](#_heading=h.is565v) Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale 254

[14.2](#_heading=h.1hx2z1h) Step-by-step procedure to calculate the food security indicators 255

[14.2.1](#_heading=h.41wqhpa) Prevalence of moderate and severe hunger in the population, based on the Household Hunger Scale 255

[14.2.2](#_heading=h.w7b24w) Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale 257

[15.](#_heading=h.4fbwdob) Children’s and women’s nutrition indicators 267

[15.1](#_heading=h.2uh6nw4) Guidelines to construct the indicators 267

[15.1.1](#_heading=h.19mgy3x) Children’s nutritional status indicators 267

[15.1.2](#_heading=h.nwp17c) Prevalence of exclusive breastfeeding of children under 6 months of age 269

[15.1.3](#_heading=h.37wcjv5) Percent of children 6-23 months of age receiving a minimum acceptable diet 269

[15.1.4](#_heading=h.1n1mu2y) Prevalence of underweight women of reproductive age 270

[15.1.5](#_heading=h.2m6kmyk) Women’s dietary diversity: Mean number of food groups consumed by women of reproductive age 271

[15.1.6](#_heading=h.11bux6d) Percent of women of reproductive age consuming a diet of minimum diversity 272

[15.2](#_heading=h.3lbifu6) Step-by-step procedures to calculate nutrition indicators 272

[15.2.1](#_heading=h.20gsq1z) Prevalence of underweight children under 5 years of age 272

[15.2.2](#_heading=h.4kgg8ps) Prevalence of stunted children under 5 years of age 274

[15.2.3](#_heading=h.3yqobt7) Prevalence of wasted children under 5 years of age 276

[15.2.4](#_heading=h.2dvym10) Prevalence of healthy weight children under 5 years of age 278

[15.2.5](#_heading=h.t18w8t) Prevalence of exclusive breastfeeding of children under 6 months of age 279

[15.2.6](#_heading=h.3d0wewm) Prevalence of children 6-23 months of age receiving a minimum acceptable diet 283

[15.2.7](#_heading=h.2rb4i01) Prevalence of underweight women of reproductive age 289

[15.2.8](#_heading=h.16ges7u) Women’s Dietary Diversity: Mean number of food groups consumed by women of reproductive age 291

[15.2.9](#_heading=h.kqmvb9) Percent of women of reproductive age consuming a diet of minimum dietary diversity 294

[Appendix A. A-WEAI summary information 299](#_heading=h.1jvko6v)

[Appendix B. A-WEAI confidence interval and design effect calculations 303](#_heading=h.2j0ih2h)

[B1. Background and guidelines 303](#_heading=h.y5sraa)

[B1.1. The confidence interval estimation 303](#_heading=h.3i5g9y3)

[B1.2. The Design effect estimation 304](#_heading=h.1xaqk5w)

[B1.3. Differences of indicators and test of significance 304](#_heading=h.4hae2tp)

[References 305](#_heading=h.1bkyn9b)

[B2. Step-by-step procedures 306](#_heading=h.3vkm5x4)

[B2.1 A-WEAI indicator with standard error estimation 306](#_heading=h.2apwg4x)

[B2.2 A-WEAI context indicator with standard error estimation 317](#_heading=h.pv6qcq)

# List of figures

[Figure 1: Setting up a Working Directory in R 258](#_heading=h.3g6yksp)

[Figure 2: Perform Equating Worksheet 262](#_heading=h.1vc8v0i)

# List of tables

[Table 1: Population-based ZOI-level Performance Indicators by Program Phase 3](#_heading=h.lnxbz9)

[Table 2: Data File Names by Units of Analysis and File Format 9](#_heading=h.2bn6wsx)

[Table 3: Special Response Categories and Their Codes, by Question Type 14](#_heading=h.3o7alnk)

[Table 4: PPP 2011 Conversion Factor, Private Consumption (LCU per international $) 143](#_heading=h.2olpkfy)

[Table 5: PPP 2005 Conversion Factor, Private Consumption (LCU per international $) 146](#_heading=h.13qzunr)

[Table 7: Unsatisfied Basic Needs Categories and Criteria 151](#_heading=h.1h65qms)

[Table 8: Reference Survey Anchoring Point Values 152](#_heading=h.415t9al)

[Table 9: Example Cases that Show Variables with 0 or 1 ‘Yes’ Reponses 163](#_heading=h.vgdtq7)

[Table 10: Frequencies and Percentages by UBN Score 169](#_heading=h.3fg1ce0)

[Table 11: Comparison of the Original WEAI and A-WEAI 189](#_heading=h.2d51dmb)

[Table 12: Activities Included in the A-WEAI Time Use Module 194](#_heading=h.33zd5kd)

[Table 13: ZOI Survey Variables and Response Options to Identify Illustrative Improved Management Practices and Technologies by Management Practice and Technology Type Category for Maize 229](#_heading=h.1yib0wl)

[Table 14: ZOI Survey Variables and Response Options to Identify Improved Aquaculture Management Practices and Technologies by Aquaculture Management Sub-category 233](#_heading=h.2xn8ts7)

[Table 15: ZOI Survey Variables and Response Options to Identify Improved Dairy Cow Management Practices and Technologies by Livestock Management Sub-category 237](#_heading=h.3ws6mnt)

[Table 16: Z-score Cutoffs for Children’s Nutritional Status Indicators 268](#_heading=h.3tm4grq)

[Table 17: Women’s Nutritional Status Category by BMI 271](#_heading=h.471acqr)

[Table 18: Seven Food Groups Used to Construct Dietary Diversity for Breastfed Children 285](#_heading=h.1s66p4f)

[Table 19: Nine Food Groups Used to Generate Women’s Dietary Diversity Score 292](#_heading=h.3qg2avn)

[Table 20: Ten Food Groups Used to Generate Women’s Food Score 295](#_heading=h.34qadz2)

[Table A1: Summary of A-WEAI Domains, Indicators, Survey Questions, Variables, Definitions, and Weights 299](#_heading=h.43v86uo)

# Abbreviations

5DE Five domains of empowerment

A-WEAI Abbreviated Women’s Empowerment in Agriculture Index

AWI Absolute wealth index

BFS Bureau for Food Security

BMI Body mass index

CPI Consumer Price Index

CSPro Census and Survey Processing System

CWI Comparative wealth index

DHS Demographic and Health Surveys

EA Enumeration area

FAO United Nations Food and Agriculture Organization

FIES Food Insecurity Experience Scale

FTFMS Feed the Future Monitoring System

GFSS Global Food Security Strategy

GNI/p Gross national income per capita

GPI Gender parity index

HAZ Height-for-age z-score

HHS Household Hunger Scale

LCU Local currency unit

P1-ZOI Phase one Zone of Influence

P2-ZOI Phase two Zone of Influence

PPP Purchasing power parity

SD Standard deviation

TP Total production

UBN Unsatisfied basic needs

UP Unit of production

USAID United States Agency for International Development

USD U.S. dollars

VCC Value chain commodity

WAZ Weight-for-age z-score

WDDS Women’s Dietary Diversity Score

WEAI Women’s Empowerment in Agriculture Index

WHO World Health Organization

WHZ Weight-for-height z-score

ZOI Zone of Influence

**Part I.**

**INTRODUCTION TO ZOI SURVEY DATA AND VARIABLES**

# Introduction

Feed the Future seeks to sustainably reduce global poverty, hunger, and malnutrition by helping partner countries boost agriculture-led growth, resilience, and nutrition. Program efforts are concentrated 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 performance indicators. The purpose of the Feed the Future ZOI Surveys is to provide U.S. Government interagency partners, the United States Agency for International Development (USAID) Bureau for Food Security (BFS), USAID Missions, host country governments, and development partners with information on the current status of the Feed the Future population-based ZOI-level performance indicators. The ZOI Survey is designed to monitor progress and determine whether there has been statistically significant change over time at the population level in key outcome and impact indicators, with the expectation that the effects of the program should spread beyond beneficiaries to the general population in the ZOI over the life of the program.

The priority of the Feed the Future ZOI Survey 2018-2019 is to collect high-quality data that are reliable and comparable across all Feed the Future countries. To achieve comparable information across the ZOI Surveys, Feed the Future developed a Feed the Future ZOI Survey Methods Toolkit[[1]](#footnote-0) that provides detailed guidance and templates covering all aspects of the planning, implementation, and analysis of the ZOI Surveys, ensuring that good survey methods are implemented across the initiative.

This *Guide to Feed the Future Statistics* is part of the Feed the Future ZOI Survey Toolkit. It contains a detailed description of the process and steps needed to analyze the ZOI Survey dataset and calculate the Feed the Future population-based ZOI-level performance indicators and related statistics. The Guide is divided into two parts. Part I provides an overview of the organization and use of the ZOI Survey dataset, describes the process for weighting and performing statistical tests on the data, defines key variables used to calculate indicators, and defines indicator disaggregates. Part II is devoted to the Feed the Future population-based 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 or statistics and then the step-by-step process to calculate these indicators and statistics using the core ZOI Survey dataset. Demographic and dwelling statistics that are not Feed the Future indicators but are important descriptors of the Feed the Future populations are also described in Part II of the guide.

This guide should be used in conjunction with other Feed the Future documents and in particular with the Feed the Future ZOI Survey Codebook,[[2]](#footnote-1) the Feed the Future ZOI Survey Data Processing Manual,[[3]](#footnote-2) and the Feed the Future Indicator Handbook.[[4]](#footnote-3) In addition, Stata program files to compute the Feed the Future population-based ZOI-level performance indicators and other key statistics included in the country reports are available as a complement to this guide.

## Feed the Future population-based ZOI-level performance indicators

This guide pertains to a subset of all Feed the Future indicators and is concerned only with the population-based, ZOI-level performance indicators.[[5]](#footnote-4) For the ZOI Survey 2018-2019, these population-based indicators are divided into three categories: (1) indicators that are included only in the results framework for Feed the Future phase one, (2) indicators that are included only in the results framework for Feed the Future phase two, and (3) indicators that are included in the results framework for both Feed the Future phase one and phase two (Table 1).

Indicators for both phase one and phase two are included in this guide because the ZOI Survey 2018-2019 combines in one survey the endline for phase one and the baseline for phase two in a number of countries. The phase one data analysis entails generating endline indicator estimates and comparing those estimates to baseline estimates for phase one indicators to determine whether there were statistically significant changes over the program period in the phase one ZOI (P1-ZOI). The phase two data analysis entails calculating baseline estimates for all phase two indicators in the phase two ZOI (P2-ZOI).[[6]](#footnote-5)

**Table 1: Population-based ZOI-level Performance Indicators by Program Phase**

| **Feed the Future phase one**[[7]](#footnote-6) | | **Feed the Future phase two** | |
| --- | --- | --- | --- |
| 1 | Prevalence of poverty: Percentage of people living on less than $1.25/day (2005 PPP) | NO LONGER REPORTED UNDER  FEED THE FUTURE PHASE TWO | |
| 2 | Depth of poverty: Mean percentage shortfall relative to the $1.25/day (2005 PPP) poverty line |
| 3 | Daily per capita expenditures (as a proxy for incomes) (2010 USD) |
| 4 | Prevalence of underweight (WAZ<-2) children under age 5 years |
| 5 | Women dietary diversity: Mean number of food groups consumed by women of reproductive age (WDDS) |
| 6 | Women’s empowerment in agriculture index (WEAI) |
| 7 | Prevalence of moderate and severe hunger in the population, based on the Household Hunger Scale (HHS) |
| 8 | Prevalence of stunted (HAZ<-2) children under age 5 years | Prevalence of stunted (HAZ<-2) children under age 5 years | 1 |
| 9 | Prevalence of wasted (WHZ<-2) children under age 5 years | Prevalence of wasted (WHZ<-2) children under age 5 years | 2 |
| 10 | Prevalence of underweight (BMI<18.5) women of reproductive age | Prevalence of underweight (BMI<18.5) women of reproductive age | 3 |
| 11 | Percent of children ages 6-23 months receiving a minimum acceptable diet | Percent of children ages 6-23 months receiving a minimum acceptable diet | 4 |
| 12 | Prevalence of exclusive breastfeeding of children under age 6 months | Prevalence of exclusive breastfeeding of children under age 6 months | 5 |
| NEW IN FEED THE FUTURE PHASE TWO | | (Abbreviated) Women’s Empowerment in Agriculture Index (A-WEAI) | 6 |
| Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale (FIES) | 7 |
| Percent of women of reproductive age consuming a diet of minimum diversity | 8 |
| Prevalence of healthy weight (WHZ ≤+2 or ≥-2) among children under age 5 years | 9 |
| Percent of households with the self-perceived ability to successfully manage future shocks and stresses | 10 |
| Percent of households with high social capital | 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 with access to a basic sanitation service | 14 |
| Percent of households with soap and water at a handwashing station commonly used by family members | 15 |
| Yield of targeted agricultural commodities within target areas | 16 |
| Percent of producers in the targeted area who have applied targeted improved management practices or technologies | 17 |
| Prevalence of poverty: Percentage of people living on less than $1.90/day (2011 PPP) | 18 |
| Depth of poverty of the poor: Mean percentage shortfall relative to the $1.90/day (2011 PPP) poverty line | 19 |
| Percent of households below the comparative threshold for the poorest quintile of the asset-based comparative wealth index | 20 |
| Percent of households that are near-poor: households that are within 100 percent and 125 percent of the poverty line [context indicator] | 21 |
| Percent of women who are empowered according to the five domains of the A-WEAI [context indicator] | 22 |

Missions may add questions to the survey questionnaire to calculate a limited number of additional population-based performance indicators. For example, a ZOI-level market-related outcome indicator could be added to further assist the Mission in linking beneficiary-level outcomes to population-level impacts and tracking progress toward higher-level impacts. Guidance to calculate Mission-specific indicators, however, is not included in this document.

## Feed the Future ZOI Survey 2018-2019 core questionnaire

The Feed the Future ZOI Survey Toolkit includes a core questionnaire and guidance on how to adapt the core questionnaire to the country context by incorporating country-specific foods, shocks and stressors, or agricultural practices. The core questionnaire for the ZOI Survey 2018-2019 includes the following modules, in addition to the household identification and informed consent information:

* Module 1: Household roster and demographics
* Module 2: Dwelling characteristics
* Module 3: Food security and resilience
* Module 4: Women’s nutrition and anthropometry
* Module 5: Children’s nutrition and anthropometry
* Module 6: Abbreviated Women’s Empowerment in Agriculture Index (A-WEAI)
* Module 7: Agricultural technologies
* Module 8: Household consumption expenditures

## Data management and results reporting

The ZOI Surveys are designed to capture statistically significant changes in indicator estimates from baseline. The data collected to report Feed the Future key indicators are used to inform programmatic and policy decisionmaking, which impact the well-being of people living in the ZOI.

The ZOI Surveys collect high-quality data in a timely manner. As part of achieving this, ZOI Survey field teams identify the most efficient ways to securely access the Internet, use reliable back-up methods, and regularly transmit data from the field. ZOI Surveys ensure quality control both during face-to-face interviews with respondents and after. During the interview, the computer-assisted interviewing software runs internal checks to constrain value ranges, control skip patterns, preclude item-missingness, and establish consistency across variables. After interviews are completed, quality control reports called field check tables are generated on a weekly basis to allow an assessment of the quality of the data received from the field. These reports identify patterns in missing data, outliers, heaping, and age displacement, and calculate response rates.

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, weighted N, standard deviation (SD) (if applicable), 95 percent confidence interval, and design effect. Statistical tests of differences are conducted among disaggregate categories. For phase one endline reports, statistical tests of difference are also conducted between baseline and endline values. ZOI Survey findings are presented in country reports. Standardized templates for the phase one endline report and phase two baseline report can be found in the Survey Toolkit.[[8]](#footnote-7)

ZOI Survey indicator values are also entered in the Feed the Future Monitoring System (FTFMS), which consolidates U.S. Government reporting on Feed the Future. Indicator data in the FTFMS constitute the official Feed the Future results. Multiple groups contribute data to the system. The Feed the Future Indicator Handbook specifies the information that must be entered into the FTFMS for each indicator in the Reporting Notes section of the indicator reference sheets. The data provide the foundation for the Feed the Future Progress Report and inform decisions on future programming and budget allocations.

After publication of the country reports, datasets are made available for public use. Prior to release, Feed the Future ensures protection for the privacy and confidentiality of respondents by removing from the datasets personally identifiable information, such as name and geographic identifiers below the second-level administrative division.

# Organization of ZOI Survey data

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

Feed the Future ZOI Surveys use a public domain software package called Census and Survey Processing System (CSPro) to enter, edit, and tabulate data during fieldwork. This software is developed and supported by the U.S. Census Bureau and ICF to support collection and data 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 come in three formats: hierarchical, rectangular, and 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 a flat file or a rectangular file for statistical analysis. The ZOI Survey CSPro data export program creates flat files, but the program can be edited to also create rectangular files. Note that the ZOI Survey template syntax files assume flat data files.

### 2.1.1 Hierarchical files

Data collection using CSPro software is hierarchical, with all data stored in a single data file. The base unit for data collection in the Feed the Future ZOI Surveys is the household, and each household has a questionnaire in the data file. A cluster and household number uniquely identify a household. Some data are also associated with household members, so the number of records at the individual level depends on the composition of the household. Within a household, information is collected from primary adult male and female decisionmakers, all women of reproductive age (15-49 years of age), and all children under 5 years of age. In the Feed the Future phase two baseline surveys, individual-level data are also collected from every person who has responsibility over the production of the selected value chains (crop, livestock, or fish). For individual-level data, such as in the nutrition and anthropometry modules, an additional identifying variable exists—the line number of the person in the household roster—which, in combination with the cluster 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 the Feed the Future ZOI Surveys. Note that households may exist in the data file without individual-level data if, for example, household members refused to participate or were not home, or if the dwelling was destroyed, vacant, or not found.



PADM=primary adult decisionmaker

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 (e.g., dwelling characteristics, which are unique to the household) or multiple-occurring (e.g., household roster and anthropometry information, which may have multiple entries per household—one per household member [roster] or eligible household member [anthropometry]). 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 decisionmakers. As mentioned above, hierarchical files cannot be read by statistical software such as Stata and SPSS, however; they must be transformed into flat or rectangular 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 at the highest level. In Feed the Future ZOI Surveys, there are two flat files: one for household-level data and another for individual-level data. There is not a 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. For example, Module 7.91, Plot Area, is administered to producers to collect information on each plot of agricultural land that they responsible for. All the variables related to a producer’s first plot have a “\_1” appended to the end of the name. All the variables related to a producer’s second plot have a “\_2” appended to the end of the name, and so on. This creates a simple data file with one record per household or individual, and each record has a fixed length that contains the maximum number of cases possible for repeating sections of the questionnaire. For Module 7.91, the CSPro data collection program allows interviewers to collect information for up to eight plots of land per producer. If a producer does not have eight plots, the variables for all eight plots still exist, but the variables for any plots not applicable to the producer will be blank in the data file.

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

Feed the Future datasets are made available as file types used for R, SAS, SPSS, and Stata. Datasets include survey documentation and syntax files that describe the data and allow users to easily import the data into the statistical software program. The naming conventions for ZOI Survey data files exported from CSPro are presented in Table 2.

**Table 2: Data File Names by Units of Analysis and File Format**

| **Analysis Unit** | **File Format** | |
| --- | --- | --- |
| **Hierarchical**  **(CSPro)** | **Flat**  **(R, SAS, SPSS, Stata)** |
| Household | CFTF\_HOUSEHOLD\_Date | CFTF\_HOUSEHOLD\_Date |
| Individuala | CFTF\_PERSONS\_Date | CFTF\_PERSONS\_Date |

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

a The individual-level file includes records for women 15-49 years of age, children 0-59 months of age, and producers.

All data files exported from CSPro begin with “CFTF,” where “C” is a placeholder for the first letter of the country where the ZOI Survey was implemented. The file name also contains the year of data collection (Date) and indicates whether it is the household-level or individual-level data file. All associated documentation and syntax are included when downloading data files. More information on exporting data from CSPro can be found in the Feed the Future ZOI Survey Data Processing Manual.[[9]](#footnote-8)

### 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 Feed the Future 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 decisionmaker and the primary adult female decisionmaker 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 decisionmakers and another file of primary adult female decisionmakers, 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 decisionmaker is matched to exactly one primary adult female decisionmaker, 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 do the following actions:

* Determine the identifier variables on which to match data files.
  + In the 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.
  + In a one-to-one merge, typically, although not always, the base data file is the one with the fewest number of records.
  + In a one-to-many merge, typically, although not always, 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.

Using proper syntax, depending on the software package, data will be merged according to the inputs and matching identifier variables.

## Variable name conventions

The ZOI Survey variables created in CSPro use standard naming conventions that are described in this section. Please see the Feed the Future Codebook for detailed information about all ZOI Survey 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 Survey variables are named according to the following conventions:

* Household identification (Cover page)[[10]](#footnote-9)

ah+question identifier (ahyear, ahvisits, ahinterv, …)

* Household roster (Module 1)

v+question number (v101, v102, v103, …)

* Household dwelling characteristics (Module 2)

v+question number (v201, v202, v203, …)

* Food security and resilience (Module 3)

v+question number (v301, v302, v303, …)

* Women’s nutrition and anthropometry (Module 4)

v+question number (v401, v402, v403, …)

* Children’s nutrition and anthropometry (Module 5)

v+question number (v501, v502, v503, …)

* Women’s Empowerment in Agriculture Index (WEAI), female respondents (Module 6)

Part 1 (Respondent information) v6+question number (v6101, v6102, v6103, …, v61xx)

Part 2 (Role in decisionmaking) v6+question number (v6201, v6202, v6203, …, v62xx)

Part 3 (Access to capital and credit) v6+question number (v6301, v6302, v6303, …, v63xx)

Part 4 (Group membership) v6+question number (v6404, v6405…, v64xx)

Part 6 (Time allocation) v6+question number (v6601)

For questions that are asked about multiple items (e.g., economic activities, sources of credit, assets) in the WEAI module, 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., where v6201\_01 corresponds to the first economic activity (a – food crop farming) and v6202\_02 corresponds to the second economic activity (b – 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 for every 15 minutes during the day preceding the survey, beginning at 4 a.m. Separate variables are created for the primary activities performed during these 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. Following is an example using hour 1 (4:00am-4:59am):

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

* Women’s Empowerment in Agriculture, male respondent (Module 6M)

Part 1 (Respondent information) m6+question number (m6101, m6102, m6103, …, m61xx)

Part 2 (Role in decisionmaking) m6+question number (m6201, m6202, m6203, …, m62xx)

Part 3 (Access to capital and credit) m6+question number (m6301, m6302, m6303, …, m63xx)

Part 4 (Group membership) m6+question number (m6401, m6402, m6403, …, m64xx)

Part 6 (Time allocation) m6+question number (m6601, m6602, m6603, …, m66xx)

The time allocation variables and questions that are asked about multiple items are captured using variables that are named in a similar way as that explained above for the female respondents.

* Agricultural technologies (Module 7)

Part 1 (maize) v7+question number (v7101, v7102, v7103, …, v71xx)

Part 2 (beans) v7+question number (v7201, v7202, v7203, …, v72xx)

Part 3 (coffee) v7+question number (v7301, v7302, v7303, …, v73xx)

Part 4 (groundnuts) v7+question number (v7401, v7402, v7403, …, v74xx)

Part 5 (wheat) v7+question number (v7501, v7502, v7503, …, v75xx)

Part 6 (soybeans) v7+question number (v7601, v7602, v7603, …, v76xx)

Part 7 (paddy rice) v7+question number (v7701, v7702, v7703, …, v77xx)

Part 8 (cow pea) v7+question number (v7801, v7802, v7803, …, v78xx)

Part 9 (chick pea) v7+question number (v7901, v7902, v7903, …, v79xx)

Part 10 (millet) v7+question number (v71001, v71002, v71003, …, v710xx)

Part 50 (dairy cows) v7+question number (v75001, v75002, v75003, …, v750xx)

Part 51 (goats) v7+question number (v75101, v75102, v75103, …, v751xx)

Part 80 (fish) v7+question number (v78001, v78002, v78003, …, v780xx)

Part 90 (land map) v7+question number (v79000a, v79000b, v79000c, …, v79000xxx)

Part 91 (plot area) v7+question number (v79101, v79102, v79103, …, v791xx)

Part 92 (crop yield) v7+question number (v79201, v79202, v79203, …, v792xx)

Part 95 (pond map) v7+question number (v79501)

Part 96 (pond area) v7+question number (v79601, v79602, v79603, …, v791xx)

* Household consumption expenditure (Module 8)

Part 1 (food consumption) v8+question number (v8101, v8102, v8103, …, v81xx)

Part 2 (non-food past 7 days) v8+question number (v8201, v8202, v8203, …, v82xx)

Part 3 (non-food past 1 month) v8+question number (v8301, v8302, v8303, …, v83xx)

Part 4 (non-food past 3 months) v8+question number (v8401, v8402, v8403, …, v84xx)

Part 5 (non-food past 12 months) v8+question number (v8501, v8502, v8503, …, v85xx)

Part 6 (housing expenditures) v8+question number (v8601, v8602, v8603, …, v86xx)

Part 7 (durable goods) v8+question number (v8701, v8702, v8703, …, v87xx)

For most Module 8 sections, the respondent is asked a set of questions for every item on a list—similar to sections of the WEAI module in which the same questions are asked about multiple items. The list of items constitutes the first variable, and each item represents a value attached to the variable. For example, v8101 corresponding to question 8.101 in Food Consumption has a value set that corresponds to all food items included in the ZOI Survey questionnaire. For example, the variable for the first food item (v8101\_001) has the value “8001. normal maize flour,” and the variable for the variable for the second food item (v8101\_002) has the value “8002. fine maize flour.” This is true for all sections of Module 8 in which a set of questions is asked for a list of items. The list of items in each section can be found in the final country-customized ZOI Survey questionnaire used for data collection and in the associated country-customized codebook.

### 2.2.2 Select-all questions

For questions in which interviewers can select all applicable responses a respondent may provide, response options are indicated by letters in the questionnaire.[[11]](#footnote-10) In the data files exported from CSPro, these questions are coded in two ways: in alphanumerical and 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 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 217 in Module 2, *Dwelling characteristics,* has multiple response options: the respondent can indicate more than one method used by the household to make water safer to drink.



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 sample 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 multiple-response question is missing, all corresponding binary variables are also coded as missing, or if a multiple-response question is not applicable (i.e., it is not asked because of the skip patterns set up in the CSPro data collection program), all corresponding binary variables are also coded as not applicable.

### 2.2.3 Special response options and variables that are not applicable

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 are clearly noted in the codebook and variable labels. **Special response options and missing observations should be excluded when statistics such as means and medians are calculated.** 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” (and 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 the ZOI Survey, 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 3 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 3: Special Response Categories and Their Codes, by Question Type**

| **Special Response Option** | **Response Option Code** | |
| --- | --- | --- |
| **Single Response Question** | **Multiple Response Question** |
| Missing values | 9, 99, 999, 9999, … | ? |
| Don’t know, unsure, unknown | 8, 98, 998, 9998, … | Z |
| Refusals | 7, 97, 997, 9997, … | Y |
| Other (specify) | 6, 96, 996, 9996, … | X |

# Data analysis information

## 3.1 Sampling weights

Sample weights are constructed to adjust results of tabulations based on the probability of selection for the survey and the response rates of the survey. The final sample of completed interviews will not be representative of the entire population by itself; the sample weights must be applied to the data to create representative results. In some cases, to obtain a sufficient number of cases for certain indicators or subgroups of the population, the sample size is increased 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.

Before calculating the Feed the Future indicators and other statistics presented in the country reports, the data should be weighted using weights relevant to the ZOI Survey (with the exception of the A-WEAI). Feed the Future ZOI Surveys have at least six sample weights: household, children under 5 years of age, children under 2 years of age, women of reproductive age, primary adult female decisionmakers, and primary adult male decisionmakers. Phase two baseline ZOI Surveys will also include producer weights. The number of weights for producers will depend 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. If the survey does not include targeted VCCs but does include a module for crop or livestock producers, a weight will be created for each producer sample (for example, crop producers or livestock producers).

Sample weights are needed to calculate representative statistics, such as means and proportions, but they are not used to estimate relationships among variables, such as through correlations or regressions. The sampling statistician for the ZOI Survey should calculate the relevant weights for data analysis, taking into consideration any overlap between the P1-ZOI and the P2-ZOI if the survey is both a phase one endline survey and phase two baseline survey. If the ZOI Survey is both a phase one endline survey and a phase two baseline survey, and the ZOIs for the two phases are not identical, separate phase one and phase two 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 refer to the *Feed the Future Data Processing Manual* for detailed information on adding the weights to the data files that are exported from CSPro.

Please note that sample weights should not be created so that they scale to the ZOI population, but they should be normalized so that the weighted numbers of observation are on the same scale as the sample (i.e., the summation of the weight will equal the total sample size and the average of the weight will be 1).

## 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.[[12]](#footnote-11)

Statistics presented in the ZOI Survey reports for each Feed the Future ZOI Survey indicator and its standard disaggregates (i.e., estimates presented in Table A1.1)[[13]](#footnote-12) include the unweighted and weighted number of observations, the sample-weighted point estimate (for example, a mean or percentage) and its 95 percent confidence interval, and the design effect.[[14]](#footnote-13) Standard deviations are also presented for means. Result tables throughout the body of descriptive reports generally include the sample-weighted point estimates and the unweighted number of observations for the ZOI Survey indicators and their disaggregates. Result tables throughout the body of comparative reports also include confidence intervals and differences in estimates over time, including the associated p-values and level of significance designation,[[15]](#footnote-14) for the ZOI Survey indicators and their disaggregates. Results tables in descriptive reports 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 (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 endline) 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 prevalence of stunting for children under 5 years of age at endline is compared to baseline, and then the prevalence of stunting for male and for female children under 5 years of age at endline are compared to baseline separately.

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 confidence intervals 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 primary sampling unit, 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:

* hhs: household hunger scale score, continuous variable (scale: 0-6)
* hhs\_cat: household hunger scale score, dichotomous variable (0=little to none, 1=moderate to severe)
* survey: survey round, dichotomous variable (1=baseline, 2=endline)
* 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

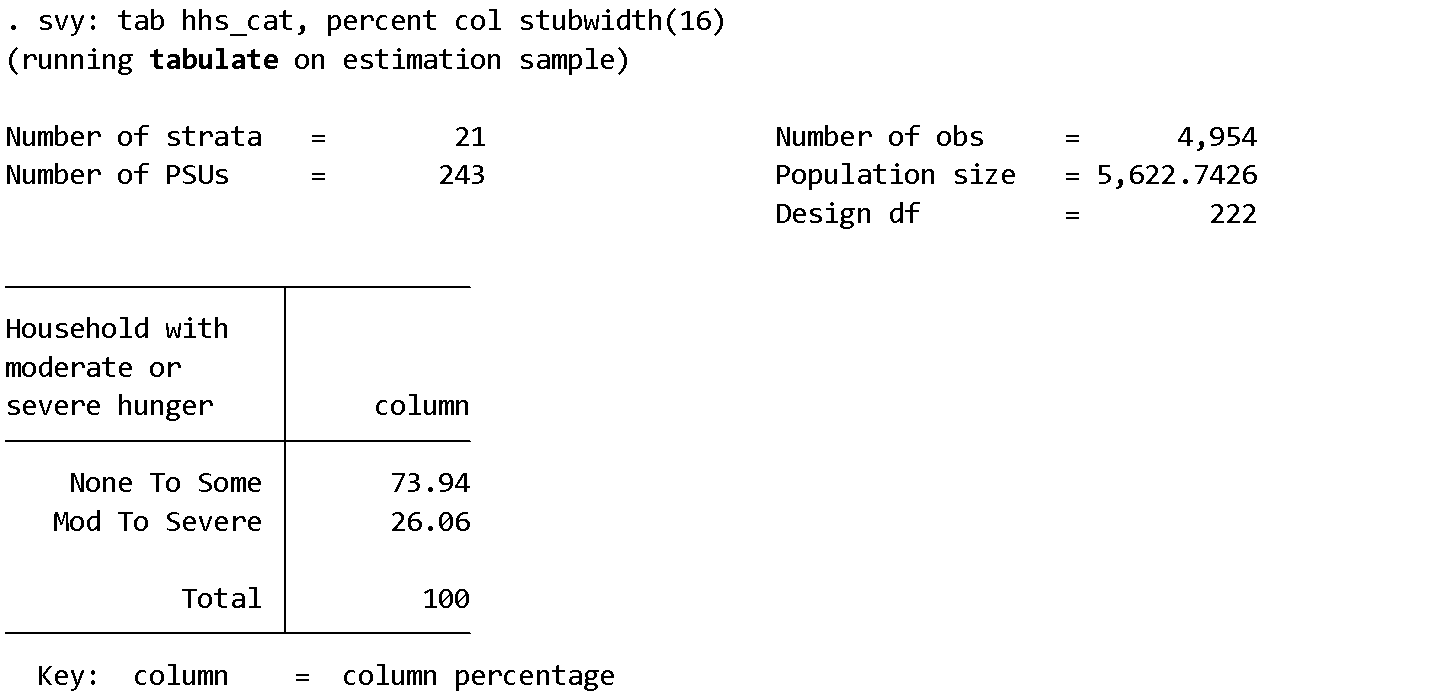
Descriptive estimates for an indicator of proportions can be calculated using the following syntax:[[16]](#footnote-15),[[17]](#footnote-16)

*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.

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



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 estmate, can be calculated using the following syntax:

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

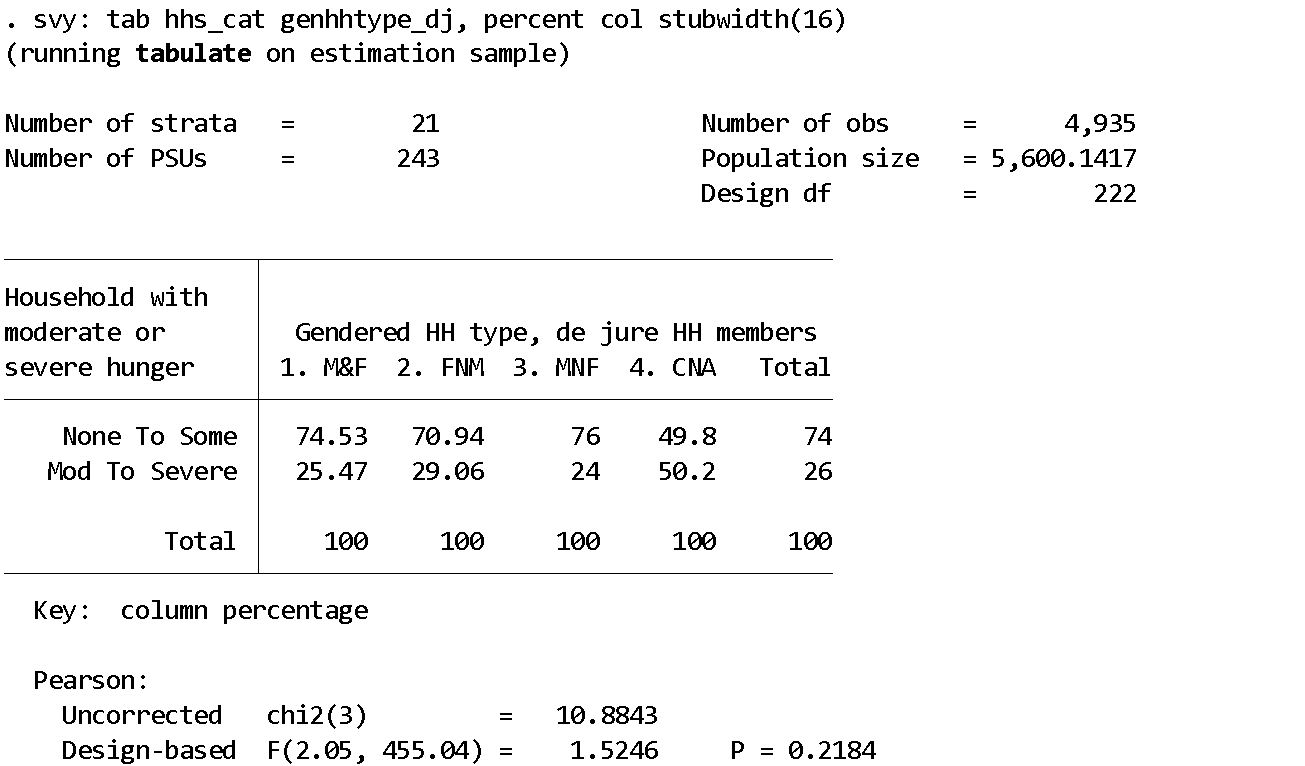
where: ***outcome*** is the variable for the indicator of proportions

***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 disaggegate categories) and the corresponding total population size (i.e., the weighted sample size across all disaggegate 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 household hunger). The p-value indicates whether a difference exists between categories if the disaggregate has only two categories (e.g., poverty status—poor or non-poor). 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 household hunger 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.[[18]](#footnote-17)

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 25.47 percent of male and female adult households experienced moderate-to-severe hunger. 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 HHS 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

****

M&F=male and female adult households, FNM=female adult only households (no adult males), MNF=male adult only households (no adult females), CNA=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*

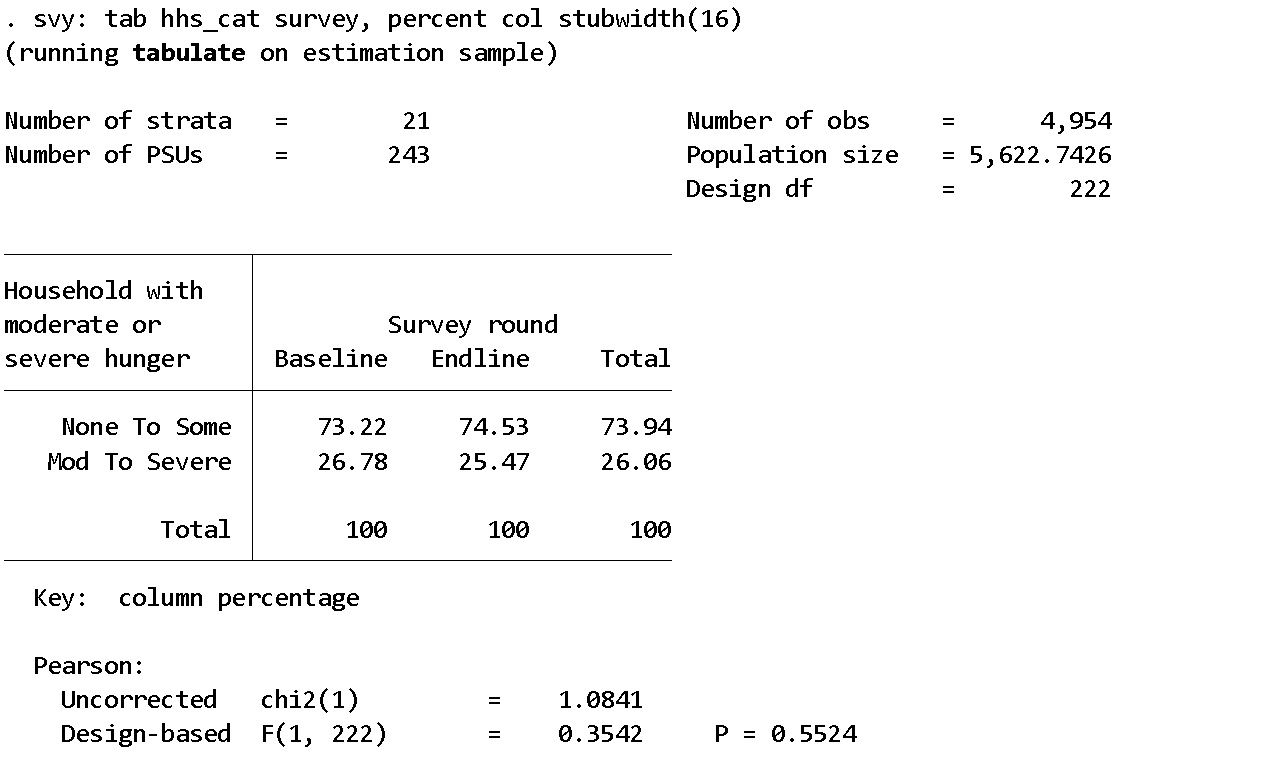
where: ***outcome*** is the variable for the indicator of proportions

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

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 26.78 percent of households experienced moderate-to-severe hunger at baseline and 25.47 percent of households experienced moderate-to-severe hunger at endline. 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 HHS 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



#### 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 endline)

***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)[[19]](#footnote-18)

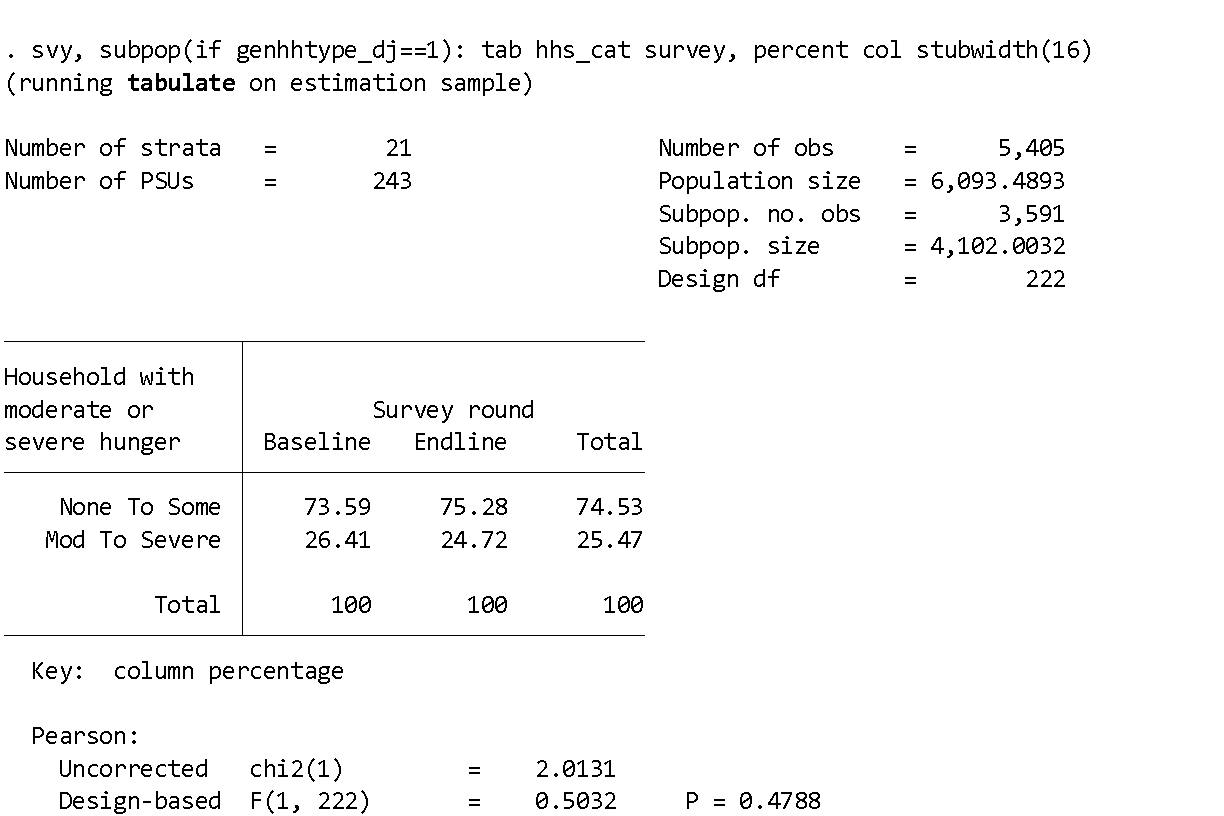
Subpopulations are created for each disaggregate category, and *svy, subpop(****sub\_X****): tab* ***outcome round,*** *percent col* is run separately for each disaggregate category.

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 subpopulations (e.g., ***sub\_1*** and ***sub\_2***), the subpopulation variable is set to 1 for the observations included in the disaggegate 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 subpopulation variable to missing if the observation is missing a value for the selected disaggregate (e.g., if the child’s sex is missing or if the household is missing a poverty status value). Subpopulations can be set up by creating a dichotomous variable or by using an IF statement; see the subpopulation estimation section of the Stata manual for more information.[[20]](#footnote-19)

Example 4 presents the output for a comparative analysis of the mean household hunger score variable over time for and 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 (ie., *if genhhtype\_dj==1*).

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



#### 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.

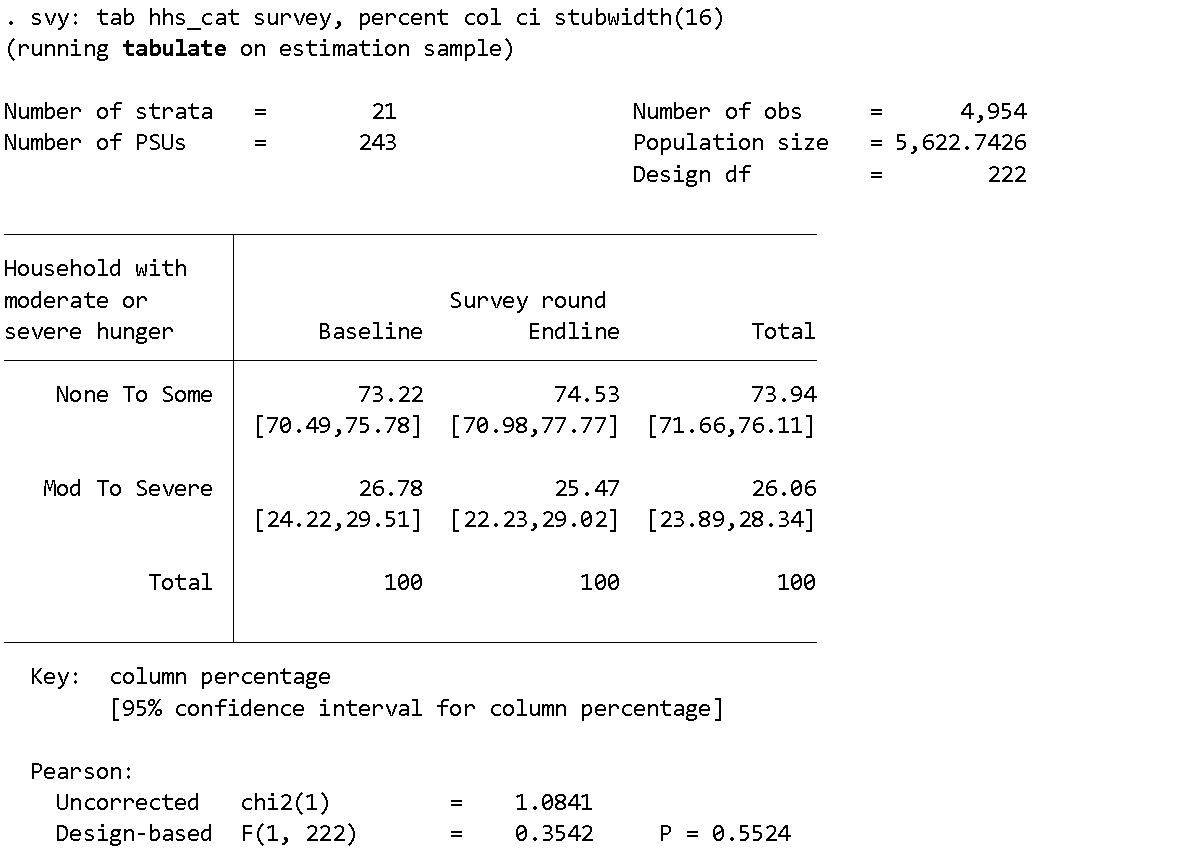
**Confidence intervals:** Confidence intervals are included in the output when the ‘*ci*’ option is specified. For example:

*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, confidence intervals

****

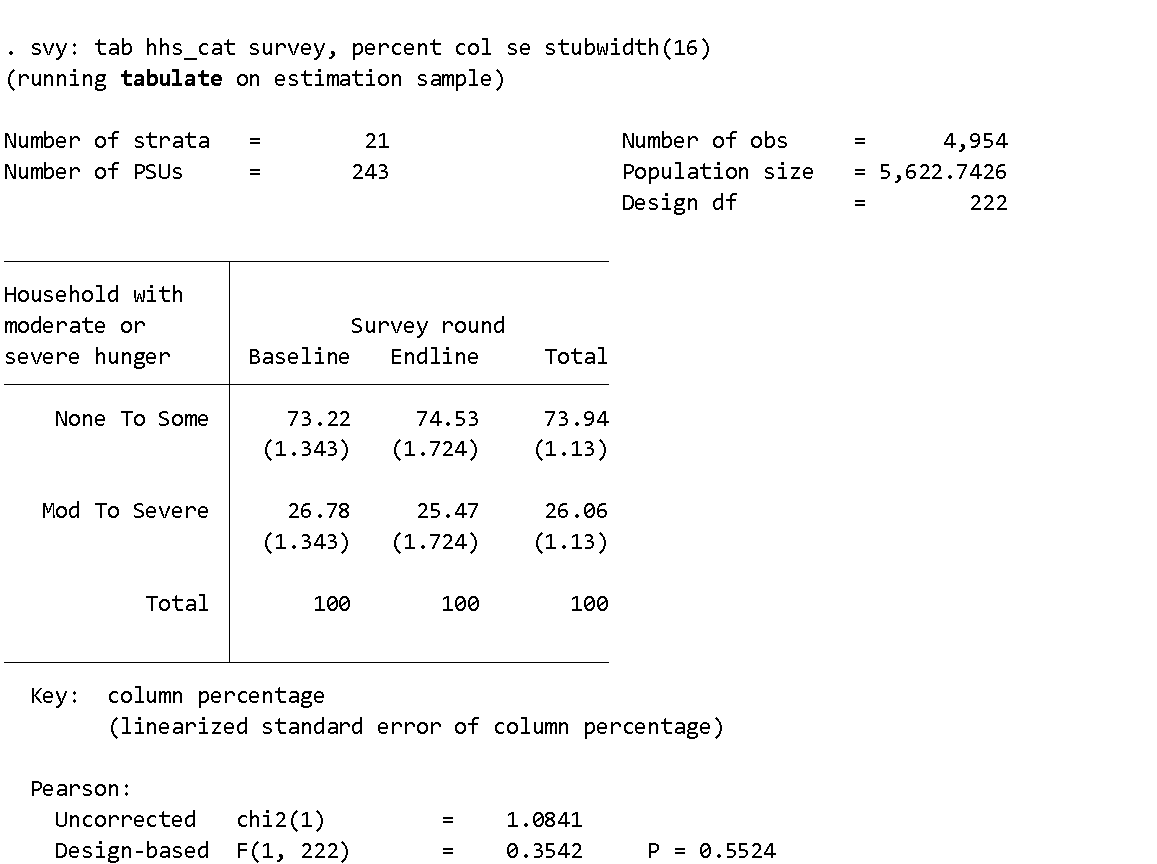
**Standard error:** Standard errors are included in the output when the ‘*se*’ option is specified. For example:

*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

****

**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:

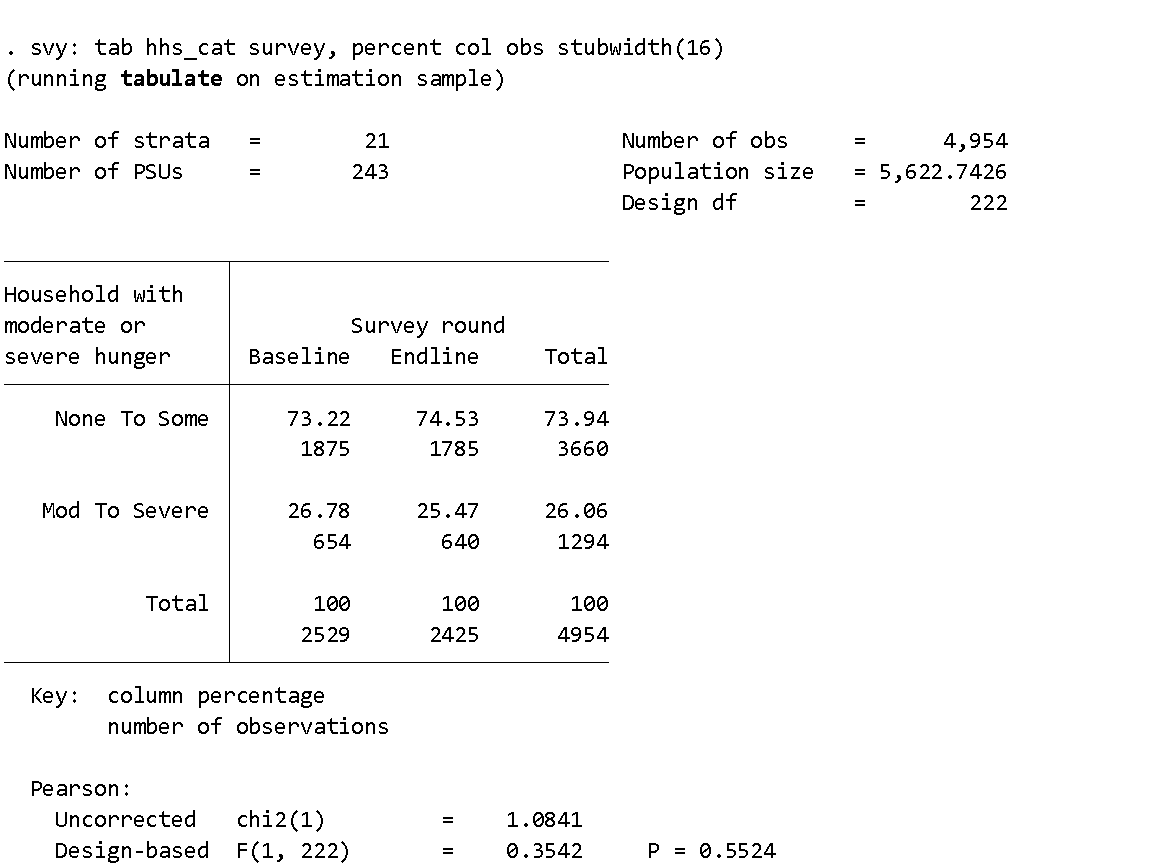
*svy: tab* ***outcome,*** *percent obs*  or

*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

****

**Population size (weighted n):** The population size, or the weighted sample size, for each disaggregate category or survey round is included in the output when the ‘*count*’ option is specified. For example:

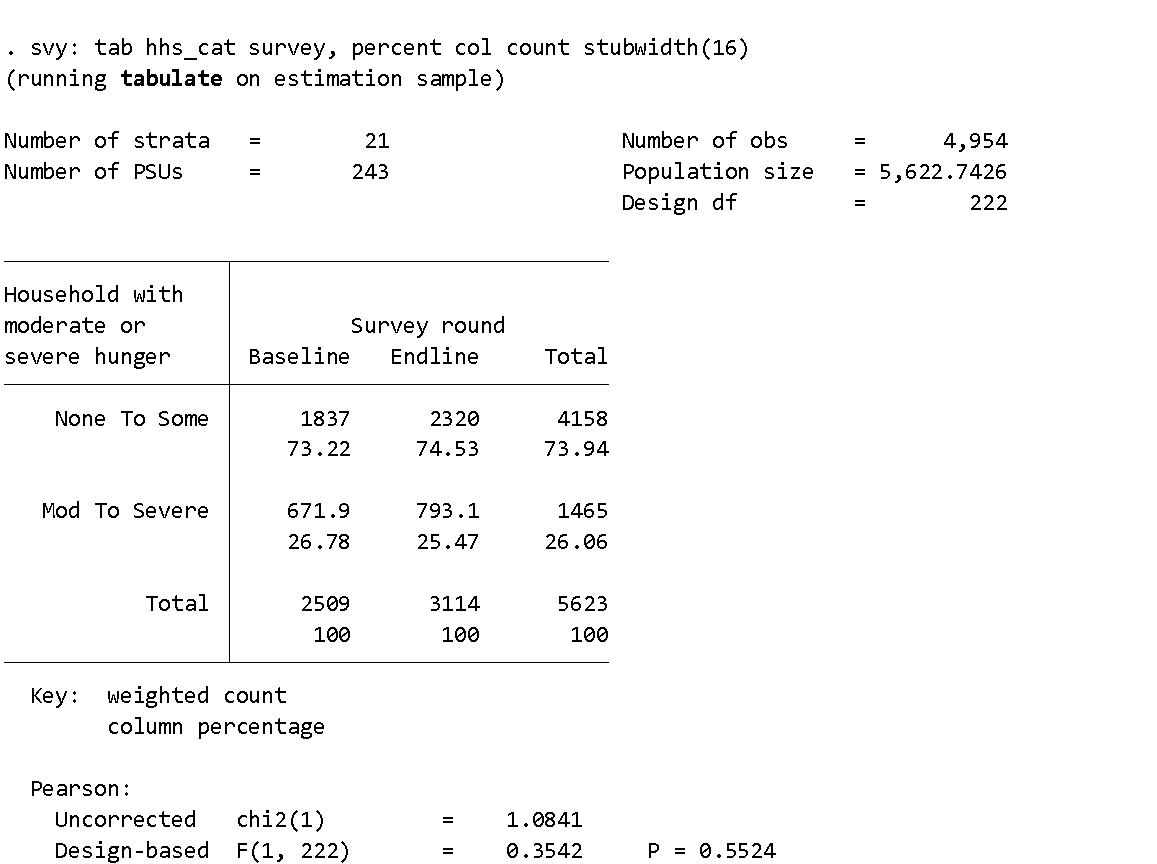
*svy: tab* ***outcome,*** *percent count*  or

*svy: tab* ***outcome disaggregate,*** *percent count* or

*svy: tab* ***outcome round,*** *percent count*

Note that this is different from the population size that is included in the output of all *svy: tab* commands, which presents only the total population size included in the calculation across all disaggregate categories or rounds; the ‘*count*’ option presents the population size for each outcome value and disaggregate category or survey round.

##### Example 8: Indicators of proportions: Supplementary statistics output, population size by survey round

****

**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

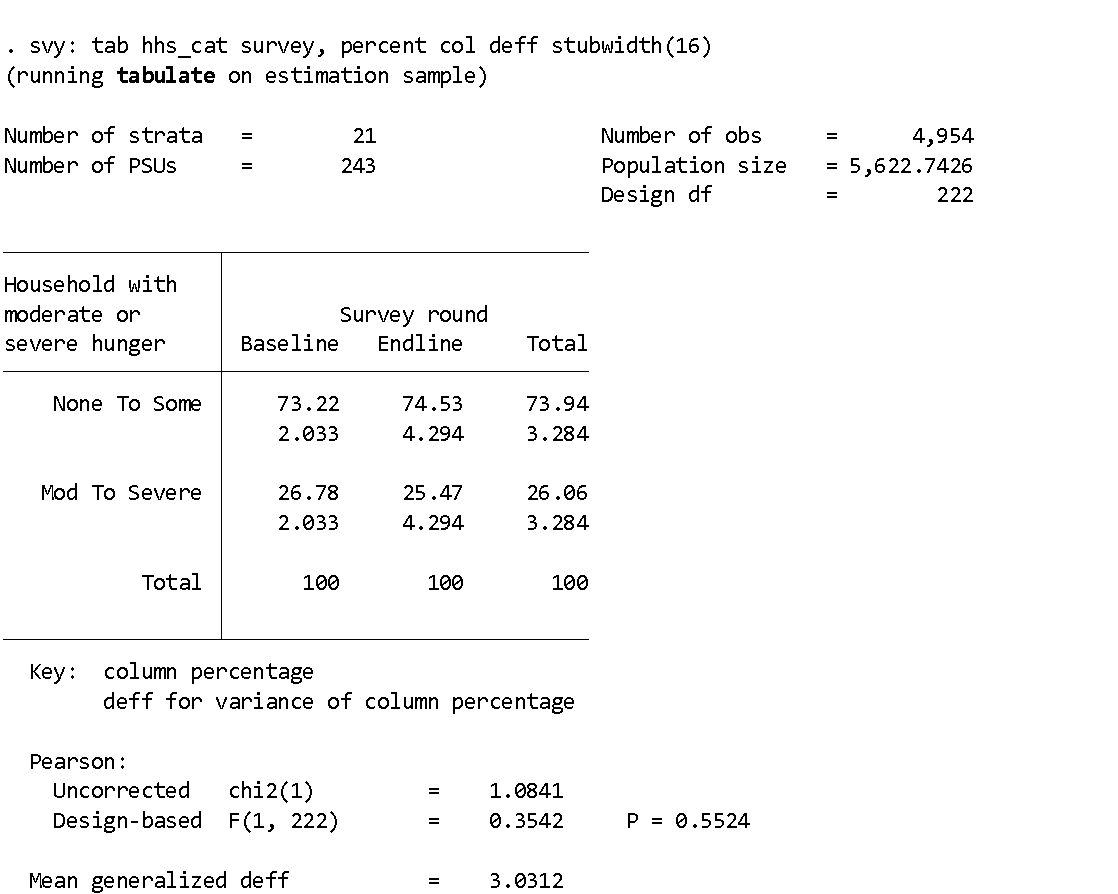
*svy: tab* ***outcome round,*** *percent deff*

Note that multiple options can be specified in the same command, but depending on the options selected, there are limitations on the number 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 9: Indicators of proportions: Supplementary statistics output, design effects



### 3.2.2 Analyzing indicators of means

#### Descriptive, overall

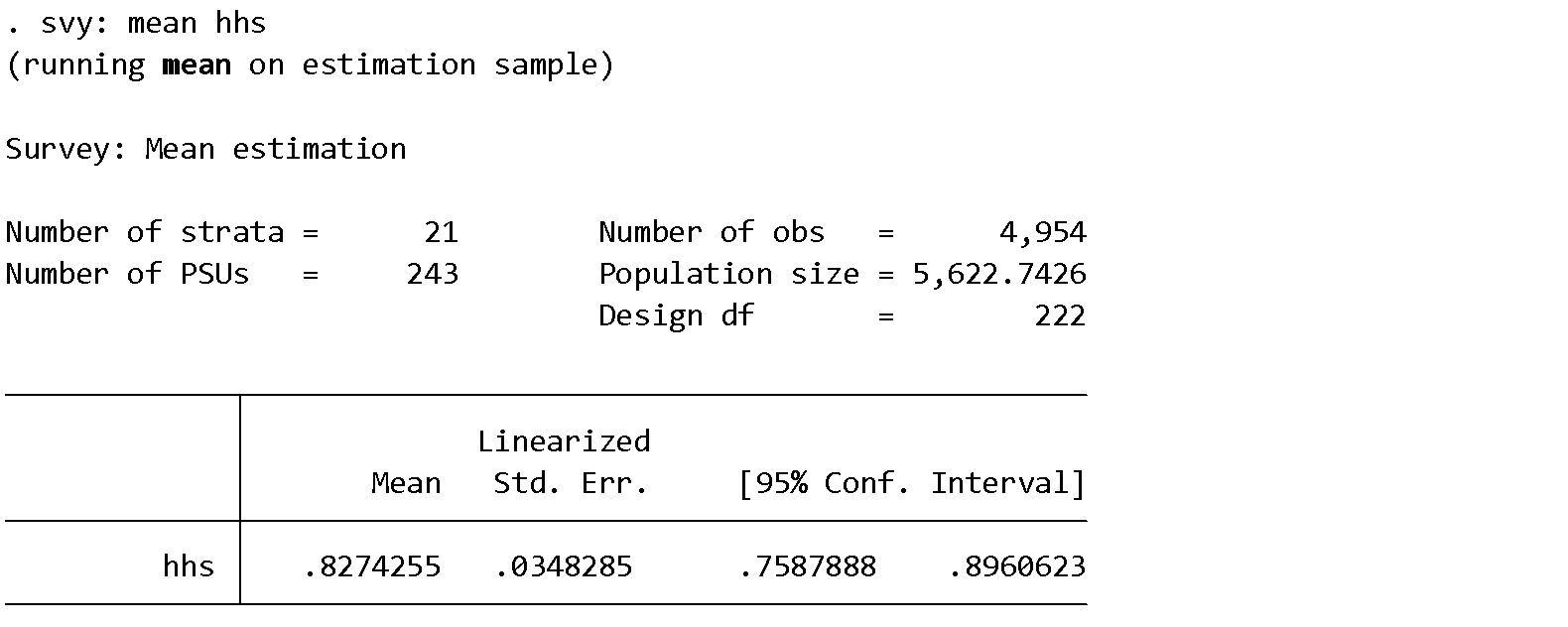
Descriptive estimates for an indicator of means can be calculated using the following syntax:

*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 confidence interval, 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 10).

##### Example 10: Indicators of means: Descriptive, overall output



#### Descriptive, by selected disaggregates

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

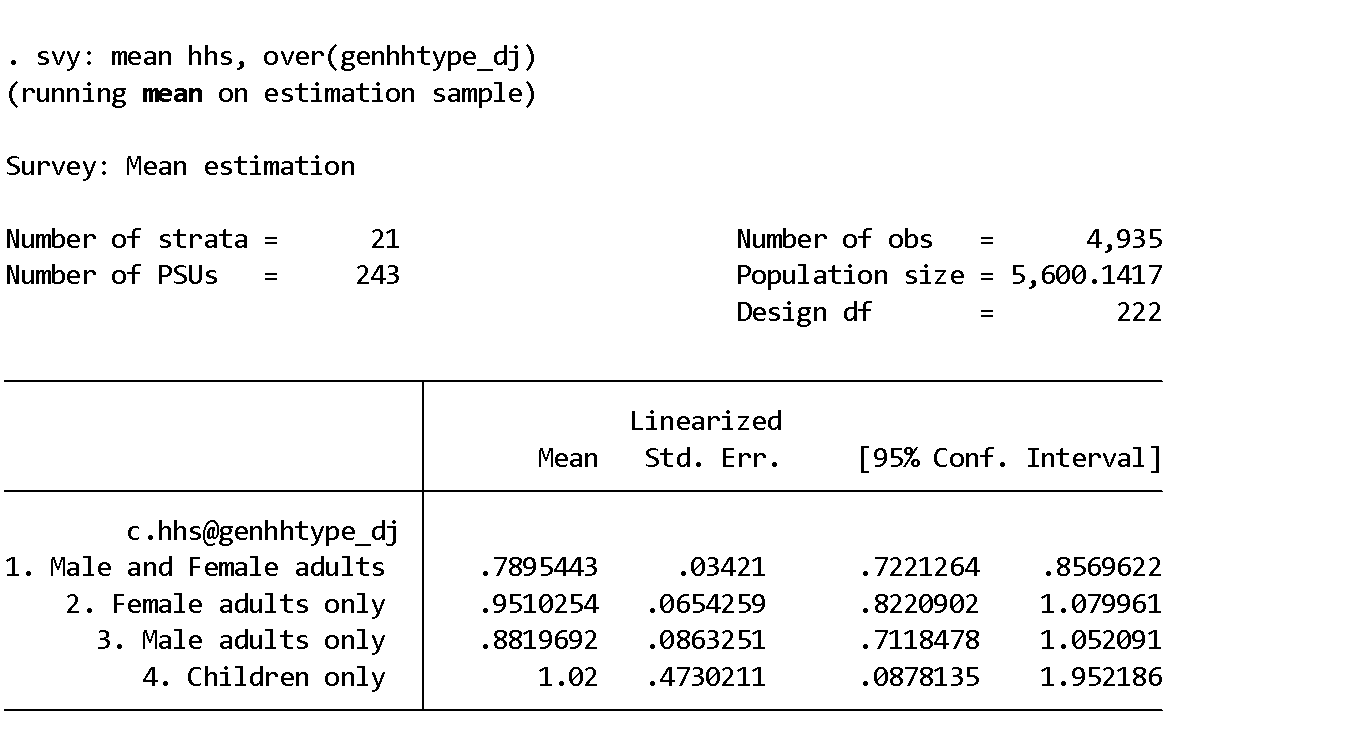
*svy: mean* ***outcome,*** *over(****disaggregate****)*

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 confidence intervals, 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 11).

##### Example 11: Indicators of means: Descriptive, by selected disaggregates output #1

****

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

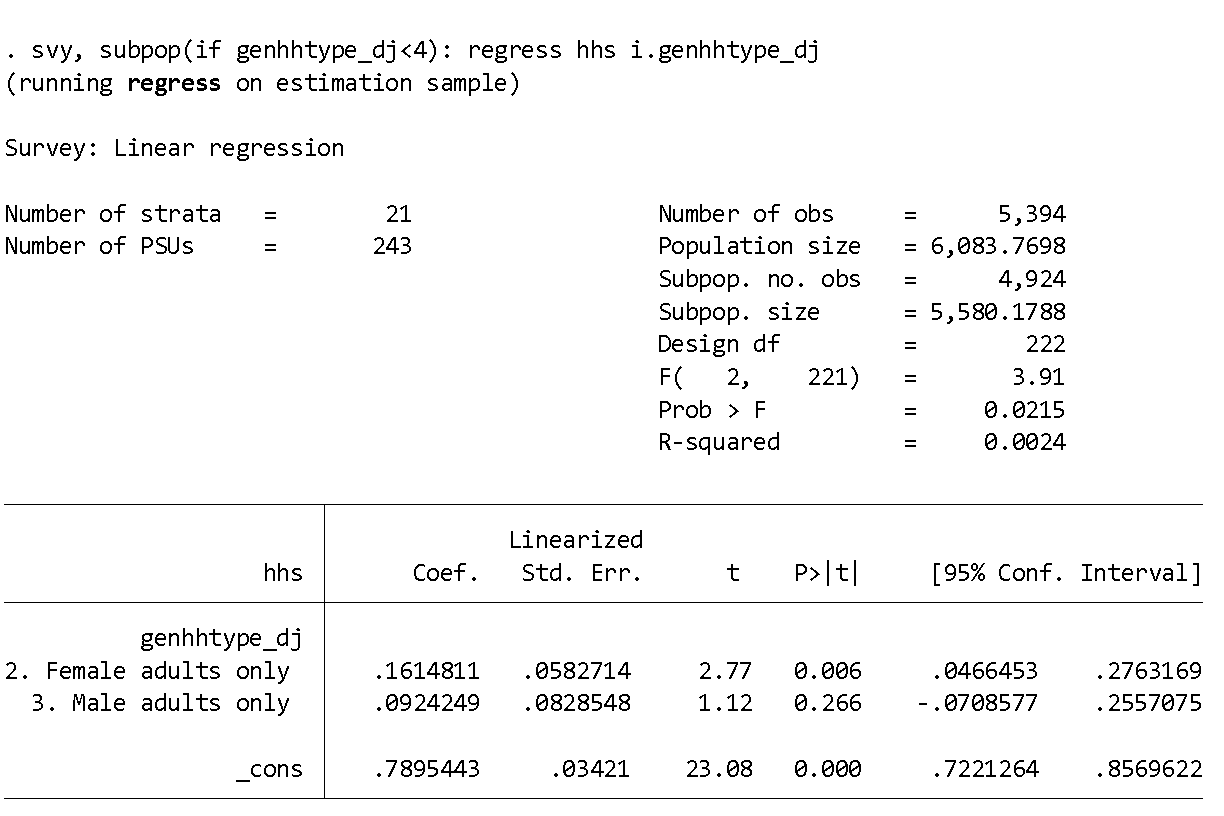
In general, a t-test is used to determine if there is a significant difference between the means of two groups, and analysis of variance (ANOVA) is used to determine if there is a significant difference among the means of three or more groups. In Stata, however, t-test and ANOVA commands that can be 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 12, and the post-estimation command ‘*lincom,*’ which is used in Example 14 to compare means over time.[[21]](#footnote-20)

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, the ‘*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 12). 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.

Note that a subpopulation of households is specified in Example 12; 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 endline (n=3) to produce valid estimates.

##### Example 12: Indicators of means: Descriptive, by selected disaggregates output #2

****

#### 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 endline)

Using the syntax above, the sample-weighted means are presented for each survey round, along with the means’ standard errors and 95 percent confidence intervals, 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 13).

##### Example 13: Indicators of means: Comparative, overall output #1



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[****outcome****1]-\_b[****outcome****0]*

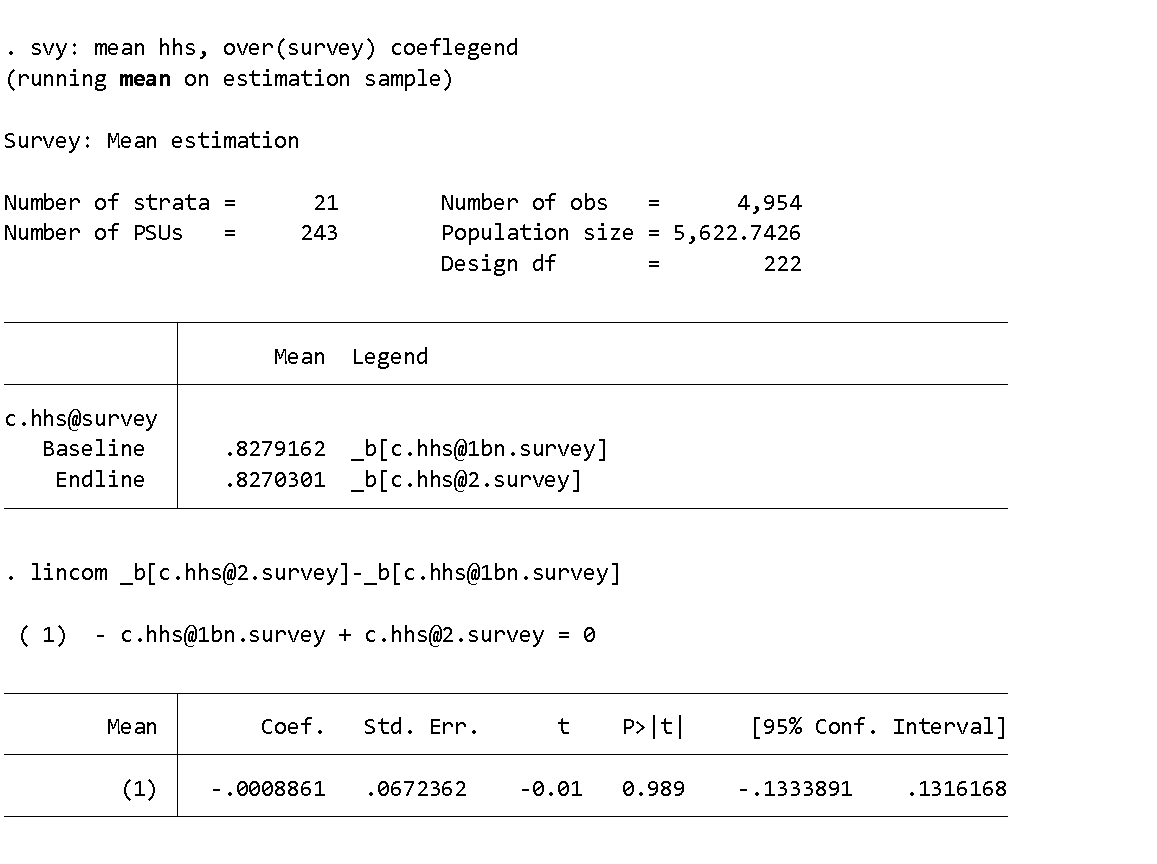
where: *\_b[****outcome****0]* is the coefficient for the indicator of means in the first survey round

*\_b[****outcome****1]* 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 14).

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 14: Indicators of means: Comparative, overall output #2



#### Comparative, by selected disaggegates

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

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

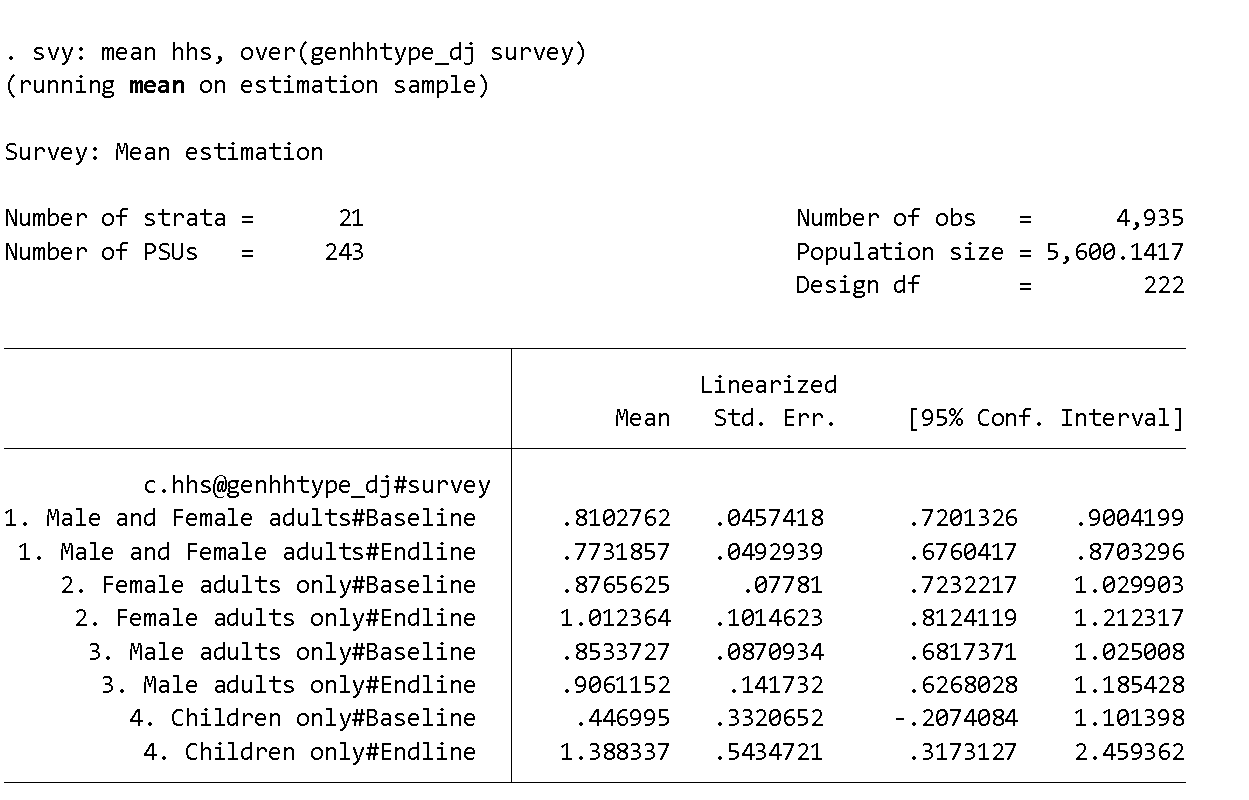
where: ***outcome*** is the variable for the indicator of means

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

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

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 confidence intervals, the total number of observations included in the calculation (i.e., the unweighted sample size across all disaggregate categories and both survey rounds), and the total population size (i.e., the weighted sample size across all disaggregate categories and both survey rounds) (see Example 15).

##### Example 15: Indicators of Means: Comparative, by selected disaggregates, output #1

****

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[****outcome****X.1]-\_b[****outcome****X.0]*

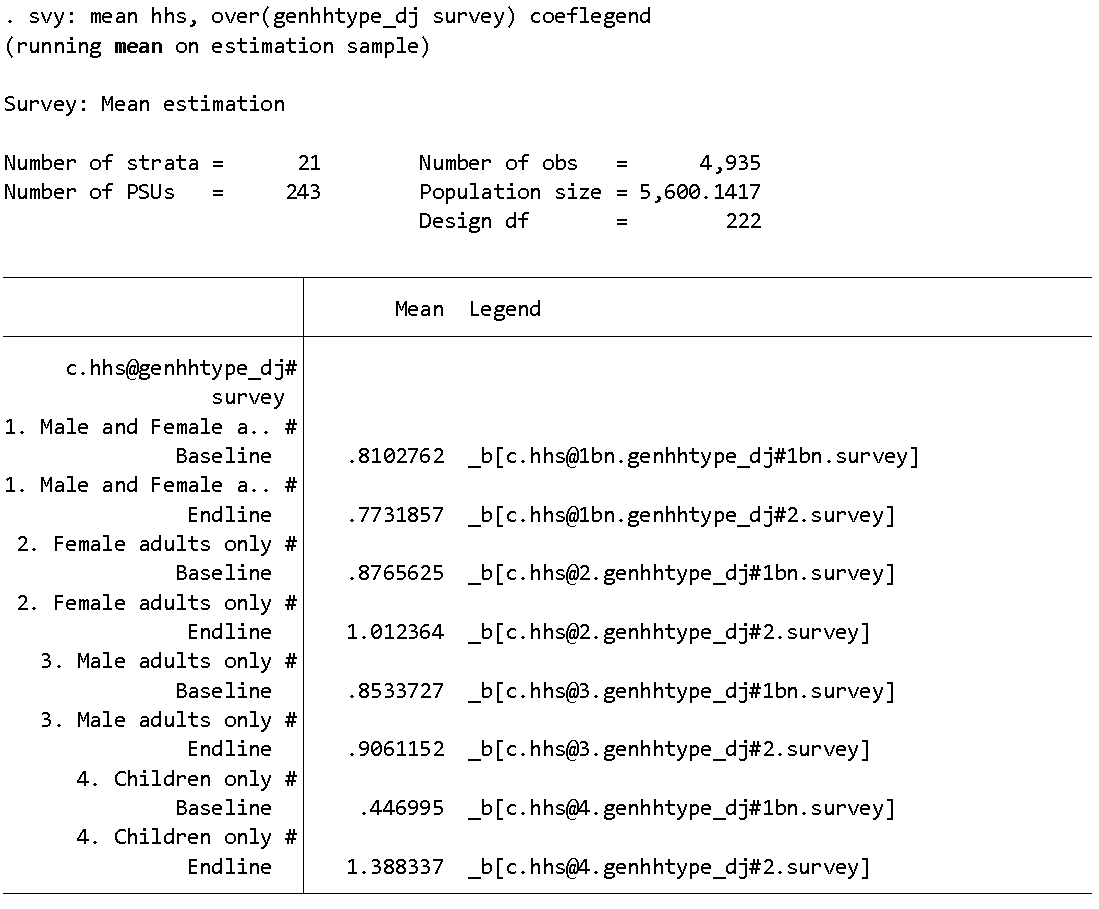
where: *\_b[****outcome****]1* is the coefficient for the indicator of means for disaggregate category X in the second survey round

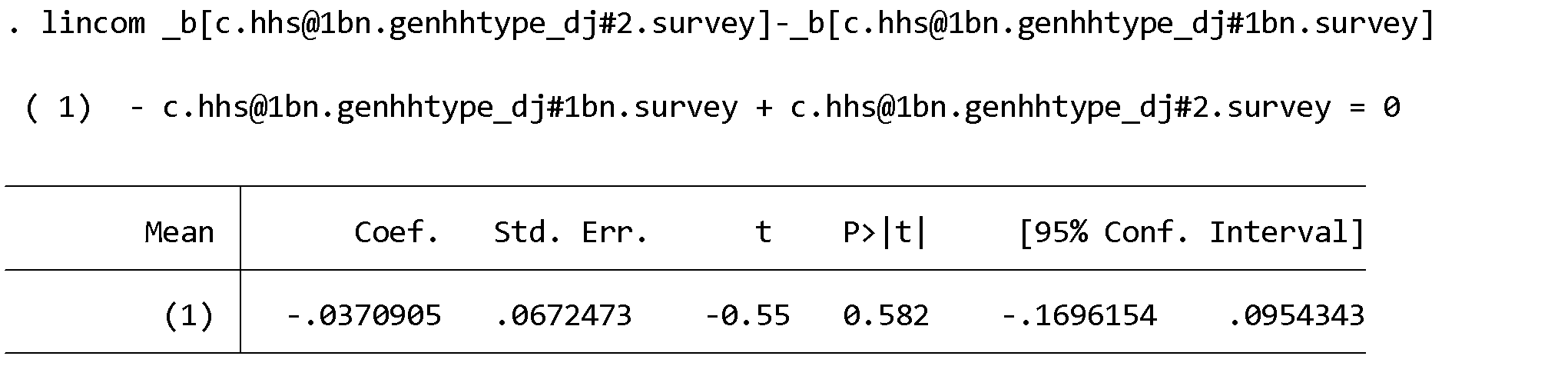
*\_b[****outcome****]0* is the coefficient for the indicator of means for disaggregate category X in the first survey round

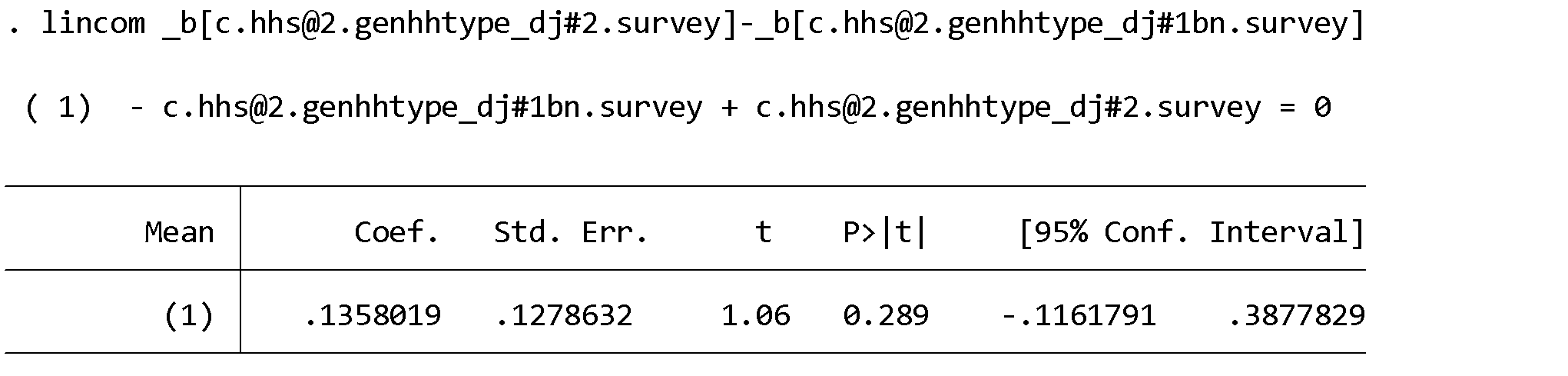
The resulting p-value indicates whether the difference the indicator estimates for each disaggregate between the two time points is statistically significant (see Example 16).

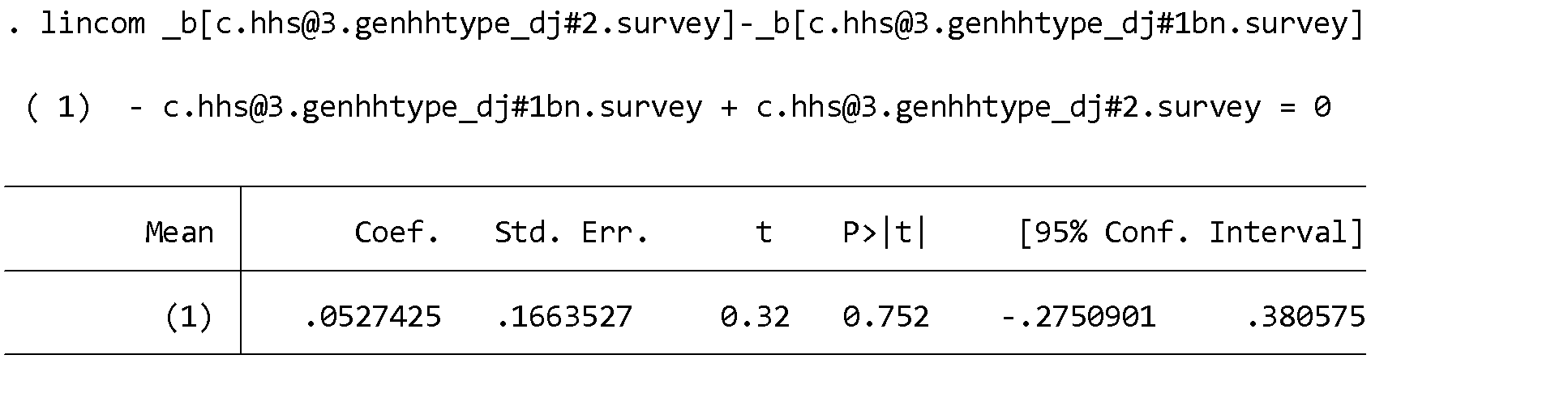
Note that to obtain the names of the coefficients to be used in the *lincom* command, the ‘*coeflegend*’ option can be added to the *svy: mean* estimation command.

##### Example 16: Indicators of means: Comparative, by selected characteristics, output #2









#### 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 17).

**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*

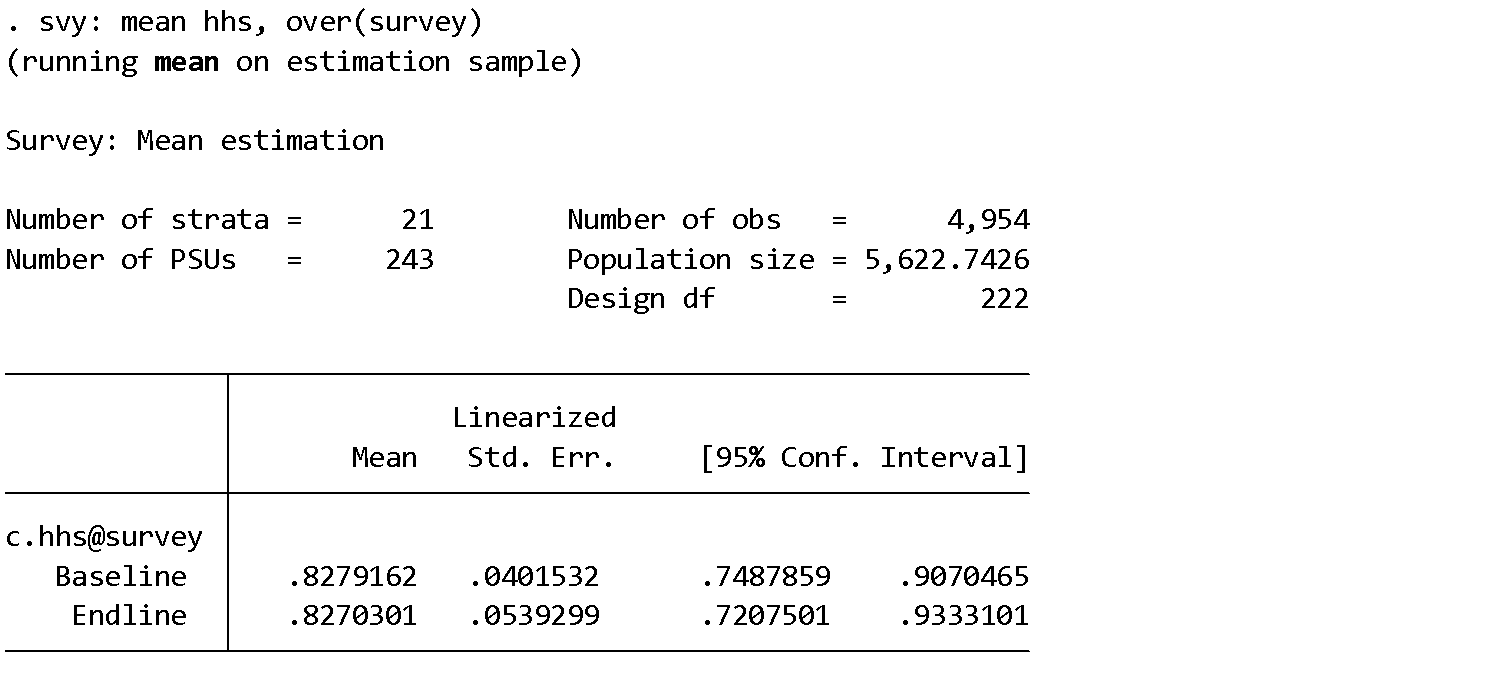
**Standard deviation:** The standard deviation for indicators of means can be obtained by running the following command after running a *svy: mean* estimation command:

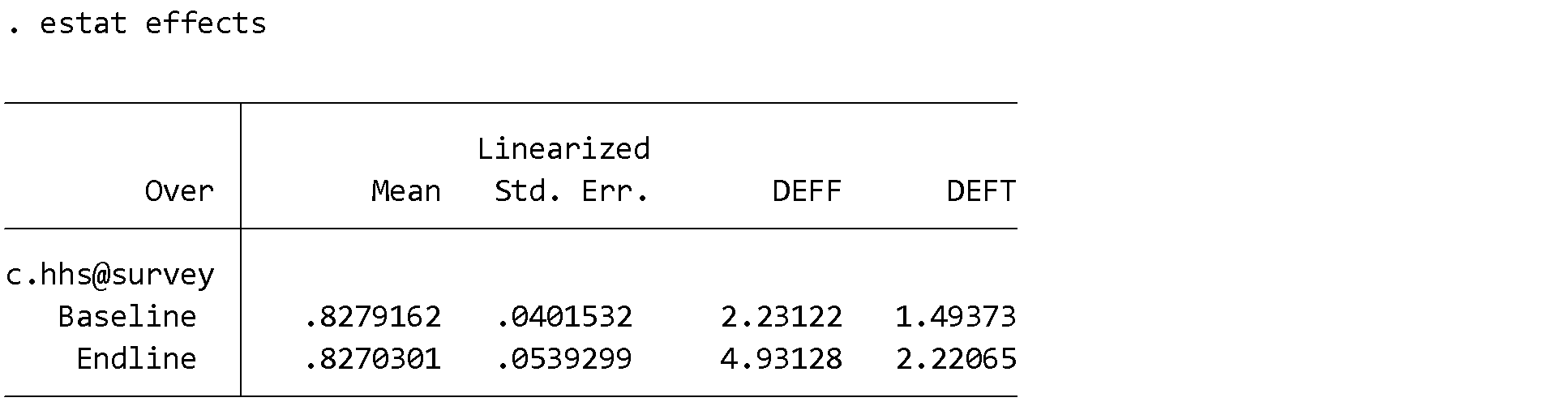
*estat sd*

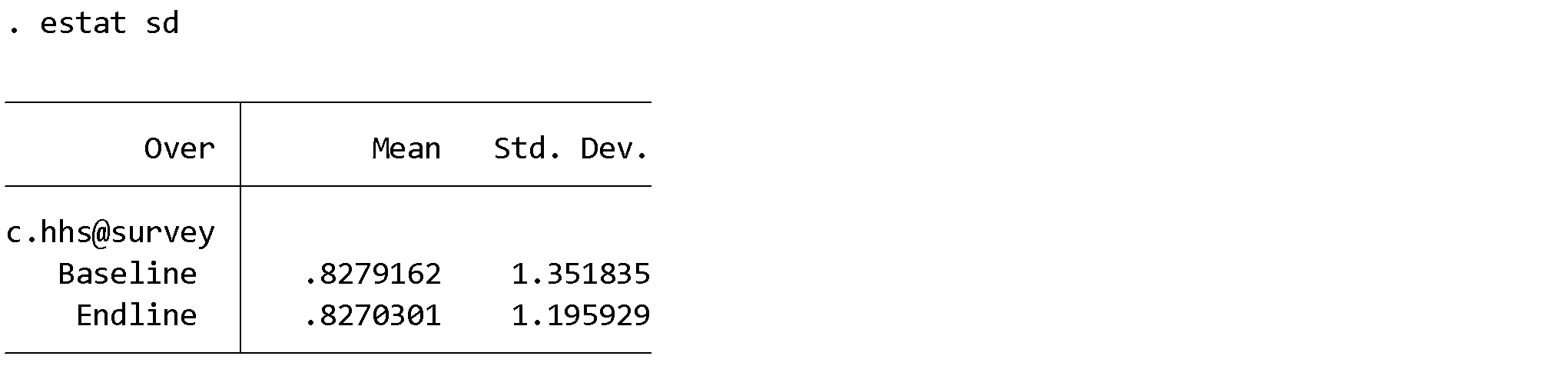
**Number of observations (unweighted n)** and **population size (weighted n):** The number of observations included in the mean calculations by disaggregate category and survey round, and the corresponding population sizes, can be obtained by running the following command after running a *svy: mean*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

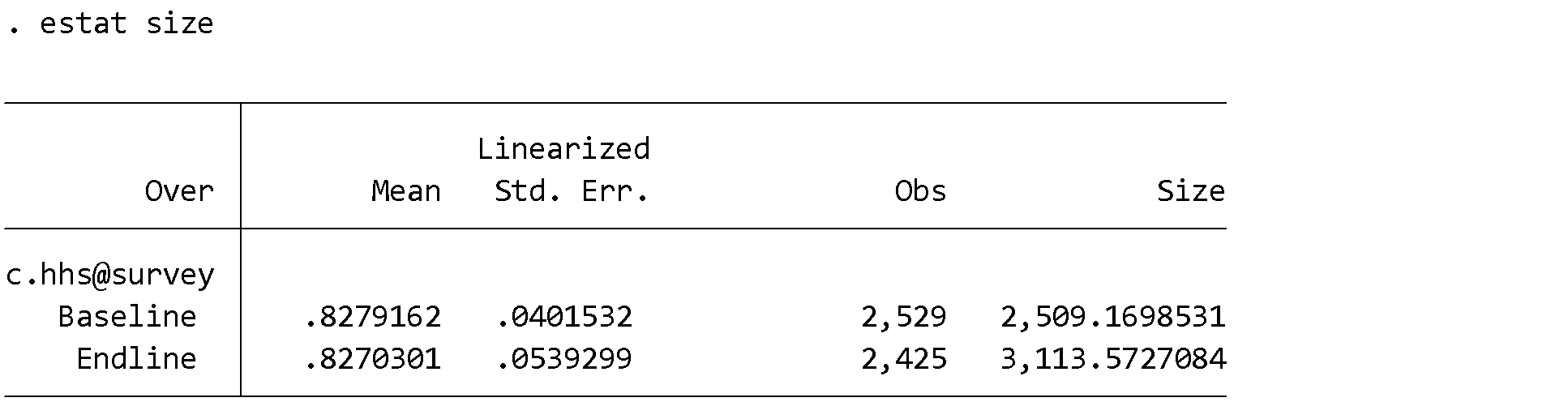
*estat size*

##### Example 17: Indicators of means, supplementary statistics output









### 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 confidence intervals 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 confidence intervals associated with the estimates, the approach is limited by the precision of the point estimates and statistics available. 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.[[22]](#footnote-21) 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 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 endline or group 1 and group 2) follow the standard normal distribution,[[23]](#footnote-22) let denote the estimate computed for baseline (or group 1), and denote the estimate computed for endline (or group 2). To test the null hypothesis, *No*: *1*=*2*, versus the alternative hypothesis, *No*: *1*≠*2*, the test statistic (Z-score) can be calculated as follows:[[24]](#footnote-23)

where SE1 is the standard error for

SE2 is the standard error for

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

where is the estimate computed for baseline (or group 1)

is the estimate computed for endline (or group 2)

is the upper confidence interval bound for baseline (or group 1)

is the upper confidence interval bound for endline (or group 2)

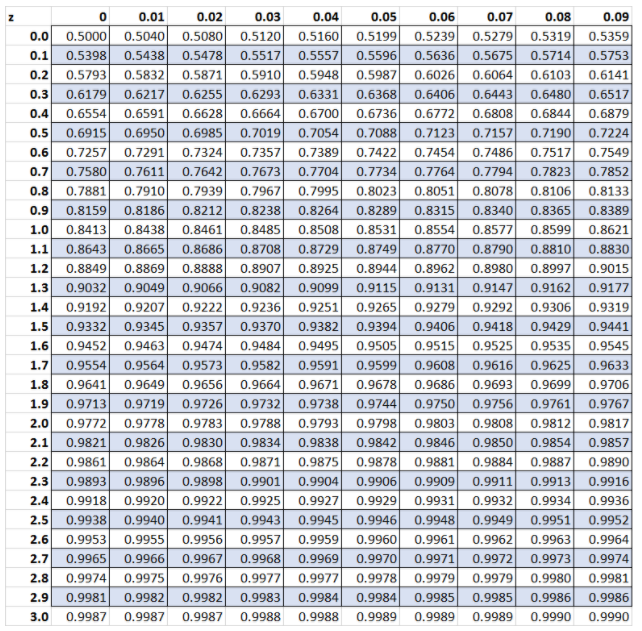
Using the standard normal tables, the associated two-tailed p-value of the Z-score can be found as 2\**P(Z>)* where is the critical Z value for the chosen α level significance (e.g., 1.96 for the 95 percent confidence level and 1.645 for the 90 percent confidence level). The p-value is then assessed 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 indicator estimates being compared are truly different.

The information from Example 13 in Section 3.2.2 that would be available in a comparative report shows that the mean household hunger score is 82.8 percent at baseline, with an SE of 0.04, a 95 percent CI lower bound of 74.9 percent, and a 95 percent CI upper bound of 90.7 percent, and is 82.7 percent at endline, with an SE of 0.05, a 95 percent CI lower bound of of 72.1 percent, and a 95 percent CI upper bound of 93.3 percent. Using the estimates and SEs, the Z-score can be calculated as follows:

If the SEs are not available, the upper bounds of the confidence interval can be used to calculate the SE. Note that due to rounding, reported SEs will not match exactly the SEs obtained using the confidence intervals. Likewise, due to rounding, SEs obtained using reported confidence intervals will not match exactly the SEs available in statistical software programs.

By plugging the Z-score into an online p-value from z-score calculator[[25]](#footnote-24) and specifying a two-tailed hypothesis and significance level of 0.05, a p-value of 0.99 is obtained, which indicates that the result is not statistically significant. Alternately, a Z table could be used. To use a Z table, first round the z-score to two decimal places (0.01 in this example). Then find the z-score to the first decimal (0.0 in this example) in the first column of the table below and then look across that row until you find the cell with the second decimal of the z-score (0.01 in this example). The value in that cell is 0.5040, which is the area under the standard normal curve to the left of the z-value (0.01). To obtain the area to the right of the z-value, subtract 0.5040 from 1; the p-value for a one-sided hypothesis test is therefore 0.4960. To obtain the p-value for a two-side hypothesis test, multiple the p-value for the one-sided hypothesis test by 2. This yields 0.992 (or 0.99 when rounded to two decimal places).

**Z Table**



Notes:

The table shows the area under the standard normal curve to the left of z.

Note that because the Z-score formula uses the absolute value of the difference between two estimates, it is not possible to obtain negative z-scores, so the z table is not presented for negative z-scores.

Source: https://www.statology.org/z-table/

## 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"

The Syntax folder has several topical sub-folders that contain the relevant core Stata template syntax files. 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 CSPro export files, the raw data files (i.e., the data files created from the CSPro export files), and all the analytic files. 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. [X] should be replaced by a term that identifies the indicator, group of indicators, or component of indictor 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 post-analysis dataset and identifier variables to enable merging the data.

Raw data files: FTF ZOI Survey [COUNTRY] [YEAR] persons data raw.dta

FTF ZOI Survey [COUNTRY] [YEAR] household data raw.dta

Analytic data files: FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic.dta

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

# Key analytic variables

In this chapter, key analytic variables used to calculate indicators are defined and their calculation described. The chapter is divided into two sections; the first describes person-level analytic variables, and the second describes household-level analytic variables.

## Person-level variables

This section describes person-level analytic variables used to calculate Feed the Future indicators. The variables in this section are all created in the individual-level data file.

### De jure household member

This variable identifies all de jure household members—that is, usual household members according to information 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.

#### 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. |
| --- | --- |
| 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”*

### 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 in the household roster. De facto household members include guests of the household, but they will not include regular 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. |
| --- | --- |
| Survey variables used | *v105b* |
| Analytic variables used | n/a |
| Analytic variables created | *hhmem\_df* |

#### Calculations

**Step 1.** Create a binary de jure 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)”*

### Sex

This variable identifies the sex of all individuals included in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. Missing sex values are not allowed in the household roster. |
| --- | --- |
| Survey variables used | *v102* |
| Analytic variables used | n/a |
| Analytic variables created | *sex* |

#### Calculations

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

*Set sex=1 if v102=1*

*Replace sex=2 if v102=2*

*Label values 1 “Male” 2 “Female”*

*Label variable “Sex of HH member”*

### 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

#### Definitions

| Treatment of missing data | Not applicable. The *age* analytic variable cannot contain missing values. |
| --- | --- |
| Survey variables used | *v104* |
| Analytic variables used | n/a |
| Analytic variables created | *age* |

#### Calculations

**Step 1.** Create a continuous age analytic variable (*age*).

*Set age=v104*

*Label variable “Age of HH member”*

### Age by 5-year categories

This variable identifies the age of all individuals in the household roster by 5-year age categories. In this variable, all household members who are 60 years or older are assigned to a 60 years of age or older (60+) category.

#### Definitions

| Treatment of missing data | Not applicable. The *age* analytic variable cannot contain missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *age* |
| Analytic variables created | *agegrp* |

#### Calculations

**Step 1.** Create a categorical variable that categorizes household members who are 0–59 years in 5-year age categories (*agegrp*) and household members who are 60 years or older in a final category using age data from the household roster.

*Set agegrp=missing*

*Replace agegrp=1 if age≥0 and age≤4*

*Replace agegrp=2 if age≥5 and age≤9*

*Replace agegrp=3 if age≥10 and age≤14*

*Replace agegrp=4 if age≥15 and age≤17*

*Replace agegrp=5 if age≥18 and age≤24*

*Replace agegrp=6 if age≥25 and age≤29*

*Replace agegrp=7 if age≥30 and age≤34*

*Replace agegrp=8 if age≥35 and age≤39*

*Replace agegrp=9 if age≥40 and age≤44*

*Replace agegrp=10 if age≥45 and age≤49*

*Replace agegrp=11 if age≥50 and age≤54*

*Replace agegrp=12 if age≥55 and age≤59*

*Replace agegrp=13 if age≥60 and age≤95*

*Label the age groups: 1 “0-4”, 2 “5-9”, 3 “10-14”, 4 “15-19”, 5 “20-24”, 6 “25-29”, 7 “30-34”, 8 “35-39”, 9 “40-44”, 10 “45-49”, 11 “50-54”, 12 “55-59”, 13 “60+”*

*Label variable “HH member age in years by age category”*

### Adults, total and by sex

These variables identify all adults—that is, individuals who are 18 years of age or older according to information in the household roster—in total and by sex.

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *sex* analytic variables cannot contain missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *age, sex, hhmem\_dj* |
| Analytic variables created | *adult, adult\_m, adult\_f, adult\_fdj, adult\_mdj* |

#### Calculations

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

*Set adult=0*

*Replace adult=1 if age≥18*

*Label values 0 “No” 1 “Yes”*

*Label variable “Adult: 18+ years”*

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

*Set adult\_m=0*

*Set adult\_m=1 if adult=1 and sex=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Male adult: 18+ years”*

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

*Set adult\_f=0*

*Set adult\_f=1 if adult=1 and sex=2*

*Label values 0 “No” 1 “Yes”*

*Label variable “Female adult: 18+ years”*

### Youth, total and by sex

These variables identify all youth—that is, individuals who are 15-29 years of age according to information in the household roster—in total and by sex.

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *sex* analytic variables cannot contain missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *age, sex* |
| Analytic variables created | *age15\_29y, age15\_29ym, age15\_29yf* |

#### 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”*

**Step 2.** Create a binary variable that identifies all adult males 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”*

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

*Set age15\_29yf=0*

*Set age15\_29yf=1 if adult=1 and sex=2*

*Label values 0 “No” 1 “Yes”*

*Label variable “Female youth: 15-29 years”*

### Women of reproductive age

This variable identifies all women of reproductive age—that is, women who are between 15 and 49 years of age according to information in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *sex* analytic variables cannot contain missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *sex, age* |
| Analytic variables created | *wra* |

#### 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 sex=2 and age≥15 and age≤49*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman of reproductive age”*

### Children under 2 years of age, total and by sex

These variables identify all children under 2 years of age according to information in the household roster—in total and by sex.

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *sex* analytic variables cannot contain missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *sex, age* |
| Analytic variables created | *cu2, cu2m, cu2f* |

#### Calculations

**Step 1.** Create a binary variable that identifies all children under 2 years of age in the household roster (*cu2*).

*Set cu2=0*

*Replace cu2=1 if age≤1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child under 2 – HH roster”*

**Step 2.** Create a binary variable that identifies all male children under 2 years of age in the household roster (*cu2m*).

*Set cu2m=0*

*Set cu2m=1 if cu2=1 and sex=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Male child under 2 – HH roster”*

**Step 3.** Create a binary variable that identifies all female children under 2 years of age in the household roster (*cu2f*).

*Set cu2f=0*

*Set cu2f=1 if cu2=1 and sex=2*

*Label values 0 “No” 1 “Yes”*

*Label variable “Female child under 2 – HH roster”*

### Children under 5 years of age, total and by sex

These variables identify all children under 5 years of age according to information in the household roster—in total and by sex.

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *sex* analytic variables cannot contain missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *sex, age* |
| Analytic variables created | *cu5, cu5m, cu5f* |

#### Calculations

**Step 1.** Create a binary variable that identifies all children under 5 years of age in the household roster (*cu5*).

*Set cu5=0*

*Replace cu5=1 if age≤4*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child under 5 – HH roster”*

**Step 2.** Create a binary variable that identifies all male children under 5 years of age in the household roster (*cu5m*).

*Set cu5m=0*

*Set cu5m=1 if cu5=1 and sex=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Male child under 5 – HH roster”*

**Step 3.** Create a binary variable that identifies all female children under 5 years of age in the household roster (*cu5f*).

*Set cu5f=0*

*Set cu5f=1 if cu5=1 and sex=2*

*Label values 0 “No” 1 “Yes”*

*Label variable “Female child under 5 – HH roster”*

### Children 5 years of age or older, total 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 information in the household roster—in total and by sex.

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *sex* analytic variables cannot contain missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *sex, age* |
| Analytic variables created | *c5\_17y, c5\_17ym, c5\_17yf* |

#### 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 of age – 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 Set c5\_17y=1 and sex=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Male child 5-17 years of age – 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 Set c5\_17y=1 and sex=2*

*Label values 0 “No” 1 “Yes”*

*Label variable “Female child 5-17 years of age – HH roster”*

### Child age in days

This variable identifies the age in days of all children under 5 years of age whose primary caregiver consented to respond to the children’s nutrition module. Child age in days information comes from the children’s nutrition module.

#### Definitions

| Treatment of missing data | Children whose day of birth is missing or unknown are assigned day 15 as their birth day. The *cage\_days* analytic variable is set to missing for children who are missing a valid month and year of birth. |
| --- | --- |
| 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 the birth day variable 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 was valid—that is, 31 or less.

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

**Step 3.** Set the birth day variable 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 only be set to 15 if the child’s caregiver consented to respond to the children’s nutrition module (that is, if *v500f=1*) and the child had a valid birth month and birth year. 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.

*Replace bday=15 if v502d≥98 and v500f=1 and (v502m≤12 or v506m≤12) and (v502y≤2020 or v506y≤2020) and bday=missing*

**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*

**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=v502y if v502y≤2020*

*Replace byear=v506y if v505=1 and v506y≤2020 and byear=missing*

**Step 6.** Concatenate the birth day, month, and year variables into a birthdate variable (*bdate*).

*Set bdate=date(bmon, bday, byear)*

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

**Step 8.** Concatenate the final interview day, month, and year variables into a date variable (*intdate*).

*Set intdate=date(ahintm, ahintd, ahinty)*

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

*Set cage\_days=intdate–bdate if bdate≠missing*

*Label variable “Child age in days”*

### Child age in months

This variable identifies the age in months of all children under 5 years of age whose primary caregiver consented to respond to the children’s nutrition module. Child age information in months comes from the children’s nutrition module.

#### Definitions

| Treatment of missing data | Use children's age in months reported by the caregiver interviewed for children who are missing a valid month and year of birth—that is, missing a value for the *cage\_days* analytic variable. Set the *cage\_mon* and *cage\_mon\_int* variables to missing for children who are also missing a caregiver-reported age in months. |
| --- | --- |
| Survey variables used | *v508* |
| Analytic variables used | *cage\_days* |
| Analytic variables created | *cage\_mon, cage\_mon\_int* |

#### Calculations

**Step 1.** Create a variable that holds the age in months (*cage\_mon*) 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 age in months, including decimal"*

**Step 2.** Create an integer version of the child’s age in months variable (*cage\_months\_int*).

*Set cage\_months\_int=integer(cage\_months)*

*Label variable cage\_months\_int "Child age in months, excluding decimal"*

**Step 3.** Compare the analytic child age in months variable (*cage\_mon\_int*) to the age in months variable collected in the survey (*v508*).

*Count if v508≠cage\_months\_int and v508≠missing*

*Count if v508=cage\_months\_int and v508≠missing*

**Step 4.** Replace the analytic child age in months variable with the survey age in months variable—that is, the child’s age as reported by the child’s caregiver in the nutrition module—if the caregiver-reported age is under 60 months.

*Replace cage\_months\_int=v508 if cage\_months\_int=missing and v508<60*

### Child age by age group

This variable identifies all children under 5 years of age whose primary caregiver consented to respond to the children’s nutrition module by specified age categories in months. Child age information in months comes from the children’s nutrition module.

#### Definitions

| Treatment of missing data | Set the child age by age group variables to missing for children who are missing a value for the *cage\_mon\_int* analytic variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *cage\_mon\_int* |
| Analytic variables created | *c0\_5m, c6\_23m, c6\_23m, c9\_23m, c0\_23m, c0\_59m* |

#### Calculations

**Step 1.** Create a binary variable to flag children 0-5 months of age (*cage0-5m*).

*Set c0\_5m=0*

*Replace c0\_5m=1 if cage\_mon\_int≥0 and cage\_mon\_int≤5*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is 0-5 months of age – Child nutrition module”*

**Step 2.** Create a binary variable to flag children 6-23 months of age (*cage6-23m*).

*Set c6\_23m=0*

*Replace c6\_23m=1 if cage\_mon\_int≥6 and cage\_mon\_int≤23*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is 6-23 months of age – Child nutrition module”*

**Step 3.** Create a binary variable to flag children 6–8 months of age (*cage6-8m*).

*Set c6\_8m=0*

*Replace c6\_8m=1 if cage\_mon\_int≥6 and cage\_mon\_int≤8*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is 6-8 months of age – Child nutrition module”*

**Step 4.** Create a binary variable to flag children 9-23 months of age (*cage9-23m*).

*Set c9\_23m=0*

*Replace c9\_23m=1 if cage\_mon\_int≥9 and cage\_mon\_int≤23*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is 9-23 months of age – Child nutrition module”*

**Step 5.** Create a binary variable to flag children 0-23 months of age (*cage0-23m*).

*Set c0\_23m=0*

*Replace c0\_23m=1 if cage\_mon\_int≥0 and cage\_mon\_int≤23*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is 0-23 months of age – Child nutrition module”*

**Step 6.** Create a binary variable to flag children 0–59 months of age (*cage0-59m*).

*Set c0\_59m=0*

*Replace c0\_59m=1 if cage\_mon\_int≥0 and cage\_mon\_int≤59*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is 0-59 months of age – Child nutrition module”*

### Household member completed primary education

This variable identifies all individuals who are 10 years of age or older according to the household roster who completed primary education. Note that the number of grades included in primary schooling varies country to country, so be sure to adjust the variable creation syntax to reflect primary schooling in the country where the ZOI Survey was conducted. The template syntax was written assuming 6 years of primary schooling.

#### Definitions

| Treatment of missing data | The *age* analytic variable cannot contain missing values. If individuals are missing information on the highest level of education completed or if the highest level completed is “don’t know,” set the variable to 0, indicating that the individuals did not complete primary education. If individuals are missing information on the highest grade of education completed or if the highest grade completed is “don't know,” and the highest level of education is primary school or lower (e.g., preschool), set the variable to 0, indicating that the individuals did not complete primary education. If individuals are missing information on the highest grade of education completed or if the highest grade completed is “don't know,” and the highest level of education is higher than primary school (e.g., secondary school), set the variable to 1, indicating that the individuals completed primary education. |
| --- | --- |
| Survey variables used | *v111a, v111b* |
| Analytic variables used | *age* |
| Analytic variables created | *edu\_prim* |

#### Calculations

**Step 1.** Create a binary variable that identifies individuals who are 10 years of age or older according to the household roster who completed primary education (*edu\_prim*). The template syntax is written such that primary schooling includes six grades.

*Set edu\_prim=missing*

*Replace edu\_prim=0 if age≥10 and age≤95*

*Replace edu\_prim=1 if age≥10 and age≤95 and ((v111a=2 or 3) or (v111a=1 and v111b=6))*

*Label values 0 “No” 1 “Yes”*

*Label variable “Individual (10+ years of age) completed primary education”*

### Household member was attending school at the time of the survey

This variable identifies all individuals who are between 5 and 24 years of age according to the household roster who were attending school at the time of the survey.

#### Definitions

| Treatment of missing data | The *age* analytic variable cannot contain missing values. If individuals are missing information about current school attendance or have a “don’t know” response, consider them to be not currently attending school and set the school attendance variable to 0. |
| --- | --- |
| Survey variables used | *v110* |
| Analytic variables used | *age* |
| Analytic variables created | *edu\_attend* |

#### Calculations

**Step 1.** Create a binary variable that identifies individuals who are between 5 and 24 years of age according to the household roster who were attending school at the time of the survey (*edu\_att*).

*Set edu\_attend=missing*

*Replace edu\_attend=1 if age≥5 and age≤24 and v110=1*

*Replace edu\_attend=0 if age≥5 and age≤24 and v110≠1*

*Label values 0 “No” 1 “Yes”*

*Label variable “HH member is currently attending school”*

### 4.1.17 Household member’s highest level of educational attainment

This variable identifies the educational attainment of each individual who is 5 years of age or older according to the household roster by category: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, (5) higher, or (6) other. Education grades and levels differ from country to country. The ZOI Survey questionnaire, therefore, must be adapted to capture the education system where the ZOI Survey is being implemented. Be sure to adjust the variable creation syntax to reflect the education system in the ZOI Survey country. The template syntax was written assuming 6 years of primary schooling and 6 years of secondary schooling.

#### Definitions

| Treatment of missing data | The *age* analytic variable cannot contain missing values. If individuals are missing information on the highest level of education completed, set the variable to missing. If individuals' highest ***level*** of education is “don't know,” and they cannot be assigned to a country-specific education category, set them to “other.” If individuals are missing information on the highest ***grade*** of education completed or if the highest grade completed is “don't know,” and they cannot be assigned to a country-specific education category, set them to the appropriate country-specific education category based on the highest level of education completed. |
| --- | --- |
| Survey variables used | *v109, v111* |
| Analytic variables used | *age* |
| Analytic variables created | *edulevel* |

#### Calculations

**Step 1.** Create a variable that indicates the educational attainment of each household member (*edulevel*).

*Set edulevel=missing*

*Replace edulevel=1 if age≥5 and age≤95 and v109=2 [No education]*

*Replace edulevel=2 if age≥5 and age≤95 and (v111a=0 or (v111a= 1 and (v111b<6 or v111b=98))) [Less than completed primary]*

*Replace edulevel=3 if age≥5 and age≤95 and (v111a=1 and v111b=6) or (v111a=2 and (v111b<6 or v111b=98))] [Completed primary]*

*Replace edulevel=4 if age≥5 and age≤95 and ((v111a=2 and v111b=6) or (v111a=3 and v111b=0)) [Completed secondary]*

*Replace edulevel=5 if age≥5 and age≤95 and v111a=3 and v111b>0 [Higher]*

**Step 2.** Create an “other” category, if needed, to capture any individuals who cannot be placed in an education category—that is, individuals whose highest education level is “don’t know.”

*Replace edulevel=6 if age≥5 and age≤95 and v111a=8*

*Label values: 1 “No education”, 2 “Less than primary”, 3 “Completed primary”, 4 “Completed secondary”, 5 “Higher”, 6 “Other”*

*Label variable “Education attained by HH member”*

### 4.1.18 Primary adult decisionmaker, by sex

These variables identify primary adult decisionmakers—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 decisionmakers is not allowed in the household roster. |
| --- | --- |
| Survey variables used | *v101a, v101b, m1\_line, vtype1,2* |
| Analytic variables used | n/a |
| Analytic variables created | *mdm, fdm, pdm* |

#### Calculations

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

*Set mdm=missing*

*Replace mdm=1 if m1\_line=1*[[26]](#footnote-25)

*Label values 1 “Yes”*

*Label variable "Primary adult male decisionmaker"*

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

*Set fdm=missing*

*Replace fdm=1 if m1\_line=2*[[27]](#footnote-26)

*Label values 1 “Yes”*

*Label variable "Primary adult female decisionmaker"*

**Step 3.** Create a variable that indicates the household member is either a primary adult male decisionmaker or a primary adult female decisionmaker (*pdm*).

*Set pdm=missing*

*Replace pdm=1 if mdm=1 or fdm=1*

*Label values 1 “Yes”*

*Label variable "Primary adult decisionmaker"*

### Primary caregiver of child under 5 years of age

This variable identifies primary caregivers of children who are under 5 years of age according to information in the household roster and children’s nutrition module.

#### Definitions

| Treatment of missing data | If the line number of a child’s caregiver is missing in the children’s nutrition module, set the *caregiver* variable to missing. |
| --- | --- |
| Survey variables used | *v500e, m1\_line, hhea, hhnum* |
| Analytic variables used | *cage\_mon\_int* |
| Analytic variables created | *caregiver* |

#### Calculations

**Step 1.** Save the person-level analytic data file with the analytic variables already created.

*Save "$analytic\FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic"*

**Step 2.** Create a variable that flags children whose primary caregiver consented to respond to the children’s nutrition module and who are between 0 and 59 months of age (caregiver).

*Set caregiver=missing*

*Replace caregiver=1 if v500e≠missing and cage\_mon\_int<60*

*Label values “1” Yes*

*Label variable "Primary caregiver of a child under 5 years of age"*

**Step 3.** Keep only records that were flagged in the previous step, and for each record, keep only the cluster number, household number, caregiver line number, and caregiver flag variables.

*Keep record if caregiver=1*

*Keep hhea hhnum v500e caregiver*

**Step 4.** Sort the data by cluster number, household number, and caregiver line number, keeping only the first record for each unique caregiver—that is, eliminate multiple records for caregivers who have more than one child between 0 and 59 months of age. Rename the caregiver line number variable (*v500e*) to be *m1\_line* so that it can be used as a key matching variable in Step 6.

*Sort hhea hhnum v500e*

*By hhea hhnum v500e: keep if \_n=1*

*Rename v500e m1\_line*

**Step 5.** Save the data as a temporary data file.

*Save "$analytic\Temp\temp\_caregiver", replace*

**Step 6.** Add the caregiver variable in the temporary caregiver data file created in the previous step to the person-level data file saved in Step 1 using cluster number, household number, and caregiver line number as the key matching variables.

*Use “$analytic\FTF ZOI Survey [COUNTRY] [YEAR] persons data analytic”*

*Add variable caregiver from $analytic\Temp\temp\_caregiver”, using hhea hhnum m1\_line as key matching variables*

### 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, according to information in the targeted VCC modules. The template syntax includes maize, millet, okra, and sheep 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 consent field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module. |
| --- | --- |
| Survey variables used\* | *v7100d, v71000d, v71100d, v75200d* |
| Analytic variables used | n/a |
| Analytic variables created | *vcc* |

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

#### Calculations

**Step 1.** Create a binary variable to flag household members who were responsible for cultivating or raising at least one targeted VCC (*vcc*).

*Set vcc=0*

*Replace vcc=1 if v7100d≠missing or v71000d≠missing or v71100d≠missing or v75200d≠missing*

*Label variable “Producer of at least one targeted VCC”*

### 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, according to information in the targeted VCC modules. The template syntax includes maize, millet, okra, and sheep 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 consent field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module. |
| --- | --- |
| Survey variables used\* | *v7100d, v71000d, v71100d, v75200d* |
| Analytic variables used | n/a |
| Analytic variables created\* | *vcc\_maize, vcc\_millet, vcc\_okra, vcc\_sheep* |

\* 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.

*Set vcc\_maize2=0*

*Replace vcc\_maize2=1 if v7100d≠missing*

*Label variable "Maize producer"*

**Step 2.** Create a variable to indicate if the household member was responsible for cultivating millet.

*Set vcc\_millet2=0*

*Replace vcc\_millet2=1 if v71000d≠missing*

*Label variable "Millet producer"*

**Step 3.** Create a variable to indicate if the household member was responsible for cultivating okra.

*Set vcc\_okra2=0*

*Replace vcc\_okra2=1 if v71100d≠missing*

*Label variable "Okra producer"*

**Step 4.** Create a variable to indicate if the household member was responsible for raising sheep.

*Set vcc\_sheep2=0*

*Replace vcc\_sheep2=1 if v75200d≠missing*

*Label variable "Sheep producer"*

### Maize farming decisionmakers

These variables identify individuals who made key decisions about cultivating maize during the year preceding the survey. The template syntax includes maize as the targeted VCC, but the syntax should be adapted for all targeted VCCs included in the ZOI Survey.[[28]](#footnote-27) Note that although this variable is defined and described in this guide, it is not included in the Stata template syntax files.

#### Definitions

| Treatment of missing data | Not applicable. The consent field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module. |
| --- | --- |
| Survey variables used\* | *v7107b, v7112d, v7123a* |
| Analytic variables used | n/a |
| Analytic variables created\* | *makesseeddec\_maize, makesfertdec\_maize, makesirrigdec\_maize* |

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

#### Calculations

**Step 1.** Create a categorical variable to capture who made decisions about what maize seed to use (*makesseeddec\_maize*).

*Set makesseeddec\_maize=missing*

*Replace makesseeddec\_maize=1 if v7107b='A'*

*Replace makesseeddec\_maize=2 if v7107b='B'*

*Replace makesseeddec\_maize=3 if v7107b='AB'*

*Replace makesseeddec\_maize=4 if v7107b='ABC' or v7107b='ABD' or v7107b='ABCD' or v7107b='AC' or v7107b='ACD' or v7107b='AD'*

*Replace makesseeddec\_maize=5 if v7107b='BC' or v7107b='BD' or v7107b='BCD' or v7107b='C' or v7107b='CD' or v7107b='D'*

*Label values 1 “Self alone” 2 “Partner/spouse alone” 3 “Self and partner/spouse together” 4 “Self and others (may also include partner)” 5 “Others”*

*Label variable “Who made decision on seed to use”*

**Step 2.** Create a categorical variable to capture who made decisions about whether to use fertilizer when cultivating maize (*makesfertdec\_maize*).

*Set makesfertdec\_maize=missing*

*Replace makesfertdec\_maize=1 if v7112d='A'*

*Replace makesfertdec\_maize=2 if v7112d='B'*

*Replace makesfertdec\_maize=3 if v7112d='AB'*

*Replace makesfertdec\_maize=4 if v7112d='ABC' or v7112d='ABD' or v7112d='ABCD' or v7112d='AC' or v7112d='ACD' or v7112d='AD'*

*Replace makesfertdec\_maize=5 if v7112d='BC' or v7112d='BD' or v7112d='BCD' or v7112d='C' or v7112d='CD' or v7112d='D'*

*Label values 1 “Self alone” 2 “Partner/spouse alone” 3 “Self and partner/spouse together” 4 “Self and others (may also include partner)” 5 “Others”*

*Label variable “Who made decision on whether to use fertilizer”*

**Step 3.** Create a categorical variable to capture who made decisions about whether to irrigate when cultivating maize (*makesfertdec\_maize*).

*Set makesirrigdec\_maize=missing*

*Replace makesirrigdec\_maize=1 if v7123a='A'*

*Replace makesirrigdec\_maize=2 if v7123a='B'*

*Replace makesirrigdec\_maize=3 if v7123a='AB'*

*Replace makesirrigdec\_maize=4 if v7123a='ABC' or v7123a='ABD' or v7123a='ABCD' or v7123a='AC' or v7123a='ACD' or v7123a='AD'*

*Replace makesirrigdec\_maize=5 if v7123a='BC' or v7123a='BD' or v7123a='BCD' or v7123a='C' or v7123a='CD' or v7123a='D'*

*Label values 1 “Self alone” 2 “Partner/spouse alone” 3 “Self and partner/spouse together” 4 “Self and others (may also include partner)” 5 “Others”*

*Label variable “Who made decision on whether to irrigate”*

## 4.2 Household-level variables

This section describes household-level analytic variables used to calculate Feed the Future indicators. The first four variables in this section are all created using information in the individual-level data file that is summed by household and then added to the household-level data file. The last two variables in this section are created directly in the household-level data file using analytic variables that are created in Sections 4.1.1 and 4.1.2 and added to the household-level data file.

### Number of de jure adults in household, total and by sex

These variables identify the number of de jure adult household members per household—in total and by sex according to information in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. The analytic variables *adult*, *adult\_m*, *adult\_f*, and *hhmem\_dj* cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *adult, adult\_m, adult\_f, hhmem\_dj* |
| Analytic variables created | *nadult\_dj, nadult\_mdj, nadult\_fdj* |

#### Calculations

**Step 1.** Create a continuous variable that counts the total number of de jure adults in each household (*nadult\_dj*) by summing the adult variable (*adult)* created in Section 4.1.6 for de jure household membersby household.

*By cluster and household: set nadult\_dj=sum(adult) if hhmem\_dj=1*

*Label variable “Number of de jure adult HH members”*

**Step 2.** Create a continuous variable that counts the total number of de jure adult males in each household (*nadult\_mdj*) by summing the adult variable (*adult\_m)* created in Section 4.1.6 for de jure household members by household.

*By cluster and household: set nadult\_mdj=sum(adult\_m) if hhmem\_dj=1*

*Label variable “Number of de jure adult male HH members”*

**Step 3.** Create a continuous variable that counts the total number of de jure adult females in each household (*nadult\_f*) by summing the adult variable (*adult\_f)* created in Section 4.1.6 for de jure household members by household.

*By cluster and household: nadult\_fdj=sum (adult\_f) if hhmem\_dj=1*

*Label variable “Number of de jure adult female HH members”*

### Number of de facto adults in household, total and by sex

These variables identify the number of de facto adult household members per household—in total and by sex according to information in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. The analytic variables *adult*, *adult\_m*, *adult\_f* and *hhmem\_df* cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *adult, adult\_m, adult\_f, hhmem\_df* |
| Analytic variables created | *nadult\_df, nadult\_mdf, nadult\_fdf* |

#### Calculations

**Step 1.** Create a continuous variable that counts the total number of de facto adults in each household (*nadult\_df*) by summing the adult variable (*adult)* created in Section 4.1.6 for de facto household membersby household.

*By cluster and household: set nadult\_df=sum(adult) if hhmem\_df=1*

*Label variable “Number of de facto adult HH members”*

**Step 2.** Create a continuous variable that counts the total number of de facto adult males in each household (*nadult\_mdf*) by summing the adult variable (*adult\_m)* created in Section 4.1.6 for de facto household members by household.

*By cluster and household: set nadult\_mdf=sum(adult\_m) if hhmem\_df=1*

*Label variable “Number of de facto adult male HH members”*

**Step 3.** Create a continuous variable that counts the total number of de facto adult females in each household (*nadult\_f*) by summing the adult variable (*adult\_f)* created in Section 4.1.6 for de facto household members by household.

*By cluster and household: nadult\_fdf=sum (adult\_f) if hhmem\_df=1*

*Label variable “Number of de facto adult female HH members”*

### Number of de jure women of reproductive age in household

This variable identifies the number of de jure women of reproductive age in each household according to information in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. The analytic variables *wra* and *hhmem\_dj* 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 de jure women of reproductive age in each household (*nwra\_dj*) by summing the woman of reproductive age variable (*wra)* created in Section 4.1.8 for de jure household membersby household.

*By cluster and household: set nwra\_dj=sum(wra) if hhmem\_dj=1*

*Label variable “Number of de jure WRA in HH”*

### Number of de jure children under 2 years of age in household

This variable identifies the de jure number of children under 2 years of age in each household according to information in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. The analytic variables *cu2* and *hhmem\_dj* cannot have missing values |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *cu2, hhmem\_dj* |
| Analytic variables created | *ncu2\_dj* |

#### Calculations

**Step 1.** Create a continuous variable that counts the total number of de jure children under 2 years of age in each household (*ncu2\_dj*) by summing the children under 2 years of age variable (*cu2)* created in Section 4.1.9 for de jure household membersby household.

*By cluster and household: set ncu2\_dj=sum(cu2) if hhmem\_dj=1*

*Label variable “Number of de jure children under 2 years in HH”*

### Number of de jure children under 5 years of age in household

This variable identifies the number of de jure children under 5 years of age in each household according to information in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. The analytic variables *cu5* and *hhmem\_dj* cannot have missing values |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *cu5, hhmem\_dj* |
| Analytic variables created | *ncu5\_dj* |

#### Calculations

**Step 1.** Create a continuous variable that counts the total number of de jure children under 5 years of age in each household (*ncu5\_dj*) by summing the children under 5 years of age variable (*cu5)* created in Section 4.1.9 for de jure household membersby household.

*By cluster and household: set ncu5\_dj=sum(cu5) if hhmem\_dj=1*

*Label variable “Number of de jure children under 5 years in HH”*

### Number of de jure children 5-17 years of age in household

This variable identifies the number of de jure children who are between 5 and 17 years of age in each household according to information in the household roster.

#### 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 de jure children 5-17 years of age, inclusive, in each household (*nc5\_17y\_dj*) by summing the children 5-17 years of age variable (*c5\_17y)* created in Section 4.1.11 for de jure household membersby household.

*By cluster and household: set nc5\_17y\_dj=sum(c5\_17y) if hhmem\_dj=1*

*Label variable “Number of de jure children 5-17 years in HH”*

### Number of de jure youth in household

This variable identifies the number of de jure youth (15-29 years of age) in each household according to information in the household roster.

#### Definitions

| Treatment of missing data | Not applicable. The analytic variables *c15\_29y* and *hhmem\_dj* cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *c15\_29y, hhmem\_dj* |
| Analytic variables created | *nc15\_29y\_dj* |

#### Calculations

**Step 1.** Create a continuous variable that counts the total number of de jure youth in each household (*nc15\_29y\_dj*) by summing the youth variable (*c15\_29y)* created in Section 4.1.7 for de jure household membersby household.

*By cluster and household: set nc5\_17y\_dj=sum(c5\_17y) if hhmem\_dj=1*

*Label variable “Number of de jure children 5-17 years in HH”*

### Number of de jure producers of targeted VCCs in household

This variable identifies the number of de jure producers of targeted VCCs in each household according to information in the household roster and Module 2, *Dwelling Characteristics*. The template syntax includes maize, millet, okra, and sheep 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 | Missing de jure values are not allowed in the household roster, and the *vcc* analytic variable cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used\* | *vcc, vcc\_maize2, vcc\_millet2, vcc\_okra2, vcc\_sheep2, hhmem\_dj* |
| Analytic variables created\* | *nvcc\_dj, nvcc\_maize\_dj, nvcc\_millet\_dj, nvcc\_okra\_dj, nvcc\_sheep\_dj* |

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

#### Calculations

**Step 1.** Create a continuous variable that counts the total number of de jure producers of any targeted VCC included in the ZOI Survey (*nvccc\_dj*) by summing the producer variable (*vcc)* created in Section 4.1.20 for de jure household membersby household.

*By cluster and household: set nvccc\_dj=sum(vcc) if hhmem\_dj=1*

*Label variable “Number of de jure producers of any targeted VCC in HH”*

**Step 2.** Create a continuous variable that counts the total number of de jure producers of maize (*nvccc\_maize\_dj*) by summing the maize producer variable (*vcc\_maize)* created in Section 4.1.21 for de jure household membersby household.

*By cluster and household: set nvccc\_maize\_dj=sum(vcc\_maize) if hhmem\_dj=1*

*Label variable “Number of de jure maize producers in HH”*

**Step 3.** Create a continuous variable that counts the total number of de jure producers of millet (*nvccc\_millet\_dj*) by summing the millet producer variable (*vcc\_millet)* created in Section 4.1.21 for de jure household membersby household.

*By cluster and household: set nvccc\_millet\_dj=sum(vcc\_millet) if hhmem\_dj=1*

*Label variable “Number of de jure millet producers in HH”*

**Step 4.** Create a continuous variable that counts the total number of de jure producers of okra (*nvccc\_okra\_dj*) by summing the okra producer variable (*vcc\_okra)* created in Section 4.1.21 for de jure household membersby household.

*By cluster and household: set nvccc\_okra\_dj=sum(vcc\_okra) if hhmem\_dj=1*

*Label variable “Number of de jure okra producers in HH”*

**Step 5.** Create a continuous variable that counts the total number of de jure producers of sheep (*nvccc\_sheep\_dj*) by summing the sheep producer variable (*vcc\_sheep)* created in Section 4.1.21 for de jure household membersby household.

*By cluster and household: set nvccc\_sheep\_dj=sum(vcc\_sheep) if hhmem\_dj=1*

*Label variable “Number of de jure sheep producers in HH”*

### 4.2.9 Household size—de jure

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 and share cooking arrangements.[[29]](#footnote-28) The variable counts all de jure individuals listed in the household roster in Module 1.

#### Definitions

| Treatment of missing data | Not applicable. The *hhmem\_dj* analytic variable cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *hhmem\_dj* |
| Analytic variables created | *hhsize\_dj* |

#### Calculations

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

*By cluster and household: set hhsize\_dj=sum (hhmem\_dj)*

*Label variable “Number of de jure HH members”*

### Household size—de facto

This variable counts the number of de facto household members in each household. De facto household members are those individuals who reported sleeping in the household the night preceding the survey. The variable counts all de facto individuals listed in the household roster in Module 1.

#### Definitions

| Treatment of missing data | Not applicable. The *hhmem\_df* analytic variable cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *hhmem\_df* |
| Analytic variables created | *hhsize\_df* |

#### Calculations

**Step 1.** Create a variable that contains the number of de facto household members in each household (*hhsize\_df*)by summing the de facto household member variable created in Section 4.1.2.

*By cluster and household: set hhsize\_df=sum (hhmem\_df)*

*Label variable “Number of de facto HH members”*

### 4.2.11 Household size category—de jure

This variable identifies the household size category (small, medium, or large) into which the household falls, based on the number of de jure household members.

#### Definitions

| Treatment of missing data | Not applicable. The *hhmem\_dj* analytic variable cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *hhsize\_dj* |
| Analytic variables created | *hhsizegrp\_dj* |

#### Calculations

**Step 1.** In the household-level data file, create a categorical variable for household size (*hhsizegrp\_dj*) based on de jure household members.

*Set hhsizegrp\_dj=missing*

*Replace hhsizegrp\_dj=1 if hhsize\_dj≥1 and hhsize\_dj≤5*

*Replace hhsizegrp\_dj=2 if hhsize\_dj≥6 and hhsize\_dj≤10*

*Replace hhsizegrp\_dj=3 if hhsize\_dj≥11 and hhsize\_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 – de jure members (categorical)”*

### Household size category—de facto

This variable identifies the household size category (small, medium, or large) into which the household falls, based on the number of de facto household members.

#### Definitions

| Treatment of missing data | Not applicable. The *hhmem\_df* analytic variable cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *hhsize\_df* |
| Analytic variables created | *hhsizegrp\_df* |

#### Calculations

**Step 1.** In the household-level data file, create a categorical variable for household size (*hhsizegrp\_df*) based on de facto household members.

*Set hhsizegrp=missing*

*Replace hhsizegrp\_df=1 if hhsize\_df≥1 and hhsize\_df≤5*

*Replace hhsizegrp\_df=2 if hhsize\_df≥6 and hhsize\_df≤10*

*Replace hhsizegrp\_df=3 if hhsize\_df≥11 and hhsize\_df≠missing*

*Label values 1 “1-5 de facto members”, 2 “6-10 de facto members”, 3 “11+ de facto members”*

*Label variable “Household size – de facto members (categorical)”*

# Disaggregate variables

In this chapter, variables used as disaggregates for Feed the Future indicators are defined and their calculation described. The chapter is divided into five sections: household, primary adult decisionmakers, women of reproductive age, children under 5, and producers. Some of the disaggregates are required for Feed the Future reporting—that is, they must be entered into the FTFMS, and others are used only in the Feed the Future baseline or endline/baseline report tables.

## Household

This section describes household-level disaggregates that are used to report Feed the Future indicators. All variables are created in the household-level data file, although the last disaggregate variable, VCC production, requires first creating variables in the individual-level data file and then adding them to the household-level data file.

### 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, adult male-only, or child-only—based on the sex of de jure adult household members.

#### Definitions

| Treatment of missing data | Not applicable. The *nadult\_mdj* and *nadult\_fdj* analytic variables cannot be missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *nadult\_mdj, nadult\_fdj* |
| Analytic variables created | *genhhtype\_dj* |

#### 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*).

*Set 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 n\_adult\_fdj=0*

*Replace genhhtype\_dj=4 if nadult\_fdj=0 and nadult\_mdj=0*

*Label values: 1 “De jure male and female adults” 2 “De jure female adults only, no de jure male adults” 3 “De jure male adults only, no de jure female adults” 4 “De jure children only, no de jure adults”*

*Label variable “Gendered household type – de jure household members”*

### Gendered household type—de facto household members

This variable identifies each household’s gendered household type—that is, adult male and female, adult female-only, adult male-only, or child-only—based on the sex of de facto household members.

#### Definitions

| Treatment of missing data | Not applicable. The *nadult\_mdf* and *nadult\_fdf* analytic variables cannot be missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *nadult\_mdf, nadult\_fdf* |
| Analytic variables created | *genhhtype\_df* |

#### Calculations

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

*Set genhhtype\_df=1 if nadult\_fdf≥1 and nadult\_mdf≥1*

*Replace genhhtype\_df=2 if nadult\_fdf≥1 and nadult\_mdf=0*

*Replace genhhtype\_df=3 if nadult\_mdf≥1 and n\_adult\_mdf=0*

*Replace genhhtype\_df=4 if nadult\_fdf=0 and nadult\_mdf=0*

*Label values: 1 “De facto male and female adults” 2 “De facto female adults only, no de facto male adults” 3 “De facto male adults only, no de facto female adults” 4 “De facto children only, no de facto adults”*

*Label variable “Gendered household type – de facto household members”*

### Household education—de jure household members

This variable identifies the highest completed level of education category into which the household falls based on de jure household members’ education. The categories are: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, or (5) higher, but these may be adapted based on the survey questionnaire and education system in the ZOI Survey country. See the *edulevel* variable in Section 4.1.17; any adaptations made to that person-level variable are also made to this disaggregate variable.

#### Definitions

| Treatment of missing data | The *hhmem\_df* variable cannot have missing values. If the highest level of educational attainment for all de jure household members is missing, or if the educational attainment for all de jure household members is “other,” set the variable to missing. If the highest level of educational attainment is missing or “other” for a subset of de jure household members, determine the household’s highest level of educational attainment using the available information for de jure household members. Please see Section 4.1.17 for information on how missing values are handled when creating the *edulevel* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *edulevel, hhmem\_dj* |
| Analytic variables created | *edulevel\_dj, edulevel\_hh\_dj* |

#### Calculations

**Step 1.** In the person-level analytic data file, create a variable that indicates the highest level of educational attainment of each de jure household member (*edulevel\_dj*).

*Set edulevel\_dj=missing*

*Replace edulevel\_dj=edulevel if hhmem\_dj=1*

**Step 2.** Create a variable that contains the highest educational attainment of any de jure household member in each household *(edulevel\_hh\_dj)* by determining the highest educational attainment category among de jure household members, not including the “other” category, and then add the household-level variable to the household-level data file.

*By cluster and household: set edulevel\_hh\_dj=maximum(edulevel\_dj) if edulevel<6*

*Merge edulevel\_hh\_dj into “$analytic/FTF ZOI Survey [COUNTRY] [YEAR] household data analytic”*

*Label values: 1 “No education”, 2 “Less than primary”, 3 “Completed primary”, 4 “Completed secondary”, 5 “Higher”*

*Label variable “Highest education attainted by any de jure HH member”*

### Household education—de facto household members

This variable identifies the highest completed level of education category into which the household falls based on de facto household members’ education. The categories are: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, or (5) higher, but they may be adapted based on the survey questionnaire and education system in the ZOI Survey country. See the *edulevel* variable in the first section of Chapter 4; any adaptations made to that person-level variable are also made to this disaggregate variable.

#### Definitions

| Treatment of missing data | The hhmem\_df variable cannot have missing values. If the highest level of educational attainment for all de facto household members is missing, or if the educational attainment for all de facto household members is “other,” set the variable to missing. If the highest level of educational attainment is missing or “other” for a subset of de facto household members, determine the household’s highest level of educational attainment using the available information for de facto household members. Please see Section 4.1.17 for information on how missing values are handled when creating the *edulevel* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *edulevel, hhmem\_df* |
| Analytic variables created | *edulevel\_df, edulevel\_hh\_df* |

#### Calculations

**Step 1.** In the person-level analytic data file, create a variable that indicates the highest level of educational attainment of each de facto household member (*edulevel\_df*).

*Set edulevel\_df=missing*

*Replace edulevel\_df=edulevel if hhmem\_df=1*

**Step 2.** Create a variable that contains the highest educational attainment of any de facto household member in each household *(edulevel\_hh\_df)* by determining the highest educational attainment category among de facto household members, not including the “other” category, and then add the household-level variable to the household-level data file.

*By cluster and household: set edulevel\_hh\_df=maximum(edulevel\_df) if edulevel<6*

*Merge edulevel\_hh\_df into “$analytic/FTF ZOI Survey [COUNTRY] [YEAR] household data analytic”*

*Label values: 1 “No education”, 2 “Less than primary”, 3 “Completed primary”, 4 “Completed secondary”, 5 “Higher”*

*Label variable “Highest level of education by any de facto HH member”*

### Wealth quintile

The wealth quintile variable (a*wiquint*) is created as part of the comparative wealth index indicator. Please see Chapter 10 for step-by-step instructions on how to create the variable.

### Poverty status

The poverty status variables (*poor125* and *poor190*) are created to calculate the prevalence of poverty indicators. The variable *poor125* is the poverty status variable used to calculate the prevalence of poverty using the USD $1.25 (2005 PPP) poverty threshold and should be used as the disaggregate for Feed the Future phase one analyses. The variable *poor190* is the variable used to calculate the prevalence of poverty using the USD $1.90 (2011 PPP) poverty threshold and should be used as the disaggregate for Feed the Future phase two analyses. Please see Chapter 9 for step-by-step instructions on how to create the variables.[[30]](#footnote-29)

### Food insecurity

The food insecurity variable (*fies*) is created as part of the food insecurity inexperience scale (FIES) indicator. Please see Chapter 14 for step-by-step instructions on how to create the variable.

### Shock exposure index

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 (ARSSI) indicator. Please see Chapter 11 for step-by-step instructions on how to create the SEI variable, which is the starting point to create the shock severity variable.

#### Definitions

| Treatment of missing data | If a household has any missing or refused responses for the questions used to calculate the SEI, set the *shock\_sev* variable to missing. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *sei* |
| Analytic variables created | *shock\_sev* |

#### Calculations

**Step 1.** Using the SEI variable (*sei*) created in Chapter 11, 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 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= 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= 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”*

### Farm size

The farmsize disaggregate (*farmsize*) is a binary variable that classifies households as smallholders or non-smallholders based on the total amount of agricultural land that they own or have rights over.

#### Definitions

| Treatment of missing data | Set the variable to missing for households that have missing or “don’t know” values for both the amount of land and the amount of land used, for households that do not own land and have a missing or “don’t know” value for the amount of land that they use, and for households that do not use land and have a missing or “don’t know” value for the amount of land that they own. |
| --- | --- |
| Survey variables used | *v240a, v240b, v241a, and v241b* |
| Analytic variables used | n/a |
| Analytic variables created | *agland\_own, agland\_use, farmsize* |

#### Calculations

**Step 1.** Create a continuous variable that indicates the amount of agricultural land each household owns *(agland\_own)*.

*Set agland\_own=missing*

*Replace agland\_own=0 if v240a=2*

*Replace agland\_own=v240b if v240b≤95.0 and land\_own=missing*

*Label variable “Ag land HH owns (hectares)”*

**Step 2.** Create a continuous variable that indicates the amount of agricultural land each household has the right to use but does not own (*agland\_use*).

*Set agland\_use=missing*

*Replace agland\_use=0 if v241a=2*

*Replace agland\_use=v241b if v241b≤95.0 and land\_use=missing*

*Label variable “Ag land HH has rights to use (hectares)”*

**Step 3.** Calculate the total amount of agricultural land available to a household by summing the amount of land owned and the amount of land the household has use rights over (*agland\_tot*). Set the variable to missing for households that are missing values for both analytic variables that are being summed, for households that do not own land and are missing the amount of land that they use, and for households that do not use land and are missing the amount of land that they own.

*Set agland\_tot=missing*

*Replace agland\_tot=land\_own+land\_use*

*Replace agland\_tot=missing if land\_own=missing and land\_use=missing*

*Replace agland\_tot=missing if land\_own=missing and land\_use=0*

*Replace agland\_tot=missing if land\_own=0 and land\_use=missing*

*Label variable “Ag land available to HH (own+use) (hectares)”*

**Step 4.** Create a binary variable that flags smallholder and non-smallholder households.

*Set farmsize=0 if land\_tot≤5 and land\_tot≠missing*

*Replace farmsize=1 if land\_tot>5 and land\_tot≠missing*

*Label values 0 “Smallholder” 1 “Non-smallholder”*

*Label variable “Farm size”*

### Agricultural land ownership, by category

The agricultural land ownership variable (*agland\_ownc*) is a categorical variable that indicates the amount of agricultural land a household owns, according to information in Module 2, *Dwelling Characteristics.*

#### Definitions

| Treatment of missing data | If a household is missing information on whether it owns agricultural land and the amount of agricultural land owned, or if the amount of agricultural land owned is reported as “don’t know,” set the variable to missing. |
| --- | --- |
| Survey variables used | *240a, 240b* |
| Analytic variables used | n/a |
| Analytic variables created | *agland\_ownc* |

#### Calculations

**Step 1.** Create a categorical variable that indicates the amount of agricultural land each household owns (*agland\_ownc*).

*Set agland\_ownc=missing*

*Replace agland\_ownc=1 if v240a=2*

*Replace agland\_ownc=2 if v240b<5*

*Replace agland\_ownc=3 if v240b≥5 and v240b<10*

*Replace agland\_ownc=4 if v240b≥10 and v240b≤95*

*Label variable "Hectares of agricultural land household owns (categorical)"*

*Label values 1 "0 ha" 2 "<5 ha" 3 "5 to <10 ha" 4 "10+ ha"*

### Livestock ownership

These variables indicate if the household owns at least one of each type of animal, according to information in Module 2, *Dwelling Characteristics.* These variables are not true disaggregate variables because the household could have owned more than one type of animal. These variables include poultry but exclude fish. They should be adjusted to align with the animals included in the country-customized ZOI Survey questionnaire.

#### Definitions

| Treatment of missing data | If a household is missing information on the number of a certain type of animal owned, consider the household to not own any and set the variable to 0 (no). If a household reported the number of a certain type of animal owned as “don’t know,” consider the household to own at least one of that type of animal and set the variable to 1 (yes). |
| --- | --- |
| Survey variables used | *v225, v226a, v226b, v226c, v226d, v226e, v226f* |
| Analytic variables used | n/a |
| Analytic variables created | *own\_cow, own\_cattle, own\_horse, own\_goat, own\_sheep, own\_poultry, own\_none* |

#### Calculations

**Step 1.** Create a variable that indicates whether the household owns any cows or bulls (*own\_cow*).

*Set own\_cow=missing*

*Replace own\_cow=0 if v225=2 or v226a=0 or v226a=missing*

*Replace own\_cow=1 if v226a>0 and v226≤98*

*Label values 0 “None” 1 “1+”*

*Label variable "Household owns at least 1 cow or bull"*

**Step 2.** Create a variable that indicates whether the household owns any other cattle (*own\_cattle*).

*Set own\_cattle=missing*

*Replace own\_cattle=0 if v225=2 or v226b=0 or v226b=missing*

*Replace own\_cattle=1 if v226b>0 & v226b≤98*

*Label variable "Household owns at least 1 other cattle"*

**Step 3.** Create a variable that indicates whether the household owns any horses, donkeys, or mules (*own\_horse*).

*Set own\_horse=missing*

*Replace own\_horse=0 if v225=2 or v226c=0 or v226c=missing*

*Replace own\_horse=1 if v226c>0 & v226c≤98*

*Label values 0 “None” 1 “1+”*

*Label variable "Household owns at least 1 horse, donkey or mule"*

**Step 4.** Create a variable that indicates whether the household owns any goats (*own\_goat*).

*Set own\_goat=missing*

*replace own\_goat=0 if v225=2 or v226d=0 or v226d=missing*

*replace own\_goat=1 if v226d>0 & v226d≤98*

*Label values 0 “None” 1 “1+”*

*Label variable "Household owns at least 1 goat"*

**Step 5.** Create a variable that indicates whether the household owns any sheep (*own\_sheep*).

*Set own\_sheep=missing*

*replace own\_sheep=0 if v225=2 or v226e=0 or v226e=missing*

*replace own\_sheep=1 if v226e>0 & v226e≤98*

*Label values 0 “None” 1 “1+”*

*Label variable "Household owns at least 1 sheep"*

**Step 6.** Create a variable that indicates whether the household owns any poultry (*own\_poultry*).

*Set own\_poultry=missing*

*Replace own\_poultry=0 if v225=2 or v226f=0 or v226f=missing*

*Replace own\_poultry=1 if v226f>0 & v226f≤98*

*Label values 0 “None” 1 “1+”*

*Label variable "Household owns at least 1 chicken or other poultry"*

**Step 7.** Create a variable that indicates whether the household does not own any of the animals included in Steps 1-6 (*own\_none*). Note that fish and “other” animals are not included, but the syntax can be customized to include or exclude animals.

*Set own\_none=missing*

*Replace own\_none=0 if v225≠missing*

*Replace own\_none=1 if v225=2*

*Replace own\_none=1 if own\_cow=0 and own\_cattle=0 and own\_horse=0 and own\_goat=0 and own\_sheep=0 and own\_poultry=0 and own\_none=0*

*Label values 0 “No” 1 “Yes”*

*Label variable "Household does not own any farm animals"*

### VCC production

These variables indicate if any de jure household members cultivated or raised targeted VCCs during the year prior to the survey, according to information in the household roster and targeted VCC modules. These variables are not true disaggregate variables because a household could have cultivated or raised more than one VCC. The template syntax includes maize as the targeted VCC. Adapt the syntax to reflect the VCCs included in the ZOI Survey being analyzed.

#### Definitions

| Treatment of missing data | Not applicable. The consent field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module. |
| --- | --- |
| Survey variables used\* | *v7100d* |
| Analytic variables used | *hhmem\_dj* |
| Analytic variables created\* | *vcc\_maize2, ncvcc\_maize\_dj, vcchh\_maize, vcchh\_none* |

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

#### Calculations

**Step 1.** In the individual-level data file, create a binary variable that indicates whether individuals cultivated maize during the year preceding the survey (*nvcc\_maize2*), regardless of whether they completed the maize ZOI Survey module.

*Set vcc\_maize2=0*

*Replace vcc\_maize2=1 if v7100d≠missing*

*Label values 0 “No” 1 “Yes”Label variable “Maize producer”*

**Step 2.** Create a continuous variable that counts the total number of de jure household members who cultivated maize during the year preceding the survey (*nvcc\_maize*) and add the variable to the household-level data file.

*By cluster and household: set nvcc\_maize\_dj=sum(vcc\_maize2) if hhmem\_dj=1*

*Label variable “Number of de jure HH members who cultivated maize”*

**Step 3.** In the household-level data file, create a binary variable indicates if any de jure household members cultivated maize in the year preceding the ZOI Survey (*vcchh\_maize*).

*Set vcchh\_maize=0*

*Replace vcchh\_maize=1 if nvcc\_maize\_dj>0*

*Label values 0 “No” 1 “Yes”*

*Label variable “HH cultivated maize (de jure HH members)”*

**Step 4.** In the household-level data file, create a binary variable indicates if any de jure household members cultivated maize in the year preceding the ZOI Survey (*vcchh\_maize*).

*Set vcchh\_none=0*

*Replace vcchh\_crop\_none=1 if vcchh\_maize=0*

*Label values 0 “No” 1 “Yes”*

*Label variable “De jure HH member(s) did not cultivate any crop VCCs”*

## Primary adult decisionmakers

This section describes primary adult decisionmaker-level disaggregates that are used to report the Future the Future A-WEAI indicators. The Feed the Future A-WEAI indicators are calculated using data for only de jure primary adult decisionmakers. The indicator disaggregates are also calculated using data for only de jure primary adult decisionmakers because the data to calculate the A-WEAI indicators are in the household-level data file, but the disaggregate information is in the individual-level data file.

### De jure primary adult decision makers, by sex

These variables identify de jure primary adult decisionmakers in the household roster by sex.

#### Definitions

| Treatment of missing data | Not applicable. Missing information on usual household members and the presence of primary adult male and female decisionmakers is not allowed in the household roster. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *pdm, mdm, fdm, hhmem\_dj* |
| Analytic variables created | *mdm\_dj, fdm\_dj* |

#### Calculations

**Step 1.** Create a variable that indicates whether the primary adult female decisionmaker 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 “De jure primary adult female decisionmaker”*

**Step 2.** Create a variable that indicates whether the primary adult male decisionmaker 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 “De jure primary adult male decisionmaker”*

### Age category, by sex (de jure only)

These variables identify de jure primary adult decisionmakers in the household roster by age categories and by sex. In these variables, de jure primary adult decisionmakers under 18 years of age are assigned to an under 18 years of age category, and those 60 years of age or older are assigned to a 60+ years of age category.

#### Definitions

| Treatment of missing data | Not applicable. Missing information on usual household members and the presence of primary adult male and female decisionmakers is not allowed in the household roster. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *mdm\_dj, fdm\_dj, hhmem\_dj, pdm, age* |
| Analytic variables created | *agegrp\_pdm\_dj, agegrp\_mdm\_dj, agegrp\_fdm\_dj* |

#### Calculations

**Step 1.** Create a variable that indicates de jure primary adult decisionmakers’ age by age category (*agegrp\_pdm\_dj*).

*Set agegrp\_pdm\_dj=1 if age≥18 and age≤24 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=2 if age≥25 and age≤29 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=3 if age≥30 and age≤34 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=4 if age≥35 and age≤39 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=5 if age≥40 and age≤44 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=6 if age≥45 and age≤49 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=7 if age≥50 and age≤54 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=8 if age≥55 and age≤59 and pdm=1 and hhmem\_dj=1*

*Replace agegrp\_pdm\_dj=9 if age≥60 and age≤95 and pdm=1 and hhmem\_dj=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 “50-54” 9 “55-59” 10 “60+”*

*Label variable "Primary adult decisionmaker’s age (categorical)"*

**Step 2.** Create a variable that indicates de jure primary adult male decisionmakers’ age by age category (*agegrp\_mdm\_dj*).

*Set agegrp\_mdm\_dj=missing*

*Replace agegrp\_mdm\_dj=agegrp\_pdm\_dj if mdm=1*

*Label variable "De jure male PDM’s age (categorical)"*

**Step 3.** Create a variable that indicates de jure primary adult female decisionmakers’ age by age category (*agegrp\_fdm\_dj*).

*Set agegrp\_fdm\_dj=missing*

*Replace agegrp\_mdm\_dj=agegrp\_pdm\_dj if fdm=1*

*Label variable "De jure female PDM’s age (categorical)"*

### Youth, by sex (de jure only)

These variables identify de jure primary adult decisionmakers in the household roster by whether they are youth (under 30 years of age) or non-youth (30 years of age or older) and by sex.

#### Definitions

| Treatment of missing data | Not applicable. Missing information on usual household members, age, sex, and the presence of primary adult male and female decisionmakers is not allowed in the household roster. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *mdm\_dj, fdm\_dj, age\_pdm\_dj, hhmem\_dj, pdm* |
| Analytic variables created | *youth\_pdm\_dj, youth\_mdm\_dj, youth\_fdm\_dj* |

#### Calculations

**Step 1.** Create a variable that indicates whether de jure primary adult decisionmakers are youth (*youth\_pdm\_dj*).

*Set youth\_pdm\_dj=missing*

*Replace youth\_pdm\_dj=0 if pdm=1 and hhmem\_dj=1*

*Replace youth\_pdm\_dj=1 if pdm=1 and hhmem\_dj=1 and age\_pdm<30*

*Label values 0 “No” 1 “Yes”*

*Label variable “De jure PDM is under 30 years of age"*

**Step 2.** Create a variable that indicates whether de jure primary adult male decisionmakers are youth (*youth\_mdm\_dj*).

*Set youth\_mdm\_dj=missing*

*Replace youth\_pdm\_dj=1if mdm\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable "De jure male PDM is under 30 years of age"*

**Step 3.** Create a variable that indicates whether de jure primary adult female decisionmakers are youth (*youth\_fdm\_dj*).

*Set youth\_fdm\_dj=missing*

*Replace youth\_pdm\_dj=1if fdm\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable "De jure female PDM is under 30 years of age"*

### Education, by sex (de jure only)

These variables identify the highest educational attainment of each de jure primary adult decisionmaker by category: (0) no education, (1) less than primary, (2) completed primary, (3) completed secondary, or (4) higher. Education grades and levels differ from country to country. The ZOI Survey questionnaire, therefore, must be adapted to capture the education system in each country. Be sure to adjust the code to appropriately capture the education system in the country. The template syntax was written assuming 6 years of primary schooling and 6 years of secondary schooling.

#### Definitions

| Treatment of missing data | Missing information on usual household members and the presence of primary adult male and female decisionmakers is not allowed in the household roster. If individuals are missing information on the highest level of education completed, set the variable to missing. If individuals’ highest level of education is “don't know,” and they cannot be assigned to a country-specific education category, set them to “other.” If individuals are missing information on the highest grade of education completed or if the highest grade completed is “don't know,” and they cannot be assigned to a country-specific education category, set them to the appropriate country-specific education category based on the highest level of education completed. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *edulevel, pdm, hhmem\_dj mdm\_dj, fdm\_dj* |
| Analytic variables created | *edu\_pdm\_dj, edu\_mdm\_dj, edu\_fdm\_dj* |

#### Calculations

**Step 1.** Create a variable that indicates the educational attainment of each de jure primary adult decisionmaker (*edu\_pdm\_dj*).

*Set edu\_pdm\_dj=missing*

*Replace edu\_pdm\_dj=edulevel if pdm=1 and hhmem\_dj=1*

*Label values: 1 “No education”, 2 “Less than primary”, 3 “Completed primary”, 4 “Completed secondary”, 5 “Higher”, 6 “Other”*

*Label variable "Education attained by de jure PDM"*

**Step 2.** Create a variable that indicates the educational attainment of each de jure primary adult male decisionmaker (*edu\_mdm\_dj*).

*Set edu\_mdm\_dj=missing*

*Replace edu\_mdm\_dj=edulevel if mdm\_dj=1*

*Label variable "Education attained by de jure male PDM"*

**Step 3.** Create a variable that indicates the educational attainment of each de jure primary adult female decisionmaker (*edu\_fdm\_dj*).

*Set edu\_fdm\_dj=missing*

*Replace edu\_fdm\_dj=edulevel if fdm\_dj=1*

*Label variable "Education attained by de jure female PDM"*

## Women of reproductive age

This section describes women of reproductive age-level disaggregates that are used to report the Future the Future women’s nutrition indicators. The Feed the Future women’s nutrition indicators are calculated using data for only de facto women of reproductive age. The indicator disaggregates are calculated using data for all women of reproductive age; therefore, the de facto variable (*hhmem\_df*) must be specified when calculating the indicators by their disaggregates.

### 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.

#### Definitions

| Treatment of missing data | Not applicable. If a woman’s self-reported age is missing in the women’s nutrition module, use the woman’s age in the household roster. Missing information on age, sex, and whether individuals are de facto household members is not allowed in the household roster. |
| --- | --- |
| Survey variables used | *v402* |
| Analytic variables used | *age, wra, hhmem* |
| Analytic variables created | *agegrp\_wra* |

#### Calculations

**Step 1.** Create a continuous variable that indicates the age of women of reproductive age according to their self-reported age in the women’s nutrition module (*age\_wra*). If their self-reported age in the women’s nutrition module is missing, use the woman’s age in the household roster.

*Set age\_wra=missing*

*Replace age\_wra=v402 if wra=1 and v402<98*

*Replace age\_wra=age if wra=1 and v402=98 or v402=missing*

*Label variable “Age of woman of reproductive age”*

**Step 2.** Create a categorical variable that indicates the age of women of reproductive age by age category (*agegrp\_wra*).

*Set agegrp\_wra=1 if age≥15 and age\_wra≤19 and wra=1*

*Replace agegrp\_wra=2 if age≥20 and age\_wra≤24 and wra=1*

*Replace agegrp\_wra=3 if age≥25 and age\_wra≤29 and wra=1*

*Replace agegrp\_wra=4 if age≥30 and age\_wra≤34 and wra=1*

*Replace agegrp\_wra=5 if age≥35 and age\_wra≤39 and wra=1*

*Replace agegrp\_wra=6 if age≥40 and age\_wra≤44 and wra=1*

*Replace agegrp\_wra=7 if age≥45 and age\_wra≤49 and wra=1*

*Label variable "Age of woman of reproductive age (categorical)"*

### Age category—15-18 years and 19-49 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: 15-18 years of age and 19-49 years of age.

#### Definitions

| Treatment of missing data | Not applicable. If a woman’s self-reported age is missing in the women’s nutrition module, use the woman’s age in the household roster. Missing information on age, sex, and whether individuals are de facto household members is not allowed in the household roster. |
| --- | --- |
| Survey variables used | *v402* |
| Analytic variables used | *age, wra* |
| Analytic variables created | *wra\_cage* |

#### Calculations

**Step 1.** Create a categorical variable that categorizes women of reproductive age into two age categories: 15-18 years of age and 19-49 years of age. If their self-reported age in the women’s nutrition module is missing, use the woman’s age in the household roster.

*Set wra\_cage =1 if age≥15 and age\_wra≤18 and wra=1*

*Replace wra\_cage=2 if age≥19 and age\_wra≤49 and wra=1*

*Label variable "Age of woman of reproductive age (15-18, 19-49 years old)"*

### Education

This variable identifies the highest educational attainment of all women of reproductive age in the household roster by category: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, or (5) higher. Education grades and levels differ from country to country. The ZOI Survey questionnaire, therefore, must be adapted to capture the education system in each country. Be sure to adjust the code to appropriately capture the education system in the country. The template syntax was written assuming 6 years of primary schooling and 6 years of secondary schooling.

#### Definitions

| Treatment of missing data | The *wra* analytic variable cannot have missing values. Please see Section 4.1.17 for information on how missing values are handled when creating the *edulevel* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *edulevel, wra* |
| Analytic variables created | *edu\_wra* |

#### Calculations

**Step 1.** Create a variable that indicates the educational attainment of each woman of reproductive age (*edu\_wra*).

*Set edu\_wra=edulevel if wra=1*

*Label values: 1 “No education”, 2 “Less than primary”, 3 “Completed primary”, 4 “Completed secondary”, 5 “Higher”, 6 “Other”*

*Label variable “Education attained by woman of reproductive age"*

### Pregnancy status

This variable indicates whether women of reproductive age were pregnant at the time of the survey according to information in the women’s nutrition module.

#### Definitions

| Treatment of missing data | The *wra* analytic variable cannot have missing values. For women missing information about whether they are pregnant, the *preg\_stat* variable is set to missing. Women who respond that they do not know if they are pregnant are considered to not be pregnant. |
| --- | --- |
| Survey variables used | *an405* |
| Analytic variables used | *wra* |
| Analytic variables created | *preg\_stat* |

#### Calculations

**Step 1.** Create a variable that indicates the pregnancy status of each woman of reproductive age (*preg\_stat*).

*Set preg\_stat=missing*

*Replace preg\_stat=0 if wra=1 and an405=2 or an405=8*

*Replace preg\_stat=1 if wra=1 and an405=1*

*Label values 0 “No” 1 “Yes”*

*Label variable "Woman is pregnant"*

## Children under 5

This section describes children-level disaggregates that are used to report the Future the Future children’s nutrition indicators. The Feed the Future children’s nutrition indicators are calculated using data for only de facto children. The indicator disaggregates are calculated using data for all children; therefore, the de facto variable (*hhmem\_df*) must be specified when calculating the indicators by their disaggregates.

### Sex

Please see the sex variable (*sex*) created in Section 4.1.3. This variable is used for the child sex disaggregate.

### Age category—12 months

This variable categories children under 5 into 12-month age categories using information from the children’s nutrition module.

#### Definitions

| Treatment of missing data | For children who are missing a value for the *cage\_months\_int* variable, the *cage\_6m* variable is set to missing. Please see Section 4.1.13 for information on how missing values are handled when creating the *cage\_months\_int* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *cage\_months\_int* |
| Analytic variables created | *cage\_12m* |

#### Calculations

**Step 1.** Create a variable that identifies the age of all children under 5 in the children’s module by 12-month age categories (*cage\_12m*).

*Set cage\_12m=missing*

*Replace cage\_12m=1 if cage\_months\_int<12*

*Replace cage\_12m=2 if cage\_months\_int≥12 and cage\_months\_int<24*

*Replace cage\_12m=3 if cage\_months\_int≥24 and cage\_months\_int<36*

*Replace cage\_12m=4 if cage\_months\_int≥36 and cage\_months\_int<48*

*Replace cage\_12m=5 if cage\_months\_int≥48 and cage\_months\_int<60*

*Label values 1 “0-11 months” 2 “12-17 months” 3 “18-23 months” 4 “24-35 months” 5 “36-59 months”*

*Label variable “Child age category 0-59 months (12m interval)”*

### Age category—6 months

This variable categorizes all children 6-23 months of age by 6-month age categories using information from the children’s nutrition module.

#### Definitions

| Treatment of missing data | For children who are missing a value for the *cage\_months\_int* variable, the *cage\_6m* variable is set to missing. Please see Section 4.1.13 for information on how missing values are handled when creating the *cage\_months\_int* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *cage\_months\_int* |
| Analytic variables created | *cage\_6m* |

#### 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\_6m*).

*Set cage\_6m=missing*

*Replace cage\_6m=1 cage\_months\_int≥6 and cage\_months\_int<12*

*Replace cage\_6m=3 if cage\_months\_int≥12 and cage\_months\_int<18*

*Replace cage\_6m=4 if cage\_months\_int≥18 and cage\_months\_int<24*

*Label variable “Child age category 6-23 months (6m interval)”*

*Label values 1 “6-11 months” 2 “12-17 months” 3 “18-23 months”*

### Age category—0-23 months and 24-59 months

This variable categorizes children 0-59 months of age into two age categories: 0-23 months of age and 24-49 months of age using information from the children’s nutrition module.

#### Definitions

| Treatment of missing data | For children who are missing a value for the *cage\_months\_int* variable, the *cnut\_age* variable is set to missing. Please see Section 4.1.13 for information on how missing values are handled when creating the *cage\_months\_int* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *cage\_months\_int* |
| Analytic variables created | *cnut\_age* |

#### 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 24-49 months of age (*cnut\_age*).

*Set cnut\_age=missing*

*Replace cnut\_age=1 cage\_months\_int≥0 and cage\_months\_int<24*

*Replace cnut\_age=2 if cage\_months\_int≥24 and cage\_months\_int<60*

*Label variable “Child age category 0-59 months”*

*Label values 1 “0-23 months” 2 “23-59 months”*

### Caregiver’s education

This variable identifies the highest educational attainment of each primary caregiver of children under 5 included in the children’s nutrition module by category: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, (5) higher, or (6) other. Education grades and levels differ from country to country. The ZOI Survey questionnaire, therefore, must be adapted to capture the education system in each country. Be sure to adjust the code to appropriately capture the education system in the country. The template syntax was written assuming 6 years of primary schooling and 6 years of secondary schooling.

#### Definitions

| Treatment of missing data | Please see Section 4.1.17 for information on how missing values are handled when creating the *edulevel* variable. Please see Section 4.1.19 for information on how missing values are handled when creating the *caregiver* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *edulevel, caregiver* |
| Analytic variables created | *edu\_cg* |

#### Calculations

**Step 1.** Create a variable that indicates the educational attainment of each primary caregiver of children under 5 included in the children’s nutrition module (*edu\_cg*).

*Set edu\_cg=missing*

*Replace edu\_cg=edulevel if caregiver=1*

*Label variable "Education attained by primary caregiver of children under 5"*

## Targeted VCC producers

This section describes producer-level disaggregates that are used to report the Future the Future agriculture indicators. The Feed the Future agriculture indicators are calculated using data for only de jure producers. The indicator disaggregates are calculated using data for all producers; therefore, the de jure variable (*hhmem\_dj*) must be specified when calculating the indicators by their disaggregates.

### Sex

Please see the sex variable (*sex*) created in Section 4.1.3. This variable is used for the producer sex disaggregate.

### Age category

This variable identifies the age of all producers interviewed for the agriculture module by 5-year age categories. In this variable, all producers 60 years of age or older are assigned to a 60 years of age or older category.

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *vcc* analytic variables cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *age, vcc* |
| Analytic variables created | *agegrp\_vcc* |

#### Calculations

**Step 1.** Create a variable that indicates the producers’ age by age category (*agegrp\_vcc*).

*Set agegrp\_vcc=1 if age≥15 and age≤19 and vcc=1*

*Replace agegrp\_vcc=2 if age≥20 and age≤24 and vcc=1*

*Replace agegrp\_vcc=3 if age≥25 and age≤29 and vcc=1*

*Replace agegrp\_vcc=4 if age≥30 and age≤34 and vcc=1*

*Replace agegrp\_vcc=5 if age≥35 and age≤39 and vcc=1*

*Replace agegrp\_vcc=6 if age≥40 and age≤44 and vcc=1*

*Replace agegrp\_vcc=7 if age≥45 and age≤49 and vcc=1*

*Replace agegrp\_vcc=7 if age≥50 and age≤54 and vcc=1*

*Replace agegrp\_vcc=7 if age≥55 and age≤59 and vcc=1*

*Replace agegrp\_vcc=7 if age≥60 and age≤95 and vcc=1*

*Label variable "Producer’s age (categorical)"*

### Youth

This variable identifies all famers in the household roster by whether they are youth (under 30 years of age) or non-youth (30 years of age or older).

#### Definitions

| Treatment of missing data | Not applicable. The *age* and *vcc* analytic variables cannot have missing values. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *vcc, age* |
| Analytic variables created | *youth\_vcc* |

#### Calculations

**Step 1.** Create a variable that indicates whether producers are youth (*youth\_vcc*).

*Set youth\_vcc=missing*

*Replace youth\_vcc=0 if vcc=1*

*Replace youth\_vcc=1 if vcc=1 and age<30*

*Label values 0 “No” 1 “Yes”*

*Label variable “Producer is under 30 years of age"*

### Education

This variable identifies the highest educational attainment of each producer in the household roster by category: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, (5) higher, or (6) other. Education grades and levels differ from country to country. The ZOI Survey questionnaire, therefore, must be adapted to capture the education system in each country. Be sure to adjust the code to appropriately capture the education system in the country. The template syntax was written assuming 6 years of primary schooling and 6 years of secondary schooling.

#### Definitions

| Treatment of missing data | The *vcc* analytic variable cannot have missing values. Please see Section 4.1.17 for information on how missing values are handled when creating the *edulevel* variable. |
| --- | --- |
| Survey variables used | n/a |
| Analytic variables used | *edulevel, vcc* |
| Analytic variables created | *edu\_vcc* |

#### Calculations

**Step 1.** Create a variable that indicates the educational attainment of each producer (*edu\_vcc*).

*Set edu\_vcc=edulevel if vcc=1*

*Label values: 1 “No education”, 2 “Less than primary”, 3 “Completed primary”, 4 “Completed secondary”, 5 “Higher”, 6 “Other”*

*Label variable "Education attained by producer"*

### Commodity

These variables identify individuals in the household roster who were primarily responsible for cultivating or raising each targeted VCC included in the ZOI Survey and who completed the VCC module. The template syntax includes maize, millet, okra, and okra 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 consent field in a targeted VCC module cannot have a missing value if the individual was the eligible respondent for the module, and the *vcc* variable cannot have missing values. |
| --- | --- |
| Survey variables used\* | *v7100d, v71000d, v71100d, v75200d* |
| Analytic variables used | vcc |
| Analytic variables created\* | *vcc\_maize, vcc\_millet, vcc\_okra, vcc\_sheep* |

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

#### Calculations

**Step 1.** Create a binary variable to indicate if individuals cultivated maize in the year preceding the ZOI Survey (*vcc\_maize*).

*Set vcc\_maize=0*

*Replace vcc\_maize=1 if v7100d=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Maize producer, completed maize module”*

**Step 2.** Create a binary variable to indicate if individuals cultivated millet it in the year preceding the ZOI Survey (*vcc\_millet*).

*Set vcc\_millet=0*

*Replace vcc\_millet=1 if v71000d=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Millet producer, completed millet module”*

**Step 3.** Create a binary variable to indicate if individuals cultivated okra in the year preceding the ZOI Survey (*vcc\_okra*).

*Set vcc\_okra=0*

*Replace vcc\_okra=1 if v71100d=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Okra producer, completed okra module”*

**Step 4.** Create a binary variable to indicate if individuals raised sheep in the year preceding the ZOI Survey (*vcc\_sheep*).

*Set vcc\_sheep=0*

*Replace vcc\_sheep=1 if v75200d=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Sheep producer, completed sheep module”*

**Part II.**

**INDICATORS**

# 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 three sections: household demographic characteristics, characteristics of de jure primary adult decisionmakers, and education of de facto household members.

## Household demographic characteristics

This section describes the following household-level demographic indicators:

* Mean household size
* Mean number of adult male members
* Mean number of adult female members
* Mean number of women of reproductive age (15-49 years)
* 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)
* Percent of adults who are male
* Percent of adults who are female
* Percent distribution of households by household size category
* Percent distribution of households by highest educational attainment

### Mean household size

This indicator is the sample-weighted mean number of de jure household members per household.

#### 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 | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *hhsize\_dj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, hhsize\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *hhsize\_*dj variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean hhsize\_dj*

*Svy: mean hhsize\_dj, over(genhhtype\_dj)*

### Mean number of adult male household members

This indicator is the sample-weighted mean number of de jure adult male household members per household. It is based on information reported in the household roster.

#### 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 | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *adult\_mdj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, nadult\_mdj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *nadult\_mdj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean nadult\_mdj*

*Svy: mean nadult\_mdj, over(genhhtype\_dj)*

### Mean number of adult female household members

This indicator is the sample-weighted mean number of de jure adult female household members per household. It is based on information reported in the household roster.

#### 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 | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *nadult\_fdj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, nadult\_fdj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *nadult\_fdj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean nadult\_fdj*

*Svy: mean nadult\_fdj, over(genhhtype\_dj)*

### Mean number of women of reproductive age

This indicator is the sample-weighted mean number of de jure adult women of reproductive age (15-49 years) per household. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of de jure adult women of reproductive age in surveyed households |
| --- | --- |
| Denominator | Total number of surveyed households |
| Unit of measure | Number |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *nwra\_dj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, nwra\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *nwra\_dj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean nwra\_dj*

*Svy: mean nwra\_dj, over(genhhtype\_dj)*

### Mean number of children under 2 years

This indicator is the sample-weighted mean number of de jure children under 2 years of age per household. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of de jure children under 2 years of age in surveyed households |
| --- | --- |
| Denominator | Total number of surveyed households |
| Unit of measure | Number |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *ncu2\_dj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, ncu2\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *ncu2\_dj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean ncu2\_dj*

*Svy: mean ncu2\_dj, over(genhhtype\_dj)*

### Mean number of children under 5 years

This indicator is the sample-weighted mean number of de jure children under 5 years of age per household. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of de jure children under 5 years of age in surveyed households |
| --- | --- |
| Denominator | Total number of surveyed households |
| Unit of measure | Number |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *ncu5\_dj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, ncu5\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *ncu5\_dj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean ncu5\_dj*

*Svy: mean ncu5\_dj, over(genhhtype\_dj)*

### Mean number of children 5-17 years of age

This indicator is the sample-weighted mean number of de jure children between 5 and 17 years of age, inclusive, per household. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of de jure children between 5 and 17 years of age |
| --- | --- |
| Denominator | Total number of surveyed households |
| Unit of measure | Number |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *nc5\_17\_dj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, nc5\_17\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *nc5\_17y\_dj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. Repeat by gendered household type category. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean nc5\_17y\_dj*

*Svy: mean nc5\_17y\_dj, over(genhhtype\_dj)*

### Mean number of youth 15-29 years of age

This indicator is the sample-weighted mean number of de jure youth (15-29 years of age) per household. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of de jure youth between 15 and 29 years of age |
| --- | --- |
| Denominator | Total number of surveyed households |
| Unit of measure | Number |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *nc15\_29\_dj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, nc15\_29\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the mean value of the *nc15\_29y\_dj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean nc15\_29y\_dj*

*Svy: mean nc15\_29y\_dj, over(genhhtype\_dj)*

### Mean number of producers of any targeted commodity

This indicator is the sample-weighted mean number of de jure producers of any targeted commodity per household. It is based on information reported in the household roster and Module 2, *Dwelling characteristics*. Targeted commodities will vary by country. Each country will specify up to four livestock or crop commodities about which it would like to collect information in the ZOI Survey. Producers are considered to be those who are responsible for making management decisions about one or more plots of targeted crop commodities or about raising one or more types of targeted livestock commodities. The template syntax includes maize, millet, okra, and sheep as the targeted VCCs. Be sure to adapt the syntax to reflect the VCCs included in the ZOI Survey being analyzed.

#### Definitions

| Numerator | Number of de jure producers of any targeted commodity |
| --- | --- |
| Denominator | Total number of surveyed households |
| Unit of measure | Number |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *nvcc* analytic variables used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used\* | *genhhtype\_dj, nvccc\_maize\_dj, nvccc\_millet\_dj, nvccc\_okra\_dj, nvccc\_sheep\_dj* |
| 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 *nvccc* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean nvccc\_dj*

*Svy: mean nvccc, over(genhhtype\_dj)*

**Step 2.** Repeat Step 1 for each targeted VCC included in the ZOI Survey.

### Percent of adults who are male

This indicator is the sample-weighted percentage of de jure adult household members who are male. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of de jure adult male household members in surveyed households |
| --- | --- |
| Denominator | Number of de jure adult members in surveyed households |
| Unit of measure | Percentage |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *nadult\_mdj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, nadult\_mdj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** Ensure that the *nadult\_mdj* variable is set to missing for all household members under 18 years of age.

*Replace nadult\_mdj=missing if age<18*

**Step 2.** After applying the household sampling weight, calculate the percentage of de jure adults who are male using the *nadult\_mdj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab nadult\_mdj*

*Svy: tab nadult\_mdj genhhtype\_dj, col*

### Percent of adults who are female

This indicator is the sample-weighted percentage of de jure adult household members who are female. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of de jure adult female household members in surveyed households |
| --- | --- |
| Denominator | Number of de jure adult members in surveyed households |
| Unit of measure | Percentage |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *nadult\_fdj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, nadult\_fdj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** Ensure that the *nadult\_fdj* variable is set to missing for all household members under 18 years of age.

*Replace nadult\_fdj=missing if age<18*

**Step 2.** After applying the household sampling weight, calculate the percentage of de jure adults who are female using the *nadult\_fdj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab nadult\_fdj*

*Svy: tab nadult\_fdj genhhtype\_dj, col*

### Percent distribution of households by household size category—de jure

This indicator is the sample-weighted percentage distribution of households among three household size categories (small, medium, and large), based on the number of de jure household members. It is based on information reported in the household roster. 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.

#### 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 | Gendered household type (de jure household members) |
| Treatment of missing data | Not applicable. The *hhsizegrp\_dj* analytic variable used cannot be missing values. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, hhsizegrp\_dj* |
| 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 *nhhsize\_dj* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab hhsize\_dj*

*Svy: tab hhsize\_dj genhhtype\_dj, col*

### Percent distribution of households by highest level of educational attainment category

This indicator is the sample-weighted percentage of households in each of four educational attainment categories. Each household is classified by the highest level of education attainment of any de facto household member: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, (5) higher, or (6) other. It is based on information reported in the household roster.

#### Definitions

| Numerator | Number of households in each educational attainment category |
| --- | --- |
| Denominator | Total number of surveyed households |
| Unit of measure | Number |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de facto household members) |
| Treatment of missing data | Households missing information about the highest level of educational attainment for all de facto household members are excluded from the numerator and denominator. Please see Section 5.1.4 for information on how missing values are handled when creating the *edulevel\_hh\_df* variable. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_df, edulevel\_hh\_df* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the household sampling weight, calculate the percentage of households that have no education, have less than primary education, completed primary education, completed secondary education, or completed higher education using the *edulevel\_hh\_df* variable. Household education is determined by the highest educational attainment of any de facto household member. Repeat using the gendered household type disaggregate constructed using de facto household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab edulevel\_hh\_df*

*Svy: tab edulevel\_hh\_df genhhtype\_df, col*

## Characteristics of primary adult male and female decisionmakers

This section describes indicators that characterize primary adult male and female decisionmakers.

* Percent distribution of primary adult male decisionmakers by age group
* Percent distribution of primary adult female decisionmakers by age group
* Percent distribution of primary adult male decisionmakers by current marital status
* Percent distribution of primary adult female decisionmakers by current marital status
* Percent distribution of primary adult male decisionmakers by highest educational attainment
* Percent distribution of primary adult female decisionmakers by highest educational attainment
* Percentage of primary adult male decisionmakers who participate in any economic activity
* Percentage of primary adult female decisionmakers who participate in any economic activity
* Percent distribution of primary adult male decisionmakers by type of economic activity
* Percent distribution of primary adult female decisionmakers by type of economic activity

### Percent distribution of primary adult male decisionmakers by age group

This indicator is the percentage distribution of de jure primary adult male decisionmakers (18 years of age or older) by age group in years: 18-24, 25-29, 30-39, 40-49, 50-59, and 60+. Primary adult male decisionmakers are adult males (18 years of age or older) who make the most important decisions in the household. The indicator is created using information reported in the household roster.

#### Definitions

| Numerator | Number of de jure primary adult male decisionmakers in each specified age group |
| --- | --- |
| Denominator | Number of de jure primary adult male decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult male decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | The *agegrp\_mdm\_dj* analytic variable cannot have missing values. |
| Survey variables used | *hhea, wgt\_mpdm, samp\_stratum* |
| Analytic variables used | *agegrp\_mdm\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the primary adult male decisionmaker sampling weight, calculate the percentage of de jure primary adult male decisionmakers by age category using the *agegrp\_mdm\_dj* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=mpdm\_wgt], strata(samp\_stratum)*

*Svy: tab agegrp\_mdm\_dj*

### Percent distribution of primary adult female decisionmakers by age group

This indicator is the percentage distribution of de jure primary adult female decisionmakers (18 years of age or older) by age group in years: 18-24, 25-29, 30-39, 40-49, 50-59, and 60+. Primary adult female decisionmakers are adult females (18 years of age or older) who make the most important decisions in the household. The indicator is created using information reported in the household roster.

#### Definitions

| Numerator | Number of de jure primary adult female decisionmakers in each specified age group |
| --- | --- |
| Denominator | Number of de jure primary adult female decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult female decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | The *agegrp\_fdm\_dj* analytic variable cannot have missing values. |
| Survey variables used | *hhea, wgt\_fpdm, samp\_stratum* |
| Analytic variables used | *agegrp\_fdm\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the primary adult female decisionmaker sampling weight, calculate the percentage of de jure primary adult female decisionmakers by age category using the *agegrp\_fdm\_dj* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_fpdm], strata(samp\_stratum)*

*Svy: tab agegrp\_fdm\_dj*

### Percent distribution of primary adult male decisionmakers by their current marital status

This indicator is the percentage distribution of de jure primary adult male decisionmakers by their current marital status. It is based on information reported in the household roster and Module 6, *Women’s Empowerment in Agriculture*.

#### Definitions

| Numerator | Number of de jure primary adult male decisionmakers with a marital status in each specified marital status category |
| --- | --- |
| Denominator | Number of de jure primary adult male decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult male decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | If information on marital status is missing, set the *marstat\_mdm\_dj* analytic variable to missing and exclude the individual from the indicator numerator and denominator. |
| Survey variables used | *v6105, v6106, v6107, hhea, wgt\_mpdm, samp\_stratum* |
| Analytic variables used | *mdm\_dj* |
| Analytic variables created | *marstat\_mdm\_dj* |

#### Calculations

**Step 1.** Create a categorical variable for the marital status of primary adult male decisionmakers (*marstat\_mdm*).

*Set marstat\_mdm\_dj\_dj=missing*

*Replace marstat\_mdm\_dj=1 if (m6105=1 and mdm\_dj=1)*

*Replace marstat\_mdm\_dj=2 if (m6105=2 and mdm\_dj=1)*

*Replace marstat\_mdm\_dj =3 if (m6105=3 and m6107=1 and mdm\_dj=1)*

*Replace marstat\_mdm\_dj =4 if (m6105=3 and (m6107=2 or m6107=3) and mdm\_dj=1)*

*Replace marstat\_mdm\_dj =5 if (m6105=3 and m6106=3 and mdm\_dj=1)*

*Label values 1 “Currently married” 2 “Living in union” 3 “Widowed” 4 “Divorced/separated” 5 “Never married or in union”*

**Step 2.** After applying the primary adult male decisionmaker sampling weight, calculate the percentage of de jure primary adult male decisionmakers by marital status category using the *marstat\_mdm\_dj* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_mpfm], strata(samp\_stratum)*

*Svy: tab marstat\_mdm\_dj*

### Percent distribution of primary adult female decisionmakers by their current marital status

This indicator is the percentage distribution of de jure primary adult female decisionmakers by their current marital status. The indicator is calculated using information reported in the household roster and Module 6, *Women’s Empowerment in Agriculture*.

#### Definitions

| Numerator | Number of de jure primary adult female decisionmakers with a marital status in each specified marital status category |
| --- | --- |
| Denominator | Number of de jure primary adult female decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult female decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | If information on marital status is missing, set the *marstat\_fdm\_dj* analytic variable to missing and exclude the individual from the indicator numerator and denominator. |
| Survey variables used | *v6105, v6106, v6107, hhea, wgt\_fpdm, samp\_stratum* |
| Analytic variables used | *fdm\_dj* |
| Analytic variables created | *marstat\_fdm\_dj* |

#### Calculations

**Step 1.** Create a categorical variable for the marital status of de jure primary adult female decisionmakers (*marstat\_fdm\_dj*).

*Set marstat\_fdm\_dj=missing*

*Replace marstat\_fdm=1 if (v6105=1 and fdm\_dj=1)*

*Replace marstat\_fmd=2 if (v6105=2 and fdm\_dj=1)*

*Replace marstat\_fmd =3 if (v6105=3 and v6107=1 and fdm\_dj=1)*

*Replace marstat\_fmd =4 if v6105=3 and (v6107=2 or v6107=3) and fdm\_dj=1)*

*Replace marstat\_fmd =5 if (v6105=3 and v6106=3 and fdm\_dj=1)*

*Label values 1 “Currently married” 2 “Living in union” 3 “Widowed” 4 “Divorced/separated” 5 “Never married or in union*

**Step 2.** After applying the primary adult female decisionmaker sampling weight, calculate the percentage of de jure primary adult female decisionmakers by marital status category using the *marstat\_fdm\_dj* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_fpdm], strata(samp\_stratum)*

*Svy: tab marstat\_fdm\_dj*

### Percent distribution of primary adult male decisionmakers by their highest completed level of education

This indicator is the percentage distribution of de jure primary adult male decisionmakers (18 years of age or older) by their highest completed level of education: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, and (5) higher. The indicator is created using information reported in the household roster.

#### Definitions

| Numerator | Number of de jure primary adult male decisionmakers in each specified educational attainment category |
| --- | --- |
| Denominator | Number of all de jure primary adult male decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult male decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | Individuals missing a value for the *edu\_mdm\_dj* analytic variable are excluded from the numerator and denominator. Please see Section 5.2.4 for information on how missing values are handled when creating the *edu\_mdm\_dj* variable. |
| Survey variables used | *hhea, wgt\_mpdm, samp\_stratum* |
| Analytic variables used | *edu\_mdm\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the primary adult male decisionmaker sampling weight, calculate the percentage of de jure primary adult male decisionmakers by educational attainment category using the *edu\_mdm\_dj* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_mpdm], strata(samp\_stratum)*

*Svy: tab edu\_mdm\_dj*

### Percent distribution of primary adult female decisionmakers by their highest completed level of education

This indicator is the percentage distribution of de jure primary adult female decisionmakers (18 years of age or older) by their highest completed level of education: (1) no education, (2) less than primary, (3) completed primary, (4) completed secondary, and (5) higher. The indicator is created using information reported in the household roster.

#### Definitions

| Numerator | Number of de jure primary adult female decisionmakers in each specified educational attainment category |
| --- | --- |
| Denominator | Number of all de jure primary adult female decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult female decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | Individuals missing a value for the *edu\_fdm\_dj* analytic variable are excluded from the numerator and denominator. Please see Section 5.2.4 for information on how missing values are handled when creating the *edu\_fdm\_dj* variable. |
| Survey variables used | *hhea, wgt\_fpdm, samp\_stratum* |
| Analytic variables used | *edu\_fdm\_dj* |
| Analytic variables created | n/a |

#### Calculations

**Step 1.** After applying the primary adult female decisionmaker sampling weight, calculate the percentage of de jure primary adult female decisionmakers by educational attainment category using the *edu\_fdm\_dj* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_fpdm], strata(samp\_stratum)*

*Svy: tab edu\_fdm\_dj*

### Percent of primary adult male decisionmakers by participation in economic activity

This indicator is the percentage of primary adult male decisionmakers 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 based on information reported in the household roster and Module 6, *Women’s Empowerment in Agriculture*.

#### Definitions

| Numerator | Number of de jure primary adult male decisionmakers who participate in any economic activity |
| --- | --- |
| Denominator | Number of de jure primary adult male decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult male decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | The analytic variable *mdm\_dj* cannot have missing values.If all information on participation in economic activity is missing, exclude the individual from the analysis. If information about a subset of activities is missing, assume that the individual did not participate in the missing activities. |
| Survey variables used | *m6201\_01-m6201\_06, hhea, wgt\_mpdm, samp\_stratum* |
| Analytic variables used | *mdm\_dj* |
| Analytic variables created | *mdm\_anymiss, mdm\_acttot, mdm\_anyact* |

#### Calculations

**Step 1.** Create a variable that counts the number of economic activities for which data are missing for each primary adult male decisionmaker (*mdm\_actmiss,* value range: 0-6).

*Set mdm\_actmiss=0 if mdm\_dj=1*

*Replace mdm\_actmiss=count if mdm\_dj=1 and (m6201\_01=missing or m6201\_02=missing or m6201\_03=missing or m6201\_04=missing or m6201\_05=missing or m6201\_06=missing)*

**Step 2.** Create a variable that counts the number of economic activities in which the primary adult male decisionmaker participated (*mdm\_acttot*).

*Set mdm\_acttot=0 if mdm=1 and mdm\_actmiss<6*

*Replace mdm\_acttot=count if mdm\_dj=1 and (m6201\_01=1 or m6201\_02=1 or m6201\_03=1 or m6201\_04=1 or m6201\_05=1 or m6201\_06=1) and mdm\_actmiss<6*

**Step 3.** Create a binary variable indicating whether the primary adult male decisionmaker participated in any economic activity (*mdm\_anyact*).

*Set mdm\_actany=0 if mdm\_acttot≠missing*

*Replace mdm\_actany=1 if mdm\_dj=1 and mdm\_acttot>1 and mdm\_acttot≠missing*

*Label values 0 “No” 1 “Yes”*

*Label variable “De jure primary adult male decisionmaker participated in 1+ economic activity”*

**Step 4.** After applying the primary adult male decisionmaker sampling weight, calculate the percentage of de jure primary adult male decisionmakers who participated in at least one economic activity using the *mdm\_actany* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_mpdm], strata(samp\_stratum)*

*Svy: tab mdm\_actany\_dj*

### Percentage of primary adult female decisionmakers by participation in economic activity

This indicator is the percentage of primary adult female decisionmakers 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 based on information reported in the household roster and Module 6, *Women’s Empowerment in Agriculture*.

#### Definitions

| Numerator | Number of de jure primary adult female decisionmakers who participate in any economic activity |
| --- | --- |
| Denominator | Number of de jure primary adult female decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult female decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | The analytic variable *fdm\_dj* cannot have missing values.If all information on participation in economic activity is missing, exclude the individual from the analysis. If information about a subset of activities is missing, assume that the individual did not participate in the missing activities. |
| Survey variables used | *v6201\_01-v6201\_06, hhea, wgt\_fpdm, samp\_stratum* |
| Analytic variables used | *fdm\_dj* |
| Analytic variables created | *fdm\_anymiss, fdm\_acttot, fdm\_anyact* |

#### Calculations

**Step 1.** Create a variable that counts the number of economic activities for which data are missing for each primary adult female decisionmaker (*fdm\_actmiss,* value range: 0-6).

*Set fdm\_actmiss=0 if fdm\_dj=1*

*Replace fdm\_actmiss=count if fdm\_dj=1 and (v6201\_01=missing or v6201\_02=missing or v6201\_03=missing or v6201\_04=missing or v6201\_05=missing or v6201\_06=missing)*

**Step 2.** Create a variable that counts the number of economic activities in which the primary adult female decisionmaker participated (*fdm\_acttot*).

*Set fdm\_acttot=0 if fdm\_dj=1 and fdm\_actmiss<6*

*Replace fdm\_acttot=count if fdm\_dj=1 and (m6201\_01=1 or m6201\_02=1 or m6201\_03=1 or m6201\_04=1 or m6201\_05=1 or m6201\_06=1) and fdm\_actmiss<6*

**Step 3.** Create a binary variable indicating whether the primary adult female decisionmaker participated in any economic activity (*fdm\_anyact*).

*Set fdm\_actany=0 if fdm\_acttot≠missing*

*Replace fdm\_actany=1 if fdm\_dj=1 and fdm\_acttot>1 and fdm\_acttot≠missing*

*Label values 0 “No” 1 “Yes”*

*Label variable “De jure primary adult female decisionmaker participated in 1+ economic activity”*

**Step 4.** After applying the primary adult female decisionmaker sampling weight, calculate the percentage of de jure primary adult female decisionmakers who participated in at least one economic activity using the *fdm\_actany variable*. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_fpdm], strata(samp\_stratum)*

*Svy: tab fdm\_actany\_dj*

### Percent distribution of primary adult male decisionmakers by type of economic activity

This indicator is the percentage distribution of primary adult male decisionmakers by type of economic activity in which they participate. 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 based on information reported in the household roster and Module 6, *Women’s Empowerment in Agriculture*. Note that the percentages will not add up to 100 percent because individuals can engage in more than one type of economic activity.

#### Definitions

| Numerator | Number of primary adult male decisionmakers who participate in each type of economic activity |
| --- | --- |
| Denominator | Number of primary adult male decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult male decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | The analytic variable *fdm\_dj* cannot have missing values.If all information on participation in economic activity is missing, set the *mdm\_dj* variable to missing. If information about a subset of activities is missing, assume that the individual did not participate in the missing activities. |
| Survey variables used | *m6201\_01-m6201\_06, hhea, wgt\_mpdm, samp\_stratum* |
| Analytic variables used | *mdm\_dj, mdm\_miss* |
| Analytic variables created | *mdm\_farm, mdm\_nonfarm, mdm\_wage* |

#### Calculations

**Step 1**. Create a binary variable indicating whether de jure primary adult male decisionmakers participated in farm work (*mdm\_farm*).

*Set mdm\_farm=0 if mdm\_dj=1 and mdm\_anymiss<6*

*Replace mdm\_farm=1 if mdm\_dj=1 and (m6201\_01=1 or m6201\_02=1 or m6201\_03=1 or m6201\_04=1)*

**Step 2.** Create a binary variable indicating whether de jure primary adult male decisionmakers participated in non-farm work (*mdm\_nonfarm*).

*Set mdm\_nonfarm=0 if mdm\_dj=1 and mdm\_anymiss<6*

*Replace mdm\_farm=1 if mdm\_dj=1 and m6201\_05=1*

**Step 3.** Create a binary variable indicating whether de jure primary adult male decisionmakers participated in wage or salaried work (*mdm\_wage*).

*Set mdm\_wage=0 if mdm\_dj=1 and mdm\_anymiss<6*

*Replace mdm\_wage=1 if mdm\_dj=1 and m6201\_06=1*

**Step 4.** After applying the primary adult male decisionmaker sampling weight, calculate the percentage of de jure primary adult male decisionmakers who participated in farm work using the *mdm\_farm* variable. Repeat for non-farm work using the *mdm\_nonfarm* variable and for wage or salaried work using the *mdm\_wage* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_mpdm], strata(samp\_stratum)*

*Svy: tab mdm\_farm*

*Svy: tab mdm\_nonfarm*

*Svy: tab mdm\_wage*

### Percent distribution of primary adult female decisionmakers by type of economic activity

This indicator is the percentage distribution of primary adult female decisionmakers by type of economic activity in which they participate. 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 based on information reported in the household roster and Module 6, *Women’s Empowerment in Agriculture.* Note that the percentages will not add up to 100 percent because individuals can engage in more than one type of economic activity.

#### Definitions

| Numerator | Number of primary adult female decisionmakers who participate in each type of economic activity |
| --- | --- |
| Denominator | Number of primary adult female decisionmakers in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Primary adult female decisionmaker |
| Disaggregation levels | n/a |
| Treatment of missing data | The analytic variable *fdm\_dj* cannot have missing values.If all information on participation in economic activity is missing, set the *fdm\_dj* variable to missing. If information about a subset of activities is missing, assume that the individual did not participate in the missing activities. |
| Survey variables used | *v6201\_01-v6201\_06, hhea, wgt\_fpdm, samp\_stratum* |
| Analytic variables used | *fdm\_dj, fdm\_miss* |
| Analytic variables created | *fdm\_farm, fdm\_nonfarm, fdm\_wage* |

#### Calculations

**Step 1.** Create a binary variable indicating whether de jure primary adult female decisionmakers participated in farm work (*fdm\_farm*).

*Set fdm\_farm=0 if fdm\_dj=1 and fdm\_anymiss<6*

*Replace fdm\_farm=1 if fdm\_dj=1 and (v6201\_01=1 or v6201\_02=1 or v6201\_03=1 or v6201\_04=1)*

**Step 2.** Create a binary variable indicating whether de jure primary adult female decisionmakers participated in non-farm work (*fdm\_nonfarm*).

*Set fdm\_nonfarm=0 if fdm\_dj=1 and fdm\_anymiss<6*

*Replace fdm\_farm=1 if fdm\_dj=1 and v6201\_05=1*

**Step 3.** Create a binary variable indicating whether de jure primary adult female decisionmakers participated in wage or salaried work (*fdm\_wage*).

*Set fdm\_wage=0 if fdm\_dj=1 and fdm\_anymiss<6*

*Replace fdm\_wage=1 if fdm\_dj=1 and v6201\_06=1*

**Step 4.** After applying the primary adult female decisionmaker sampling weight, calculate the percentage of de jure primary adult female decisionmakers who participated in farm work using the *fdm\_farm* variable. Repeat for non-farm work using the *fdm\_nonfarm* variable and for wage or salaried work using the *fdm\_wage* variable. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_fpdm], strata(samp\_stratum)*

*Svy: tab fdm\_farm*

*Svy: tab fdm\_nonfarm*

*Svy: tab fdm\_wage*

## 6.3 Education of household members

The section describes the following educational attainment characteristics of de facto household members:

* Percent distribution of de facto household members (total, male, and female) 5-24 years of age who are currently attending school, by age category
* Percent distribution of de facto household members (total, male, and female) 10 years of age or older who completed a primary-level education, by age category

### 6.3.1 Percent distribution of de facto household members who are currently attending school, by age category

This indicator is the percentage of de facto household members who are 5-24 years of age who were attending school at the time of the survey, by age category: 5-9, 10-14, 15-19, and 20-24 years. It is based on information in the household roster.

#### Definitions

| Numerator | Number of de facto household members 5-24 years of age reported as attending school at the time of the survey in each age category |
| --- | --- |
| Denominator | Number of de facto household members 5-24 years of age in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Household |
| Disaggregation levels | Sex |
| Treatment of missing data | The *agegrp, hhmem\_df,* and *sex* analytic variables cannot contain missing values. If a de facto household member is missing a value for the *edu\_attend* variable, set the corresponding *edu\_attend* analytic variables for de facto household members to missing. Please see Section 4.1.16 for information on how missing values are handled when creating the *edu\_attend* variable. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *agegrp, hhmem\_df, sex, edu\_attend* |
| Analytic variables created | *edu\_attend\_df, edu\_attend\_mdf, edu\_attend\_fdf* |

#### Calculations

**Step 1.** Create a binary variable indicating whether de facto household members who are 5-24 years of age are currently attending school (*edu\_attend\_df*).

*Set edu\_attend\_df=missing*

*Replace edu\_attend\_df=0 if age≥5 and age≤24 and hhmem\_df=1*

*Replace edu\_attend\_df=1 if age≥5 and age≤24 and hhmem\_df=1 and edu\_attend=1*

**Step 2.** Create a binary variable indicating whether de facto male household members who are 5-24 years of age are currently attending school (*edu\_attend\_mdf*).

*Set edu\_attend\_mdf=missing*

*Replace edu\_attend\_mdf=0 if edu\_attend\_df≠missing and sex=1*

*Replace edu\_attend\_mdf=1 if edu\_attend\_df=1 and sex=1*

**Step 3.** Create a binary variable indicating whether de facto female household members who are 5-24 years of age are currently attending school (*edu\_attend\_df*).

*Set edu\_attend\_fdf=missing*

*Replace edu\_attend\_mdf=0 if edu\_attend\_df≠missing and sex=2*

*Replace edu\_attend\_mdf=1 if edu\_attend\_df=1 and sex=2*

**Step 4.** After applying the household sampling weight, calculate the percentage of de facto household members who are 5-24 years of age who were attending school at the time of the survey using the *edu\_attend\_df* variable. Repeat for de facto male and female household members separately. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab edu\_attend\_df*

*Svy: tab edu\_attend\_mdf*

*Svy: tab edu\_attend\_fdf*

### 6.3.2 Percent distribution of de facto household members 10 years of age or older who completed a primary-level education, by age group

This indicator is the percentage distribution of de facto household members 10 years of age or older who completed a primary-level education, by age category: 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 54-59, 60 or older. It is based on information in the household roster. The years of schooling required to complete a primary-level education differ from country to country. The ZOI Survey questionnaire, therefore, must be adapted to capture the education system in each country. If children are usually older than 10 years of age when they complete primary school, the indicator can be adjusted to reflect the country context. The indicator should also be adjusted to reflect the number of grades that individuals must complete to complete a primary-level education. Any adjustments made should be footnoted in the country report.

#### Definitions

| Numerator | Number of de facto household members 10 years of age or older who completed primary-level schooling in each age category |
| --- | --- |
| Denominator | Number of de facto household members 10 years of age or older in surveyed households |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Household |
| Disaggregation levels | Sex |
| Treatment of missing data | The *agegrp, hhmem\_df,* and *sex* analytic variables cannot contain missing values. If a de facto household member is missing a value for the *edu\_attend* variable, set the corresponding *edu\_attend* analytic variables for de facto household members to missing. Please see Section 4.1.15 for information on how missing values are handled when creating the *edu\_attend* variable. |
| Survey variables used | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *agegrp, hhmem\_df, sex, edu\_prim* |
| Analytic variables created | *edu\_prim\_df, edu\_prim\_mdf, edu\_prim\_fdf* |

#### Calculations

**Step 1.** Create a binary variable to flag de facto household members who are 10 years of age or older and completed a primary-level education (*edu\_prim\_df*).

*Set edu\_prim\_df=missing*

*Replace edu\_prim\_df=0 if age≥10 and hhmem\_df=1 and edu\_prim≠missing*

*Replace edu\_prim\_df=1 if age≥10 and hhmem\_df=1 and edu\_prim=1*

**Step 2.** Create a binary variable to flag de facto male household members who are 10 years of age or older and completed a primary-level education (*edu\_prim\_mdf*).

*Set edu\_prim\_mdf=missing*

*Replace edu\_prim\_mdf=0 if age≥10 and hhmem\_df=1 and edu\_prim≠missing and sex=1*

*Replace edu\_prim\_mdf=1 if age≥10 and hhmem\_df=1 and edu\_prim=1 and sex=1*

**Step 3.** Create a binary variable to flag de facto female household members who are 10 years of age or older and completed a primary-level education (*edu\_prim\_fdf*).

*Set edu\_prim\_fdf=missing*

*Replace edu\_prim\_fdf=0 if age≥10 and hhmem\_df=1 and edu\_prim≠missing and sex=2*

*Replace edu\_prim\_fdf=1 if age≥10 and hhmem\_df=1 and edu\_prim=1 and sex=2*

**Step 4.** After applying the household sampling weight, calculate the percentage of de facto household members who are 10 years of age or older who completed primary schooling using the *edu\_prim\_df* variable. Repeat for de facto male and female household members separately. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab edu\_prim\_df*

*Svy: tab edu\_prim\_mdf*

*Svy: tab edu\_prim\_fdf*

# Water, sanitation, and hygiene indicators

In this chapter, indicators related to water, sanitation, and hygiene are defined and their calculation described. The following seven 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

## 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, a public tap or standpipe, a tube well or borehole, a protected dug well, a protected spring, and rainwater. Improved water sources may also include tanker truck, cart with small tank, or bottled water, depending on the country. Before calculating this indicator, confirm what are considered to be improved water sources in the country, and adjust the indicator calculation accordingly. Regular availability of the water source means that it is (1) available all year round and (2) available 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 | Gendered household type (de jure household members)  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 | *v211, v214, v215, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *h2o\_improved, h2o\_regular, h2o\_imp\_reg* |

#### 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, or 51*

*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=0*

*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=0*

*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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab h2\_imp\_reg*

*Svy: tab h2\_imp\_reg genhhtype\_dj, 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

UNICEF and World Health Organization. 2019. *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/water_sanitation_health/publications/jmp-report-2019/en/>

World Health Organization. 2019. *Water sanitation hygiene. Key terms.* Available at: <https://www.who.int/water_sanitation_health/monitoring/jmp2012/key_terms/en/>

## 7.2 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, based on information collected in Module 2, *Dwelling Characteristics*. 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 | Gendered household type (de jure household members)  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, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *h2o\_corrtreat* |

#### Calculations

**Step 1.** Create a binary variable to flag households using an effective water treatment practice or technology (*h2o\_corrtreat*).

*Set h2o\_corrtreat=0*

*Replace h2o\_corrtreat=1 if (v2171a=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\_corrtreat* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab h2o\_corrtreat*

*Svy: tab h2o\_corrtreat genhhtype\_dj, 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

UNICEF and World Health Organization. 2019. *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/water_sanitation_health/publications/jmp-report-2019/en/>

World Health Organization and UNICEF. 2017. *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.* License: CC BY-NC-SA 3.0 IGO. Available at: <http://www.who.int/water_sanitation_health/publications/jmp-2017/en/>

## 7.3 Percent of households with soap and water at a handwashing station commonly used by family members (Feed the Future phase two indicator)

This indicator, a Feed the Future P2-ZOI-level indicator, measures the percentage of households with soap and water at a handwashing station commonly used by family members. Enumerators collect information for this indicator by physically observing the handwashing station during the household visit.

#### Definitions

| Numerator | Number of households in which soap and water are found at the commonly used handwashing station |
| --- | --- |
| Denominator | Number of households |
| Unit of measure | Percentage |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type\*  Residence (urban, rural)\* |
| Coverage | P2-ZOI |
| 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 | *v205, v206, v207, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *handwash* |

\*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*).

*Set handwash=0*

*Replace handwash=1 if v205=1 and v206=1 and v207=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “HH has a handwashing station with soap and water”*

**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 disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab handwash*

*Svy: tab handwash genhhtype\_dj, 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

World Health Organization. 2018. *Global reference list of 100 core health indicators*. Available at: <http://www.who.int/healthinfo/indicators/2018/en/>

World Health Organization. 2019. *Water sanitation hygiene. Monitoring hygiene.* Available at: <https://www.who.int/water_sanitation_health/monitoring/coverage/monitoring-hygiene/en/>

World Health Organization and UNICEF. 2017. *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.* License: CC BY-NC-SA 3.0 IGO. Available at: [http://www.who.int/water\_sanitation\_health/publications/jmp-2017/en/](http://www.who.int/water_sanitation_health/publications/jmp-2017/en/%20)

## Percent of households with access to a 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, based on information in Module 2, *Dwelling Characteristics.* Improved sanitation facilities include flush or pour-flush toilets that empty into a piped sewer system, a septic tank, or a pit latrine; 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” or “don’t know where,” 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 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 (de jure household members)\*  Residence (urban, rural)\* |
| Coverage | P2-ZOI |
| Treatment of missing data | Households that report using an improved toilet facility, but that are missing information about sharing the toilet facility, are considered to be not using an improved, non-shared facility. They are included in the denominator but 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 | *v208, v209, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *san\_improved, san\_shared, san\_impnotshared* |

\*Standard Feed the Future disaggregate

#### 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 or 12 or 13 or 21 or 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=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\_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 4.** 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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab san\_impnotshared*

*Svy: tab san\_impnotshared genhhtype\_dj, 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

UNICEF and World Health Organization. 2019. *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/water_sanitation_health/publications/jmp-report-2019/en/>

World Health Organization and UNICEF. 2017. *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.* License: CC BY-NC-SA 3.0 IGO. Available at: <http://www.who.int/water_sanitation_health/publications/jmp-2017/en/>

## 7.5 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, based on information in Module 2, *Dwelling Characteristics.* Improved sanitation facilities include flush or pour-flush toilets that empty into a piped sewer system, a septic tank, or a pit latrine; 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” or “don’t know where,” open pits or a 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. Households with an improved sanitation facility that is shared are considered to not have access to a basic sanitation service.

#### 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 | Gendered household type (de jure household members)  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 | *hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *san\_improved, san\_shared, genhhtype\_dj* |
| Analytic variables created | *san\_impshared* |

#### Calculations

**Step 1.** 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*

*Label values 0 “No” 1 “Yes”*

*Label variable “HH uses an improved, shared sanitation facility”*

**Step 2.** 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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab san\_impshared*

*Svy: tab san\_impshared genhhtype,\_dj 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

UNICEF and World Health Organization. 2019. *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/water_sanitation_health/publications/jmp-report-2019/en/>

World Health Organization and UNICEF. 2017. *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.* License: CC BY-NC-SA 3.0 IGO. Available at: <http://www.who.int/water_sanitation_health/publications/jmp-2017/en/>

World Health Organization. 2019. *Water sanitation hygiene. Key terms.* Available at: <https://www.who.int/water_sanitation_health/monitoring/jmp2012/key_terms/en/>

## 7.6 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” or “don’t know where,” 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.

#### 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 | Gendered household type (de jure household members)  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 not considered to be using an unimproved facility; they are included in the denominator but excluded from the numerator. |
| Survey variables used | *v208, hhea, hh\_wgt, samp\_stratum, ahtype* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *san\_notimp* |

#### Calculations

**Step 1.** Create a binary variable to flag households using an unimproved sanitation facility (*san\_notimp*). Other (response code 96) is considered to be unimproved.

*Set san\_notimp =0*

*Replace san\_notimp=1 if (v208=14 or 15 or 23 or 41 or 51 or 61 or 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\_unimp* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab san\_unimp*

*Svy: tab san\_unimp genhhtype\_dj, 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

UNICEF and World Health Organization. 2019. *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/water_sanitation_health/publications/jmp-report-2019/en/>

World Health Organization and UNICEF. 2017. *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.* License: CC BY-NC-SA 3.0 IGO. Available at: <http://www.who.int/water_sanitation_health/publications/jmp-2017/en/>

## 7.7 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 | Gendered household type (de jure household members)  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 | *v208, hhea, hh\_wgt, samp\_stratum, ahtype* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *san\_opendef* |

#### Calculations

**Step 1.** Create a binary variable to flag households practicing open defecation (*san\_opendef*).

*Set san\_opendef=missing*

*Replace san\_opendef=1 if (v208=61)*

*Replace san\_opendef=0 if v208=missing*

*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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab san\_opendef*

*Svy: tab san\_opendef genhhtype\_dj, col*

#### References

UNICEF and World Health Organization. 2019. *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Available at: <https://www.who.int/water_sanitation_health/publications/jmp-report-2019/en/>

World Health Organization and UNICEF. 2017. *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines.* License: CC BY-NC-SA 3.0 IGO. Available at: <http://www.who.int/water_sanitation_health/publications/jmp-2017/en/>

# Dwelling and household characteristic indicators

In this chapter, indicators related to dwelling and household characteristics are defined and their calculation described. The following five 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

## 8.1 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 | Gendered household type (de jure household members)  Residence (urban, rural) |
| Treatment of missing data | Households missing information on their main cooking fuel or with a response of “other” are considered to be not using a solid fuel. They are included in the denominator but excluded from the numerator. Households that report that food is not cooked in the house are excluded from the numerator and denominator. |
| Survey variables used | *v219, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *dw\_solidfuel* |

#### Calculations

**Step 1.** Create a binary variable that flags households using solid fuels for cooking (*dw\_solidfuel*).

*Set dw\_solidfuel=missing*

*Set dw\_solidfuel=1 if (v219=6 or 7 or 8 or 9 or 10 or 11)*

*Set 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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab dw\_solidfuel*

*Svy: tab dw\_solidfuel genhhtype\_dj, col*

## 8.2 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 | Gendered household type (de jure household members)  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 | *v204, hhea, hh\_wgt, samp\_stratum, ahtype* |
| Analytic variables used | *hhsize\_dj, genhhtype\_dj* |
| Analytic variables created | *nroom, memsleep\_dj* |

#### Calculations

**Step 1.** Create a variable that indicates the number of rooms in each household’s dwelling that are used for sleeping (*nroom*). 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 nroom=v204*

*Replace nroom=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=(hhsize\_dj÷nroom)*

*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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: mean memsleep\_dj*

*Svy: mean memsleep\_dj, over(genhhtype\_dj)*

## 8.3 Percent distribution of households by dwelling roof materials

This indicator is the percentage distribution of households by specified categories of the main dwelling roof materials: (1) natural roofing (no roof, thatch, sod, and bamboo); (2) rudimentary roofing (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 | Gendered household type (de jure household members)  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, hh\_wgt, samp\_stratum, ahtype* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *dw\_roof* |

#### Calculations

**Step 1.** Create a categorical variable that flags the household that matches each of the specified main dwelling roof materials (*dw\_roof*). Note that the types of roof 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\_roof=1 if v201=11 or 12 or 13 or 14*

*Replace dw\_roof=2 if v201=21 or 22*

*Replace dw\_roof=3 if v201=31 or 32 or 33 or 34 or 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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab dw\_roof*

*Svy: tab dw\_roof genhhtype\_dj, col*

## 8.4 Percent distribution of households by dwelling exterior wall materials

This indicator is the percentage distribution of households by specified categories of main exterior dwelling wall materials: (1) natural wall materials (no walls, dirt, cane, palm or tree trunks, bamboo with mud, and stone with mud); (2) rudimentary wall materials (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 | Gendered household type (de jure household members)  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 | *v203, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *dw\_wall* |

#### Calculations

**Step 1.** Create a categorical variable that flags the households in each specified main wall material category (*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 or 12 or 13 or 14 or 15*

*Replace dw\_wall=2 if v203=21 or 22 or 23 or 24*

*Replace dw\_wall=3 if v203=31 or 32 or 33 or 34 or 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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab dw\_wall*

*Svy: tab dw\_wall genhhtype\_dj, col*

## 8.5 Percent distribution of households by dwelling floor materials

This indicator is the percentage 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 | Gendered household type (de jure household members)  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, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *dw\_floor* |

#### Calculations

**Step 1.** Create a categorical variable that flags the households that match each of the main floor materials (*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 or 12 or 13*

*Replace dw\_floor=2 if v202=21 or 22*

*Replace dw\_floor=3 if v202=31 or 32 or 33 or 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 gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=hh\_wgt], strata(samp\_stratum)*

*Svy: tab dw\_floor*

*Svy: tab dw\_floor genhhtype\_dj, 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

# The consumption aggregate and the poverty indicators

In this chapter, the Feed the Future poverty indicators are defined and their calculation described. The 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. Please note that the second section of this chapter (Section 9.2) has been removed while the step-by-step procedures to calculate the Feed the Future ZOI-level poverty indicators are updated. Please contact [RFS.ALD.DMA@USAID.GOV](mailto:RFS.ALD.DMA@USAID.GOV) for further information.

## 9.1 Guidelines to construct the poverty indicators

This section includes two subsections. The first section provides information on how to create the consumption aggregate, which is required to calculate the Feed the Future poverty indicators, and the second section provides information on how to construct the poverty indicators using the consumption aggregate.

### 9.1.1 Creating 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, and it may be 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 because 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)[[31]](#footnote-30) and Grosh & Muñoz (1996),[[32]](#footnote-31) 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 the ZOI Survey questionnaire, consumption expenditures are grouped into four categories in six sub-modules, based on characteristics and likely frequency of purchase or recall periods:

* 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. 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.

#### 9.1.1.1 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 will be missing observations and outliers remaining in the dataset. Respondents may not be able to provide a quantity or value for an item that has been consumed by the household, either because they do not remember or do not know. 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.

Analysts need to carefully document all cases they found missing or determined to be outliers, what protocol was followed to check for outliers, and how imputation was done. The frequency and geographical distribution of missing values and outliers by item should be noted, because they may reveal issues with data collection or data entry, which should then be investigated further and corrected if possible. Analysts also need to keep 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 paramount.

The same general procedure is used to replace an outlier or a missing value. We first describe below how to impute a value to replace a missing observation or an outlier and then provide general guidelines on checking for and verifying outliers.

#### 9.1.1.2 Procedures for imputed values

If an observation is missing or determined otherwise to be an outlier (see Section 9.9.2, Treatment of outliers), it should be replaced (imputed) by the median[[33]](#footnote-32) per capita value within the smallest geographic unit (starting with the enumeration area or cluster) that has at least five observations.

The steps to do this are as follows:

* Calculate the per capita daily expenditures per household at the item or total level, depending on where imputation is needed, by dividing valid consumption expenditures data by the number of household members and then by the number of days in the recall period.
* For unit value in the food consumption sub-module, sum the monetary value by total quantity consumed, using valid household-level data by item.
* If there are five or more observations in a cluster, calculate the median value at the cluster level.
* If there are less than five observations in a cluster, calculate the median value at the next lowest administrative unit[[34]](#footnote-33) level.
* Use the median per capita daily value or the unit value at the lowest administrative level possible to impute a missing value or a confirmed outlier.

If the total number of observations in the entire sample is less than five, no imputation should be attempted. All observations should be reviewed together, and a determination should be made as to whether to include the item and if so, what per capita median value is plausible.

The per capita median value calculated should be compared with that from other (comparable on some level) administrative units to make sure that it is not markedly different. If it seems too large or too small compared to other clusters, it should be checked for plausibility. If it does not seem plausible, the per capita median values of the next administrative level should be used instead.

Note that in the food sub-module, the respondent is asked to provide both volume and values consumed overall and by sources. To the extent possible, the analyst should replace a missing value or an outlier using unit values calculated at the household or the cluster level. See the section on the food sub-module below.

#### 9.1.1.3 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 range as a total but not as a per capita.

As a general rule, if a value is more than three SD from its 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 UDS $500 on a smart phone and Y spent just UDS $30 on a basic flip phone). There is no established protocol for further examination, and analysts should develop their own protocol and document every step carefully. Other questions can be used to triangulate a particular response or provide context. For instance, at the end of the Food Consumption over the Past 7 Days sub-module, (questions 8.108 to 8.111), the respondent is asked about non-members of the household eating meals in the household. 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 will have a large impact on the survey statistics.

If, after further examination, an observation is determined to be an outlier, it should be coded as missing and then replaced with an imputed value following the general rules described in Section 9.2.1.

In the following sections, we describe some specific procedures to follow for each of the four expenditure categories (food, non-food, durables, and housing) in computing the consumption aggregate, including additional, category-specific outliers and missing value instructions for the food consumption data.

#### 9.1.1.4 Food consumption

The food item sub-module records consumption of a comprehensive list of country-specific foods. In this sub-module, both volume and monetary values are recorded. The interviewer asks whether the item was consumed and, if so, what quantity did the household consume 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. In each case, the respondent is asked to give a corresponding value, in local currency units (LCUs). 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 (8.103b, 8.104b, 8.106b, and 8.107b). 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 (8.103a=8.104a+8.106a+8.107a).[[35]](#footnote-34) If the sum is different from the total quantity consumed, adjust the total quantity consumed (8.103a) to the sum.
3. Compute the unit value for an item by dividing the amount spent on purchases by the quantity consumed from purchases (8.105 divided by 8.104a), if they both have positive values. If either one is 0 or blank, 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).

*Missing food item observations[[36]](#footnote-35)*

Because the respondent is asked to provide both volume and values consumed overall and by sources, unit value can be calculated and used to impute missing values. There are different cases of missing values possible in this sub-module.

1. *The respondent reports that an item was consumed (8.102=1), but monetary values are missing (8.105, 8.106c, and 8.107c are missing).*

In this case, a unit value cannot be calculated at the household level. Instead, a unit value for the item must be obtained at the cluster level or at the lowest administrative unit level possible. To calculate a cluster-level unit value, use the following steps:

* Calculate per capita total monetary value per household for the items for which imputation is needed.
* Calculate per capita total quantity consumed per household for the same items.
* Divide monetary value by quantity consumed per household to obtain the household-level unit value.
* If there are five or more observations in a cluster, estimate the median unit value at the cluster level.
* If there are less than five observations in a cluster, estimate the median value at the next lowest administrative unit[[37]](#footnote-36) level.
* To impute a value for a particular household in which a reported value was found to be an outlier or missing, multiply the cluster-level (or the lowest administrative level possible) median unit value by the quantity reported to obtain the missing monetary value or divide the total monetary value by the cluster-level median unit value to obtain the missing quantity values.

1. *The respondent reports consuming a food item (8.102=1), but no information is available on quantities or monetary value from any source (all other questions are blank).*

In this case, the median per capita monetary value across all households in the smallest geographic unit should be calculated for the item in question and used to impute the missing observation following the general procedure outlined in Section 9.2.1.

*Further testing for extreme values in the food consumption sub-module*

After missing observations and confirmed outliers have been replaced as appropriate (see Sections 9.2.1 and 9.2.2), it is useful to compute some overall statistics from the food consumption sub-module to help identify additional outliers.

* For each household, sum monetary values by source for each food item.
* Sum across food items to obtain the total value of food consumption by household.
* Divide the total value by the number of household members to obtain the per capita food consumption by household.
* Divide the per capita food consumption by 7 days to obtain the daily per capita food consumption by household.
* Calculate the sample mean and SD of daily per capita food consumption across all households.
* Flag household-level values that are less than -3 SD or greater than 3 SD from the mean as potential outliers.[[38]](#footnote-37)
* Look at the frequency distribution of the flagged cases to understand the pattern of outliers.
* Calculate the median value of daily per capita food consumption across all households, after removing outliers at various geographic levels.
* Replace confirmed outliers with the relevant median values, following the decision rule described in Section 9.2.1.

After remaining outliers are smoothed, the sample distribution should look relatively normal. It is also useful at this point to calculate the relative importance of food items consumed from the different sources. We should not expect, for instance, many households consuming staple food items obtained only from “gifts and other sources,” unless the sample population is impacted by disaster or other hardships. We can expect rural poor households to consume a significant proportion of the main food crops from home production, however, and to observe that in general the proportion of food purchased to food non-purchased goes up as income goes up. If these expectations are not met, the data are not necessarily faulty, but further investigation is recommended.

#### 9.1.1.5 Non-food, non-durable goods

Consumption expenditures information for non-food, non-consumer goods is collected for different recall period: 7 days, 1 month, 3 months, and 12 months. Items such as charcoal or kerosene, candles, and public transport are regular purchases and are better recalled on a weekly basis. Soap and cosmetics, petrol, wages to servants, utilities, and cell phone charges, 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 are collected on a 12-month recall basis. Because of these different time periods, calculating total consumption of non-food items requires first converting the values reported to a common reference period.

Not all non-food items for which a household has incurred expenses contribute to household consumption in the sense of enhancing its welfare. Expenditures on items such as clothing, footwear, beauty, or recreation items increase household consumption and therefore should be included in the consumption aggregate. However, expenditures on items such as debt payment, fines, or legal fees do not lead to higher consumption and thus should be excluded. Expenditures that are made as productive investments or as part of the household production activities, such as purchase of a bus, fertilizers, or agriculture seeds, do not improve household consumption at present and should be excluded.

Finally, it is recommended to 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 rest house or hotel (item 8275)
* Fines, legal costs, marriage and funeral costs (items 8299 to 8303)

Following recommendations in Deaton & Zaidi (2002), education expenditures (items 8310 to 8318) should be included.

In the non-food, non-durable sub-module, the interviewer asks whether the respondent or someone in the household has bought the item over the recall period. If the respondent answers yes, then the interviewer asks how much was bought in total. Because of the heterogeneity of the goods included here, data on quantity are not collected. The only exceptions are a few items that could be gathered in the wild as well as, or instead of, purchased, such as wood poles, bamboo, or grass for thatching roofs (items 8319 to 8321). For these items, 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 or the cost for any amount purchased.

#### 9.1.1.6 Consumer durables

Purchases of durable goods 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, the overall reference period of the survey. Deaton & Zaidi (2002) argue that the value of services that the household receives from the 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:

Where: is the current value of the item, is the real rate of interest, and is the depreciation rate for the durable goods. Each of these components is computed separately as follows:

* The current value of the item (: This is the value of the item in its current shape (second-hand). If the household owns more than one item (question 8.705), the respondent is asked to report an average value.
* The real rate of interest (): 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[[39]](#footnote-38) 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 Global Food Security Strategy (GFSS) target countries,[[40]](#footnote-39) time series are available for all but three countries.[[41]](#footnote-40) If a continuous real interest rate series is not available for at least 10 years (and up to 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 (using the Consumer Price Index [CPI]) for each year should be subtracted from the nominal interest rate () and then averaged over as many years as possible:
* The rate of depreciation ( is calculated for each item[[42]](#footnote-41) as:

Where is the current average price (average value as reported by the respondent) of the item, is the average price of the item when purchased, and is the average age of the item in years.[[43]](#footnote-42)

Note that it is possible for a depreciation rate to be negative if a good has gained value over time (as in the case of investment goods, such as art or antiques). This, however, 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, after being examined for plausibility, are replaced by the consumer durable-specific median value of the smallest geographic unit (see Section 9.2.1). 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 (owned by the household, if more than one, as reported in question 8.703.

The total value of annual consumption of durable goods ( consumed by household is calculated as:

The per capita daily rental equivalent of durable goods consumed by the household is calculated by dividing the annual value by 365 days and by the number of household members.

#### 9.1.1.7 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 should be treated as an investment. However, imputing the value of housing services is not straightforward,**[[44]](#footnote-43)** especially if the housing market is not well developed in the areas where the survey is implemented.

If the household pays a non-subsidized rent for its dwelling, rent payment (8.605) 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 (8.601=1–4), a “rental equivalent” needs to be estimated.

If the household lives in an employer-provided house or lives in a house for free, the respondent is asked to provide an estimate of how much monthly rent could be charged, if the house were to be rented today (8.604). This rent estimate should be checked for plausibility,[[45]](#footnote-44) because it will be used to impute a value for housing consumption.

Questions 8.606 to 8.608 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 that some expenses 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.[[46]](#footnote-45) This model is developed by regressing available observations and estimates of rental value on a series of dwelling characteristics, including location, 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 the most commonly used functional form and typically performs better than a linear form.[[47]](#footnote-46)

Where:

is the rent paid or rent estimate of housing unit i

is a set of characteristics or attributes of housing unit i

and are the coefficients to estimate in the hedonic model, and

is the error term

There are several characteristics or attributes (the set of independent variables ) of the dwelling that can be fitted in the equation:

* Structural attributes: material of the exterior walls (question 203), floors (question 202), and roof (question 201); number of rooms (question 204[[48]](#footnote-47)); type of toilet (question 208); source of drinking water (question 211); access to electricity (question 222a); age of the house (question 8.603[[49]](#footnote-48)); etc.
* Location: district (question 05) or region (question 06); rural or urban[[50]](#footnote-49)
* If available, distance from an improved road; distance from a market[[51]](#footnote-50)

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 . 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 too much of a concern.

The equation with the estimated regression coefficients () 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 8.609 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 for renovation incurred in the past month that would otherwise be infrequent. In such a case, it may be better to replace the reported value with the median value of the other households within the smallest geographical unit with data on monthly expenditures on repairs and maintenance (see Section 9.2.1 on the general procedure to impute a value).

#### 9.1.1.8 Consumption aggregate

The consumption aggregate is obtained by adding daily expenditures across all food and non-food categories, to obtain total daily expenditures and total daily per capita expenditures for each household in the sample. For each household, total daily consumption expenditures are divided by the number of household members to obtain the per capita consumption level. In this approach, every household member is assumed to have an equal share of total consumption, regardless of age and other household member characteristics.

Where:

is the daily consumption aggregate of household

is the daily consumption expenditure on item by household

is the daily per capita consumption aggregate of household

is the number of household members in household

### 9.1.2 Poverty indicators

After the aggregation is completed, the various poverty indicators can be estimated. We are presenting how to calculate both the set of poverty indicators for the Feed the Future phase one endline survey and the set of poverty indicators for the Feed the Future phase two baseline survey.[[52]](#footnote-51)

The Feed the Future phase two poverty indicators are:

* Prevalence of poverty at $1.90 per capita per day 2011 purchasing power parity (PPP)
* Depth of poverty of the poor at $1.90 per capita per day 2011 PPP
* Prevalence of people who are “near-poor” at $1.90 per capita per day 2011 PPP

The Feed the Future phase one poverty indicators are:

* Prevalence of poverty at $1.25 per capita per day 2005 PPP
* Depth of poverty at $1.25 per capita per day 2005 PPP
* Average per capita expenditure in constant 2010 $US

#### 9.1.2.1 Prevalence of poverty: Percent of people living on less than $1.90 per day 2011 PPP

The prevalence of poverty, or poverty headcount ratio, is the proportion of the population in the survey area living below the international extreme poverty line, defined as a daily per capita consumption of less than U.S. dollars (USD) $1.90 at 2011 PPP prices.

The steps to calculate the prevalence of poverty indicator are as follows:

1. Convert the USD $1.90 line into the LCU, using the 2011 PPP conversion factor of private consumption based on the International Comparison Program 2011.[[53]](#footnote-52) The PPP 2011 conversion factors can be obtained from the World Development Indicator database (World Bank: <http://databank.worldbank.org/data/source/world-development-indicators>) and are provided in Table 4 for the GFSS target countries.
2. Adjust the poverty line in LCU for the inflation from 2011 to the year and month of the survey. In all cases, the official source for the CPI should be used. The formula to adjust the poverty line for inflation is:

Where the subscript refers to the month and year of the ZOI Survey.

1. Calculate the prevalence of poverty:

* Calculate the numerator as the sum of sample-weighted number of household members in sampled households in which the per capita daily consumption is less than the poverty line (less than ).
* Calculate the denominator as the sum of sample-weighted number of household members in the sampled households with consumption data.
* Multiply the ratio by 100 to obtain a percentage.

Where:

is the number of household members in household

is the sample weight of household

is the number of households in the sample

is a parameter that takes on the following values:

if

if .

**Table 4: PPP 2011 Conversion Factor, Private Consumption (LCU per international $)**

| **GFSS Target Countries** | **PPP 2011** |
| --- | --- |
| Bangladesh | 24.849 |
| Ethiopia | 5.439 |
| Ghana | 0.788 |
| Guatemala | 3.873 |
| Honduras | 10.080 |
| Kenya | 35.430 |
| Mali | 221.868 |
| Nigeria | 79.531 |
| Niger | 228.753 |
| Nepal | 25.759 |
| Senegal | 246.107 |
| Uganda | 946.890 |

*Source: World Bank, World Development Indicators, Series PPP conversion factor, private consumption (LCU per international $)*

#### 9.1.2.2 Depth of poverty of the poor: Mean percent shortfall of the poor relative to the $1.90/day 2011 PPP poverty line

The depth of poverty of the poor measures how far below the USD $1.90/day 2011 PPP poverty line poor people are. This indicator differs from the depth of poverty indicator that is used by the World Bank and was used under Feed the Future phase one (see Section 9.1.2.5). The difference is that this indicator only tracks the depth of poverty of households that are under the poverty line. In other words, it **does not include all households:** households whose 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 (), 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 household members and then sum across all poor households.
3. Sum the sample-weighted number of 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.

With:

Where:

is the proportional shortfall of poor household

is the daily per capita consumption aggregate (in LCU) of poor household

is the number of household members in poor household

is the weight assigned to household

is the USD $1.90 poverty threshold converted to local currency at current prices for the year and month of the survey

is the number of poor households in the sample

#### 9.1.2.3 Prevalence of people who are “near-poor,” living on 100 percent to less than 125 percent of the $1.90 2011 PPP poverty line

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. The applicable poverty line is USD $1.90 per person per day at 2011 PPP. Therefore, the near-poor individuals are those with daily consumption expenditures equal or greater than USD $1.90 2011 PPP but less than USD $2.38 per day at 2011 PPP**.**

The steps to calculate this indicator are essentially the same as for the prevalence of poverty at USD $1.90 2011 PPP, except for the numerator, which is the sum of individuals that are defined as near poor. The upper limit of the near-poor range (USD $2.38 2011 PPP) also needs to be converted to LCU and adjusted for inflation for the year and month of the survey, using the approach outlined in Section 9.4.1.

Where:

is the number of household members in household

is the sample weight of household

is the number of households in the sample

is a parameter that takes on the following values:

if

if .or

#### 9.1.2.4 Prevalence of poverty: Percent of people living on less than $1.25/day 2005 PPP

The baseline surveys of Feed the Future phase one were conducted before the international poverty line was adjusted to USD $1.90/day following the realignments of the purchasing power parities in 2010. Until then, the international extreme poverty line was USD $1.25/day at the 2005 PPP. This indicator is essentially the same, except for the threshold.

The steps to calculate the prevalence of poverty indicator are as follows.

1. Convert the USD $1.25 line into the LCU, using the 2005 PPP conversion factor of private consumption based on the International Comparison Program 2005. The PPP 2005 conversion factors cannot be obtained from the World Development Indicator database (World Bank: <http://databank.worldbank.org/data/source/world-development-indicators>) and therefore are provided in Table 5 for the former focus countries.
2. Adjust the poverty line in LCU for the inflation from 2005 to the year and month of the survey, using the official CPI source:

Where the subscript refers to the year and month of the ZOI Survey.

1. Calculate the prevalence of poverty:

* Calculate the numerator as the sum of sample-weighted number of household members in sampled households in which the per capita daily consumption is less than the poverty line (less than ).
* Calculate the denominator as the sum of the sample-weighted number of household members in the sampled households with consumption data.
* Multiply the ratio by 100 to obtain a percentage.

Or:

Where:

is the number of household members in household

is the sample weight of household

is the number of households in the sample

is a parameter that takes on the following values:

if

if .

**Table 5: PPP 2005 Conversion Factor, Private Consumption (LCU per international $)**

| **Feed the Future**  **Former Focus Countries** | **2005 PPP** |
| --- | --- |
| Bangladesh | 25.494 |
| Cambodia | 1615.298 |
| Ethiopia | 2.751 |
| Ghana | 0.448 |
| Guatemala | 4.540 |
| Haiti | 19.365 |
| Honduras | 9.662 |
| Kenya | 32.684 |
| Liberia | 0.511 |
| Malawi | 56.922 |
| Mali | 289.679 |
| Mozambique | 11.626 |
| Nepal | 26.467 |
| Rwanda | 236.745 |
| Senegal | 298.245 |
| Tajikistan | 0.927 |
| Tanzania | 482.451 |
| Uganda | 744.618 |
| Zambia | 2.830 |

*Source: World Bank, World Development Indicators, Series PPP conversion factor, private consumption (LCU per international $)*

#### 9.1.2.5 Depth of poverty: Mean percent shortfall relative to the $1.25/day 2005 PPP poverty line

The depth of poverty, also referred to as the poverty gap index, measures the average gap between the consumption levels of the population and the poverty line, with the gap for non-poor set equal to zero. The depth of poverty is expressed as a percentage of the poverty line. This indicator can be interpreted as the per capita cost across the entire population required to lift everyone below the poverty line to the poverty line, if resources were perfectly targeted to the poor.

The applicable poverty line for this indicator is USD $1.25 2005 PPP.

To calculate this indicator using the ZOI Survey dataset, use the following steps:

1. Using the poverty line converted to local currency and adjusted for inflation (), 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 household members and then sum across all poor households.
3. Sum the sample-weighted number of household members in sampled households with consumption data.
4. Divide (2) by (3) and multiply by 100 to obtain the depth of poverty expressed as a percentage of the USD $1.25 per person per day poverty line.

With:

Where:

is the proportional shortfall of poor household ;

is the daily per capita consumption aggregate (in LCU) of poor household ;

is the number of household members in poor household ;

is the weight assigned to household ;

is the USD $1.25 poverty threshold converted to local currency at current prices for the year and month of the survey;

is the number of poor households in the sample; and

is the total number of households in the sample.

#### 9.1.2.6 Average daily per capita expenditures in constant 2010 USD

For this indicator, we need to convert the daily per capita consumption expenditures data that are collected in LCU into constant 2010 USD.

The steps are as follows:

1. Multiply the per capita consumption of each sampled household to 2005 prices, by multiplying by the ratio of the 2005 CPI and the CPI for the year and month of the survey (time=) to bring these to 2005 prices:
2. Convert them into 2005 USD by dividing by the 2005 PPP conversion rate.[[54]](#footnote-53)
3. Bring them into constant 2010 USD prices by multiplying by the CPI ratio between 2010 and 2005: [[55]](#footnote-54)
4. Estimate the average daily per capita expenditures by summing sample-weighted per capita daily consumption expressed in constant 2010 USD across sampled households with consumption data, and then dividing by the sum of sample-weighted number of household members in the sample:

Where

is the sample weight of household , and

is the number of household members in household .

## Step-by-step procedure to calculate the poverty indicators

This section of the poverty chapter has been removed while the step-by-step procedures to calculate the Feed the Future ZOI-level poverty indicators are updated. Please contact [RFS.ALD.DMA@USAID.GOV](mailto:RFS.ALD.DMA@USAID.GOV) for further information.

## References

Balcazar, C.F., Ceriani, L., Olivieri, S., and 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., and 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., and 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>

# The comparative wealth index

Although the Demographic and Health Surveys (DHS) Wealth Index, an absolute 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 comparative wealth index.

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 asset-based wealth index. 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.

## 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.

1. **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 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, BFS examined the World Bank’s 2018 gross national income per capita estimates in current U.S. dollars calculated using the Atlas method[[56]](#footnote-55) for countries classified as low or lower-middle-income countries by the World Bank.[[57]](#footnote-56) The median gross national income per capita (GNI/p) (Atlas method, current USD) for these countries was determined to be USD $1,580. BFS then determined when countries with a GNI/p close to this median GNI/p last conducted a DHS survey. BFS chose the Senegal 2017 DHS as the reference survey because the country’s 2018 GNI/p (Atlas method, current USD) was USD $1,1410—not far below the median—and because the survey was one of the most recent DHS surveys for which data were publicly available. The Senegal 2017 DHS will be used as the reference survey to calculate the Feed the Future CWI indicator across all ZOI Surveys.

1. **Calculation of the AWI for the reference survey:** The AWI for the Senegal 2017 DHS was calculated following the method developed by DHS.[[58]](#footnote-57),[[59]](#footnote-58) 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.[[60]](#footnote-59) The steps to calculate the AWI for the reference survey are the same steps used to calculate 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 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 AW1 Scores** | **Quintile Cutoffs** |
| --- | --- | --- |
| Lowest | -1.706648 - -0.8628537 | -0.86276845 |
| Second | -0.8626832 - 0.0010609 | 0.00108905 |
| Middle | 0.0011172 - 0.8036242 | 0.80370025 |
| Fourth | 0.8037763 - 1.287166 | 1.2871975 |
| Highest | 1.287229 - 2.907386 |  |

1. **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,[[61]](#footnote-60) 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.

*Logit(p)=ln(p÷(1-p))=a+b\*AWI*

Where p is the probability that the household owns the respective item, p=prob(Y=1), and AWI is the wealth index score. Y is the 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 α and β:

*ln[0.5÷1-0.5]=α+β\*AWI*

*ln(1)=α+β\*AWI*

*0=α+β\*AWI*

*AWI=-α÷β*

The UBN anchoring points are obtained using the UBN framework, which defines a set of basic needs and assigns 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 7 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 7: Unsatisfied Basic Needs Categories and Criteria**

| **UBN Category** | **UBN Point Assigned to Household If:** |
| --- | --- |
| **Inadequate dwelling construction** | The household resides in a dwelling with inadequate walls or floors. Inadequate walls are made of natural or rustic materials, and inadequate floors are made of earth, sand, dung, manure, or dirt. |
| **Inadequate sanitation** | 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, or sharing a facility—even an improved facility—with other households. Inadequate drinking water sources include unprotected wells, unprotected springs, rainwater, tanker trucks, small carts with a tank, and surface water. |
| **Overcrowded housing** | There are four or more de jure household members per sleeping room. |
| **High economic dependencya** | (a) There are no de jure working-age adults (age 15–64) 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 decisionmakerb has not completed primary education. |

a 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.

b 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 decisionmaker and the primary adult female decisionmaker in each household, so these data were used instead. Note that a household may have only a female primary adult decisionmaker, only a male primary adult decisionmaker, both, or neither.

Eight anchoring points—four asset and four UBN—were calculated using the reference survey data. These anchoring points, shown below in Table 8, will be used in the Feed the Future CWI indicator calculation across all ZOI Surveys.

**Table 8: Reference Survey Anchoring Point Values**

| **Type** | **Anchoring Point** | **Value for Reference Survey** |
| --- | --- | --- |
| Asset | Television | 0.1166113 |
| Refrigerator | 1.107843 |
| Computer | 1.591285 |
| Car or truck | 6.366876 |
| UBN | Inadequate dwelling construction | 1.284437 |
| Inadequate sanitation | 0.6740997 |
| Overcrowded housing | -0.6691726 |
| High economic dependency | -1.329942 |

1. **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.[[62]](#footnote-61),[[63]](#footnote-62) The step-by-step instructions to calculate the AWI for the ZOI Survey being analyzed are included in the step-by-step procedures.
2. **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.
3. **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 (*basecuti*) as the dependent variables and the ZOI Survey anchoring points as the independent variables (*compcuti*) to obtain α and β:

*basecuti=α+β\*compcuti*

Where *basecuti* is the value of the anchoring point on the baseline AWI of item I, and *compcuti* is the value of the anchoring point on the ZOI Survey AWI.

The constant α represents the amount of adjustment of the ZOI Survey AWI relative to the baseline survey AWI, and β represents the dispersion of the ZOI Survey AWI relative to the baseline survey AWI.

The regression coefficient β and constant α 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 β, plus the constant:

*CWI=α+β\*AWI*

1. **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.

## Step-by-step procedures to calculate the CWI indicator

#### 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 (de jure household members)\*  Household education  Poverty status  Shock exposure index |
| Coverage | P2-ZOI |
| 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 and CWI.

##### Part 1. AWI

| Survey variables used | *hhea, hhnum, ahtype, v103, v201, v202, v203, v208, v209, v211, v219, v222a-v222f, v223a-v223g, v224, v225, v226a-v226g, v240a, v240b, v8601* |
| --- | --- |
| Analytic variables used | *memsleep\_dj* |
| Analytic variables created | *memsleep, land, landarea, house, domestic, bankacct, water\_11-water\_96, toilet\_111-toilet\_962, foor\_11-floor\_96, roof\_11-roof\_96, wall\_11-wall\_96, cookfuel\_1-cook\_fuel\_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\_9, cat\_fish10\_29, cat\_fish30, v222ax-v222fx, v223ax-v223gx, landarea\_mean, com, urb, rur, rur\_const, rur\_coeff, urb\_const, urb\_coeff, combscor, awi, hhmemwgt, awiquint* |

**Step 1.** Create the variables needed to construct the AWI in the household-level analytic data file.

**Step 1a.** 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*)*.[[64]](#footnote-63)*

*Set memsleep=integer(memsleep\_dj)*

**Step 1b.** Create a variable that indicates whether the household owns agricultural land (*land*). Set missing responses to “no.”[[65]](#footnote-64)

*Set land=0*

*Replace land=1 if v240a=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Household owns agricultural land”*

**Step 1c.** 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 mean in the later 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 1d.** Create a variable that indicates whether the household owns its dwelling (*house*). If a household is missing this information or the response is “don’t know,” consider the household to not own its dwelling.[[66]](#footnote-65)

*Set house=0*

*Replace house=1 if v8601=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Household owns dwelling”*

**Step 1e.** 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 the list 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 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”*

*Label variable water\_91 “Bottled water”*

*Label variable water\_96 “Other water source”*

**Step 1f.** 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 the list 11-15 21-23 31 41 51 96:*

*For each value (l) of the list 1 2:*

*Set toilet\_`i'\_`l'=0*

*Replace toilet\_`i'\_`l =1 if v208=`i' and v209=`l'*

*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 DK 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 1g.** 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 the list 11-13 21 22 31-35 96:*

*Set floor\_`i'=0*

*Replace floor\_`i'=1 if v202=`i'*

*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 1h.** 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 the list 11-14 21 22 31-36 96:*

*Set roof\_`i'=0*

*Replace roof\_`i'=1 if v201=`i’*

*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 1i.** 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 the list 11-15 21-24 31-36 96:*

*Set wall\_`i'=0*

*Replace wall\_`i'=1 if v203=`i'*

*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”*

*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 1j.** 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 the list 1-11 95 96:*

*Set cookfuel\_`i'=0*

*Replace cookfuel\_`i'=1 if v219=`i'*

*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 1k.** Create a categorical variable for each farm animal that indicates the number that the household owns. Set “don’t know” responses to be 95 (95 or more animals). Set missing responses to be 0. In the template syntax, large animals are categorized into four categories (0, 1–4, 5–9, 10+), and small animals are categorized into four categories (0, 1–9, 10–29, 30+). 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.

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*

*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*

*Set cat\_cattle1\_4=0*

*Replace cat\_cattle1\_4=1 if num\_cattle≥1 and num\_cattle≤4*

*Label variable “HH owns 1-4 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 cattle”*

*Set cat\_cattle10=0*

*Replace cat\_cattle10=1 if num\_cattle≥10 and num\_cattle≤98*

*Label variable “HH owns 10+ cattle”*

*Label values 0 “No” 1 “Yes”*

*Set num\_horse=0 if v225=2 or v226c=missing*

*Replace num\_horse=v226c if v226c≤98*

*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*

*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*

*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 v226e=missing*

*Replace num\_poultry=v226e if v226e≤98*

*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 v226e=missing*

*Replace num\_fish=v226e if v226e≤98*

*Set cat\_fish1\_9=0*

*Replace cat\_fish1\_9=1 if num\_fish≥1 and num\_fish≤9*

*Label variable “HH owns 1-9 fish”*

*Set cat\_fish10\_29=0*

*Replace cat\_fish10\_29=1 if num\_fish≥10 and num\_fish≤29*

*Label variable “HH owns 10-29 fish”*

*Set cat\_fish30=0*

*Replace cat\_fish30=1 if num\_fish≥30 and num\_fish≤98*

*Label variable “HH owns 30+ fish”*

*Label values 0 “No” 1 “Yes”*

**Step 1l.** Create new binary variables for each asset included in the survey (*v222ax–v222fx* and *v223ax–v223gx*) so that “no” responses (2), “don’t know” responses (8 or 98), “missing” responses (9 or 99), and missing values all have a value of 0.

*For each variable (var) in the variable list v222a v222b v222c v222d v222e v222f v223a v223b v223c v223d v223e v223f v223g:*

*Set `var'x=0 if `var'=2 or 8 or 9 or 99 or 95 or 98 or missing*

*Replace `var'x=1 if `var'=1*

*Label variable v222ax “Household has electricity”*

*Label variable v222bx “Household has a radio”*

*Label variable v222cx “Household has a television”*

*Label variable v222dx “Household has a non-mobile phone”*

*Label variable v222ex “Household has a computer”*

*Label variable v222fx “Household has a refrigerator”*

*Label variable v223ax “Household member has a watch”*

*Label variable v223bx “Household member has a mobile phone”*

*Label variable v223cx “Household member has a bicycle”*

*Label variable v223dx “Household member has a motorcycle or motorscooter”*

*Label variable v223ex “Household member has an animal-drawn cart”*

*Label variable v223fx “Household member has a car or truck”*

*Label variable v223gx “Household member has a boat with motor”*

**Step 1m.** Create a binary variable that indicates whether any household member has a bank account (*bankacct*) so that “no” responses (2), “don’t know” responses (8), “missing” responses (9), and missing values all have a value of 0.

*Set bankacct=0 if v224=2 or 8 or 9 or missing*

*Replace bankacct=1 if v224=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Household member has bank account”*

**Step 2.** Create a binary variable that indicates whether the household had a maid or servant (*domestic*).[[67]](#footnote-66)

**Step 2a.** 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 2b.** 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*

**Step 2c.** 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 household members*

*By cluster and household: replace domestic=0 if domestic=0 for all household members*

*Save “temp\_domestic”*

**Step 2d.** Load the temporary AWI data file created in Step 2a, and add the *domestic* variable from the temporary data file created in Step 2c.

*Load “temp\_awi.dta”*

*Add domestic from “temp\_domestic.dta”, matching cluster (hhea) and household number (hhnum) as key variables*

**Step 3.** Determine the mean land area value and substitute the mean value for any missing land area values.

*Set landarea\_mean=mean(landarea)*

*Replace landarea=landarea\_mean if landarea=missing*

**Step 4.** Determine which variables created in Steps 1–3 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. To do this, run the frequencies for all the created indicator variables and flag any variables that have no variation.[[68]](#footnote-67)

**Step 4a.** Exclude all variables identified as having no variation—that is, exclude all binary variables that have a value of 0 for all households or that have a value of 1 for all households. For example, for water source, if the variable for unprotected spring (*water\_42*) has no ‘Yes’ responses, be sure to exclude the variable from the factor analysis.

**Step 4b.** 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 9), 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 9).

**Table 9: Example Cases that Show Variables with 0 or 1 ‘Yes’ Reponses**

| **Case 1** | | **Case 2** | |
| --- | --- | --- | --- |
| **Variable** | **# ‘Yes’ Responses** | **Variable** | **# ‘Yes’ Responses** |
| *water\_32* (unprotected well ) | 1 | *water\_32* (unprotected well ) | 0 |
| *water\_42* (unprotected spring) | 1 | *water\_42* (unprotected spring) | 1 |
| *water\_96* (other) | 0 | *water\_96* (other) | 1 |
| Action: combine *water\_32* and *water\_42* | | Action: combine *water\_42* and *water\_96* | |

*For each variable (var) in variable list domestic memsleep\_dj land landarea house bankacct cat\_cow1 cat\_cow2 cat\_cow3 cat\_cattle1 cat\_cattle2 cat\_cattle3 cat\_horse1 cat\_horse2 cat\_horse3 cat\_goat1 cat\_goat2 cat\_goat3 cat\_sheep1 cat\_sheep2 cat\_sheep3 cat\_poultry1 cat\_poultry2 cat\_poultry3 cat\_fish1 cat\_fish2 cat\_fish3 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\_91 water\_96 toilet\_111 toilet\_11 toilet\_121 toilet\_12 toilet\_131 toilet\_13 toilet\_141 toilet\_14 toilet\_151 toilet\_15 toilet\_211 toilet\_21 toilet\_221 toilet\_22 toilet\_231 toilet\_23 toilet\_311 toilet\_31 toilet\_411 toilet\_41 toilet\_511 toilet\_51 toilet\_961 toilet\_96 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\_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 v223gx*

*Summarize `var' [examine number of observations, mean, standard deviation, minimum,*

*maximum]*

*Tabulate `var' [examine number of observations and variable values]*

**Step 5.** Create a list of all variables that will be included in the wealth index, after excluding or combining variables in Step 4.

**Step 6.** Createa list that includes all variables thought to have the same relationship with the underlying economic status in both urban and rural areas. 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 from the list of variables created in Step 5. 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 7.** Run the common factor analysis using the variable list created in Step 6, and save the component scores as a new variable (*com*).[[69]](#footnote-68) (Note that ‘factor, pcf’ and ‘predict’ are Stata commands used in the template syntax file.)

*Factor variable list created in Step 6, pcf factors(1)*

*Predict com, no rotation*

**Step 8.** Createa list that includes all variables relevant to the economic status in urban areas (*ahtype*=1). 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 SD 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 all variables in the list created in Step 5:*

*Summarize `var' if ahtype=1*

*Tabulate `var' if ahtype=1*

**Step 9.** Run the common factor analysis using the variable list created in Step 6, and save the component scores as a new variable (*urb*).[[70]](#footnote-69) (Note that ‘factor, pcf’ and ‘predict’ are Stata commands used in the template syntax file.)

*Factor variable list created in Step 8 if ahtype=1, pcf factors(1)*

*Predict urb if ahtype=1, no rotation*

**Step 10.** Createa list that includes all variables relevant to the economic status in rural areas (*ahtype*=2). 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, but with a different relationship to wealth. If a variable has no SD 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 all variables in the list created in Step 5:*

*Summarize `var' if ahtype=2*

*Tabulate `var' if ahtype=2*

**Step 11.** Run the common factor analysis using the variable list created in Step 6, and save the component scores as a new variable (rur).[[71]](#footnote-70) (Note that ‘factor, pcf’ and ‘predict’ are Stata commands used in the template syntax file.)

*Factor variable list created in Step 10 if ahtype=2, pcf factors(1)*

*Predict rur if ahtype=2, no rotation*

**Step 12.** 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 13.** Run a regression with the common factor score (*com*) as the dependent variable and the rural area factor score (r*ur*) as the independent variable. Save the constant term (r*ur\_const*) and the coefficient (*rub\_coeff*).

*Regress com rur if ahtype=2*

*Set rur\_const=\_b[\_cons]*

*Set rur\_coeff=\_b[urb]*

**Step 14.** Create a variable for the combined score (*combscor*) equal to 0. Then calculate the combined score using the appropriate urban or rural factor scores, constant and coefficient obtained in steps 8-13.

*Set combscor=0*

*Replace combscor=urb\_const+(urb\_coeff\*urb) if ahtype=1*

*Replace combscor=rur\_const+(rur\_coeff\*rur) if ahtype=2*

**Step 15.** Create a variable for the household member weight (*hhmemwgt*), if not already created, by multiplying the number of de jure household members (*hhsize\_dj*) by the household weight (*hh\_wgt*). Note: Adjust the name of the household weight variable to match the name of the variable in your dataset if it is different.

*Set hhmemwgt=(hhsize\_dj\*hh\_wgt)*

**Step 16.** Create wealth quintiles (*wquint*) using the AWI scores and applying the household member weight.[[72]](#footnote-71) 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 6 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 cut points (e.g., 20 percent, 40 percent, 60 percent) because a single AWI score may increase the cumulative percentage by several percentage points.

*Rename awi=combscore*

*Set wquint=I 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*

**Step 17.** Create a null variable to use for adding variables in a later step and save the data file to use in Part 2.

*Set null=1*

*Save “FTF ZOI Survey [COUNTRY] [YEAR] wealthindex AWI”*

##### Part 2. CWI

| Survey variables used | *hhea, hhnum, wgt\_hh, samp\_stratum, v110, v202, v203, v208, v209, v211, v222c, v222e, v222f, v223f* |
| --- | --- |
| Analytic variables used | *memsleep\_dj, age, edu\_prim, awi, awiquint* |
| Analytic variables created | *ubn, ubn1, ubn2, ubn3, ubn4, wadult\_dj, wadult\_noprim, wadult\_att, wadult\_edu, null, ptile1, ptile2, ptile3, ptile4, sumwts, cut1, cut2, cut3, cut4, freq, cumfreq, percent, cumpercent, compcut1, compcut2, compcut3, compcut4, basecut1, basecut2, basecut3, basecut4, tv, computer, fridge, car, const, coeff, cwi, cwiquint, comp\_poor* |

**Step 1.** Create the first three UBN variables using household-level data.

**Step 1a.** Load the data file created in Part 1.

*Load “FTF ZOI Survey [COUNTRY] [YEAR] wealthindex AWI”*

**Step 1b.** Create the first UBN variable (*ubn1*) that indicates inadequate dwelling construction—dirt floors or natural/rustic walls.

*Set ubn1=0*

*Replace ubn1=1 if (v202=11 or 12 or 13 or 96) or [inadequate floor]*

*(v203=11 or 12 or 13 or 14 or 15 or 96) [inadequate wall]*

**Step 1c.** Create the second UBN variable (*ubn2*) that indicates inadequate sanitation or drinking water source.

*Set ubn2=0*

*Replace ubn2=1 if (v208=14, 15, 23, 41, 51, 61, or 96) or (v209=1) or [inadequate sanitation]*

*(v211=32, 42, 51, 61, 71, 81, or 96) [inadequate water source]*

**Step 1d.** Create the third UBN variable (*ubn3*) that indicates household crowding—four or more de jure household members per sleeping room.[[73]](#footnote-72)

*Set ubn3=0*

*Replace ubn3=1 if memsleep>3*

**Step 2.** Create the fourth UBN variable (*ubn4*) that indicates high economic dependency using individual-level data.

**Step 2a.** Save the current data in a temporary file and load the individual-level analytic data file to create the final UBN variable.

*Save data as “temp\_ubn\_hh”*

*Load “FTF ZOI Survey [COUNTRY] [YEAR] persons analytic data”*

**Step 2b.** Create a variable to flag working-age adults who are de jure household members (*wadult*).

*Set wadult\_dj=0*

*Replace wadult\_dj=1 if (age≥15 and age≤64) and hhmem\_dj=1*

**Step 2c.** Create a variable to flag working-age adults who are de jure household members but did not complete primary education (*wadult\_nopri*).

*Set wadult\_noprim=missing*

*Replace wadult\_noprim=0 if wadult\_dj=1*

*Replace wadult\_noprim=1 if wadult\_dj=1 and edu\_prim=0*

**Step 2d.** 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*).

*Set wadult\_att=missing*

*Replace wadult\_att=0 if wadult\_dj=1*

*Replace wadult\_att=1 if v110=1 and wadult\_dj=1*

**Step 2e.** Create household-level variables by transforming the individual-level variables created so far in Step 2 as well as *edu\_prim\_pdm\_dj* (binary variable that flags de jure primary adult decisionmakers who completed primary school, which was created in an earlier chapter of the guide and is already included in the persons-level analytic data file), so that there is one record for each per household.

*By hhea and hhnum: set wadult=sum(wadult\_dj)*

*By hhea and hhnum: set wadult\_noprim=sum(wadult\_noprim)*

*By hhea and hhnum: set wadult\_att=sum(wadult\_att)*

*By hhea and hhnum: set edu\_prim\_pdm\_dj=max(edu\_prim\_pdm\_dj)*

**Step 2f.** 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 adult in the household is still in school, OR (c) there are no working-age adults in the household or neither a primary male adult decisionmaker nor a primary adult decisionmaker has completed primary education.

*Set ubn4=0*

*Replace ubn4=1 if wadult\_dj>0 and wadult\_dj≠missing and wadult\_dj=wadult\_noprim*

*Replace ubn4=1 if wadult\_dj>0 and wadult\_dj≠missing and wadult\_dj=wadult\_att*

*Replace ubn4=1 if wadult\_dj=0 and edu\_prim\_pdm\_dj≠1*

**Step 2g.** 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 2a using *hhea* and *hhnum* as the key matching variables.

*Keep variables hhea hhnum ubn4*

*Add ubn4 to “temp\_ubn\_hh”, using hhea and hhnum as key matching variables*

**Step 2h.** 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.

*Set ubn=ubn1+ubn2+ubn3+ubn4*

*Set null=1*

*Save data as “temp\_ubn”*

**Step 3.** Calculate the UBN cutpoint values.

**Step 3a.** Create a new dataset that includes five observations (0–4) and four variables, as shown in Table 10. (In Stata, the ‘contract’ command will produce the desired dataset of frequencies and percentages.)

*Set a=1*

*Contract ubn a, freq(freq) cfreq(cumfreq) percent(percent) cpercent(cumpercent)*

**Table 10: Frequencies and Percentages by UBN Score**

| **UBN Score** | **Frequency**  **(*freq*)** | **Cumulative Frequency**  **(*cumfreq*)** | **Percentage**  **(*percent*)** | **Cumulative Percentage**  **(*cumpercent*)** |
| --- | --- | --- | --- | --- |
| 0 |  |  |  |  |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |

**Step 3b.** Create four percentile variables (*ptile1*, *ptile2*, *ptile3*, and *ptile4*).

*Set ptile1=0*

*Replace ptile1=(100-cumpercent) if ubn=0*

*Label variable “Percentage of households with at least one unmet basic need”*

*Set ptile2=0*

*Replace ptile2=(100-cumpercent) if ubn=1*

*Label variable “Percentage of households with at least two unmet basic needs”*

*Set ptile3=0*

*Replace ptile3=(100-cumpercent) if ubn=2*

*Label variable “Percentage of households with at least three unmet basic needs”*

*Set ptile4=0*

*Replace ptile4=(100-cumpercent) if ubn=3*

*Label variable “Percentage of households with all four unmet basic needs”*

**Step 3c.** Create a null variable (*null*) equal to 1, sum the frequency across all records to get a total frequency variable (freq), and add all *ptile* variables and the *freq* variable to the *temp\_ubn* data file. Use the null variable, which was created in both datasets, as the matching variable so that the variable values are added to all household records, and then save the file.

*Set null=1*

*Replace freq=sum(freq)*

*Add ptile1 ptile2 ptile3 ptile4 freq to “temp\_ubn”, using null as the key matching variable*

*Save “temp\_cwi\_ptile”*

**Step 3d.** Load the *FTF ZOI Survey [Country] [Year] wealthindex AWI* data file and add the *awi* variableto the *temp\_cwi\_ptile* data file. Use the null variable, which was created in both datasets, as the matching variable so that the variable values are added to all household records, and then save the file. Note that before performing the next step it is necessary to ensure that the households are sorted by their awi score (lowest to highest).

*Use “FTF ZOI Survey [COUNTRY] [YEAR] wealthindex AWI”*

*Merge awi into “temp\_cwi\_ptile”, using null as the key variable*

*Save “temp\_cwi\_ptile”*

*Rename awi to be windex\_awi*

*Sort windex\_awi*

**Step 3e.** 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.)

*Set sumwts=0*

*Replace sumwts=wgt\_hh if case number=1*

*Replace sumwts=wgt\_hh of previous case+wgt\_hh if case number>1*

*Set tot\_wgt=missing*

*Replace tot\_wgt=total(wgt\_hh)*

**Step 3f.** Create UBN cutpoint values for each household using the ZOI Survey AWI scores, the percentile variables (*ptile1-4*), the cumulative sampling weight (*sumwts*), and the weighted number of household (*tot\_wgt*) variables.

*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 3g.** Create UBN cutpoints by determining the minimum value of each of the four UBN cutpoint values.

*Set compcut1=minimum(cut1)*

*Set compcut2=minimum(cut2)*

*Set compcut3=minimum(cut3)*

*Set compcut4=minimum(cut4)*

**Step 3h.** 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*.

**Step 4.** Create the UBN cutpoint variables for the reference survey. (Note that the values of these variables were determined previously and will be used in the CWI indicator calculation across all ZOI Surveys.)

*Set basecut1=1.284437*

*Set basecut2=0.6740997*

*Set basecut3=(-0.6691726)*

*Set basecut4=(-1.329942)*

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 to be in the following format and save the data to a temporary file.[[74]](#footnote-73)

| **compcut** | **basecut** |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

where: *basecut=*UBN cutpoint values for the baseline survey

*compcut=*UBN cutpoint values for the ZOI Survey

*Reshape long compcut basecut, i(id) j(num)*

*Save “temp\_UBNcutpoint”*

**Step 6.** Calculate asset cutpoint values.

**Step 6a.** Load household-level data *household\_analytic*.

**Step 6b.** Create binary variables for the four asset items to be used as anchoring points (*tv*, *computer*, *fridge*, and *car*).

*Set tv=0; Replace tv=1if v222c=1*

*Set computer=0; Replace computer=1 if v222e=1*

*Set fridge=0; Replace fridge=1if v222f=1*

*Set car=0; Replace car=1if v223f=1*

**Step 6c.** 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, which is equal to the negative value of the regression constant divided by the AWI score coefficient.

*Svyset hhea [pweight=wgt\_hh], strata(samp\_strat)*

*Svy: logit tv awi; Set compcut1=(-constant÷coeff) [tv]*

*Svy: logit fridge awi; Set compcut2=(-constant÷coeff) [fridge]*

*Svy: logit computer awi; Set compcut3=(-constant÷coeff) [computer]*

*Svy: logit car awi; Set compcut4=(-constant÷coeff) [car]*

**Step 6d.** 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.

*Drop if \_n≠1*

*Keep compcut1 compcut2 compcut3 compcut4*

**Step 7.** Create variables that contain the four asset *basecut* variables for the reference survey. (The values for these variables were calculated previously and included in the syntax file.)

*Set basecut1=0.1166113 [television]*

*Set basecut2=1.107843 [fridge]*

*Set basecut3=1.591285 [computer]*

*Set basecut4=6.366876 [car]*

The data file now contains only one observation and eight variables: *compcut1, compcut2, compcut3, compcut4, basecut1, basecut2, basecut3,* and *basecut4.*

**Step 8.** Reshape the data the same as the UBN data in Step 4 and save the asset cutpoint value data file as *temp\_ASSETcutpoint.*

**Step 9.** Add the UBN and ASSET cutpoint value data together. The appended dataset will have two variables (*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 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 1. Save the data to a temporary file. Then add these two variables to the data file created in Step 8 of Part 1.

*Set null=1*

*Save “temp\_regress\_result”*

*Load “FTF ZOI Survey [COUNTRY] [YEAR] wealthindex AWI”*

*Add const coeff from “temp\_regress\_result” using null as the key matching variable*

**Step 12.** Calculate the comparative wealth index scores for each household (*cwi*) using the ZOI Survey wealth index scores (*awi*) and the regression constant and coefficient obtained in Step 9.

*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<-0.86276845*

*Label values 0 “No” 1 “Yes”*

*Label variable “HH falls below threshold for the poorest quintile of CWI“*

**Step 14.** After applying the household sampling weight, calculate the comparative wealth index indicator using the *comp\_poor* analytic variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.).

*Svyset hhea [pweight=wgt\_hh] strata(samp\_stratum)*

*Svy: tab comp\_poor*

*Svy: tab comp\_poor, over(genhhtype)*

**Step 15.** Assign all surveyed households to a CWI quintile.

*Set cwiquint=missing*

*Replace cwiquint=1 if cwi<(-0.86276845)*

*Replace cwiquint=2 if cwi<0.00108905 and cwiquint=missing*

*Replace cwiquint=3 if cwi<0.80370025 and cwiquint=missing*

*Replace cwiquint=4 if cwi<1.2871975 and cwiquint=missing*

*Replace cwiquint=5 if cwi≥ 1.2871975 and cwiquint=missing and cwi≠missing*

## References

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>

Rutstein, S., and 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>

# 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 three 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 three resilience indicators above and the guidelines to construct the indicators in general, and the second describes the step-by-step procedures to calculate each indicator.

## Guidelines to construct the resilience indicators

### 11.1.1 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 after asking about the households’ exposure to and the severity of 16 types of shocks and stressors the household might have experienced during the year before the survey:

Question 359: “Would you say that right now, your household’s ability to meet your food needs is:”

The response options are:

Better than before these difficult times=1

The same as before these difficult times=2

Worse than before these difficult times=3

Question 360: “Looking ahead over the next year, do you believe your household’s ability to meet your food needs will be:”

The response options are:

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 359 and 360 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 359+recoded response to question 360

The next step in calculating the ARSSI is to calculate a shock exposure index (SEI) using the responses to ZOI Survey questions posed about households’ exposure to 16 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: too much rain (309); too little rain (312); erosion of land (315); loss of land (318); sharp increase in the price of food (321); someone stealing or destroying belongings (324); not being able to access inputs for crops (328); disease affecting crops (331); pests affecting crops (334); theft of crops (337); not being able to access inputs for livestock (340); disease affecting livestock (343); someone stealing animals (346); not being able to sell crops, livestock, or other products at a fair price (349); severe illness in the family (352); and death in the household (355).

Because each surveyed household did not experience the same types of shocks and stressors of the same severity, it is necessary to create the SEI as a measure of the household’s ability to recover from the shocks and stressors that it experienced. It is a correction factor that is applied to the ATR index that accounts for the shocks and stressors to which a household was exposed out of the total number of shocks or stresses included in the ZOI Survey (i.e., 16), and the perceived severity of the shock or stressor on household income and food consumption.

Perceived severity is measured using two variables: (1) impact of the shock or stress on income security, and (2) 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 your household’s economic situation?”   
“How severe was the impact on household food consumption?”

The possible responses are:

Not severe=1

Somewhat severe=2

Severe=3

Extremely severe=4

The responses to the two questions are combined into one severity variable that has a minimum value of 2 and a maximum value of 8 for each shock and stressor experienced.

The SEI is then calculated as the sum of the incidence of experience of each shock (a variable equal to 1 if the shock or stressor was experienced and 0 otherwise) multiplied by the perceived severity of the shock. The SEI ranges from 0 (if the household did not experience any shocks or stressors) to 128 (if the household experienced all 16 shocks or stressors at the highest level of severity).

Finally, the shock exposure-corrected ARSSI is calculated to create a comparable measure of ability to recover that corrects for any differences among households in their shock and stressor exposure. To create the ARSSI, a linear regression of the base ATR index on the SEI is run across all households with ART 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 stressor 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.

### 11.1.2 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 361: 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.

361a. During difficult times, will the household be able to lean on relatives living in your community?

If the respondent answers “YES,” the subsequent question is asked:

361b. Will the same relatives living in your community that you will be able to lean on during your difficult times also be able to lean on you 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 (361a and 361b) are repeated for:

Relatives living outside your community? (361c and 361d)

Non-relatives living in your community? (361e and 361f)

Non-relatives living outside your community (361g and 361h)

The bonding social capital index considers responses to questions 361a, 361b, 361e, and 361f. The bridging social capital index considers responses to questions 361c, 361d, 361g, and 361h.

For both bonding and bridging social capital, an additive index ranging from 0 to 4 is calculated, with a score of 0 for no responses to all question sub-items included in the index, and 1 for each yes response to question sub-items included in the index (361a, 361b, 361e, 361f, and 361c, 361d, 361g, 361h).

Questions that are skipped because of skip patterns in the questionnaire are also set to 0. That is, question 361b is set to 0 if the response to question 361a is No; question 361d is set to 0 if the response to question 361c is No; question 361f is set to 0 if the response to question 361e is No; question 361h is set to 0 if the response to question 361g is No.

The values are normalized and scaled to a 0–100 scale by dividing by 4, and then multiplying by 100. The index of social capital indicator is the average of the two sub-indices.

The indicator is calculated in two steps. First, the individual bonding social capital sub-index and the bridging social capital sub-index are calculated as the following:

Bonding sub-index for each household=(Number of yes=1 responses to questions 361a, 361b, 361e, 361f÷4)\*100

Bridging sub-index for each household=(Number of yes=1 responses to questions 361c, 361d, 361g÷4)\*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

The third step is to divide the sum of the index of social capital score by the total number of households:

Index of social capital score for all households=(Sum of index of social capital score÷Total number of households)

### 11.1.3 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 stressors. 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 (362):

“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; and No, support not needed=3.

### 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.

Note that because the data for this indicator are collected in Module 6, *A-WEAI*, of the core ZOI Survey questionnaire to reduce respondent burden, the indicator does not capture households without a primary adult female decisionmaker or households composed of only children. Furthermore, the information collected is on whether the household accessed credit through a group-based source or is an active member of a credit or microfinance group—not on whether the household saved money.

## 11.2 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.*

### 11.2.1 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 Module 3, *Food Security and Resilience*.

#### Definitions

| Numerator | Sum of the ARSSI scores for all surveyed households that experienced at least one shock or stressor during the 12 months preceding the survey |
| --- | --- |
| Denominator | Number of surveyed households that experienced at least one shock or stressor 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 stressors the household experienced |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members)\*  Household education (de jure household members)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P2-ZOI |
| 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, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *v309x-v357x, v359x, v360x, atr, perceived\_sev1-perceived\_sev16, sei, anymissing, ability\_recover, b\_atr* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Calculate the base ATR index.

**Step 1a.** Identify all the variables needed to construct the ARSSI indicator and check their frequencies to ensure that none of the questions has 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 1b.** 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=7*

*Replace v359x=1 if v359=3*

*Replace v359x=3 if v359=1*

*Set v360x=missing if v360=7*

*Replace v360x=1 if v360=3*

*Replace v360x=3 if v360=1*

**Step 1c.** Combine 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 (ATR)”*

**Step 2.** Calculate the SEI.

**Step 2a.** Create an analytic variable *(v310x)* for the question asking how too much rain affected the household’s economic situation *(v310)* for which all refused responses are recoded to missing.

*Set v310x=v310*

*Replace v310x=missing if v310=7*

**Step 2b.** Repeat step 2a for all variables that capture how a shock affected the household’s economic situation (*v313, v316, v319, v322, v325, v329, v332, v335, v338, v341, v344, v347, v350, v353, v356)* and food consumption (*v311, v314, v317, v320, v323, v326, v330, v333, v336, v339, v342, v345, v348, v351, v354, v357)*.

**Step 2c.** For each shock or stressor a household experienced, combine the variables for the impact of the shock on the household’s economic situation and the household’s food consumption to create one severity variable (perceived\_sev#) that has a minimum value of 2 and a maximum value of 8.

*Set:*

*perceived\_sev1=(v310x+v311x) [too much rain]*

*perceived\_sev2=(v313x+v314x) [too little rain]*

*perceived\_sev3=(v316x+v317x) [land erosion]*

*perceived\_sev4=(v319x+v320x) [lost land]*

*perceived\_sev5=(v322x+v323x) [sharp increase in food prices]*

*perceived\_sev6=(v325x+v326x) [belongings stolen or destroyed]*

*perceived\_sev7=(v329x+v330x) [lack of access to crop inputs]*

*perceived\_sev8=(v332x+v333x) [crops affected by disease]*

*perceived\_sev9=(v335x+v336x) [crops affected by pests]*

*perceived\_sev10=(v338x+v339x) [crops stolen]*

*perceived\_sev11=(v341x+v342x) [lack of access to livestock inputs]*

*perceived\_sev12=(v344x+v345x) [livestock affected by disease]*

*perceived\_sev13=(v347x+v348x) [animals stolen]*

*perceived\_sev14=(v350x+v351x) [unable to sell products for fair price]*

*perceived\_sev15=(v353x+v354x) [household member severely ill]*

*perceived\_sev16=(v356x+v357x) [family member death]*

**Step 2d.** Create an analytic variable *(v309x)* for the question asking whether the household experienced too much rain during the year prior to the survey *(v309)* in which all “no” responses are recoded from 2 to 0.

*Set v309x=v309*

*Replace v309x=0 if v309=2*

**Step 2e.** Repeat Step 2d for all variables that capture whether a household experienced a shock (*v312, v315, v318, v321, v324, v328, v331, v334, v337, v340, v341, v346, v349, v352, v355)*.

**Step 2f.** Recode the shock variables only applicable to households that cultivate crops or own livestock, which are currently missing for households that do not cultivate crops or own livestock, to be 0.

*Set v328x=0 if v328x=missing and v327=2*

*Set v331x=0 if v331x=missing and v327=2*

*Set v334x =0 if v334x=missing and v327=2*

*Set v337x=0 if v337x=missing and v327=2*

*Set v340x=0 if v340x=missing and v340a=2*

*Set v343x=0 if v343x=missing and v340a=2*

*Set v346x=0 if v346x=missing and v340a=2*

**Step 2g.** Set each shock experience variable (for example, *v309*, *v312*, *v315*) equal to 0 if the household did not experience the shock. For example, set *v309=0 if v309=2 (No)*. Then multiply each shock experience variable by the perceived severity of the shock variable and add all products into one variable, the weighted SEI variable (*sei*). The *sei* ranges from 0 if the household did not experience any shocks to 128 if the household experienced all 16 shocks at an extremely severe level.

*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)*

*Label variable “Shock exposure index (SEI)”*

**Step 2h.** Set the SEI variable to missing for any households with a missing or refused response to a question included in the calculation of the SEI variable.

*Set anymissing=0*

*Replace anymissing=1 if v309x=missing or (v309x=1 and (v310x=missing or v311x=missing))*

*Replace anymissing=1 if v312x=missing or (v312x=1 and (v313x=missing or v314x=missing))*

*Replace anymissing=1 if v315x=missing or (v315x=1 and (v316x=missing or v317x=missing))*

*Replace anymissing=1 if v318x=missing or (v318x=1 and (v319x=missing or v320x=missing))*

*Replace anymissing=1 if v321x=missing or (v321x=1 and (v322x=missing or v323x=missing))*

*Replace anymissing=1 if v324x=missing or (v324x=1 and (v325x=missing or v326x=missing))*

*Replace anymissing=1 if v328x=missing or (v328x=1 and (v329x=missing or v330x=missing))*

*Replace anymissing=1 if v331x=missing or (v331x=1 and (v332x=missing or v333x=missing))*

*Replace anymissing=1 if v334x=missing or (v334x=1 and (v335x=missing or v336x=missing))*

*Replace anymissing=1 if v337x=missing or (v337x=1 and (v338x=missing or v339x=missing))*

*Replace anymissing=1 if v340x=missing or (v340x=1 and (v341x=missing or v342x=missing))*

*Replace anymissing=1 if v343x=missing or (v343x=1 and (v344x=missing or v345x=missing))*

*Replace anymissing=1 if v346x=missing or (v346x=1 and (v347x=missing or v348x=missing))*

*Replace anymissing=1 if v349x=missing or (v349x=1 and (v350x=missing or v351x=missing))*

*Replace anymissing=1 if v352x=missing or (v352x=1 and (v353x=missing or v354x=missing))*

**Step 2i.** Replace anymissing=1 if v355x=missing or (v355x=1 and (v356x=missing or v357x=missing))2h. Finally, calculate the average SEI across households.

*Set mean\_sei=sum of sei if anymissing=0 across all surveyed households÷number of surveyed households if anymissing=0*

*Label variable “Mean SEI”*

**Step 3.** Calculate the shock exposure-corrected ARSSI.

**Step 3a.** Run a linear regression of the base ATR index on the SEI to obtain “*b,*” the SEI coefficient.

*Regress atr sei*

*Set b\_atr=coefficient of variable sei*

*Label variable “SEI regression coefficient”*

**Step 3b.** Calculate the ARSSI *(ability\_recover)* for each household.

*Set ability\_recover=atr+b\_atr\*(mean\_sei–sei)*

*Label variable “ARSSI”*

**Step 4.** After applying the household sampling weight, calculate the mean ARSSI using the *ability\_recover* analytic variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_hh], strata(samp\_stratum)*

*Svy: mean ability\_recover*

*Svy: mean ability\_recover, over(genhhtype\_dj)*

### 11.2.2 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 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 0–100 (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 (de jure household members)\*  Household education (de jure household members)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P2-ZOI |
| Treatment of missing data | If v361b has a value of missing because v361a has a value of no, set v361b to no. If v361d has a value of missing because v361c has a value of no, set v361d to no. If v361f has a value of missing because v361e has a value of no, set v361f to no. If v361h has a value of missing because v361g has a value of no, set v361h to no. After recoding the values for v361b, v361d, v361f, and v361h, households with any missing responses are excluded from the numerator and denominator. |
| Survey variables used | *v361a, v361b, v361c, v361d, v361e, v361f, v361g, v361h, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *v361ax, v361bx, v361cx, v361dx, v361ex, v361fx, v361gx, v361hx, scap, scap\_bond, scap\_bridge* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Identify and prepare the required variables. Create new variables 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.

*Set v361ax=1 if v361a=1; Replace v361ax=0 if v361a=2*

*Set v361cx=1 if v361c=1; Replace v361cx=0 if v361c=2*

*Set v361ex=1 if v361e=1; Replace v361ex=0 if v361e=2*

*Set v361gx=1 if v361g=1; Replace v361gx=0 if v361g=2*

*Set v361bx=1 if v361b=1; Replace v361bx=0 if v361b=2 or v361b=3 or v361a=2*

*Set v361dx=1 if v361d=1; Replace v361dx=0 if v361d=2 or v361d=3 or v361c=2*

*Set v361fx=1 if v361f=1; Replace v361fx=0 if v361f=2 or v361f=3 or v361e=2*

*Set v361hx=1 if v361h=1; Replace v361hx=0 if v361h=2 or v361h=3 or v361h=2*

*Replace v361bx=0 if v361ax=2 and v361bx=missing*

*Replace v361dx=0 if v361cx=2 and v361dx=missing*

*Replace v361fx=0 if v361ex=2 and v361fx=missing*

*Replace v361hx=0 if v361gx=2 and v361hx=missing*

**Step 2.** Calculate bonding social capital *(scap\_bond).*

*Set scap\_bond=((v361ax+v361bx+v361ex+v361fx)÷4)\*100*

*Replace scap\_bond=missing if v361ax=missing or 361bx=missing or v361ex=missing or v31fx=missing*

*Label variable “Bonding social capital sub-index”*

**Step 3.** Calculate bridging social capital *(scap\_bridge).*

*Set scap\_bridge=((v361cx+v361dx+v361gx+v361hx)÷4)\*100*

*Replace scap\_bridge=missing if v361cx=missing or 361dx=missing or v361gx=missing or v31hx=missing*

*Label variable “Bridging social capital sub-index”*

**Step 4.** Calculate index of social capital *(scap)* by averaging the bonding and bridging social capital.

*Set scap=missing*

*Replace scap=(scap\_bond+scap\_bond)÷2*

*Label variable “Index of social capital”*

**Step 5.** Calculate the sample-weighted mean index of social capital after applying the household weight to the data using the *scap* analytic variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_hh], strata(samp\_stratum)*

*Svy: mean scap*

*Svy: mean scap, over(genhhtype\_dj)*

**Step 6.** Repeat Step 5 for the bridging and bonding social capital sub-indices.

*Svy: mean scap\_bridge*

*Svy: mean scap\_bridge, over(genhhtype\_dj)*

*Svy: mean scap\_bond*

*Svy: mean scap\_bond, over(genhhtype\_dj)*

### 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 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 (de jure household members)\*  Household education (de jure household members)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P2-ZOI |
| 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, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *locgov\_resp* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Calculate the indicator variable (*locgovt\_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 disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_hh], strata(samp\_stratum)*

*Svy: prop locgov\_resp*

*Svy: prop locgov\_resp, over(genhhtype\_dj)*

### 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. It is based on information collected in the core ZOI Survey Module 6, *Women’s Empowerment in Agricultural Index*.

#### Definitions

| Numerator | Number of households with a primary adult female decisionmaker that 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 decisionmaker |
| Unit of measure | Percentage |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members)\*  Household education (de jure household members)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P2-ZOI |
| Treatment of missing data | All “don’t know” and missing responses are considered to be “no” responses when calculating this indicator. |
| Survey variables used | *v6605, m6605, v6308\_5, m308\_5, v6308\_6, m6308\_6, v6404\_04, m6404\_04, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj* |
| Analytic variables created | *access\_finance* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create a binary variable that indicates primary adult decisionmakers who completed the A-WEAI module and 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\_finance*). The variable indicates whether anyone in the household took a loan from a group-based microfinance or lending program (*v6308\_5*, *m6308\_5*) or from an informal credit or savings group (*v6308\_5*, *m6308\_5*) or whether a primary adult decisionmaker in the household is an active member of a savings, credit, or microfinance group (*v6405\_04*, *m6405\_04*).

*Set access\_finance=missing*

*Replace access\_finance=0 if v6605=1 or m6605=1*

*Replace access\_finance=1 if v6308\_5≤3 I m6308\_5≤3*

*Replace access\_finance=1 if v6308\_6≤3 I m6308\_6≤3*

*Replace access\_finance=1 if v6405\_04=1 or m6405\_04=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 the *access\_finance analytic* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_hh], strata(samp\_stratum)*

*Svy: prop access\_finance*

*Svy: prop access\_finance, over(genhhtype\_dj)*

## References

Frankenberger, T., Mueller, M., Spangler, T., and 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>

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_ShocksandStressesMsmt_Final.pdf>

United States Agency for International Development Center for Resilience. 2018. *Resilience evidence forum report*. Available at: <https://www.usaid.gov/sites/default/files/documents/1867/0717118_Resilience.pdf>

# Abbreviated Women’s Empowerment in Agriculture Index

## 12.1 Background

The WEAI is a composite index of women’s empowerment and gender equality in agriculture developed by USAID, the International Food Policy and Research Institute, and the Oxford Poverty and Human Development Initiative. USAID’s Feed the Future Initiative commissioned the development of the WEAI to monitor changes in women’s empowerment and gender equality resulting from its programs.

The WEAI captures empowerment, agency, and inclusion of women in agriculture, as well as gender differentials in empowerment in five domains: (1) decisionmaking over production, (2) access to resources, (3) control over income, (4) group participation and leadership, and (5) time allocation. The WEAI is calculated using data collected from the household’s primary adult male decisionmaker and the primary adult female decisionmaker in a household survey.

The WEAI comprises two sub-indices: the five domains of empowerment (5DE) and the gender parity index (GPI). The domains of empowerment contribute to improved economic status of poor households, improved nutrition and health, and increased agricultural productivity at the household level.[[75]](#footnote-74) The 5DE indicate empowerment in the five domains, and the GPI reflects the gap in empowerment between the primary female decisionmaker and primary male decisionmaker in the same household. The primary female decisionmaker is the adult female, 18 years of age or older, who makes more social and economic decisions than other women in the household. The primary male decisionmaker is the adult male, 18 years of age or older, who makes more social and economic decisions than other men in the household. A household may have both a primary female decisionmaker and a primary male decisionmaker, only a primary female decisionmaker, only a primary male decisionmaker, or neither—if there are no adults in the household. A household, however, can have at most only one primary female decisionmaker and one primary male decisionmaker.

In November 2013, the International Food Policy and Research Institute and USAID revised the WEAI based on feedback from USAID implementing partners, field teams that had implemented the data collection for WEAI, and researchers and other stakeholders who had worked with the tool. The original WEAI was streamlined and simplified to make it less time-consuming and expensive to collect the required data, and to improve some problematic and difficult-to-understand modules. The A-WEAI retains the five domains from the original WEAI but includes only 6 of the 10 indicators. Table 11 summarizes the differences between the A-WEAI and the original WEAI.

**Table 11: Comparison of the Original WEAI and A-WEAI**

| **Domains** | **Indicators** | |
| --- | --- | --- |
| **Original WEAI** | **A-WEAI** |
| Decisionmaking over production | 1. Input in productive decisions Autonomy in production | 1. Input in productive decisions |
| Access to resources | 1. Ownership of assets 2. Purchase, sale, or transfer of assets Access to and decisions on credit | 1. Ownership of assets 2. Access to credit and decisions on it |
| Control over income | 1. Control over use of income | 1. Control over use of income |
| Group participation and leadership | 1. Group membership Public speaking | 1. Group membership |
| Time allocation | 1. Workload Leisure | 1. Workload |

## 12.2 Guidelines to construct the A-WEAI indicators

### 12.2.1 A-WEAI

This section provides instructions for constructing the A-WEAI 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 A-WEAI. Part I lists the steps for data preparation and the calculation of the six indicators corresponding to the 5DE. Part 2 lists the steps for computing the A-WEAI. Part 3 discusses ways to decompose the A-WEAI to better understand the drivers of women’s and men’s empowerment and disempowerment in rural and agriculture-based communities.

The A-WEAI is computed using data collected from primary adult female and male decisionmakers through the ZOI Survey who were de jure residents (question 105a=Yes). The data needed to compute the A-WEAI come from Module 6 on decisionmaking and income generation, found in the household data file.

Constructing the A-WEAI involves preparing the data and calculating indictors that measure the empowerment of women and men in the 5DE scores and gender differences in empowerment at the household level in the GPI scores. The 5DE and GPI scores are subsequently combined to calculate the overall A-WEAI score at the ZOI level. The sample for calculating the 5DE includes only respondents who have complete information on all five domains. The sample for computing the GPI includes only households that have a pair of female and male decisionmakers—both with complete information on all five domains. The A-WEAI module is not fielded to male adult-only households because the A-WEAI requires data from at least a female respondent in each household. Appendix B, Table 1 summarizes the A-WEAI domains and indicators, relevant survey questions and Feed the Future variables, the adequacy and inadequacy criteria, and weight assigned to each indicator to compute the overall A-WEAI index.[[76]](#footnote-75)

#### Preparing the data and calculating the 5DE indicators

The data preparation steps include performing checks for consistent skip patterns, out-of-range values that do not correspond with the survey, outliers (extreme or implausible responses), missing responses, and duplicate observations, and then recoding the variables to create six indicators of achievement in the 5DE. 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. Because data for the primary female decisionmaker and the primary male decisionmaker of each household are collected in separate modules and included in the household-level data file, the data must first be transformed to individual respondent-level data before calculating the 5DE indicators.

**Domain 1: Decisionmaking over production**

To capture the individual’s decisionmaking 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 he or she cannot make his or her own decisions about the activity to a medium or high extent, or 1 if the respondent feels that he or she can make his or 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 they have a total score of 1 or higher. The survey asks respondents:

“Did you yourself participate in [ACTIVITY] in the past 12 months?” (question 6.201)

The activities are food crop farming (6.2A), cash crop farming (6.2B), livestock raising (6.2C), and fishing or fishpond culture (6.2F).

Respondents who participated in an activity are then asked:

“When decisions are made regarding [ACTIVITY], who is it that normally takes the decision?” (question 6.202)

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 one person.

Respondents who do not decide alone—those who report making decisions with others—are asked two follow-up questions:

Question 6.203: “How much input do you have in making decisions about [ACTIVITY]?”

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*.

Question 6.204: “To what extent do you feel you can make your own personal decisions regarding [ACTIVITY] if you want(ed) to?”

The options are: (1) *not at al1*, (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. Responses are scored 0 or 1, where 1 connotes adequacy. Respondents are considered adequate in asset ownership if they own at least two types of small assets, such as chickens, handtools for farming, non-mechanized farming equipment, and small consumer durables, or one large asset, such as assets in all other categories in the ZOI questionnaire,[[77]](#footnote-76) jointly or alone, assuming that the household owns the item. 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, assuming that they belong to a household that used a source of credit in the past year.

***Ownership of assets***

An asset ownership score is calculated to reflect whether the respondent owns—alone or jointly—the assets included in the survey. The survey asks:

“Does anyone in your household currently have any [ITEM]?” (question 6.301)

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 6.303)

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.

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 large asset. Small asset types include chickens, ducks, turkeys, and pigeons; handtools, 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 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 large asset or two types of small assets, or if the household owns no assets.

***Access to credit and decisions about credit***

This indicator on access to credit and decisionmaking 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: nongovernment organizations; formal lenders, such as a bank or financial institution; informal lenders; friends or relatives; group-based microfinance 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.

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 6.308)

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 (6.3bA) Non-governmental organization, (6.3bB) Informal lender, (6.3bC) Formal lender, (6.3bD) Friends or relatives, (6.3bE) Group-based micro-finance or lending, and (6.3bF) 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 6.309)

“Who makes the decision about what to do with the money/item borrowed from [SOURCE]?” (question 6.310)

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 activity, or 1 if the respondent has input into at least some decisions about the activity. There are six relevant activities: (a) food crop farming, (b) cash crop farming, (c) livestock production, (d) non-farm activities, I wage and salary work, and (f) 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 he or she cannot make his or her own decisions about the activity to a medium or high extent, or 1 if the respondent feels that he or she can make his or her own decisions about the activity to a medium or high extent. There are four relevant activities for the second sub-indicator: non-farm economic activities, wage and salary employment, major household expenditures, and minor household expenditures. The second sub-indicator, therefore, has a possible score between 0 and 4. The two sub-indicators are aggregated to get a total score between 0 and 10 for the domain. Respondents are considered to have adequate achievement in this domain if they have a total score of 1 or higher, except for in one instance; if respondents have a total score of 1—for only having input into at least some decisions regarding minor household expenditure—then they are considered to have inadequate achievement in this domain.

Three of the same questions used to calculate the Domain 1 indicator are also used to calculate this indicator—questions 6.201, 6.202, and 6.203—although additional activities are considered for this indicator. One additional question is used to calculate this indicator:

Question 6.205: “How much input did you have in decisions on the use of income generated from [ACTIVITY]?”

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 his or her community: membership in an economic or social group.

The survey asks respondents:

“Is there a [GROUP] in your community?” (question 6.404)

If respondents indicate that a certain group exists, they are asked a follow-up question:

“Are you an active member of this [GROUP]?” (question 6.405)

Options are membership in the following types of groups: (a) agricultural, livestock, or fisheries producers’ groups; (b) water users’ group; (c) forest users’ group; (d) credit or microfinance group; (e) mutual help or insurance group; (f) trade or business association; (g) civic group; (h) local government; (i) religious group; (j) other women’s or men’s group; and (k) other specified 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 reported in 15-minute intervals (question 6.601). 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 12.

**Table 12: Activities Included in the A-WEAI Time Use Module**

| **Work** **Activity Categories** | **Non-work Activity Categories** |
| --- | --- |
| School, including homework  Work as employed  Own business work  Food crop farming  Cash crop farming  Livestock raising  Shopping or getting service (including health services)  Fishing or fishpond culture  Weaving, sewing, textile care  Cooking  Domestic work, including fetching water and collecting fuel  Caring for children and adults, including sick or elderly  Commuting to or from work or school | Sleeping and resting  Eating and drinking  Personal care  Traveling, but not for work or school  Watching TV, listening to radio, or reading  Exercising  Social activities and hobbies  Religious activities  Other |

The individual is defined as adequate on workload if the number of hours he or she worked the day preceding the survey was less than the time poverty line of 10.5 hours in 24 hours.

The survey also asks respondents:

“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 6.602)

Respondents who reported the 24 hours preceding the survey as being an atypical workday are excluded.

#### Computing the A-WEAI

The A-WEAI is computed by combining the 5DE scores and the GPI into one overall score, using the following steps:

1. Calculate individual empowerment and disempowerment scores.
2. Calculate censored headcounts for the disempowered and empowered.
3. Compute the index of disempowerment and the index of empowerment (5DE index).
4. Calculate the GPI.
5. Combine the 5DE index and GPI into the overall A-WEAI.

##### Step 1. Calculate individual inadequacy and adequacy scores.

The six 5DE indicators are generated by reverse-coding the adequacy scores calculated in Part 1 so that a value of 1 is inadequate and a value of 0 is adequate. An inadequacy score (ci) is then calculated for each respondent by summing the weighted scores for the six 5DE indicators, as shown in the following calculation:

*ci=w1I1+w2I2+w3I3+w4I4+w5I5+w6I6*

Where *ci* refers to the weighted inadequacy score of respondent *i*, and *w* refers to the weight assigned to indicator *I* so that the sum of the weights across all indicators equals 1.

The score can have a value between 0 and 1 and reflects the percentage of weighted indicators in which respondents have inadequate achievement. A score of 0 implies that respondents are adequate on all indicators. A score of 1 implies that respondents are inadequate on all indicators. A respondent with a score of 0.4 has inadequate achievement in two of five domains.

An individual adequacy score is calculated by subtracting the respondents’ inadequacy score from 1. The 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 implies that respondents are not adequate on any indicator. A score of 1 implies that respondents are adequate on all indicators. A respondent with a score of 0.6 is has adequate achievement in his or her empowerment in three of five domains.

##### Step 2. Calculate censored headcounts for the disempowered and empowered and the censored inadequacy score.

**Step 2.1.** Apply a disempowerment cut-off to create censored disempowerment scores and identify the disempowered. A disempowerment threshold or cut-off is applied to identify the disempowered and create a censored inadequacy score. The cut-off is set so that baseline indices could result in improvement over time. A cut-off that is too low could overstate the prevalence of empowered respondents and result in an indicator lacking sensitivity. A cut-off 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 disempowerment cut-off is set to 0.2 so that respondents with an inadequacy score of less than or equal to 0.2—those who are inadequate in 20 percent or less of the weighted indicators—are excluded from the censored inadequacy headcount even if their score is not zero. An individual is considered to be empowered if he or she has adequacy in at least four of the five domains, or has an adequacy score of 80 percent or more.[[78]](#footnote-77)

The scores of respondents with an inadequacy score of less than 0.2 are recoded as 0 and excluded from the headcount. Respondents who do not have complete information on the 5DE indicators are also excluded from the count. The censored headcount of the disempowered (H) is the proportion of respondents who are disempowered, or the proportion of respondents with complete 5DE indicator information who have a non-zero recoded individual inadequacy score.

**Step 2.2.** Identify the empowered.

The censored headcount of the empowered (1–H) is the proportion of respondents with complete 5DE information who have a recoded individual inadequacy score equal to zero. It is obtained by subtracting the censored headcount of the disempowered from one.

**Step 2.3**. Calculate the average inadequacy score among the disempowered (A) by obtaining the average of the inadequacy scores among those who are disempowered—that is, among those who have an inadequacy score greater than 0.2.

##### Step 3. Compute the index of disempowerment (M0) and the index of empowerment (5DE).

**Step 3.1.** Calculate the index of disempowerment. The index of disempowerment (M0) is calculated by multiplying the censored disempowered headcount (H) by the depth of disempowerment, or average inadequacy score (A), shown in the following calculation:

*M0=Hp x Ap*

Where the censored disempowered headcount (H) is defined as:

Hp=

*q*=the number of disempowered respondents in the ZOI (respondents whose disempowerment score is greater than0.2)

*n*=the total population of respondents in the ZOI with complete 5DE indicator data

And where the depth of disempowerment is the average inadequacy score (A), defined as:

Ap=

*ci(k)*=censored inadequacy score of respondent *i*

*q*=number of disempowered respondents in the ZOI

The percentage of respondents who are disempowered and the percentage of domains in which respondents are inadequately empowered are calculated to identify critical indicators, which are used to focus program efforts.

**Step 3.2.** Calculate the index of empowerment (5DE) by subtracting the index of disempowerment (M0)from 1:

*5DE=1–M0*

##### Step 4. Calculate the GPI

The GPI measures the extent of inequality in empowerment in a household between the primary male decisionmaker and the primary female decisionmaker. The GPI excludes households that lack both a primary male decisionmaker and a primary female decisionmaker. A household is considered to lack gender parity if the primary female decisionmaker is disempowered and her inadequacy score is higher than that of the primary male decisionmaker. The GPI comprises two components: (1) proportion of gender parity-inadequate households, and (2) average percentage gap in the censored score of women and men in households that are gender-parity inadequate.

The disempowerment cut-off *(k)* for calculating the GPI is 20 percent. The scores of respondents with an inadequacy score of less than or equal to the cut-off are recoded as 0.2. A new censored inadequacy score, denoted by c′i *(k),* is calculated as:

If ci>k, then c′i(k)=ci

If ci≤k, then c′i(k)=k=0.2

The proportion of gender parity-inadequate households is denoted by HGPI and calculated as:

HGPI=

Where:

*h*=the number of households that are gender-parity inadequate

*m*=total number of households with complete information for both the primary adult male decisionmaker and primary adult female decisionmaker

The average percentage empowerment gap (IGPI) in the censored score of women and men in households that are gender parity-inadequate is calculated as:

IGPI=

Where:

c′i *(k)W*=censored inadequacy score for the female decisionmaker of the household’s primary male-female decisionmaker pair

c′i *(k)M*=censored inadequacy score for the male decisionmaker of the household’s primary male-female decisionmaker pair

The GPI score is calculated as:

GPI=1–(HGPI\*IGPI)

Where:

*HGPI*=percentage of women without gender parity

*IGPI*=average empowerment gap

##### Step 5. Combine the 5DE and the GPI.

The A-WEAI is computed by summing the weighted value of the 5DE score and the GPI score, shown in the following calculation. The 5DE index receives a weight of 0.9 (90 percent), and the GPI receives a weight of 0.1 (10 percent).

*A-WEAI=(0.9\*5DE)+(0.1 GPI)*

#### Decomposing the “disempowered but adequate” component of the 5DE index

The 5DE index is composed of two components: (1) the percentage of empowered individuals and (2) the percentage of disempowered individuals multiplied by the mean adequacy score of the disempowered. The second component—disempowered but adequate—can be analyzed by indicator to illustrate which indicators present the greatest constraints to empowerment among men and women identified as disempowered by their adequacy score. For example, an indicator with a percent contribution to the “disempowered but adequate” component of the 5DE index that falls below its assigned weight indicates an area on which to focus program efforts.

The contribution of each indicator to the “disempowered but adequate” component of the 5DE index is calculated as:

Where:

*DA20p* is the proportion of disempowered individuals multiplied by the mean adequacy score of the disempowered, by sex

*wi*is the weight of indicator *i*

*CAHi* is the censored adequacy headcount of indicator *i*, or the count of individuals who are disempowered but achieved adequacy in indicator *i* divided bythe total number of individuals with complete A-WEAI indicator data (i.e., the count of individuals who are empowered based on their adequacy score + the count of individuals who are disempowered based on their adequacy score and inadequate in indicator *i* + the count of individuals who are disempowered based on their adequacy score and adequate in indicator *i*)

### 12.2.2 Average percentage of women achieving adequacy across the six indicators of the A-WEAI

The purpose of reporting on the average percentage of women achieving adequacy across the six 5DE empowerment indicators overall and on each indicator is to focus on the composition of empowerment and disempowerment, and the individual indicators that present the greatest constraints to empowerment. This Feed the Future context indicator is calculated using the censored headcount ratio for each 5DE indicator for primary adult female decisionmakers only. The censored headcount ratio is the proportion of women in the population who are disempowered but achieve adequacy in an individual A-WEAI indicator. The censored headcounts help focus attention on those indicators that are the biggest constraints to empowerment.

## 12.3 Step-by-step instructions for computing the A-WEAI indicators

The following sections provide step-by-step guidance for computing the A-WEAI, including preparing the data, calculating the 5DE indicators, the 5DE, and the GPI, and then combining the 5DE and GPI to obtain the A-WEAI. They also provide step-by-step guidance for calculating the A-WEAI context indicator. These instructions are intended for use with Feed the Future ZOI Survey data. Appendix A, Table 1 provides a summary of the A-WEAI domains and indicators, relevant survey questions and Feed the Future variables, the method of aggregation, the inadequacy cut-off, and weights assigned to each indicator used to compute the overall A-WEAI index.[[79]](#footnote-78)

### 12.3.1 A-WEAI

The step-by-step procedures to calculate the A-WEAI indicator follows the Stata syntax in the following do files: *FTF ZOI Survey [COUNTRY] [YEAR] syntax awei1\_prep.do* and *FTF ZOI Survey [COUNTRY] [YEAR] syntax awei2\_calc.do*. Variables from Module 6 are used for the calculations.

#### Definitions

| Numerator | The weighted sum of the 5DE and the GPI among primary adult female decisionmakers |
| --- | --- |
| Denominator | N/A |
| Unit of measure | Scale from 0 to 1 (higher value indicates more empowerment) |
| Level of data | Person-level, but the data are found in the household data file |
| Sampling weight | Primary adult female decisionmaker, primary adult male decisionmaker |
| Disaggregation levels | Age category (18-29 years, 30+ years)\* |
| Coverage | P2-ZOI |
| Treatment of missing data | Primary female and male decisionmakers who do not have complete data for all six 5DE indicators are excluded from the A-WEAI calculations. In addition, households that do not have both a primary female decisionmaker and a primary male decisionmaker or that are missing complete data for all six 5DE indicators for either of the primary female and male decisionmakers are excluded from the GPI calculation. |
| Survey variables used | *All variables that begin with ‘v6’ and ‘m6’, hhea, hhnum, hh\_wgt* |
| Analytic variables used | *genhhtype* |
| Analytic variables created | *age\_cat, v6\_sex, nperhh, partact\_01-partact\_08, skip\_01-skip\_08, inputdec\_01-inputdec\_08, feelmakedec\_01-feelmakedec\_08, feelinputdecagr\_sum, feelinputdecagr, own\_01-own\_15, selfjointown\_01-selfjointown\_15, jown\_count, creditaccress\_1-creditaccess\_6, creditaccess, creditselfjointborrow\_1-creditselfjointborrow\_6, creditselfjointuse\_1-creditselfjointuse\_6, creditselfjointanydec\_1-creditselfjointanydec\_6,\_creditselfjointanydecany, incomedec\_01-incomedec-06, incomedec\_sum, indec\_count, groupmember\_01-groupmember11, groupmember\_any, work, work\_hours, z105, npoor\_z105, weight, ci, n\_missing, miss\_any, total\_w, threshold, ch\_20p, a\_20p, DAI\_20p, EAI\_20p, total\_b, pop\_shr\_before, temp, sample\_r\_before, pop\_shr\_after, sample\_r\_after, sample\_lost\_ratio, pop\_shr, MO\_20p, EA\_20p, cont\_group\_20, cont\_subgroup\_DAI, cont\_subgroup\_EAI, H\_20p, A\_20p, M0\_20p, EA\_20p, gpisub, sample5do, w\_ci\_id, m\_ci\_id, W\_ci, M\_ci, W\_cen\_ci, M\_cen\_ci, ci\_above, female, women\_n, women\_wt, inadequate, inadequate\_n, H, inadequate\_wt, H\_wt, ci\_gap, ci\_gap\_sum, ci\_average, H\_GPI, PI, GPI, AWEAI*  *For each var of variables (feelinputdecagr, jown\_count, credjanydec\_any, incdec\_count, groupmember\_any, npoor\_nz105): ‘var’\_depr, w\_‘var’, wg0\_‘var’, ‘var’\_miss, ‘var’\_ raw, ‘var’\_CH\_20p, ‘var’\_cont\_20\_EAI, ‘var’\_cont\_20\_DAI* |

\*Standard Feed the Future disaggregate

#### Calculations

**Data file preparation**

The data needed to compute the A-WEAI are contained in the household data file. The data for primary adult female decisionmakers come from Module 6 of the ZOI Survey and include variables that begin with *‘v6.’* The data for primary adult male decisionmakers come from Module 6M and include variables that begin with *‘m6.*’ So that the data for primary adult female and male decisionmakers can be processed together, the data for primary adult male decisionmakers will first be moved into a separate record, and then the variables will be renamed so that the variables for both primary adult female and male decisionmakers begin with *‘v6.’*

**A-1.** Load the household-level data file.

*Load “$analytic\FTF ZOI Survey [COUNTRY] [YEAR] household data analytic”*

**A-2.** Drop all variables for the primary adult female decisionmaker.

*Drop variables that begin with ‘v6’*

**A-3.** Drop all records that do not include data for a primary adult male decisionmaker—that is, drop all records if Module 6M was not completed.

*Drop record if m6605≠1*

**A-4.** Create a sex variable that indicates whether the record is for the primary adult male decisionmaker based on whether the male version of Module 6 was completed, which will be needed later in A-WEAI calculation.

*Set v6\_sex=1 if m6605=1*

**A-5.** Rename all primary adult male decisionmaker variables so that they start with *v6* instead of *m6*.

*Rename all variables that begin with ‘m6’ to begin with ‘v6’*

**A-6.** Save the data file to a temporary data file.

*Save “$analytic\Temp\temp\_mod6\_male.dta”*

**A-7.** Load the household-level data file again and create the same sex variable that was created in the temporary data file in Step A-4, but this time indicate whether the record is for the primary adult female decisionmaker based on whether the female version of Module 6 was completed.

*Set v6\_sex=2 if v6605=1*

**A-8.** Drop all records that do not include data for a primary adult female decisionmaker—that is, drop all records if Module 6F was not completed.

*Drop record if v6605≠1*

**A-9.** Append primary adult male decisionmaker data to the primary adult female decisionmaker data file.

*Append using "$analytic\Temp\temp\_mod6\_male.dta”*

**A-10.** Drop primary adult male decisionmaker variables that begin with *‘m6.’*

*Drop variables that begin with ‘m6’*

**A-11.** Sort data by cluster, household number, and respondent line number, and ensure that there are at most two records per household. Investigate and resolve any cases in which there are more than two records per household.

*Sort by hhea hhnum v6100c*

*By hhea hhnum: set nperhh=number of primary adult decisionmaker records*

*Label variable “Number of primary adult decisionmaker records per household”*

**A-12.** Drop any records for primary adult decisionmakers who are not de jure household members.

*Drop record if mdm=1 and mdm\_dj≠1*

*Drop record if fdm=1 and fdm\_dj≠1*

**A-13.** Save the individual respondent-level A-WEAI data file.

*Save “$analytic\Results\AWEAI\FTF ZOI Survey [COUNTRY] [YEAR] aweai\_pre”*

**B. 5DE Indicator calculations**

The following steps describe how to create adequacy scores for the six 5DE indicators across the five domains.

**B.1. Domain 1: Decisionmaking over production**

Note that only activities a, b, c, and f 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 Domain 3 indicator calculation.

**B-1.1.** Review and prepare the data.

**B-1.1A.** Review frequency distributions and the number of missing cases for items 1 to 6 for variables v6201, v6202, v6203, and v6204.

**B-1.1B.** Ensure that skip patterns are correct; if the respondent did not participate in the activity s/he should not have a response for v6202 v6203 v6204.

**B-1.1C.** Create new variables for v6201, v6203, v6204 that set missing and not applicable responses (i.e., 9, 93, and 99) to missing. Also create v6201 variables for items 7 and 8 for use in later steps.

*For each economic activity (x) of 1-6:*

*Set v6201\_`x’x=v6201\_`x’*

*Replace v6201\_`x’x=missing if v6201\_`x’=9*

*For each economic activitiey (x) of 1-8:*

*Set v6204\_`x’x=v6204\_`x’*

*Replace v6204\_`x’x=missing if v6204\_`x’=9*

*Set v6203\_`x’x=v6203\_`x’*

*Replace v6203\_`x’x=missing if v6203\_`x’=93 or v6203\_`x’=99For each economic activitiey (x) of 7 and 8:*

*Set v6201\_`x'x=missing*

*Replace v6201\_`x'x=1 if v6202\_`x'≠"" or v6203\_`x'≠missing*

**B-1.2.** Create binary variables indicating participation in economic activities. The survey collects information about respondents’ participation in eight productive activities (*v6201\_1*–*v6201\_8*). This information is used to generate eight binary participation variables (*partact\_1*–*partact\_8*), one for each activity. If respondents participated in the activity, the variable is coded as 1. If respondents 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 18:*

*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’ “Participated in economic activity `x’”*

**B-1.3.** Create a variable that counts the number of economic activities in which the respondent participated (*partact*).

*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 activities participated in”*

**B-1.4.** Create a variable that counts the number of agriculture-related activities in which the respondent participated.

*Set partactagr=partact\_1+partact\_2+partact\_3+partact\_6*

*Replace partactagr=missing if partact\_1=missing and if partact\_2=missing if partact\_3=missing and if partact\_6=missing*

*Label variable “Number of agricultural activities participated in”*

**B-1.5.** Create binary variables for each activity to indicate whether the respondent has adequate participation in decisionmaking. The survey asks respondents about their input in decisionmaking for the eight activities (*v6203\_1*–*v6203\_8*), which is used to generate eight binary decisionmaking input variables (*inputdec\_1*–*inputdec\_8*). Respondents who reported that they have input in some, most, or all decisions are coded as 1, and respondents who reported that they have no input or input in a few decisions are coded as 0. Respondents who did not participate in the activity and respondents who reported “no decisions made” are excluded from the calculation of the sub-indicator and coded as missing.

*For each value (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’ “Has some input in decisions regarding `x’”*

**B-1.6.** Create binary variables for each activity to indicate whether the respondent feels that he or she has autonomy in decisionmaking. The survey asks respondents about their autonomy in decisionmaking for the eight activities (*v6204\_1*–*v6204\_8*), which is used to generate eight binary variables on autonomy in decisionmaking (*feelmakedec\_1*–*feelmakedec\_8*). The response: “only *self*,” coded as 1 on the decisionmaking filter (*v6202aa\_1*–*v6202aa\_8*), is automatically coded as 1 for the autonomy in decisionmaking variable. Respondents who report 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 are excluded from the calculation of the sub-indicator and coded as missing.

*For each value (x) of 1 2 3 4 5 6 7 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’ “Feels can make decisions about `x’”*

**B-1.7.** Create a variable that counts the number of agricultural economic activities for which respondents have input into at least some decisions, plus the number of agriculture economic activities for which respondents feel that they can make decisions to a medium or high extent if they wanted to. This variable (*feelinputdecagr\_sum*) is a count of the variables related to agricultural economic activities created in Steps B-1.5 and B-1.6 that have a value of 1 (meaning ‘YES’). The variable can have a value between 0 and 8. Variables created in Steps B-1.5 and B-1.6 related to the following economic activities are not included in the count because they are not related to agriculture: nonfarm activity, wage and salary employment, and major and minor household expenditures.

*Set feelinputdecagr\_sum=inputdec\_1+inputdec\_2+inputdec\_3+inputdec\_6+feelmakedec\_1+ feelmakedec\_2+feelmakedec\_3+feelmakedec\_6*

*Replace feelinputdecagr=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*

*Label variable “Number of agricultural activities makes decisions or feels can”*

**B-1.8.** Create the indicator of adequate achievement in agricultural production. The indicator (*feelinputdecagr*) is coded as 1 if the variable created in Step B-1.7, *feelinputdecagr\_sum*, has a value of 1 or more. The indicator is coded as missing if the household does not engage in any agricultural activity.

*Set feelinputdecagr=1 if feelinputdecagr\_sum≥1*

*Replace feelininputdecagr=0 if feelinputdecagr\_sum<1*

*Replace feelinputdecagr=missing if feelinputdecagr\_sum=missing*

*Label values 0 “No” 1 “Yes”*

*Label variable “Has decision-making inputs/power in 1+ area”*

**B.2. Domain 2: Access to Resources #1 Ownership of Assets**

**B-2.1.** Review and prepare the data.

**B-2.1A.** Review frequency distributions and number of missing cases.

**B-2.1B:** Create new variables for v6301 that have no responses (2) recoded as 0 and missing responses (9) recoded as missing.

*For each asset (x) of 1-15:*

*Set v6301\_`x’x=v6301\_`x’*

*Replace v6301\_`x’x=0 if v6301\_`x*

*Replace v6301\_`x’x=missing if v6301\_`x’=9*

**B-2.2.** Create binary variables indicating whether someone in the household owns each asset. The survey collects information on household ownership of 15 kinds of assets (*v6301a*–*v6301n*). These variables are used to create 15 binary variables on household asset ownership (*own\_a*–*own\_n*), 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. If respondents do not provide a response, the variable is coded as missing.

*For each value 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 `x’ “Household owns `x’”*

**B-2.3.** Create variables that count the number of total assets the household owns (*own\_sum*) and the number of agricultural assets the household owns (*ownagr\_sum*).

*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”*

*Set ownagr\_sum= own\_01+own\_02+own\_03+own\_04+own\_05+own\_06+own\_07+own\_08*

*Replace own\_sum=missing if own\_01 through own\_08 are all missing values*

*Label variable “Number of agricultural asset types household owns”*

**B-2.4.** Create binary variables for each asset indicating whether respondent owns the asset alone or jointly. The variables created in Step B-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 respondents alone or jointly own the asset and as 0 if someone else owns the asset. If the household does not own the asset, the variable is coded as missing.

*For each value (x) of 0-15:*

*Set selfjointown\_`x'=1 if v6303a\_`x'=1 and own\_`x'=1*

*Replace selfjointown\_`x'=0 if v6303a\_`x'≠1 and own\_`x'=1*

*Replace selfjointown\_`x'=missing if own\_`x'≠1*

*Label values 0 “No” 1 “Yes”*

*Label variable `x’ “Owns asset alone or jointly”*

**B-2.5.** Create a variable that counts the number of assets owned by respondents. This variable (*selfjointown\_sum*) is a count of the number of assets owned solely or jointly by the respondent.

*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 own\_sum=missing if own\_01 through own\_15 are all missing values*

*Label variable “No. of assets the respondent owns alone or jointly”*

**B-2.6.** Create the indicator for adequate achievement in asset ownership. This variable (*jown\_count*) 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 and then recoding *jown\_count* to 0 if the respondent owns only one asset that is considered to be a small asset (i.e., chickens [d], handtools for farming [f], non-mechanized farm equipment [ff], or small consumer durables [k]).

*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 ownsum=0*

*Label values 0 “No” 1 “Yes”*

*Label variable “Owns 1+ large or 2 small assets alone or jointly”*

**B.3. Domain 2: Access to Resources, #2 Access to Credit**

**B-3.1.** Review and prepare the data. Review frequency distributions and number of missing cases.

**B-3.2.** Create binary variables indicating whether the household borrowed from each source of credit. The survey collects information on access to credit from six lending 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 lending source. 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 lending source, the variable is coded as 0. If the respondent does not know or refuses to answer, the variable is coded as missing.

*For each value (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  
Label values 0 “No” 1 “Yes”*

*Label variable `x’ “Household borrowed from `x’”*

**B-3.3.** Create a variable that indicates the number of credit sources the household used.

*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”*

**B-3.4.** For each source of credit, create a variable indicating whether respondents participated in decisions related to borrowing or how to use the borrowed money or item. First create six variables that indicate whether respondents participated in the decision to borrow from each lending source (*creditselfjointborrow\_1*–*creditselfjointborrow\_6*). Then create six variables that indicate whether respondents participated in decisions on what to do with the money or item borrowed from each lending source (*creditselfjointuse\_1*–*creditselfjointuse\_6*). Use the variables created to create another set of variables that indicate whether the respondent made at least one credit decision for each lending source (*creditselfjointanydec\_1*–*creditselfjointanydec\_6*). All variables are coded as 1 if respondents alone or jointly made the decision and 0 otherwise. Variables are coded as missing if a response was missing or if the response was not applicable.

*For each value (x) of 1-6:*

*Set creditselfjointborrow\_`x'=1 if v6309a\_`x'=1 and creditaccess\_`x'=1*

*Replace creditselfjointborrow\_`x'=0 if v6309a\_`x'≠1 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 `x’ “Participated in the decision to borrow from `x’”*

*Set creditselfjointuse\_`x'=1 if v6310a\_`x= 1 and creditaccess\_`x'=1*

*Replace creditselfjointuse\_`x'=0 if v6310a\_`x'≠1 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 `x’ “Participated in decision how to use money/item borrowed from `x’”*

*Set creditselfjointanydec\_`x'=1 if creditselfjointborrow\_`x'=1 or 1 creditselfjointuse\_`x'=1*

*Replace creditselfjointanydec\_`x'=0 if creditselfjointborrow\_`x'≠1 and creditselfjointuse\_`x'≠1*

*Replace creditselfjointanydec\_`x'=missing if creditselfjointborrow\_`x'=missing or creditselfjointuse\_`x'=missing*

*Label values 0 “No” 1 “Yes”*

*Label variable `x’ “Made 1+ credit decision from `x’”*

**B-3.5.** Create the indicator for adequate achievement in access to credit. If respondents can make at least one decision on at least one source of credit alone or jointly, the indicator (*credjanydec\_any*) is coded as 1. If respondents cannot make any decision or provide input in the decision to borrow from any source of credit, the indicator is coded as 0. If a household did not use any form of credit, the indicator is coded as 0.

*Set 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=0 if creditaccess=0*

*Replace credjanydec\_any=missing if creditaccess=missing*

*Label values 0 “No” 1 “Yes”*

*Label variable “Made 1+ decision regarding 1+ source of credit alone or jointly”*

**B.4. Domain 3: Control over income**

**B-4.1.** Review and prepare the data.

**B-4.1A.** Review frequency distributions and number of missing cases.

**B-4.1B.** Ensure that skip patterns are correct; if the respondent did not participate in the activity s/he should not have a response for v6205.

**B-4.1C.** Create new variables for v6205 that recode missing values (99) and no decisions made values (93) to missing

*For each economic activity (x) of 1-6:*

*Set v6205\_`x’x=v6205\_`x’*

*Replace v6205\_`x’x=missing if v6205\_`x’=99 or v6205\_`x’=93*

**B-4.2.** Create binary variables to indicate whether the respondent had input into decisions made on the use of income from each activity. The survey collects data on decisionmaking for the use of income generated from food crop farming, cash crop farming, livestock raising, nonfarm activities, wage and salary employment, and fishing (*v6205\_1*–*v6205\_6*). This information is used to generate six income decisionmaking variables (*incomedec\_1*–*incomedec\_6*), one for each activity. Responses that indicate that respondents have input into some, or most or all decisions are coded as 1, and responses that indicate that respondents haveno input or input into very few decisionsare coded as 0. Respondents who do not participate in the activity and respondents who report no decisions made are coded as missing.

*For each value (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 "Has some input in decisions regarding income from `x’”*

**B-4.3.** Create a variable that counts the number of relevant activities for which respondents have input into decisions on the use of income or feel that they could make decisions if they wanted. The variable (*incdec\_sum*) is a sum of the binary variables created in the previous step and also three of the variables created in Step B-1.6 (i.e., those related to non-farm economic activities, wage and salary work, and major household expenses). The variable can have a value between 0 and 10, depending on the number of variables with a value of 1:

*Set incomedec\_sum=missing*

*Replace incomedec\_sum=incomedec\_1+incomedec\_2+incomedec\_3+incomedec\_4+incomedec\_5+ incomedec\_6+feelmakedec\_4+feelmakedec\_5+feelmakedec\_7+feelmakedec\_8*

*Label variable “No. of areas in which has income decision-making input/power”*

**B-4.4.** Create the indicator for adequate achievement in control over income (*incdec\_count*). Responses are coded as 1 if respondents have input into some, or most or all decisions on the use of income for at least one activity or feel that they can make decisions to a medium or high extent for at least one activity. If respondents do not have input into decisions on the use of income or feel that they can make decisions to a medium or high extent for at least one activity, the indicator is recoded as 0. Also recode the indicator to 0 for respondents who can only make decisions about minor household expenditures.

*Set incdec\_count=1 if incomedec\_sum>0 and incomedec\_sum≠missing*

*Replace incdec\_count=0 if incdec\_count=1 and incomedec\_sum=1 and feelmakedec\_8=1*

*Replace incdec\_count=missing if incomedec\_sum=missing*

*Label values 0 “No” 1 “yes”*

*Label variable "Has some input/power in income decisions AND not only about minor household expenditures"*

**B.5. Domain 4: Leadership (group membership)**

**B-5.1.** Review and prepare the data. Review frequency distributions and the number of missing cases. Ensure that skip patterns are correct.

**B-5.2.** Create binary variables to indicate whether the respondent belongs to each type of community group. The survey collects information about respondents’ 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 a member, the variable is coded as 1. If a group exists in the community and the respondent is not a member, the variable is coded as 0. If a group exists in the community but the respondent refuses to answer the question, the variable is coded as missing. If a group does not exist in the community or respondents do not know whether it exists, the variable is coded as 0.

*For each value (x) of 01-11:*

*Set groupmember\_`x'=1 if v6405\_`x'=1 and v6404\_`x'=1*

*Replace groupmember\_`x'=0 if (v6405\_`x'=2 and v6404\_`x'=1) or (v6404\_`x’=2 or v6404\_`x’=8 or v6404\_`x’=missing)*

*Replace groupmember\_`x'=missing if (v6405\_`x'=missing or v6405\_`x'=7) and v6404\_`x'=1*

*Label values 0 “No” 1 “Yes”*

*Label variable `x’ “Belongs to `x’”*

**B-5.3.** Create the indicator for adequate achievement in group membership (*groupmember\_any*), which is coded as 1 if the respondent is a member of at least one group and coded as 0 if the respondent does not belong to any group.

*Set 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=0 if groupmember\_any=missing*

*Label values 0 “No” 1 “Yes”*

*Label variable “Belongs to at least one community group”*

**B.6. Domain 5: Time Allocation (Workload)**

**B-6.1.** Review and prepare the data. Review frequency distributions and the number of missing cases.

**B-6.2.** Create a variable to count the number of 15-minute increments that respondents spent performing work activities as their primary activities (*w*). 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*). Fourteen activities (D through Q) (*z*) are considered to be work activities.

*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:*

*For each value (z) in D E F G H I J K L M N O P Q:*

*Replace work=work+1 if v6601p\_`x'\_`i'="`z'"*

*Label variable “Number of 15-min. increments worked day prior to survey”*

**B-6.3.** Convert into hours the total number of 15-minute increments spent working.

*Set work\_hours=work÷4*

*Label work\_hours “Number of hours worked day prior to survey”*

**B-6.4.** Define time poverty line at 10.5 hours per day, 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 live (10.5 hours per day)”*

**B-6.5.** Create the indicator of adequate achievement in productive and domestic workload (*npoor\_z105*). If the respondents’ time in work and domestic activities (work) 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, as shown:

*Set npoor\_z105=0 if work\_hours>z105*

*Replace npoor\_z105=1 if work\_hours≤z105*

*Label values 0 “No” 1 “Yes”*

*Label variable “Worked 10.5 hours or less day before survey”*

**B-6.6.** Save the data file.

*Save “$analytic\Temp\awei prep”*

**C. A-WEAI calculation**

The remainder of the syntax for the A-WEAI calculation is contained in the “FTF ZOI Survey [COUNTRY] [YEAR] syntax AWEAI\_2 calc” Stata do file.

**C.1. Weighted inadequacy and adequacy scores**

**C-1.1.** Define a variable list (*varlist\_emp*) that includes the six indicators of achievement of adequacy created in previous steps.

*Set varlist\_emp=(1) feelinputdecagr, (2) jown\_count, (3) credjanydec\_any, (4) incdec\_count, (5) groupmember\_any, and (6) npoor\_ nz105.*

**C-1.2.** Transform the 5DE indicators so that 1 identifies those who are inadequate. In previous steps, indicators were created to identify respondents who are adequate in achievement of the six empowerment indicators. In this step, those six indicators are renamed to have the suffix “*\_ndepr*” appended to their names, and the original variables are reverse-coded so that 1 is now equal to inadequate empowerment (disempowerment) (1=inadequate, 0=adequate).

*For each (var) in varlist\_emp:*

*Rename `var' `var'\_ndepr*

*Set `var'=1 if `var'\_ndepr=0*

*Replace `var'=0 if `var'\_ndepr=1*

**C-1.3.** Create a weight variable (*weight*)equal to the primary adult female and male sampling weights.

*Set weight=missing*

*Replace weight=wgt\_mdmif v6\_sex=1*

*Replace weight=wgt\_fpdmif v6\_sex=2*

**C-1.4.** Keep only records for households with primary adult male and female decisionmakers and primary adult female decisionmaker-only households. Drop records for primary adult male decisionmaker-only households.

*Drop if vtype=2*

**C-1.6.** Define the domain indicator weights.

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.

*For each variable (var) of variable list feelinputdecagr incdec\_count groupmember\_any npoor\_z105:*

*gen w\_`var' =(1/5)*

*Set w\_jown\_count =(2/5)*

*Set w\_credjanydec\_any =(1/15)*

**C-1.7.** Apply the weights to each indicator.

*For each variable (var) of varlist\_emp:*

*Set wg0\_`var'=`var'\*w\_`var'*

**C-1.8.** Compute the frequency of missing values for each 5DE indicator.

*For each variable (var) of varlist\_emp:*

*Set `var'\_miss=1 if `var'=missing*

*Replace `var'\_miss=0 if `var'≠missing*

**C-1.9.** Calculate the weighted disempowerment inadequacy score for each respondent (*ci*).

*Set ci=wg0\_feelinputdecagr+wg0\_jown\_count+wg0\_credjanydec\_any+wg0\_incdec\_count+ wg0\_groupmember\_any+wg0\_npoor\_nz105*

*Replace ci=round(ci,.0001)*

*Label variable "Weighted inadequacy Count"*

**C-1.10.** Determine the number of 5DE indicators 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\_ nz105=missing*

**C-1.11.** Create a variable (*miss\_any*) to indicate whether the respondent is missing any 5DE indicators.

*Set miss\_any=0*

*Replace miss\_any=1 if n\_missing>0 and n\_missing≠missing*

**C-1.12.** Check for records with missing values, and drop any records with missing values.

*Drop record if miss\_any=1*

**C-1.13.** Save the data file.

*Save “$analytic\Temp\aweai depr indicators”*

**C.2. Censored inadequacy and adequacy scores**

**C-2.1.** Create a total weight variable (*total\_w*)—a constant that sums all individual sampling weights for respondents with complete 5DE indicator data.

*Set total\_w=total(weight) if miss\_any=0*

**C-2.2.** Create a binary variable that indicates whether primary adult decisionmakers should be included from the censored inadequacy headcount (*ch\_20p)*—that is, apply a threshold of 0.2 to the censored head counts to identify those with inadequate empowerment higher than the threshold. Respondents with an inadequacy score of less than or equal to 0.2 are excluded from the censored inadequacy headcount, even if their score is not 0.

*Set ch\_20p=1 if ci>0.2*

*Replace ch\_20p=0 if ci≤0.2*

*Replace ch\_20p=missing if miss\_any=1*

**C-2.3.** Calculate the individual inadequacy of those who are disempowered (*a\_20p*). The scores of respondents with an inadequacy score less than 0.2 are recoded to 0.

*Set a\_20p=ci if miss\_any=0*

*Replace a\_20p=0 if ch\_20p=0*

**C-2.4.** Calculate the disempowerment index (*DAI*) and empowerment index (*EAI=1-DAI*).

*Set DAI\_20p=total(ci\*ch\_20p\*weight÷total\_w)*

*Set EAI\_20p=1–DAI\_20p*

**C-2.5.** Calculate the raw, or uncensored, headcounts for each 5DE indicator.

*For each variable (var) in varlist\_emp:*

*Set `var'\_raw=`var'*

*Replace `var'\_raw=missing if miss\_any=1*

*Summarize `var’\_raw using individual weight=weight*

**C.3. Censored headcounts by gender**

**C-3.1.** Create a new variable for each 5DE indicator that reflects the censored headcount of disempowered individuals who are ***inadequate*** in that indicator. Set the indicator value to missing if any of the 5DE indicators are missing a value for that respondent.

*For each variable (var) in varlist\_emp:*

*Set `var'\_CH\_20p=1 if `var'=1 and ch\_20p=1*

*Replace `var'\_CH\_20p=missing if miss\_any=1*

**C-3.2.** Create a new variable for each 5DE indicator that reflects the censored headcount of disempowered individuals who are ***adequate*** in that indicator. Set the indicator value to missing if any of the 5DE indicators are missing a value for that respondent. [These variables will be used to calculate the A-WEAI context indicator, as explained in Section 12.3.2.]

*For each variable (var) in feelinputdecagr\_ndepr jown\_count\_ndepr credjanydec\_any\_ndepr incdec\_count\_ndepr groupmember\_any\_ndepr npoor\_z105\_ndepr:*

*Set `var'\_CH\_20p=1 if `var'=1 and ch\_20p=1*

*Replace `var'\_CH\_20p=missing if miss\_any=1*

**C-3.3.** Determine sample size, including all records (*before*), and include only records that have complete 5DE indicator information (*after*). Create a variable (*total\_b*) that captures the total sample. Create a variable (*pop\_shr\_before*) that captures the weighted population share of each gender before the sample reduction. Create a variable (*sample\_r\_before*) that captures the sample size of each gender before sample reduction. Create comparable variables (*pop\_shr\_after*, *sample\_r\_after*) that capture the share and sample size of each gender after the sample reduction. Create a ratio (*sample\_lost\_ratio*) that captures the relative size of the final sample after reduction in each gender.

*Set total\_b=total(weight)*

*Set pop\_shr\_before=total(weight÷total\_b), by(v6\_sex)*

*Set temp=1*

*Set sample\_r\_before=total(temp), by(v6\_sex)*

*Set pop\_shr\_after=total(weight÷total\_w) if miss\_any=0, by(v6\_sex)*

*Set sample\_r\_after=total(temp) if miss\_any=0, by(v6\_sex)*

*Set sample\_lost\_ratio=sample\_r\_after÷sample\_r\_before*

**C-3.4.** Create a third population share variable by gender (*pop\_shr*).

*Set pop\_shr=total(weight÷total\_w) if miss\_any=0, by(v6\_sex)*

**C-3.5.** Save the individual-level data file, before transforming it to a household-level data file.

*Save “$analytic\Temp\aweai individual indices”*

**C-3.6.** Calculate weighted averages of the variables created in Step C by sex. Thus far the data have been individual-level data.

*By v6\_sex: Calculate the means of the following variables: ch\_20p; a\_20p; miss\_any; DAI\_20p; EAI\_20p; all variables that contain \_CH\_20p; all variables that end with \_raw or \_miss; all variables that start with w\_, pop\_shr, or sample\_r\_; sample\_lost\_ratio [aw=weight]*

**C-3.7.** Calculate *M0* (subgroup *DAI,* which reflects disempowerment) and *EA* (subgroup *EAI,* which reflects empowerment) by gender. This is done by multiplying the condition of disempowerment (*ch\_p20*) by the individual average inadequacy (*a\_20p*) to get the sex subgroup *DAI* (*M0*), and then subtracting the result from 1 to get the sex subgroup *EAI* (*EA\_20p*)

*Set M0\_20p=ch\_20p\*a\_20p*

*Set EA\_20p=1–M0\_20p*

**C-3.8.** Rename the variable *ch\_20p* to be *H\_20p* and the variable *a\_20p* to be *A\_20p*.

*Rename ch\_20p H\_20p*

*Rename a\_20p*

**C-3.9.** Calculate the decomposed contribution of each 5DE variable to total empowerment and disempowerment.

*For each variable (var) in varlist\_emp:*

*Set `var'\_cont\_20\_EAI=(`var'\_CH\_20p\*w\_`var')÷EA\_20'p*

*Set `var'\_cont\_20\_DAI=(`var'\_CH\_20p\*w\_`var')÷M0\_20p*

**C-3.10.** Calculate decomposed contribution and gender contribution to DAI and EAI.

*Set cont\_group\_20=M0\_20p÷DAI\_20p\*pop\_shr*

*Set cont\_subgroup\_DAI\_20=M0\_20p÷DAI\_20p\*pop\_shr\_after*

*Set cont\_subgroup\_EAI\_20=EA\_20p÷EAI\_20p\*pop\_shr\_after*

**C-3.11.** Save the data file with the results by sex.

*Save “$analytic\Results\AWEAI\FTF ZOI Survey [COUNTRY] [YEAR] aweai results gender*”

**C-3.12.** Review the 5DE results by gender in the data file.

**C-3.12A.** Review the disempowered headcount (*H\_20p*), average inadequacy score (*A\_20p*), disempowerment index (*M0\_20p*), and empowerment index (*EA\_20P*) by gender:

*Browse H\_20p A\_20p M0\_20p EA\_20p if v6\_sex=2*

*Browse H\_20p A\_20p M0\_20p EA\_20p if v6\_sex=1*

**C-3.12B.** Review the censored headcounts for 5DE indicators by gender:

*Browse all variables that end with \_CH\_20p if v6\_sex=2*

*Browse all variables that end with \_CH\_20p if v6\_sex=1*

**C-3.12C.** Review the 5DE indicator contributions to disempowerment by gender:

*Browse all variables that end with cont\_20\_DAI if v6\_sex=2*

*Browse all variables that end with cont\_20\_DAI if v6\_sex=1*

**C.4. GPI**

The GPI is calculated using the following steps.

**C-4.1.** Open the temporary data file saved in Step C-1.13.

*Load “$analytic\Temp\aweai\_depr\_indicators*.

**C-4.2.** Define the sample that will be used to compute the GPI (*gpisub*). The sample will include one pair of primary adult female and male decisionmakers with complete information on all six 5DE indicators per household. The GPI excludes households that lack complete information on both a primary adult male decisionmaker and a primary adult female decisionmaker. Women in female adult-only households and men in male adult-only households do not contribute to the GPI. Keep only records included in the GPI subsample.

*Sort by hhea hhnum v6\_sex*

*Set temp=number of individual records per household*

*Set gpisub=1 if temp=2*

*Set temp2=number of individual male and female records per household*

*Replace gpisub=0 if temp2≠1*

*Keep if gpisub=1*

**C-4.3.** Create a variable that captures records that have complete information for all 5DE indicators and set *ci* to missing if the variable is 0.

*Set sample5do=1 if feelinputdecagr≠missing and jown\_count≠missing and credjanydec\_any≠missing and incdec\_count≠missing and groupmember\_any≠missing and npoor\_z105≠missing*

*Replace sample5do=0 if feelinputdecagr=missing or jown\_count=missing or credjanydec\_any=missing or incdec\_count=missing or groupmember\_any=missing or npoor\_z105=missing*

*Replace ci=missing if sample5do=0*

**C-4.4.** Compute censored inadequacy scores for men and women by household (*M\_cen\_ci* and *W\_cen\_ci*).

*Set w\_ci\_id=ci if v6\_sex=2 and gpisub=1*

*Set m\_ci\_id=ci if v6\_sex=1 and gpisub=1*

*Set W\_ci=max(w\_ci\_id), by hhnum*

*Set M\_ci=max(m\_ci\_id), by hhnum*

*Set W\_cen\_ci=W\_ci*

*Replace W\_cen\_ci=0.2 if W\_cen\_ci≤0.2 and W\_cen\_ci≠missing*

*Set M\_cen\_ci=M\_ci*

*Replace M\_cen\_ci=0.2 if M\_cen\_ci≤0.2 and M\_cen\_ci≠missing*

**C-4.5.** Create a variable (*ci\_above*) to identify households without gender parity. Set the variable to 1 if the primary female decisionmaker is more disempowered than the primary male decisionmaker in a household. Set the variable to 0 if the primary female decisionmaker is more empowered than the primary male decisionmaker, or if the two decisionmakers are equally empowered. Set the variable to missing if an individual score is missing.

*Set ci\_above=1 if W\_cen\_ci>M\_cen\_ci*

*Replace ci\_above=0 if W\_cen\_ci≤M\_cen\_ci*

*Replace ci\_above=missing if W\_cen\_ci=missing or M\_cen\_ci=missing*

**C-4.6.** Compute the GPI.

**C-4.6A.** Create the weight for primary female decisionmakers (*women\_wt*).

*Set female=I if v6\_sex=2 and ci\_above≠missing*

*Set women\_n=total(female)*

*Set women\_wt=total(female\*weight)*

**C-4.6B.** Calculate weighted and unweighted headcount ratio of parity inadequate women.

*Set inadequate=1 if ci\_above=1 and v6\_sex=2*

*Set inadequate\_n=total(inadequate)*

*Set H=inadequate\_n÷women\_n*

*Set inadequate\_wt=total(inadequate\*weight)*

*Set H\_wt=inadequate\_wt÷women\_wt*

**C-4.6C.** Calculate normalized (*average*) gender parity gap*.*

*Set ci\_gap= W\_cen\_ci – M\_cen\_ci)÷(1–M\_cen\_ci) if ci\_above=1 and v6\_sex=2*

*Set ci\_gap\_sum=total(ci\_gap\*weight)*

*Set ci\_average=ci\_gap\_sum÷inadequate\_wt*

**C-4.6D.** Calculate GPI (*GPI*).

*Set H\_GPI=inadequate\_wt÷women\_wt*

*Set P1=H\_GPI\*ci\_average*

*Set GPI=1–P1*

**C-4.7.** Summarize and extract GPI results. The variable *H\_GPI* is the percentage of women without gender parity (Hgpi). The variable *ci\_average* is the average empowerment gap (Igpi), and the variable *GPI* is the GPI.

*Summarize H\_GPI ci\_average P1 GPI*

*Count if v6\_sex=2*

*Tabulate women\_n women\_wt*

**C-4.8.** Save the data file with the GPI results.

*Save “$analytic\Results\AWEAI\aweai\_results\_GPI”*

**D. Combining the 5DE and the GPI**

The A-WEAI is computed by summing the weighted value of the women’s 5DE index (*1-M0*) and the GPI. The 5DE index receives a weight of 0.9 (90 percent), and the GPI receives a weight of 0.1 (10 percent). This can be done in Excel after extracting the needed values from the Stata data files.

**D-1.** Extract the value of *EA\_20p* for women from the A-WEAI gender results file:

*“$analytic\Results\AWEAI\FTF ZOI Survey [COUNTRY] [YEAR] aweai results gender”*

**D-2.** Extract the *GPI* value from the A-WEAI GPI results file:

*“$analytic\Results\AWEAI\aweai\_results\_GPI”*

**D-3.** Substitute the two values into the following equation:

*A-WEAI=(0.9\*EA\_20p for women)+(0.1\*GPI)*

**E. Calculating the A-WEAI score by age category**

**E-1.** Load the data file created in Step C-1.13 that has the calculated weighted inadequacy scores,

*Load “$analytic\Temp\aweai\_depr\_indicators”*

**E-2.** Create a variable that indicates whether the primary adult female decisionmaker is between 18 and 29 years of age or if she is older than 29 years of age, according to information in the A-WEAI module. If the primary adult female decisionmaker is missing an age value in the A-WEAI module, use her age in the household roster.

*Set age\_cat=missing*

*Replace age\_cat=1 if v6102<30*

*Replace age\_cat=2 if v6102≥30 and v6102<98*

*Replace age\_cat= 1 if youth\_fdm\_dj=1 and age\_cat=missing*

*Replace age\_cat=2 if youth\_fdm\_dj=0 and age\_cat=missing*

**E-3.** Keep only the records for primary adult female decisionmakers who are 18-29 years of age.

*Keep if age\_cat=1*

**E-4.** Starting with Step C.2 (*Censored inadequacy and adequacy scores)*, repeat all the steps to calculate the 5DE index, GPI, and A-WEAI) for women 18-29 years of age.

**E-5.** Keep only the records for primary adult female decisionmakers who are 30 years of age or older.

*Keep if age\_cat=2*

**E-6.** Starting with Step C.2 (*Censored inadequacy and adequacy scores)*, repeat all the steps to calculate the 5DE index, GPI, and A-WEAI for women who are 30 years of age or older.

### 12.3.2 Average percentage of women achieving adequacy across the six indicators of the A-WEAI

The step-by-step procedures to calculate the average percentage of women achieving adequacy across the six indicators of the A-WEAI follows the Stata syntax in *FTF ZOI Survey [COUNTRY] [YEAR] syntax awei3\_context.do*. It requires that the A-WEAI indicator has already been calculated usingthe following do files: *FTF ZOI Survey [COUNTRY] [YEAR] syntax awei1\_prep.do* and *FTF ZOI Survey [COUNTRY] [YEAR] syntax awei2\_calc.do*. Variables created when calculating the A-WEAI indicator are used to calculate this indicator.

#### Definitions

| Numerator | Sum across the six A-WEAI indicators: number of primary adult female decisionmakers who are disempowered according to A-WEAI but achieve adequacy in the A-WEAI indicator divided by the total number of primary adult female decisionmakers with complete A-WEAI indicator information |
| --- | --- |
| Denominator | Total number of A-WEAI indicators (i.e., 6) |
| Unit of measure | Percentage |
| Level of data | Person-level, but the data are found in the household data file |
| Sampling weight | Primary adult female decisionmaker |
| Disaggregation levels | Age category (18-29 years, 30+ years)\* |
| Coverage | P2-ZOI |
| Treatment of missing data | Female and primary decisionmakers who do not have complete data for all six A-WEAI indicators are excluded from the calculations. |
| Survey variables used | *wgt\_fpdm, hhea, samp\_stratum* |
| Analytic variables used | *v6\_sex, miss\_any, feelinputdecagr\_ndepr\_CH\_20p, jown\_count\_ ndepr\_CH\_20p, credjanydec\_any*\_ *ndepr\_CH\_20p, incdec\_count*\_ *ndepr\_CH\_20p, groupmember\_any*\_ *ndepr\_CH\_20p, npoor\_nz105\_ ndepr\_CH\_20p* |
| Analytic variables created | *mean\_overall, age\_cat, mean\_lt30, mean\_ge\_30,*  *For each variable X of (feelinputdecagr\_ ndepr\_CH\_20p, jown\_count\_ ndepr\_CH\_20p, credjanydec\_any*\_ *ndepr\_CH\_20p, incdec\_count*\_ *ndepr\_CH\_20p, groupmember\_any*\_ *ndepr\_CH\_20p, npoor\_nz105\_ ndepr\_CH\_20p): mean\_X, mean\_lt30\_X, mean\_ge30\_X* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Using the A\_WEAI individual indices data file, define a variable list (*varlist\_emp*) that includes the censored headcount variables for the six indicators of achievement of adequacy created in Step C-3.2 in calculation of the A-WEAI index.

*Load “$analytic\Temp\aweai\_individual\_indices.dta”*

*Set varlist\_emp=(1) feelinputdecagr\_ndepr\_CH\_20p, (2) jown\_count\_ndepr\_CH\_20p, (3) credjanydec\_any\_ndepr\_CH\_20p, (4) incdec\_count\_ndepr\_CH\_20p, (5) groupmember\_any\_ndepr\_CH\_20p, and (6) npoor\_nz105\_ndepr\_CH\_20p*

**Step 2.** Drop all records for primary adult male decisionmakers.

*Drop if v6\_sex=1*

**Step 3.** Apply the primary adult female decisionmaker weight and create variables that capture the average percentage of sample-weighted disempowered women who achieve adequacy in each of the six empowerment indicators.

*Svyset hhea [pweight=fpdm], strata(samp\_stratum)*

*For each `var' in varlist\_emp:*

*Svy: set mean\_`var'=average of sample-weighted `var’ if missing\_any=0*

**Step 4.** Sum the percentage variables created in the previous step and divide by 6 to obtain the average percentage of women achieving adequacy across the six A-WEAI indicators.

*Set mean\_overall=(Sum of all ‘mean\_’ variables)÷6*

**Step 5.** Create a variable that indicates if the woman is under 30 years of age or 30 years of age or older (*age\_cat*). If a woman is missing a self-reported age in the A-WEAI module, use the age information available in the household roster.

*Set age\_cat=missing*

*Replace age\_cat=1 if v6102<30*

*Replace age\_cat=2 if v6102>=30 & v6102<98*

*Replace age\_cat= 1 if youth\_fdm\_dj=1 and age\_cat=missing*

*Replace age\_cat=2 if youth\_fdm\_dj=0 and age\_cat=missing*

*Label values 1 "<30yo" 2 "30+ yo"*

*Label variable “Age category: <30 or 30+”*

**Step 6.** Apply the primary adult female decisionmaker weight and create variables that capture the average percentage of sample-weighted disempowered women under 30 years of age who achieve each adequacy in each of the six empowerment indicators.

*Svyset hhea [pweight=fpdm], strata(samp\_stratum)*

*For each `var' in varlist\_emp:*

*Svy: set mean\_`var'=average of sample-weighted `var’ if missing\_any=0 and age\_cat=1*

**Step 7.** Calculate the average percentage of women 18-29 years of age achieving adequacy across the six A-WEAI indicators.

*Set mean\_overall\_lt30=(Sum of all ‘mean\_’ variables)÷6 if age\_cat=1*

**Step 8.** Apply the primary adult female decisionmaker weight and create variables that capture the average percentage of sample-weighted disempowered women 30 years of age or older who achieve each adequacy in each of the six empowerment indicators.

*Svyset hhea [pweight=fpdm], strata(samp\_stratum)*

*For each `var' in varlist\_emp:*

*Svy: set mean\_`var'=average of sample-weighted `var’ if missing\_any=0 and age\_cat=2*

**Step 9.** Calculate the average percentage of women 30 years of age or older achieving adequacy across the six A-WEAI indicators.

*Set mean\_overall\_gte30=(Sum of all ‘mean\_’ variables)÷6 if age\_cat=2*

## References

Alkire, S., Meinzen-Dick, R., Peterman, A., Quisumbing, A., Seymour, G., and 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>

Hillesland, M. 2016. *Causal mapping of the gender integration framework*. Available at: <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., and 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., and Meinzen-Dick, R. 2017. *The abbreviated women’s empowerment in agriculture index (A-WEAI). IFPRI Discussion Paper 01647.* Available at: <http://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/131231/filename/131442.pdf>

# Agricultural productivity indicators

This chapter describes the following two Feed the Future ZOI-level agricultural sector indicators: (1) percent of producers in the targeted area who have applied targeted improved management practices or technologies, and (2) yield of targeted agricultural commodities. The indicators are calculated for selected prioritized agricultural VCCs identified by in-country Feed the Future teams. Questions about these prioritized VCCs are included in the agricultural module of the country-customized ZOI Survey questionnaire.

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. 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 two indicators for specific VCCs. The improved management practices and technologies indicator step-by-step procedures include maize, fishponds, and dairy cows. The yield indicator includes a generic crop referred to as “VCC1,” fishponds, and dairy cows.

## 13.1 Guidelines to construct the indicators

This section provides the guidelines to construct the two agricultural productivity indicators.

### 13.1.1 Percent of producers in the targeted area who have applied targeted improved management practices or technologies

This indicator measures the percentage of producers, such as farmers, ranchers, pastoralists, and other primary sector producers of food and nonfood 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 15 management practice and technology type categories.[[80]](#footnote-79) These are listed below 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 alflatoxin 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 adaptation and climate risk management that includes technologies promoted 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
* Other, such as improved mechanical and physical land preparation; nonmarket- 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 promoted by the Feed the Future program**. This means that the indicator can only be calculated if the Feed the Future team has defined the set of practices and technologies it will be promoting in the ZOI 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 yet defined the set of promoted 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 that apply any promoted improved practice or technology in the ZOI and tracking the percentage of producers that apply promoted improved management practices or technologies in specific categories.

This indicator is designed to capture the application of promoted 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 promoted improved management practices or technologies on a demonstration plot or other common plot.

When calculating the indicator:

* Count each producer in the sample that applied a promoted improved practice or technology only once in the overall indicator and the applicable sex and age disaggregate categories, regardless of the number of promoted improved practices or technologies applied, to track the percentage of producers who applied Feed the Future-promoted practices or technologies. If more than one producer in a household applied promoted 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 she or he applied one or more Feed the Future-promoted practice or technology types. For example, if a producer used promoted 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 she or he applied one or more Feed the Future-promoted practice or technology types. Producers can be counted under more than one practice and technology type category. For example:
* If a producer applied more than one promoted improved management practice or technology in the year preceding the survey that are categorized under different categories, count the producer under each relevant category.
* If Feed the Future is promoting a practice or technology for multiple benefits, the producers applying the technology should be reported **under each category** **for which the practice or technology is being promoted.** 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 promoted for both purposes. If the practice is being promoted for only one purpose, the producer should be reported under only the relevant category. Conversely, if Feed the Future is promoting a practice or technology for a single benefit even though it could be promoted 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 promoted.**
* Count a producer once, regardless of how many times she or he applied a promoted improved management practice or technology type 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-promoted 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 promoted 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.

### 13.1.2 Yield of targeted agricultural commodities

Yield is the measure of the total output of production of an agricultural commodity—that is, crop, fish, milk, eggs, and live animal offtake, divided by the total number of units in production—that is, hectares planted with crops, area in hectares for pond aquaculture, cubic meters of cage for cage aquaculture, or number of animals in the herd or flock in the year preceding the survey for live animals. Total output per hectare, per animal, and per cubic meter of cage per producer are measures of productivity from farms, fisheries, or livestock, which are then averaged across the ZOI.

Producer-level yield=Total production÷Unit of production

ZOI-level yield = Sum of producer-level yields÷Number of producers of the commodity

The units and recall period used to report on total production (TP) and unit of production (UP) vary depending on the commodity type. All TP and UP data should be converted to the same units and timeframe.

The following are the units used for TP by commodity type.

* Crops: weight in metric tons harvested in the season preceding the survey
* Pond aquaculture: weight in kilograms of fish harvested in the past 30 days
* Cage aquaculture: weight in kilograms of fish harvested in the past 30 days
* Dairy: volume in liters of milk collected in the previous day
* Eggs: number of eggs collected in the previous day
* Livestock: weight in kilograms of entire animals that were offtake in the previous year, measured as number of animals that were taken from the herd for sale, home consumption, or loans or gifts

TP is the amount that was produced, regardless of how it was ultimately used and whether there was any post-harvest loss—that is, post-harvest loss should not be subtracted from TP.

The following are the units used for UP by commodity type.

* Crops: area in hectares planted in the season preceding the survey
* Tree crop: planted area in hectares during the year preceding the survey is recommended, but number of trees can be used
* Pond aquaculture: total pond area in hectares at the time of the survey
* Cage aquaculture: total volume of cages in cubic meters at the time of the survey
* Dairy: number of producing cows during the day preceding the survey
* Eggs: number of producing hens during the day preceding the survey
* Livestock: total number of animal in herd or flock during the year preceding the survey, measured as current stock of animals plus animals that were taken from the herd for sale, home consumption, or loans or gifts, plus animals that died during the year preceding the survey

The yield indicator for crop value chains is disaggregated by farm size—smallholder and non-smallholder. Although country-specific definitions vary, the Feed the Future definition of a smallholder crop producer is one who holds five hectares or less of arable land.[[81]](#footnote-80) Land holding is measured at the household level for this purpose, and the household does not need to formerly own the land, as long as it has use rights over it.

The yield indicator for livestock value chains is disaggregated by livestock production system. Each livestock-producing household should be categorized into the system that best fits its livestock production system characteristics. There are four production systems: rangeland, rural mixed crop-livestock, urban/peri-urban, and intensive/commercial livestock production. The definition of each production system is presented below for informational purposes. The step-by-step section explains how to categorize each livestock-producing household into one of these four systems based on its responses to specific ZOI Survey questions concerning the number of animals in the herd, how mobile the herd is, how animals are housed and fed, and whether the household holds land.

**Rangelands—**pastoral, transhumant, agro-pastoral, sylvo-pastoral, and extensive grasslands

* Livestock and livestock-crop systems in which production is extensive with low stocking rates (typically<10 tropical livestock units per hectare) and there is a degree of herd mobilityin the grazing system beyond the farm for at least part of the production cycle
* Typically in arid and semi-arid zones, with rainfall-dependent (forage) growing seasons less than 180 days per year

**Rural mixed crop-livestock—**ruminants, pigs and poultry and small stock such as rabbits and guinea pigs, and animals kept principally for traction, including oxen, buffalo, and equids

* Integrated crop and livestock production in which crop and livestock systems rely on one another for inputs and exist in a fixed rural location, typically a small holding or farmstead; for example, a system in which at least some of the livestock feed comes from crop residues and by-products produced on-farm

**Urban/peri-urban—**poultry, small-scale dairy, small and large ruminants, pigs, micro-stock, small-scale fattening operations

* Livestock are kept in close proximity to human population centers. Land holdings are small or include confined, caged, and landless production systems.
* Small to medium scale, variable levels of intensification from a single animal to a mid-sized enterprise, such as a small peri-urban cow dairy or small-scale fattening operator.
* Production may target home consumption, local markets, or both.

**Intensive/commercial livestock production—**large pig and poultry production units; also includes ruminant fattening, large dairying, and large-scale dry lots

* Operate at considerable scale and highly commercialized, with significant financial investments and technical inputs in specialized housing, feeding, and animal health and marketing approaches.
* Animals are typically housed and fed formulated, nutritionally balanced rations. Scale of operation, level of technical inputs, and capital investment distinguishes from the urban/peri-urban category.

## Step-by-step procedure to calculate agricultural productivity indicators

This section describes the detailed step-by-step procedures to calculate each agricultural productivity indicator.

### 13.2.1 Percent of producers in the targeted area who have applied targeted improved management practices or technologies

Because it is essential that the *Percent of producers in the targeted area who have applied targeted improved management practices or technologies* indicator only counts the application of improved management practices and technologies that are being promoted 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, and dairy cows as the priority value chains using **illustrative** sets of promoted improved management practices and technologies in Tables 13 (maize), 14 (fishpond aquaculture), and Table 15 (dairy cows). This example **must** be adapted to each ZOI Survey’s specific context; it should never be applied as is unless the Feed the Future program is supporting the same value chains and promoting the same set of improved management practices and technologies.

If a Feed the Future country has not yet identified the set of improved management practices and technologies it will be promoting, tables detailing the percentage of producers applying practices and technologies included in the ZOI Survey should be presented, but the Feed the Future indicator should not be calculated and included in the survey report.

#### Definitions

| Numerator | Number of producers who produced selected agricultural products in the past year/season/month/day and applied targeted improved management practices or technologies |
| --- | --- |
| Denominator | Number of producers who produced selected agricultural products in the past year/season/month/day |
| Unit of measure | Percentage |
| Level of data | Individual (producer) |
| Sampling weight | Producer of any selected agricultural product, producer of each selected agricultural product[[82]](#footnote-81) |
| Disaggregation levels | Sex\*  Age (15-29, 30+)\*  Improved management practice or technology\*  Commodity\* |
| Coverage | P2-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 | *hhea; hhnum; m1\_line; samp\_stratum; all variables in maize, fishpond, and dairy cow modules\*\** |
| Analytic variables used\*\* | *Sex, age, vcc\_youth, hhmem\_dj, wgt\_v1, wgt\_fish, vcc\_maize, vcc\_fish, vcc\_dairy* |
| Analytic variables created\*\* | *imp\_genetics, imp\_culture, imp\_ecosys, imp\_pest, imp\_fert, imp\_irrig, imp\_water, imp\_adapt, imp\_market, imp\_harvest, imp\_valadd, imp\_oth, totimp\_maize, anyimp\_maize, imp\_pondprep, imp\_fingerlings, imp\_fishfeed, imp\_fishhealth, imp\_pondculture, imp\_fishcarry, imp\_fishharvest, totimp\_fish, anyimp\_fish, imp\_lsbreeds, imp\_lshealth, imp\_lshousing, imp\_lsfeed, imp\_lsgrazing, imp\_lswaste, imp\_lsfodder, imp\_lsdualcrop,* *totimp\_cow, anyimp\_cow, imptech\_tot, imptech\_any* |

\* Standard Feed the Future disaggregate

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

#### Maize calculations

In this example, the steps to calculate the percentage of maize 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\_maize.do.* Variables from Module 7.1 are used for the calculations.

**Step 1.** Review the maize section of the agriculture module in the ZOI Survey questionnaire and identify questions that relate to improved management practices and technologies being promoted in the ZOI. Determine which response options would be considered targeted improved management practices and technologies, and also determine under which category or categories the targeted improved management practices and technologies fall.

Table 13 shows the relevant variables and corresponding response options used to categorize promoted improved management practices or technologies for maize in the step-by-step procedures and Stata syntax files. Note that these are only examples and **MUST** be adapted for the country context.

If Feed the Future is promoting a management practice or technology for multiple benefits, be sure that producers applying the practice or technology are reported under **each category** for which the practice or technology is being promoted. Conversely, if Feed the Future is promoting a practice or technology for a single benefit even though it could be promoted 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 promoted.

**Table 13: ZOI Survey Variables and Response Options to Identify Illustrative Improved Management Practices and Technologies by Management Practice and Technology Type Category for Maize**

| **Management Practice and Technology Type Category (Maize)** | **Feed the Future-promoted Practices and Technologies for Maize (Illustrative)** |
| --- | --- |
| Crop genetics | v7107=B (Improved open pollinated) or v7107=C (Hybrid) or v7107a=2 (Mostly modern/improved) or v7107a=3 (About half traditional/half improved) |
| Cultural practices | v7109=1 (Plant in rows) or v7109=3 (Some in rows and some randomly broadcast) |
| Natural resource or ecosystem management | v7111b=A (Soil based organic) or v7111b=C (Foliar feeds organic) or v7119=C (Weed control mulching) or v7121=A (Terracing) or v7121=C (Soil bands/trenches) |
| Pest and disease management | v7115=1 (Chemical pest control) or v7119=B (Herbicide) or v7119=C (Mulching) |
| Soil-related fertility and conservation | v7121=A (Terracing) or v7121=B (Mulching) or v7121=C (Soil/bands/trenches) |
| Irrigation | v7123=C (Permanent hose/drip irrigation) or v7123=D (Pumps) |
| Agriculture water management (non-irrigation) | v7121=A (Terracing) or v7121=B (Mulching) or v7121=C (Soil/bands/trenches) |
| Climate adaptation | v7107=B (Improved open pollinated) or v7107=C (Hybrid) or v7123=C (Permanent hose/drip irrigation) or v7123=D (Pumps) |
| Marketing and distribution | v7106=4 (Bought from ag dealer with cash) or v7106=5 (Bought from ag dealer with voucher) or v7124a=G (Stalks harvested and sold to others) or v7124c=E (Husks sold/traded with others as animal feed) |
| Post-harvest handling and storage | v7126=H (Solar dryers) or v7126=I (Mechanized dryers) or v7129=C (Hermetic bag) or v7130= E (Stored in warehouses) |
| Value-added processing | v7127=C (Shelling by machine) |
| Other | v7105=C (Land prep with motorized tiller) or v7105=D (Tractor) or 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 improved management practices or technologies indicator for maize. Also review the *vcc\_maize* variable that should have already be created and included in the individual-level data file.

*Load “$analytic\FTF ZOI Survey [Country] [Year] persons data analytic”*

*Keep if vcc\_maize=1*

*Keep hhea, hhnum, m1\_line, samp\_stratum, sex, age, all variables that contain with ‘vcc’, and all variables that begin with ‘v71’*

**Step 3.** Create 12 binary variables for each targeted improved practice and technology type category to flag producers who applied targeted improved practices or technologies to cultivate maize (applied=1, not applied=0). These steps identify whether producers applied an improved management practice or technology to cultivate maize in the year preceding the survey by practice or technology type category.

**Step 3a.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to crop genetics (*imp\_genetics*).

*Set imp\_genetics=0*

*Replace imp\_genetics=1 if (v7107b1=1 or v7107c1=1) or (v7107aa=2 or v7107aa=3)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved crop genetics”*

**Step 3b.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to cultural practices (*imp\_culture*).

*Set imp\_culture=0*

*Replace imp\_culture=1 if (v7109=1 or v7109=3)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved cultural practices”*

**Step 3c.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to natural resource or ecosystem management (*imp\_ecosys*).

*Set imp\_ecosys=0*

*Replace imp\_ecosys=1 if (v7111ba=1 or v7111bc=1) or v7119c=1 or (v7121a=1 or v7121c=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved natural resources and ecosystem management”*

**Step 3d.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to improved pest and disease management (*imp\_pest*).

*Set imp\_pest=0*

*Replace imp\_pest=1 if v7115=1 or (v7119b=1 or v7119c=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved pest and disease management”*

**Step 3e.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to soil-related fertility and conservation (*imp\_fert*).

*Set imp\_fert=0*

*Replace imp\_fert=1 if (v7121a=1 or v7121b=1 or v7121c=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved soil related fertility and conservation”*

**Step 3f.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to irrigation (*imp\_irrig*).

*Set imp\_irrig=0*

*Replace imp\_irrig=1 if (v7123c1=1 or v7123d1=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved irrigation”*

**Step 3g.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to water management (non-irrigation) (*imp\_water*).

*Set imp\_water=0*

*Replace if imp\_water=1 if (v7121a=1 or v7121b=1 or v7121c=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved water management (non-irrigation)”*

**Step 3h.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to climate adaption (*imp\_adapt*).

*Set imp\_adapt=0*

*Replace imp\_adapt=1 if (v7107b=1 or v7107c=1) or (v7123c1=1 or v7123d1=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved climate adaption”*

**Step 3i.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to marketing and distribution (*imp\_market*).

*Set imp\_market=0*

*Replace imp\_market=1 if (v7106=4 or v7106=5) or v7124ag=1 or v7124c=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved marketing and distribution”*

**Step 3j.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to post-harvest handling and storage (*imp\_harvest*).

*Set imp\_harvest=0*

*Replace imp\_harvest=1 if (v7126h=1 or v7126i=1) or v7129c=1 or v7130e=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved post-harvest handling and storage”*

**Step 3k.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to value-added processing (*imp\_valadd*).

*Set imp\_valadd=0*

*Replace imp\_valadd=1 if v7127c1=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved value added processing”*

**Step 3l.** Create a binary variable to flag maize producers who used a targeted improved practice or technology related to other management practices or technologies (*imp\_oth*).

*Set imp\_oth=0*

*Replace imp\_oth=1 if (v7105c=1 or v7105d=1) or (v7124=2 or v7124=3)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Other improved management practices and technologies”*

**Step 4.** Create a countvariable to capture the total number of targeted improved practices or technology types to cultivate maize (*totimp\_ maize*).

*Set totimp\_maize=imp\_genetics+imp\_culture+imp\_ecosys+imp\_pest+imp\_fert+ imp\_irrig+imp\_water+imp\_adapt+imp\_market+ imp\_harvest+imp\_valadd+imp\_oth*

*Label variable “Total number of improved practice and technology types applied (Maize)”*

**Step 5.** Create a binary variable to indicate whether producers applied any targeted improved management practices or technologies to cultivate maize (*anyimp\_maize*).

*Set anyimp\_maize=0*

*Replace anyimp\_maize=1 if totimp\_ maize>0*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Applied any targeted improved practice or technology (Maize)”*

**Step 6.** After applying the maize producer sampling weight, calculate the percentage of de jure maize producers who applied at least one targeted improved management practice or technology to cultivate maize during the year preceding the survey using the *anyimp\_maize* analytic variable. Repeat using producers’ age (under 30 years of age, 30 years of age or older) and sex, and improved management practice and technology type categories as disaggregates. (Example code uses Stata syntax.)

*Svyset hhea [pw=wgt\_maize], strata(samp\_stratum)*

*Svy, subpop(hhmem\_dj): prop anyimp\_maize*

*Svy, subpop(hhmem\_dj): prop anyimp\_maize, over(vcc\_youth)*

*Svy, subpop(hhmem\_dj): prop anyimp\_maize, over(sex)*

*For each variable that starts with ‘imp\_’:*

*svy, subpop(hhmem\_dj): prop ‘variable'*

**Step 7.** 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, totimp\_maize, anyimp\_maize, and all variables that begin with ‘imp\_’*

*Save “$analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] agtech\_maize”*

#### Fishpond calculations

In this example, the steps to calculate the percentage of fishpond 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\_fishpond.do.* Variables from Module 7.8 are used for the calculations.

**Step 1.** Review the fishpond section of the agriculture module in the ZOI Survey questionnaire and identify questions that relate to improved management practices and technologies being promoted in the ZOI. Determine which response options would be considered targeted improved management practices and technologies, and also determine under which aquaculture sub-category or sub-categories the targeted improved management practices and technologies fall.

Table 14 shows the relevant variables and corresponding response options used to categorize promoted improved management practices or technologies for fishponds in the step-by-step procedures and Stata syntax files. Note that these are only examples and **MUST** be adapted for the country context.

If Feed the Future is promoting a practice or technology for multiple benefits, be sure that producers applying the practice or technology are reported under **each sub-category** for which the practice or technology is being promoted. Conversely, if Feed the Future is promoting a practice or technology for a single benefit even though it could be promoted for multiple benefits, be sure that producers applying the practice or technology are reported under **only the one sub-category** for which the practice or technology is being promoted.

**Table 14: ZOI Survey Variables and Response Options to Identify Improved Aquaculture Management Practices and Technologies by Aquaculture Management Sub-category**

| **Management Practice and Technology Type sub-category (Fishponds)** | **Feed the Future-promoted Practices and Technologies for Fishponds (Illustrative)** |
| --- | --- |
| Pond preparation | v78014>0 (Number of times pond drained) or v78017=1 (Added manure) |
| Improved fingerlings | v78002c=1 (Certified hatchery) or v78009=A (Carp) or v78009=B (Tilapia) or v78009=C (Catfish) |
| Improved feed and feeding practices | v78004=1 (Fish supplement) or v78017=1 (Added manure) |
| Fish health and disease control | v78010ba=1(Salt for disease) or v78011bb=1 (Salt for parasite) or (v78010=1 (Observed disease) and v78010bb=1(Formalin)) or (v78011=1 (Observed parasites) and v78011bc=1(Formalin)) |
| Improved pond culture | v78012=1 (Monitored water quality) or v78013=1 (Maintained water quality) |
| Management of carrying capacity | v78016a=1 (Sex separation) or v78016b=1 (Age separation) |
| Sampling and harvesting | v78018a=2 (Conducted partial harvests) |

NOTE: Improved cage culture is an additional sub-category included in the Feed the Future Guide to Statistics indicator reference sheet for this indicator; however, the core ZOI Survey questionnaire does not include any questions related to this sub-category. The sub-category is therefore not included in the step-by-step procedures or in the example Stata syntax file. Improved management practices and technologies can also be classified into the other category; however, the core ZOI Survey questionnaire does not include any questions related to this category. As applicable, add the improved cage culture and other categories to the analysis procedures.

**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 improved management practices or technologies indicator for fishponds. Also review the *vcc\_fish* variable that should have already be created and included in the individual-level data file.

*Load “$analytic\FTF ZOI Survey [Country] [Year] persons data analytic”*

*Keep record if vcc\_fish=1*

*Keep hhea, hhnum, m1\_line, samp\_stratum, sex, age, all variables that contain with ‘vcc’, and all variables that begin with ‘v78’*

**Step 3.** Create seven binary variables—one for each management practice or technology type sub-category—to flag producers who applied targeted improved management practices and technology to raise fishponds (1=applied, 0=not applied).[[83]](#footnote-82)

This step identifies producers who used improved management practices or technologies in fishpond aquaculture by practice or technology type. Almost all improved aquaculture practices and technologies are categorized under a single practice and technology type: aquaculture management.[[84]](#footnote-83) However, to increase understanding of the range of types of practices and technologies applied, the improved management practices and technologies are first categorized into more detailed practice or technology type sub-categories, which can be used to present the disaggregated data in the report tables.

**Step 3a.** Create a binary variable to flag fishpond producers who used a targeted improved management practice or technology related to pond preparation (*imp\_pondprep*).

*Set imp\_pondrep=0*

*Replace imp\_pondprep=1 if (v78014>0 and v78014≠missing) or v78017=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved pond preparation”*

**Step 3b.** Create a binary variable to flag fishpond producers who used a targeted improved management practice or technology related to fingerlings (*imp\_fingerlings*).

*Set imp\_fingerlings=0*

*Replace imp\_fingerlings=1 if v78002c=1 or (v78009\_a=1 or v78009\_b=1 or v78009\_c=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label imp\_fingerlings “Improved fish fingerlings”*

**Step 3c.** Create a binary variable to flag fishpond producers who used a targeted improved management practice or technology related to fish feed and feeding practices (*imp\_fishfeed*).

*Set imp\_fishfeed=0*

*Replace imp\_fishfeed=1 if v78004=1 or v78017=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved fish feed and feeding practices”*

**Step 3d.** Create a binary variable to flag fishpond producers who used a targeted improved management practice or technology related to fish health and disease control (*imp\_fishhealth*).

*Set imp\_fishhealth=0*

*Replace imp\_fishhealth=1 if ((v78010b\_a=1 or v78011b\_b=1) or ((v78010=1 and v78010b\_b=1) or (v78011=1 and v78011b\_c=1))*

*Label values 0 “Did not apply” 1 “Applied”*

*Label imp\_fishhealth “Improved fish health and disease control”*

**Step 3e.** Create a binary variable to flag fishpond producers who used a targeted improved management practice or technology related to pond culture (*imp\_pondculture*).

*Set imp\_pondculture=0*

*Replace imp\_pondculture=1 if v78012 =1 or v78013=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved pond culture”*

**Step 3f.** Create a binary variable to flag fishpond producers who used a targeted improved management practice or technology related to management of carrying capacity (*imp\_fishcarry*).

*Set imp\_fishcarry=0*

*Replace imp\_fishcarry if (v78016\_a=1or v78016\_b=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved management of carrying capacity”*

**Step 3g.** Create a binary variable to flag fishpond producers who used a targeted improved management practice or technology related to fish sampling and harvesting (*imp\_fishharvest*).

*Set imp\_fishharvest=0*

*Replace if imp\_fishharvest=1 if v78018a=2*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved fish sampling and harvesting”*

**Step 4.** Create a count variableto capture the total number of targeted improved practices or technology type sub-categories that producers used to raise fish in fishponds (*totimp\_fish*).

*Set totimp\_fish=imp\_pondprep+imp\_fingerlings+imp\_fishfeed+imp\_fishhealth+ imp\_pondculture+imp\_fishcarry+imp\_fishharvest*

*Label variable “Total number of improved practice and technology sub-types applied”*

**Step 5.** Create a binary variable to indicate whether the producer applied any targeted improved practices or technologies to raise fishponds (*anyimp\_fish*). This variable should align with the aquaculture management indicator category.

*Set anyimp\_fish=0*

*Replace anyimp\_fish=1 if totimp\_fish>0 and totimp\_fish≠missing*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Applied any targeted improved practice or technology (Fishpond)”*

**Step 6.** After applying the fishpond producer sampling weight, calculate the percentage of de jure fishpond producers who applied at least one targeted improved management practice or technology to raise fish in ponds during the year preceding the survey using the *anyimp\_fish* analytic variable. Repeat using producers’ age (under 30 years of age, 30 years of age or older) and sex as disaggregates. (Example code uses Stata syntax.)

*Svyset hhea [pw=wgt\_fish], strata(samp\_stratum)*

*Svy, subpop(hhmem\_dj): prop anyimp\_fish*

*Svy, subpop(hhmem\_dj): prop anyimp\_fish, over(vcc\_youth)*

*Svy, subpop(hhmem\_dj): prop anyimp\_fish, over(sex)*

*For each variable that starts with ‘imp\_’ :*

*svy, subpop(hhmem\_dj): prop ‘variable'*

**Step 7.** 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, totimp\_fish, anyimp\_fish, and all variables that begin with ‘imp\_’*

*Save “$analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] agtech\_fish”*

#### Dairy cow calculations

In this example, the steps to calculate the percentage of dairy cow 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\_dairycow.do.* Variables from Module 7.50 are used for the calculations.

**Step 1.** Review the dairy cow section of the agriculture module in the ZOI Survey questionnaire and identify questions that relate to improved management practices and technologies being promoted in the ZOI. Determine which response options would be considered targeted improved management practices and technologies, and also determine what livestock sub-category or sub-categories the targeted improved management practices and technologies fall under.

Table 15 shows the relevant variables and corresponding response options used to categorize promoted improved management practices or technologies for dairy cows in the step-by-step procedures and Stata syntax files. Note that these are only examples and **MUST** be adapted for the country context.

If Feed the Future is promoting an improved management practice or technology for multiple benefits, be sure that producers applying the practice or technology are reported under **each sub-category** for which the practice or technology is being promoted. Conversely, if Feed the Future is promoting a practice or technology for a single benefit even though it could be promoted for multiple benefits, be sure that producers applying the practice or technology are reported under **only the one sub-category** for which the practice or technology is being promoted.

**Table 15: ZOI Survey Variables and Response Options to Identify Improved Dairy Cow Management Practices and Technologies by Livestock Management Sub-category**

| **Management Practice or Technology Type Sub-category (Livestock)** | | **Feed the Future-promoted Practices and Technologies for Dairy Cows (Illustrative)** |
| --- | --- | --- |
| 1 | Improved livestock breeds | v75008=B (Artificial insemination) or v75011=3 (Bull has good body size, composition) or v75011=4 (Bull is son of high-producing cow) or v75011=5 (Bull is known to have good fertility) |
| 2 | Livestock health services and products such as vaccines | v75028=1 (Yes, obtained health services from trained provider) or v75030=1 (Yes, gave medicines ) or v75033=2 (Some cattle vaccinated) or v75033=3 (All cattle vaccinated) or v75034b=A (Wash udder to prevent mastitis) or v75034b=B (Teat dip to prevent mastitis) or v75034b=C (Somatic cell counts to prevent mastitis) |
| 3 | Improved livestock handling practices and housing | v75016=4 (Roof only, no sides) or v75016=5 (Roof and sides, dirt floor) or v75016=6 (Roof and sides, concrete floor) |
| 4 | Improved feeding practices | v75021a=1 (Fed crop by-products daily) or v75021a=2 (Fed crop by-products weekly) or v75024a=1 (Fed mixed concentrates daily) or v75024a=2 (Fed mixed concentrates weekly) or v75026a=1 (Fed vitamins or minerals daily) or v75026a=2 (Fed vitamins or minerals weekly) |
| 5 | Improved grazing practicesa |  |
| 6 | Improved waste management practices | v75027b=2 (Put in heap in covered area) or v75027b=3 (Put in pit/lagoon) or v75027b=4 (Put into a tank) or v75027b=5 (Put into a biogas-producing digester) |
| 7 | Improved fodder cropb | v75019=B (Conserved rice straw) or v75019=C (Conserved maize stover) or v75019=D (Legume haulms or stovers) or v75019=E (Forage legumes) or v75019=F (Napier grass) or v75019=G (Guinea grass) or v75019=H (Cut fresh grass) or v75019=I (Tree fodder) |
| 8 | Cultivation and dual purpose cropsc |  |
| 9 | Other | v75045=1 (Yes, kept written records) |

a Improved grazing practices questions are not included in the dairy cow module of the core questionnaire.

b Also manually check response option X (Other) to determine if any manually entered forages should be considered promoted improved fodder crops and included in the calculations.

c Cultivation of dual purpose crop questions are not included in the dairy cow module of the core questionnaire. Information on how crop waste is used can be found in the crop modules of the core questionnaire.

**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 improved management practices or technologies indicator for dairy cows. Also review the *vcc\_dairy* variable that should have already be created and included in the individual-level data file.

*Load “$analytic\FTF ZOI Survey [Country] [Year] persons data analytic”*

*Keep record if vcc\_dairy=1*

*Keep hhea, hhnum, m1\_line, samp\_stratum, sex, age, all variables that contain with ‘vcc’, and all variables that begin with ‘v750’*

**Step 3.** Create eight binary variables—one for each practice or technology type sub-category—to flag producers who applied targeted improved management practices or technologies to raise dairy cows (1=applied, 0=not applied).[[85]](#footnote-84)

This step identifies producers who used improved management practices or technologies to raise dairy cows by practice and technology types. Almost all improved management practices and technologies for raising dairy cows are categorized under a single practice and technology type: livestock management.[[86]](#footnote-85) However, to increase understanding of the range of types of practices and technologies applied, the improved management practices and technologies are first categorized into more detailed practice or technology type sub-categories, which can be used to present the more disaggregated data in the report tables.

**Step 3a.** Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to breeds (*imp\_lsbreeds*).

*Set imp\_lsbreeds =0*

*Replace imp\_lsbreeds =1 if (v75008b=1) or (v75011=3 or 4 or 5)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved breeds”*

**Step 3b.** Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to health services and products (*imp\_lshealth*).

*Set imp\_lshealth =0*

*Replace imp\_lshealth =1 if (v75028=1) or (v75030=1) or (v75033=1) or (v75034ba=1 or v75034bb=1 or v75034bc=1)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved health services and products”*

**Step 3c.** Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to improved dairy cow handling practices and housing (*imp\_lshousing*).

*Set imp\_lshousing=0*

*Replace imp\_lshousing=1 if v75016=4 or v75016=5 or v75016=6*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved housing”*

**Step 3d.** Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to feeding practices (*imp\_lsfeed*).

*Set imp\_lsfeed=0*

*Replace imp\_lsfeed=(v75021a=1 or v75021a=2) or (v75024a=1 or v75024a=2) or (v75026a=1 or v75026a=2)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved feeding”*

**Step 3e.** [Not relevant for core questionnaire. Placeholder for country-customized questionnaire, if applicable.] Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to grazing practices (*imp\_lsgrazing*).

*Set imp\_lsgrazing=missing*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved grazing practices”*

**Step 3f.** Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to waste management practices (*imp\_lswaste*).

*Set imp\_lswaste=0*

*Replace imp\_lswaste=1 if (v75027b=2 or v75027b=3 or v75027b=4 or v75027b=5)*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved waste management”*

**Step 3g.** Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to fodder crops (*imp\_lsfodder*).

*Set imp\_lsfodder=0*

*Replace if imp\_lsfodder=1 if v75019b=1 or v75019c=1 or v75019d=1 or v75019e=1 or v75019f=1 or v75019g=1 or v75019h=1 or v75019i=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved fodder crops”*

**Step 3h.** [Not relevant for core questionnaire. Placeholder for country-customized questionnaire, if applicable.] Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology related to cultivation and dual purpose crops (*imp\_lsdualcrop*).

*Set imp\_lsdualcrop=missing*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Improved cultivation and dual purpose crops”*

**Step 3i.** Create a binary variable to flag dairy cow producers who used a targeted improved management practice or technology not capture in any other category (*imp\_dc\_other*).

*Set imp\_dc\_other=0*

*Replace imp\_dc\_other=1 if v75045=1*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Other improved practice or technology”*

**Step 4.** Create a count variable to capture the total number of targeted improved management practices or technology sub-categories to raise dairy cows (*totimp\_cow*).

*Set totimp\_cow=imp\_lsbreeds+imp\_lshealth+imp\_lshousing+imp\_lsfeed+imp\_lsgrazing+imp\_lswaste+ imp\_lsfodder+imp\_lsdualcrop+imp\_dc\_other*

*Label variable “Total number of improved management practice and technology sub-types applied”*

**Step 5.** Create a binary variable to indicate whether the dairy cow producer applied any targeted improved management practices or technologies (*anyimp\_cow*).

*Set anyimp\_cow=0*

*Replace anyimp\_cow=1 if totimp\_cow>0 and totimp\_cow≠missing*

*Label values 0 “Did not apply” 1 “Applied”*

*Label variable “Applied any targeted improved management practice or technology (Dairy cows)”*

**Step 6.** After applying the dairy cow producer sampling weight, calculate the percentage of de jure dairy cow producers who applied at least one targeted improved management practice or technology to raise dairy cows during the year preceding the survey using the *anyimp\_cow* analytic variable. Repeat using producers’ age (under 30 years of age, 30 years of age or older) and sex as disaggregates. (Example code uses Stata syntax.)

*Svyset hhea [pw=wgt\_cow], strata(samp\_stratum)*

*Svy, subpop(hhmem\_dj): prop anyimp\_cow*

*Svy, subpop(hhmem\_dj): prop anyimp\_cow, over(vcc\_youth)*

*Svy, subpop(hhmem\_dj): prop anyimp\_cow, over(sex)*

**Step 7.** 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, totimp\_cow, anyimp\_cow, and all variables that begin with ‘imp\_’*

*Save “$analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] agtech\_dairycow”*

#### Overall indicator calculations[[87]](#footnote-86)

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.*

**Step 1.** Append all crop improved management practice and technology data files if there is more than one crop VCC included in the ZOI Survey, and save the data file. If there is only one crop VCC, skip this step. Maize and millet are used here as an example.

*Load “$analytic\Results\FTF ZOI Survey [Country] [Year] agtech\_maize”*

*Append using “$analytic\Results\FTF ZOI Survey [Country] [Year] agtech\_millet”*

**Step 2.** Create one record per producer if there is more than one crop VCC included in the survey and save the data file. If there is only one crop VCC, skip this step.

*By hhea hhnum m1\_line: Sum all variables that begin with ‘imp\_’*

*Save “$analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] agtech\_all”*

**Step 3.** Append all livestock improved management practice and technology data files if there is more than one livestock VCC included in the ZOI Survey. If there is only one livestock VCC, skip this step. Livestock VCCs include goats, sheep, and dairy cows. (There are no livestock VCCs in the example, so this step is a placeholder in the core Stata template syntax file using goats and dairy cows.)

*Load “$analytic\Results\FTF ZOI Survey [Country] [Year] agtech\_goats”*

*Append using “$analytic\Results\FTF ZOI Survey [Country] [Year] agtech\_dairycow”*

**Step 4.** Create one record per producer if there is more than one livestock VCC included in the survey. If there is only one livestock VCC, skip this step. (There is only one livestock VCC in the example, so this step is a placeholder in the core Stata template syntax file.)

*By hhea hhnum m1\_line: sum all variables that begin with ‘imp\_’*

*Save “$analytic\Temp\FTF ZOI Survey [COUNTRY] [YEAR] livestock\_all”*

**Step 5.** Add the livestock data to the crop data, keeping only the overall improved management practice and technology variable for livestock using cluster, household number, and roster line number as key matching variables.

*Load “$analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] agtech\_all”,clear*

*Add imp\_any\_lstock from” $analytic\Temp\FTF ZOI Survey [COUNTRY] [YEAR] livestock\_all”, using hhea hhnum and m1\_line as key matching variables*

*Save “$analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] agtech\_all”*

**Step 6.** Add the fishpond data to the crop data, keeping only the overall improved management practices and technology variable for fishponds using cluster, household number, and roster line number as key matching variables. Skip this step if fishponds are not a VCC included in the survey.

*Add imp\_any\_fish from “$analytic\Results\FTF ZOI Survey [Country] [Year] agtech\_fishpond”, using hhea hhum and m1\_line as key matching variables*

**Step 7.** Create a new binary variable for each improved management practice and technology category to indicate whether the producer used any improved management practices or technologies in that category.

*For each (var) that begins with ‘imp\_’:*

*Set ‘var’\_any=0*

*Replace `var’=1 if `var’>0 and `var’≠missing*

**Step 8.** Create a variable list (*IMP)* that includes all applicable improved management practice and technology category variables.

*Set variable list IMP=imp\_genetics\_any, imp\_culture\_any, imp\_ecosys\_any, imp\_pest\_any, imp\_fert\_any, imp\_irrig\_any, imp\_water\_any, imp\_adapt\_any, imp\_market\_any, imp\_harvest\_any, imp\_valadd\_any, imp\_pest\_any, imp\_oth\_any, imp\_any\_livestock, imp\_any\_fish*

**Step 9.** Create a variable to count the total number of categories from which producers used improved management practices or technologies.

*Set imptech\_tot=0*

*Replace imptech\_tot=sum of all variables in IMP variable list*

*Label variable “Total number of improved management practices and technology categories practiced”*

**Step 10.** Create a variable to indicate whether producers used any improved management practices or technologies.

*Set imptech\_any=0*

*Replace imptech\_any=1 if imptech\_tot≥1 and imptech\_tot≠missing*

*Label values 0 “No” 1 “Yes”*

*Label variable “Used any improved management practice/technology”*

**Step 11.** Add the cluster, household number, age, sex, producer weight, and sample stratum (if applicable) variables and 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.

*Add hhea hhnum age sex wgt\_f samp\_stratum and all variables that include ‘vcc’ in the name from “$analytic\FTF ZOI Survey [Country] [Year] persons analytic data” using hhea hhnum and m1\_line as key matching variables*

*Save “$analytic\Results\FTF ZOI Survey [COUNTRY] [YEAR] agtech\_all”*

**Step 12.** After applying the producer sampling weight to adjust for the survey design, calculate the indicator for producers who are de jure household members using the *imptech\_any* analytic variable. Repeat the calculation for the sex, age (under 30 years of age, 30 years of age or older), and improved management practices and technology category disaggregates. Note that the commodity disaggregate was already calculated in the individual VCC syntax files.

*Svyset hhea [pw=wgt\_f], strata(samp\_stratum)*

*Svy, subpop(hhmem\_dj): prop imptech\_any*

*Svy, subpop(hhmem\_dj): prop imptech\_any, over(sex)*

*Svy, subpop(hhmem\_dj): prop imptech\_any, over(youth\_vcc)*

*For each variable (x) in the variable list IMP:*

*Svy, subpop(hhmem\_dj): prop `x'*

#### References

United Nations Food and Agriculture Organization. 2019. *Animal production.* Available at: [*http://www.fao.org/animal-production/en/*](http://www.fao.org/animal-production/en/)

United Nations Food and Agriculture Organization. 2019. *Aquaculture technology.* Available at: [*http://www.fao.org/fishery/technology/aquaculture/en*](http://www.fao.org/fishery/technology/aquaculture/en)

United Nations Food and Agriculture Organization. 2018. *Harnessing the power of innovation in agriculture.* Available at: [*http://www.fao.org/publications/highlights-detail/en/c/1171251/*](http://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 10th Global Forum for Food and Agriculture. Available at: [*http://www.fao.org/3/i8384en/I8384EN.pdf*](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*](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)

### 13.2.2 Yield of targeted agricultural commodities

This section describes the procedures to calculate the yield indicator, using three commodity types to illustrate the procedure: maize for crops, fishpond for aquaculture, and milk cows for animals.

#### Crop yield

The yield indicator for crops estimates the average crop production in metric tons per hectare. Yield data are collected separately for each producer primarily responsible for cultivating the crop in a household. For simplicity, this guide explains the procedure to calculate the yield of VCC1. The procedures for calculating the yield of any crop in the ZOI Survey are as same as for VCC1. The step-by-step procedures follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax yield crop.do*. Variables from Modules 7.91 and 7.92 are used for the calculations.

##### Definitions

| Numerator | Yield of VCC1 in metric tons per hectare, for the season preceding the survey |
| --- | --- |
| Denominator | Number of producers who cultivated VCC1 in the season preceding the survey |
| Unit of measure | Mean yield in metric tons per hectare |
| Level of data | Individual (producer) |
| Sampling weight | Maize producer |
| Disaggregation levels | First level:  Farm size (smallholder, non-smallholder)\*  Second level:  Sex (female, male)\*  Age group (15-29, 30+)\* |
| Coverage | P2-ZOI |
| Treatment of missing data | If information on the plot area (*v79107),* percentage of land used for cultivation (*v79108*), yield amount (*v79201n*), or unit (*v79201u)* is missing for a VCC1 producer, exclude that producer from the numerator and denominator. |
| Survey variables used | *hhea, hhnum, m1\_line, samp\_stratum, wgt\_v1\*\** all variables in the land map module (7.90), the plot area module (7.91), and the crop yield module (7.92) |
| Analytic variables used | *farmsize, sex, vcc\_youth* |
| Analytic variables created\*\* | *plot1\_vcc1-plotn\_vcc1, area1\_vcc1-arean\_vcc1, area\_vcc1, prod\_vcc1, yield\_vcc1, mean\_yield\_vcc1, sd\_yield\_vcc1, yield\_vcc12, sholder\_vcc1\_dj, nsholder\_vcc1\_dj* |

\*Standard Feed the Future disaggregate

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

##### Calculations

**Step 1.** Prepare the data—that is, load the individual-level analytic data file and keep only variables required to calculate the indicator and its disaggregates, including the sampling weight for VCC1 producers. Keep only records for producers who cultivated VCC1 and completed both the plot area and crop yield modules.

*Load “$analytic\FTF ZOI Survey [Country] [Year] persons data analytic”*

*Keep hhea, hhnum, m1\_line, sex, age, wgt\_v1, and all variables that begin with ‘v79’ or that contain ‘vcc’*

*Keep if vcc\_vcc1=1*

*Keep if v79100x=1 and v79200x=1*

**Step 2.** Create a plot area variable for each plot in which VCC1 is grown (*plot1\_vcc1-plotn\_vcc1,* where ‘*n*’ is the maximum number of plots that a producer has in the data file), setting out-of-range values to missing. Four plots are used for illustration. (VCC1 is designated as ‘A’ in the variables *v79000f1-v79000fn*.)

*Set plot1\_vcc1=missing*

*Replace plot1\_vcc1=v79107\_1 if v79107\_1<9999.95 and v79000f1 contains ‘A’*

*Label variable “Area of plot 1 (hectares)”*

*Set plot2\_vcc1=missing*

*Replace plot2\_vcc1=v79107\_2 if v79107\_2<9999.95 and v79000f2 contains ‘A’*

*Label variable “Area of plot 2 (hectares)”*

*Set plot1\_vcc1=missing*

*Replace plot3\_vcc1=v79107\_3 if v79107\_3<9999.95 and v79000f3 contains ‘A’*

*Label variable “Area of plot 3 (hectares)”*

*Set plot4\_vcc1=missing*

*Replace plot4\_vcc1=v79107\_4 if v79107\_4<9999.95 and v79000f4 contains ‘A’*

*Label variable “Area of plot 4 (hectares)”*

**Step 3.** Calculate the area in hectares where the producer cultivated VCC1 by multiplying the area of each plot by the percentage of the plot where VCC1 was cultivated and sum all areas to calculate total area where VCC1 was cultivated *(area\_vcc1)*.

*Set area1\_vcc1=(plot1\_vcc1\*v79108a\_1÷100) if v79000f1 contains ‘A’*

*Label variable “Area of plot 1 used for [VCC1] (hectares)”*

*Set area2\_vcc1=(plot2\_vcc1\*v79108a\_2÷100) if v79000f2 contains ‘A’*

*Label variable “Area of plot 2 used for [VCC1] (hectares)”*

*Set area3\_vcc1=(plot3\_vcc1\*v79108a\_3÷100) if v79000f3 contains ‘A’*

*Label variable “Area of plot 3 used for [VCC1] (hectares)”*

*Set area4\_vcc1=(plot4\_vcc1\*v79108a\_4÷100) if v79000f4 contains ‘A’*

*Label variable “Area of plot 4 used for [VCC1] (hectares)”*

*Set area\_vcc1=(area1\_vcc1+area2\_vcc1+area3\_vcc1+area4\_vcc1)*

*Replace area\_vcc1=missing if area1\_vcc1=missing and area2\_vcc1= missing and area3\_vcc1=missing and area4\_vcc1=missing*

*Label variable “Area cultivated, all plots: [VCC1] (hectares)”*

**Step 4.** Calculate production of VCC1 in metric tons (*prod\_vcc1*). If the producer reported the amount harvested in a unit other than metric tons, convert the amount to metric tons. If there are other units included in the ZOI Survey questionnaire, add syntax to convert any responses in those units to metric tons.

*Set prod\_vcc1=(v79201\_1n\*1) if v79201\_1u=3 [ton]*

*Replace prod\_vcc1=(v79201\_1n\*0.01) if v79201\_1u=2 [convert kg to ton]*

*Replace prod\_vcc1=(v79201\_1n\*0.000028349523125) if v79201\_1u=1 [convert ounce to ton]*

*Replace prod\_vcc1=missing if v79201\_1n=missing or v79201\_1u=missing*

*Label variable “Amount produced: VCC1 (metric tons)”*

**Step 5.** Calculate the total yield per hectare for VCC1 for each VCC1 producer (*yield\_vcc1*).

*Set yield\_vcc1=(prod\_vcc1÷area\_vcc1)*

*Replace yield\_vcc1=missing if prod\_vcc1=missing or area\_vcc1=missing*

*Label variable “Yield per hectare: VCC1 (tons per hectare)”*

**Step 6.** Create a second crop yield indicator (*yield\_vcc12*) that excludes outliers—that is, values more than three SD from the mean. Review each yield value determined to be an outlier, because it could be a legitimate value depending on the circumstances. Specific events and local meteorological conditions, in particular, can significantly affect yields. Complete and detailed documentation should be kept on the process because there are no strict rules to determine plausibility; another analyst may come to a different determination. Set values determined to be outliers to missing in the *yield\_vcc12* variable. Include all yield values in the *yield\_vcc12* variable that are determined to be plausible.

*Set mean\_yield\_vcc1=mean(yield\_vcc1)*

*Label variable “VCC1 yield mean”*

*Set sd\_yield\_vcc1=standard deviation(yield\_vcc1)*

*Label variable “Mean VCC1 yield standard deviation”*

*Set yield\_vcc12=missing*

*Replace yield\_vcc12=yield\_vcc1 if yield\_vcc1≥(mean\_yield\_vcc1-(3\*sd\_yield\_vcc1))and yield\_vcc1≤(mean\_yield\_vcc1+(3\*sd\_yield\_vcc1))*

*Label variable “Yield: VCC1 (tons per hectare), excluding outliers”*

*List mean\_yield\_vcc1 sd\_yield\_vcc1 yield\_vcc1 yield\_vcc12 if yield\_vcc1≠missing and yield\_vcc12=missing*

**Step 7.** Add the farm size variable (*farmsize*)from the household data file into the current working data file using the cluster and household number variables as the key matching variables. Use the farm size variable create two binary variables—the first to flag de jure producers whose households are smallholders and the second to flag de jure producers whose households are non-small holders.

*Add farmsize from “FTF ZOI Survey [Country] [Year] household data analytic” using hhea and hhnum as the key matching variables*

*Set sholder\_vcc1\_dj=0*

*Replace sholder\_vcc1\_dj =1 if farmsize=0 and hhmem\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Smallholder: VCC1 producer, de jure HH member”*

*Set nsholder\_vcc1\_dj= 0*

*Replace nsholder\_vcc1\_dj =1 if farmsize=1 and hhmem\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Non-smallholder: VCC1 producer, de jure HH member”*

**Step 8.** After weighting the data using the VCC1 producer weight, calculate VCC1 yield estimates using the *yield\_vcc1* analytic variable. Repeat the calculation for the first- and second-level disaggregates. (Example code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_v1], strata(samp\_stratum)*

*Svy, subpop(hhmem\_dj): mean yield\_vcc12*

*Svy, subpop(hhmem\_dj): mean yield\_vcc12, over(farmsize)*

*Svy, subpop(sholder\_vcc1\_dj): mean yield\_vcc12, over(sex)*

*Svy, subpop(sholder\_vcc1\_dj): mean yield\_vcc12, over(youth)*

*Svy, subpop(nsholder\_vcc1\_dj): mean yield\_vcc12, over(sex)*

*Svy, subpop(nsholder\_vcc1\_dj): mean yield\_vcc12, over(youth)*

#### Fishpond yield

The yield indicator for fishpond aquaculture estimates the average fish production in kilograms per hectare. Yield data are collected separately for each producer primarily responsible for raising fish in fishponds in a household. The step-by-step procedures follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax yield fishpond.do*. Variables from Modules 7.8 and 7.96 are used for the calculations.

##### Definitions

| Numerator | Yield of fishponds in the month preceding the survey in kilograms per hectare |
| --- | --- |
| Denominator | Number of producers who raised fish in fishponds in the month preceding the survey |
| Unit of measure | Mean yield in kilogram per hectare |
| Level of data | Individual (producer) level |
| Sampling weight | Fishpond producer |
| Disaggregation levels | Sex\*  Age category (15-29, 30 or older)\* |
| Coverage | P2-ZOI |
| Treatment of missing data | If information on fishpond area (*v9604*) or amount of fish harvested (*v78019bn*) or unit (*v78019bu*) is missing or “don’t know,” exclude that observation from the calculation. |
| Survey variables used | *hhea, hhnum, m1\_line, wgt\_fish, samp\_stratum*, all variables from the fish module (7.8) and the pond area module (7.96) |
| Analytic variables used | *vcc\_fish* |
| Analytic variables created | *area\_pond1-area\_pondn, pond\_area, prod\_fish, yield\_fish, yield\_fish2, mean\_yield\_fish, sd\_yield\_fish* |

\*Standard Feed the Future disaggregate

##### Calculations

The following steps guide the procedures to calculate yields of fishpond aquaculture in the past month.

**Step 1.** Prepare data—that is, load the individual-level analytic data file, keeping only variables required for analysis, including the sampling weight for producers who raised fish in fishponds. Keep only records for producers who raised fish in ponds.

*Load “$analytic\FTF ZOI Survey [Country] [Year] persons data analytic”*

*Keep hhea, hhnum, m1\_line, sex, wgt\_fish, and all variables that begin with ‘v796’ or ‘v798’ or that contain ‘vcc’*

*Keep if vcc\_fish=1*

**Step 2.** For each producer who raised fish in fishponds, calculate the area of each pond in hectares (*area\_pond1-area\_pondn,* where ‘*n*’ is the maximum number of ponds that a fish producer has in the data file), setting out-of-range values to missing. Then add the area of all ponds to obtain the total pond area for each producer who raised fish in fishponds (*area\_pond*). Four ponds are used for illustration.

*Set area\_pond1=(v9604\_1\*0.0001) if v9604\_1<999.4*

*Label variable “Area of fishpond 1”*

*Set area\_pond2=(v9604\_2\*0.0001) if v9604\_2<999.4*

*Label variable “Area of fishpond 2”*

*Set area\_pond3=(v9604\_3\*0.0001) if v9604\_3<999.4*

*Label variable “Area of fishpond 3”*

*Set area\_pond4=(v9604\_4\*0.0001) if v9604\_4<999.4*

*Label variable “Area of fishpond 4”*

*Set area\_pond=(area\_pond1+area\_pond2+area\_pond3+area\_pond4)*

*Replace area\_pond=missing if area\_pond1=missing and area\_pond2=missing and area\_pond3=missing and area\_pond4=missing*

*Label variable “Total area of fishponds (hectares)”*

**Step 3.** Calculate the total weight of fish harvested in the month preceding the survey in kilograms (*prod\_fish*). If a producer reported the weight of fish harvested in pounds or another unit included in the country-adapted questionnaire, convert the weight to kilograms.

*Set prod\_fish=(v78019bn\*1) if v78019bu=1*

*Replace prod\_fish=(v78019bn\*0.45359237) if v78019bu=2*

*Replace prod\_fish=missing if v78019bn=missing or v78019bu=missing*

*Label variable “Amount of fish harvested, month before survey (kg)”*

**Step 4.** Calculate the yield of fishponds in kilograms per hectare for all producers who raised fish in fishponds in the year preceding the survey (*yield\_fish*).

*Set yield\_fish=(prod\_fish÷area\_pond)*

*Replace yield\_fish=missing if prod\_fish=missing or area\_pond=missing*

*Label variable “Yield: fish (kg per hectare)”*

**Step 5.** Create a second fish yield indicator (*yield\_fish2*) that excludes outliers—that is, values more than three SDs from the mean. Review each yield value determined to be an outlier, because it could be a legitimate value depending on the circumstances. Specific events and local meteorological conditions, in particular, can significantly affect yields. Complete and detailed documentation should be kept on the process because there are no strict rules to determine plausibility; another analyst may come to a different determination. Set values determined to be outliers to missing in the *yield\_fish2* variable. Include all yield values in the *yield\_fish2* variable that are determined to be plausible.

*Set mean\_yield\_fish=mean(yield\_fish)*

*Label variable “Fish yield mean”*

*Set sd\_yield\_fish=standard deviation(yield\_fish)*

*Label variable “Mean fish yield standard deviation”*

*Set yield\_fish2=missing*

*Replace yield\_fish2=yield\_fish if yield\_fish≥(mean\_yield\_fish-(3\*sd\_yield\_dish) and yield\_fish≤(mean\_yield\_fish+(3\*sd\_yield\_fish))*

*Label variable “Yield: fish (kg per hectare), excluding outliers”*

**Step 6.** After weighting the data using the weight for fishpond producers, calculate VCC1 yield estimates using the *yield\_vcc1* analytic variable. Repeat the calculation for the first- and second-level disaggregates. (Example code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_fish], strata(samp\_stratum)*

*Svy, subpop(hhmem\_dj): mean yield\_fish2*

*Svy, subpop(hhmem\_dj): mean yield\_fish2, over(sex)*

*Svy, subpop(hhmem\_dj): mean yield\_fis2h, over(vcc\_youth)*

#### Dairy cow yield

The yield indicator for dairy cows estimates the average amount of milk produced in liters (TP) per producing dairy cows (UP) per day. Yield data are collected separately for each producer primarily responsible for raising dairy cows in a household. The step-by-step procedures follow the Stata syntax *FTF ZOI Survey [COUNTRY] [YEAR] syntax yield dairycow.do*. Variables from Module 7.50 are used for the calculations.

##### Definitions

| Numerator | Sum of milk production in liters per milking cow per day |
| --- | --- |
| Denominator | Number of producers who have dairy cows that are milking |
| Unit of measure | Mean, in liters per milking cow per day |
| Level of data | Individual (producer) |
| Sampling weight | Dairy cow producer |
| Disaggregation levels | Sex\*  Age category (15-29, 30 or older)\* |
| Coverage | P2-ZOI |
| Treatment of missing data | Producers with missing information about the number of milking cows owned (*v75001*) or milk produced (*v75037n, v75037u,* *v75038n, v75038u*) or with a “don’t know” response for these variables are excluded from both the numerator and denominator. |
| Survey variables used | *hhea, hhnum, m1\_line, wgt\_dairy, samp\_stratum*, *v226a, v226j, v240a, v241a, v75001,* *v75016, v75018,* *v75037n, v75037u, v75038n, v75038u* |
| Analytic variables used | *vcc\_dairy, hhmem\_dj, vcc\_youth, sex* |
| Analytic variables created | *milkcow\_num, milk\_prod1, milk\_prod2, milk\_unit1\_liter milk\_unit2\_liter milk\_prod, yield\_milk, mean\_yield\_milk, sd\_yield\_milk, out\_yield\_milk, median\_yield\_milk, herd\_mob, land\_access, hhcow\_num, cow\_hous, cow\_grazing, prodsys\_cows, prodsys\_cows1\_dj, prodsys\_cows2\_dj, prodsys\_cows3\_dj, prodsys\_cows4\_dj* |

\*Standard Feed the Future disaggregate

##### Calculations

The following steps guide the procedures to calculate the yield of milk from dairy cows per day.

**Step 1.** Prepare data—that is, load the individual-level analytic data file, keeping only variables required for analysis, including the sampling weight for producers who raised dairy cows. Keep only records for producers who raised dairy cows.

*Load “$analytic\FTF ZOI Survey [Country] [Year] persons data analytic”*

*Keep hhea, hhnum, m1\_line, sex, wgt\_dairy, and all variables that begin with ‘v750’ or that contain ‘vcc’*

*Keep if vcc\_dairy=1*

**Step 2.** Add the household-level variables required to calculate the production system disaggregate in Step 7 to the individual-level data file using cluster and household number as the key matching variables.

*Add v226a, v226j, v240a, v241a from “$analytic\FTF ZOI Survey [Country] [Year] household data analytic”, using hhea and hhnum as key matching variables.*

**Step 3.** Calculate the number of milking cows owned (*milkcow\_num*).

*Set milkcow\_num=missing*

*Replace milkcow\_num=v75001 if v75001≤95*

*Label variable “Number of milking cows owned”*

**Step 4.** Calculate the total volume of milk in liters that dairy cows produced in the day preceding the survey (*milk\_prod*) by summing the volumes produced in the morning and in the evening. If the milk production is reported in local units, first convert the local units to liters by using conversion factors.

*Set milk\_prod1=missing*

*Replace milk\_prod1=v75037n\*1 if v75037n≤ 990 and v75037u=1 [in liter]*

*Replace milk\_prod1=v75037n\*[conversion factor1] if v75037n≤990 and v75037u=2 [convert from local unit-1 to liter]*

*Replace milk\_prod1=v75037n\*[conversion factor2] if v75037n≤990 and v75037u=3 [convert from local unit-2 to liter]*

*Label variable “Liters of milk produced day before survey-morning”*

*Set milk\_prod2=missing*

*Replace milk\_prod2=v75038n\*1 if v75037n≤ 990 and v75038u=1 [in liter]*

*Replace milk\_prod2=v75038n\*[conversion factor1] if v75038n≤990 and v75037u=2 [convert from local unit-1 to liter]*

*Replace milk\_prod2=v75038n\*[conversion factor2] if v75038n≤990 and v75037u=3 [convert from local unit-2 to liter]*

*Label variable “Liters of milk produced day before survey-evening”*

*Set milk\_prod=(milk\_prod1+ milk\_prod2)*

*Replace milk\_prod=missing if milk\_prod1=missing or milk\_prod2=missing*

*Label variable “Liters of milk produced day before survey-total”*

**Step 5.** Calculate the yield of dairy milk per milking cow per day in liters (*yield\_milk*).

*Set yield\_milk=(milk\_prod÷milkcow\_num)*

*Replace yield\_milk=missing if milk\_prod=missing or milkcow\_num=missing*

*Label variable “Yield: liters per cow per day”*

**Step 6.** Create a second milk yield indicator (*yield\_milk2*) that excludes outliers—that is, values more than three SDs from the mean. Review each yield value determined to be an outlier, because it could be a legitimate value depending on the circumstances. Specific events and local meteorological conditions, in particular, can significantly affect yields. Complete and detailed documentation should be kept on the process because there are no strict rules to determine plausibility; another analyst may come to a different determination. Set values determined to be outliers to missing in the *yield\_milk2* variable. Include all yield values in the *yield\_milk2* variable that are determined to be plausible.

*Set mean\_yield\_milk=mean(yield\_milk)*

*Label variable “Milk yield mean”*

*Set sd\_yield\_milk=standard deviation(yield\_milk)*

*Label variable “Mean milk yield standard deviation”*

*Set yield\_milk2=missing*

*Replace yield\_milk2=yield\_milk if yield\_milk≥(mean\_yield\_milk-(3\*sd\_yield\_milk)) and yield\_milk≤(mean\_yield\_milk+(3\*sd\_yield\_milk))*

*Label variable “Yield: liters per cow per day, excluding outliers”*

*List yield\_milk yield\_milk2 if yield\_milk≠missing and yield\_milk2=missing*

**Step 7.** Calculate the production system disaggregate categories for dairy producers.

**Step 7a.** Create a binary variable that indicates household-level herd mobility (*herd\_mob*).

*Set herd\_mob=missing*

*Replace herd\_mob=0 if v226j=2*

*Replace herd\_mob=1 if v226j=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Livestock have spent at least one night away from household for grazing”*

**Step 7b.** Create a binary variable that indicates the producer’s household agricultural land access (*land\_access*).

*Set land\_access=missing*

*Replace land\_access=0 if v240a=2 and v241a=2*

*Replace land\_access=1 if v240a=1 or v241a=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Household has access to agricultural land”*

**Step 7c.** Create a variable for the total number of dairy cows owned by the producer’s household (*hhcow\_num*).

*Set hhcow\_num=0*

*Replace hhcow\_num=v226a if v226a≤95 and v226a≠missing*

*Label variable “Total number of dairy cows owned by producer’s household”*

**Step 7d.** Create a binary variable that indicates whether the type of shelter the producer is providing for the cows is a permanent structure *(cow\_hous*).

*Set cow\_hous=missing*

*Replace cow\_hous=0 if v75016≤4*

*Replace cow\_hous=1 if v75016=5 or v75016=6*

*Label values 0 “No” 1 “Yes”*

*Label variable “Dairy cows kept in permanent structure”*

**Step 7e.** Create a categorical variable that indicates whether the dairy cows are allowed to graze (*cow\_grazing*).

*Set cow\_grazing=missing*

*Replace cow\_grazing=0 if v75018=2*

*Replace cow\_grazing=1 if v75018=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Dairy cows graze”*

**Step 7f.** Create a categorical variable for the four production system categories (*prodsys\_cows*).

*Set prodsys\_cows=missing*

*Replace prodsys\_cows=1 if herd\_mob=1*

*Replace prodsys\_cows=2 if herd\_mob=0 and land\_access=1*

*Replace prodsys\_cows=3 if herd\_mob=0 and land\_access=0*

*Replace prodsys\_cows=4 if hhcow\_num>49 and cow\_hous=1 and cow\_grazing=0*

*Label values 1 “Rangelands” 2 “Mixed crop-livestock” 3 “Urban/peri-urban” 4 “Intensive/commercial production”*

*Label variable “Production system, dairy cows”*

**Step 8.** Using the production system variable, create four binary variables to flag de jure producers by production system.

*Set prodsys\_cows1\_dj=0*

*Replace prodsys\_cows1\_dj=1 if prodsys\_cows=1 and hhmem\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Rangelands, de jure dairy cow producer”*

*Set prodsys\_cows2\_dj=0*

*Replace prodsys\_cows2\_dj=1 if prodsys\_cows=2 and hhmem\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Mixed crop-livestock, de jure dairy cow producer”*

*Set prodsys\_cows3\_dj=0*

*Replace prodsys\_cows3\_dj=1 if prodsys\_cows=3 and hhmem\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Urban/peri-urban, de jure dairy cow producer”*

*Set prodsys\_cows4\_dj=0*

*Replace prodsys\_cows4\_dj=1 if prodsys\_cows=4 and hhmem\_dj=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Intensive/commercial production, de jure dairy cow producer”*

**Step 9.** After weighting the data using the dairy cow producer weight, calculate the dairy cow milk yield indicator using the *yield\_milk* analytic variable. Repeat the calculation for the first- and second-level disaggregates. (Example code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_dairy], strata(samp\_stratum)*

*Svy, subpop(hhmem\_dj): mean yield\_milk2*

*Svy, subpop(hhmem\_dj): mean yield\_milk2, over(sex)*

*Svy, subpop(hhmem\_dj): mean yield\_milk2, over(vcc\_youth)*

*Svy, subpop(prodsys\_cows1\_dj): mean yield\_milk2*

*Svy, subpop(prodsys\_cows1\_dj): mean yield\_milk2, over(sex)*

*Svy, subpop(prodsys\_cows1\_dj): mean yield\_milk2, over(vcc\_youth)*

*Svy, subpop(prodsys\_cows2\_dj): mean yield\_milk2*

*Svy, subpop(prodsys\_cows2\_dj): mean yield\_milk2, over(sex)*

*Svy, subpop(prodsys\_cows2\_dj): mean yield\_milk2, over(vcc\_youth)*

*Svy, subpop(prodsys\_cows3\_dj): mean yield\_milk2*

*Svy, subpop(prodsys\_cows3\_dj): mean yield\_milk2, over(sex)*

*Svy, subpop(prodsys\_cows3\_dj): mean yield\_milk2, over(vcc\_youth)*

*Svy, subpop(prodsys\_cows4\_dj): mean yield\_milk2*

*Svy, subpop(prodsys\_cows4\_dj): mean yield\_milk2, over(sex)*

*Svy, subpop(prodsys\_cows4\_dj): mean yield\_milk2, over(vcc\_youth)*

#### References

Sadras, V.O., et al. 2015. *Yield gap analysis of field crops—Methods and case studies*. FAO Water Reports No. 41. Available at: <http://www.fao.org/3/a-i4695e.pdf>

# Household food insecurity indicators

This chapter describes the Household Hunger Scale (HHS) indicator, a Feed the Future phase one indicator that measures the prevalence of households with moderate or severe hunger, and the prevalence of households with moderate or severe food insecurity using the FIES, a Feed the Future phase two indicator.

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.

## Guidelines to construct the indicators

This section provides the guidelines to construct the two Feed the Future food insecurity indicators.

### Prevalence of moderate and severe hunger in the population, based on the Household Hunger Scale

This indicator measures the percentage of households experiencing moderate or severe hunger as indicated by a score of 2 or more on the HHS. To collect data for this indicator, respondents are asked about the frequency with which three events were experienced by household members in the 4 weeks/30 days preceding the survey: (1) no food at all in the house, (2) went to bed hungry, and (3) went all day and night without eating. For each question, four responses are possible: never, rarely, sometimes, or often. The responses are scored as follows: a never response receives a value of 0, rarely or sometimes receives a value of 1, and often receives a value of 2. The values for the three questions are summed for each household, producing a HHS score ranging from 0 to 6.

### Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale

The prevalence of moderate or severe insecurity is calculated from respondent’s raw score—that is, the number of affirmative responses given to the eight FIES questions. It 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 provide steps for statistical validation that must precede the calculation of prevalence estimates.

United Nations Food and Agriculture Organization (FAO) produces two FIES indicators for global monitoring: moderate and severe food insecurity and severe food insecurity. Only the moderate and severe food insecurity indicator has been selected as a monitoring indicator for Sustainable Development Goal 2 (Target 2.1). This guide provides prevalence estimate calculations for four levels of food insecurity: little to no food insecurity, moderate food insecurity, severe food insecurity, and moderate or severe food insecurity. For further guidance, the analyst should review the guidance available at <http://www.fao.org/in-action/voices-of-the-hungry/using-fies/en/>

## Step-by-step procedure to calculate the food security indicators

This section describes the detailed step-by-step procedures to calculate the food security indicators.

### Prevalence of moderate and severe hunger in the population, based on the Household Hunger Scale

This section presents the procedures to calculate the Feed the Future HHS indicator. It is based on information collected in the core ZOI Survey Module 3, *Food Security and Resilience*. The step-by-step procedures follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_HHS.do* file*.*

#### Definitions

| Numerator | Number of surveyed households with a HHS score of 2 or more |
| --- | --- |
| Denominator | Number of surveyed households |
| Unit of measure | Percentage |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members)\* |
| Coverage | P1-ZOI |
| Treatment of missing data | Households with any missing or refused responses are excluded from calculation of the indicator |
| Survey variables used | *v308a, v308b, v308c, v308d, v308e, v308f, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, edulevel\_hh\_dj* |
| Analytic variables created | *hhs1, hhs2, hhs3, hhs, hhs\_cat* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create a variable that indicates the frequency with which households did not have food in their household during the 4 weeks preceding the survey (*hhs1*).

*Set hhs1=missing*

*Replace hhs1=0 if (v308a=2)*

*Replace hhs1=1 if (v308b=1 or 2)*

*Replace hhs1=2 if (v308b=3)*

*Label values 1 “Never” 2 “Rare/y/sometimes” 3 “Often”*

*Label variable “Frequency HH did not have food, past month”*

**Step 2.** Create a variable that indicates the frequency with which any household members went to bed hungry during the 4 weeks preceding the survey (*hhs2*).

*Set hhs2=missing*

*Replace hhs2=0 if (v308c=2)*

*Replace hhs2=1 if (v308d=1 or 2)*

*Replace hhs2=2 if (v308d=3)*

*Label values 1 “Never” 2 “Rare/y/sometimes” 3 “Often”*

*Label variable “Frequency HH members went to bed hungry, past month”*

**Step 3.** Create a variable that indicates the frequency with which any household members went all day and night without eating during the 4 weeks preceding the survey (*hhs3*).

*Set hhs3=missing*

*Replace hhs3=0 if (v308e=2)*

*Replace hhs3=1 if (v308f=1 or 2)*

*Replace hhs3=2 if (v308f=3)*

*Label values 1 “Never” 2 “Rare/y/sometimes” 3 “Often”*

*Label variable “Frequency HH members did not eat all day and night, past month”*

**Step 4.** Calculate the HHS score. Sum the values of all three HHS variables for each household, producing a HHS score (*hhs*) ranging from 0 to 6. If a household is missing a value for any of the three HHS variables created in steps 1, 2, or 3, set the *hhs* variable to missing.

*Set hhs=(hhs1+hhs2+hhs3)*

*Replace hhs=missing if hhs1=missing or hhs2=missing or hhs3= missing*

*Label variable “Household hunger score”*

**Step 5.** Create a binary variable to flag households with moderate or severe hunger (*hhs\_cat*).

*Set hhs\_cat=missing*

*Replace hhs\_cat=0 if hhs=0 or hhs=1*

*Replace hhs\_cat=1 if hhs≥2 and hhs≤6*

*Label values 0 “None to some” 1 “Moderate to severe”*

*Label variable “HH with moderate to severe hunger”*

**Step 6.** After applying the household sampling weight, calculate the percentage of households that experienced moderate to severe hunger using the *hhs\_cat* variable. Repeat using the gendered household type disaggregate constructed using de jure household members. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_hh], strata(samp\_stratum)*

*Svy: prop hhs\_cat*

*Svy: prop hhs\_cat, over(genhhtype\_dj)*

#### References

Ballard, T., Coates, J., Swindale, A., and Deitchler, M. 2011. *Household hunger scale: Indicator definition and measurement guide*. Washington, DC: Food and Nutrition Technical Assistance II Project, FHI 360. Available at: <https://www.fantaproject.org/sites/default/files/resources/HHS-Indicator-Guide-Aug2011.pdf>

Deitchler, M., Ballard, T., Swindale, A., and Coates, J. 2010. *Validation of a measure of household hunger for cross-cultural use*. Washington, DC: Food and Nutrition Technical Assistance II Project (FANTA-2), FHI 360. Available at: <https://www.fantaproject.org/sites/default/files/resources/HHS_Validation_Report_May2010_0.pdf>

INDDEX Project. 2018. *Household Hunger Scale (HHS)*. Data4Diets: Building blocks for diet-related food security analysis. Tufts University, Boston, MA. Available at: <https://inddex.nutrition.tufts.edu/data4diets/indicator/household-hunger-scale-hhs?back=/data4diets/indicators>

### Prevalence of moderate or severe food insecurity in the population, based on the Food Insecurity Experience Scale

#### Definitions

| Numerator | Population with a probability of exceeding the food insecurity severity level |
| --- | --- |
| Denominator | The surveyed population |
| Unit of measure | Percentage |
| Level of data | Household |
| Sampling weight | Household |
| Disaggregation levels | Gendered household type (de jure household members)\*  Level of severity (moderate, severe)\* |
| Coverage | P2-ZOI |
| Treatment of missing data | All missing and refused responses are excluded from calculations of the indicator |
| Survey variables used | *v301, v302, v303, v304, v305, v306, v307, v308, hhea, hh\_wgt, samp\_stratum* |
| Analytic variables used | *genhhtype\_dj, hhsize\_dj* |
| Analytic variables created | *worried, healthy, fewfood, skipped, ateless, runout, hungry, whlday, xx, wt, mem* |

\*Standard Feed the Future disaggregate

#### Calculations

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 FIES prevalence rates.

**Step 1.** Download R, install required packages, and set a working directory and reading data.

**Step 1a.** Click on <https://cran.r-project.org/> to download R. After it is downloaded, R can be used to program directly in the console or through a user-friendly compiler, RStudio, which needs to be downloaded separately.

**Step 1b.** Click on <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 the required packages. Run the following code to install the “*RM.weights,*” “*survey,*” “*foreign,*” and “*dplyr*” packages:

*install.packages("RM.weights") #* for weighted Rasch modeling and extensions using CML

*install.packages("survey") #* for analysis of complex survey samples.

*install.packages("foreign") #* to read data stored in Stata,[[88]](#footnote-87) SPSS, or other software

*install.packages("dplyr") #* for data manipulation

After the packages are installed, upload the package in the working library by running the following code:

*library(RM.weights)*

*library(survey)*

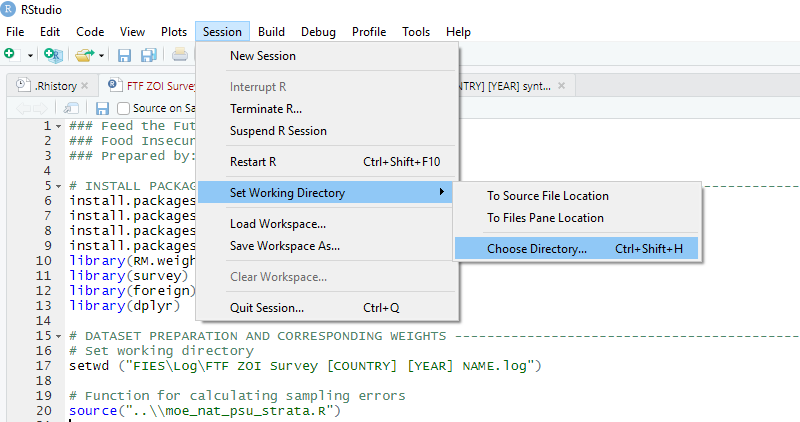
*library(foreign)*

*library(dplyr)*

**Step 1d.** Set a working directory and read function to calculate sampling errors. Set and specify a working directory where the data are stored. This will serve as the root directory *setwd ("FIES\\FTF ZOI Survey [COUNTRY] [YEAR] NAME").*

The working directory and folder name should be changed to reflect the directory path and name. Alternatively, go to RStudio console to set the working directory. Click on Session, then Set Working Directory and Choose Directory (Figure 1).

**Figure 1: Setting up a Working Directory in R**

**

**Step 1e.** Specify the function to calculate the sampling errors:

*source("moe\_nat\_psu\_strata.R")*

Read the function above for calculating the sampling errors. *moe* calculates the weighted margin of error. The code also allows the analysts to take into account the complex survey design, such as clustering and unequal probabilities of selection, by specifying the primary sampling unit, strata, and sampling weights.

**Step 1f.** Read the data file. Adding ‘*convert.factors = FALSE’* to the code means that all categorical variables are not converted to factors in R, but remain as numeric variables.

*ftf\_household = read.dta("FTF ZOI Survey [COUNTRY] [YEAR] household data analytic.dta", convert.factors = FALSE)*

**Step 2.** Prepare the data by preparing the data frame, analytic variables, recoding variables, and background characteristics.

**Step 2a.** To create a data frame, first create vectors of the FIES variables. Ensure that the list of variables is in the correct order and of the same length (check to see that you have an equal number of arguments/variables in all c( ) functions that are assigned as vectors.

*FIESvars <- c("WORRIED", "HEALTHY", "FEWFOOD", "SKIPPED", "ATELESS", "RUNOUT", "HUNGRY", "WHLDAY")*

*ftfvars <- c("v301", "v302", "v303", "v304", "v305", "v306", "v307", "v308")*

**Step 2b.** Create the data frame of the FIES variables only, and rename them to the standard FIES names. In the code below, ‘*xx’* are the eight FIES variables.

*xx <- ftf\_household[ftfvars]*

*names(xx) <- FIESvars*

**Step 2c.** Recode all variables. A response to any of the items as ‘YES’ or ‘1’ is coded as ‘1,’ and a response to any of the questions as ‘NO’ or ‘2’ is coded as ‘0.’ A response ‘REFUSED’ or “3” will be excluded from the analysis.

*xx <- xx %>% mutate\_all(*

*function(x) case\_when(*

*x == 1 ~ 1,*

*x == 2 ~ 0*

*)*

*)*

**Step 2d.** Define the weight variable and the de jure household members (i.e., the household size). The estimates represent the entire ZOI population and not only the individuals or the households interviewed in the survey, so sampling weight must be applied to account for the complex sample design.

*wt <- ftf\_household$wgt\_hh*

*mem <- ftf\_household$hhsize\_dj*

**Step 2e.** Storing the background variables and disaggregates that will be used later for analysis.

*fies <- xx*

*fies$wt <- wt*

*fies$mem <- mem*

*fies$fcluster <- ftf\_household$hhea*

*fies$hhnum <- ftf\_household$hhnum*

*fies$strata <- ftf\_household$a03c*

*fies$region <- factor(ftf\_household$region, levels=c(1,2,3), labels=c("region1","region2","region3"))*

*fies$urbanrural <- factor(ftf\_household$urban\_rural, levels=c(1,2), labels=c("urban","rural"))*

*fies$genhhtype\_dj <- factor(ftf\_household$genhhtype\_dj, levels=c(1,2,3,4), labels=c("Male and Female adults","Female adults only","Male adults only", "Children only"))*

**Step 3.** Fit weighted and unweighted Rasch at for the full sample, using household weights only and perform statistical validation of FIES data.

The Rasch model assesses the suitability of a set of survey items (questions) for scale construction (of food insecurity experience latent trait), computes parameter estimates and assessment statistics for each item, generates a scale from the items, and assesses the location of an individual or household along the continuum of the scale that uniquely reflects the food insecurity situation of that household. The model also implements procedures to calculate prevalence rates of the food insecurity (latent variable) and to equate item parameters across different contexts.

*rr <- RM.w(xx, wt, write.file = TRUE, country = "FTF ZOI SURVEY COUNTRY")*

*ru <- RM.w(xx, write.file = TRUE, country = "FTF ZOI SURVEY COUNTRY unweighted")*

*rs = rowSums(xx)*

*tab.weight(rs, wt\*mem, xx)$RS.abs.w* # to get weighted population by raw score

The Rasch output CSV files provide four main results that are used to test the quality of data collected. The four main results obtained are as follows:

* Infit (and corresponding standard errors): Assesses the data for items that did not perform well in a particular ZOI population. If the infit statistic is 1.0, 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.
* Outfit (and corresponding standard errors): Similar to infit, but sensitive to cases with unusual response patterns, even among a few respondents
* Residual correlation matrix: Looks at items that may be slightly redundant (i.e., that represent the same or closely related conditions caused by food insecurity)
* Rasch reliability: The proportion of total variance in the ZOI population that is accounted for by the measurement model. In other words, it provides information on the discriminatory power to the scale. The standard Rasch reliability statistic can be influenced by the total number of cases across the raw scores because it is weighted by the number of cases in each score. The Flat Rasch reliability statistic assumes that each score has an equal number of cases, and thus it provides a comparable measure of model fit.

In addition, the Rasch model output also provides information on the raw scores, distribution of valid responses, missing data by item, and a detailed output that shows the observed response proportion and the predicted response proportion for each raw score and items. While assessing quality of the FIES data, an analyst needs to consider the following:

* An adequate fit to the Rasch model is indicated by infit and outfit statistics of 0.7-1.3 for each item. Values between 1.3 and 1.5 are still acceptable, but items with *infit* values more than 1.5 should not be used for scoring.
* An outfit of >2 is useful to flag the presence of outliers and hence is considered high.
* If a high percentage of missing responses or a high infit affecting only one or two questions is observed, then these items may be dropped from the scale, and an analysis should be performed on the remaining six or seven items.

NOTE. Although a scale with fewer than eight items may be used for the analysis, **a minimum of six items** must be retained to produce an acceptable measure of food insecurity.

* If an item has more than 10 percent missing responses, an analyst may decide to drop it from the scale.
* A residual correlation between a pair of items is considered high if it is >|0.4|. This shows that each item is not able to capture the different aspect of food insecurity. If two items have high residual correlation, it may be necessary to drop one of them and rerun the model.
* A Rasch reliability value above 0.7 for an eight-item FIES scale is considered acceptable.
* A Rasch reliability value above 0.6 for a seven-item FIES scale is considered acceptable.

Open the example spreadsheet[[89]](#footnote-88) (provided by FAO) and fill in the highlighted cells as described in the spreadsheet using the output file produced by the RM function. For the weighted cases in cells H6:H14, use the weighted population from the output of the *tab.weights* function. Next, use the ‘perform equating’ worksheet to equate the survey results to the global standard and produce prevalence estimates.

**Figure 2: Perform Equating Worksheet**

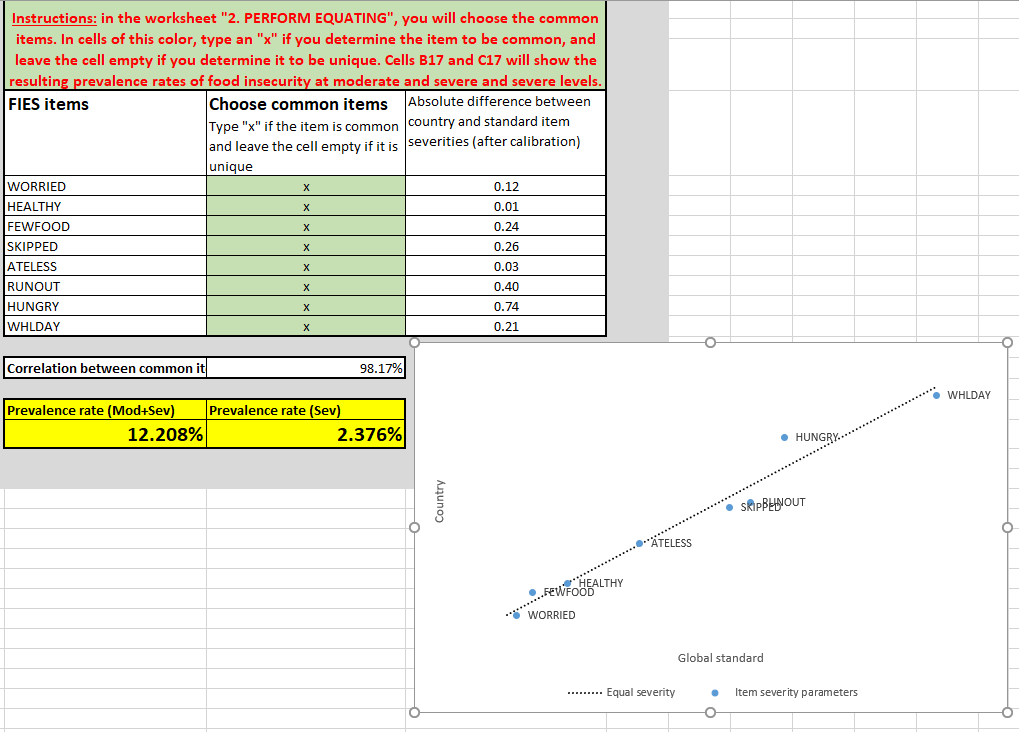
****

Figure 2 shows an example of the Perform Equating worksheet. For each of the eight items, it provides absolute differences between the item severities of the two scales and a column to choose common items (shaded in green). It also provides a plot of adjusted item severities of the two scales. The analyst needs to determine which items are common between two scales and which are unique items (the parameters differ too much between the two scales). For example, in Figure 2, by observing the absolute difference and by examining the plot, it can be observed that the item HUNGRY is discrepant. When this happens, an analyst may choose to consider an item as unique and omit it from the equating and repeat the equating. To consider an item as unique, remove the ‘x’ in the green cell for the item. Dropping an item will change the correlation between common items, which is displayed in the cells above the prevalence estimates. Figure 2 shows a correlation of 98.17 percent.

**NOTE. It is important to note that equating decisions have an effect on prevalence rates.**

The choices may involve a level of informed judgment, but they should be based on and guided by the principles described above. Further, it is important to describe the criteria used to drop the items.

Use equating output from R for measuring prevalence of food security. This is the final output that provides information on common and unique items, correlation between common items, adjusted thresholds on survey metrics, and prevalence rates (percentage) at adjusted thresholds for moderate or severe food insecurity and severe food insecurity.

**Step 3a.** Calculating prevalence estimates and standard errors.

**Step 3b.**  Equating function is used to calibrate the measure derived from the scale to the global reference scale. It then performs a probabilistic assignment of households to levels of severity along the latent trait. Next, it estimates the weighted probability of being beyond the threshold for moderate and severe food insecurity.

*ee <- equating.fun(rr, wt.spec=wt\*mem, write.file=TRUE)*

*ee$prevs\*100* # displays the prevalence estimates

Note that the weights are multiplied by the number of de jure household members to produce population estimates rather than household estimates.

**Step 3c.**  Extract the probabilities for each score for moderate and severe for each case for use in the calculation of the sampling errors.

*prob.rs=ee$probs.rs*

*prob.mod=prob.rs[rs+1,1]*

*prob.sev=prob.rs[rs+1,2]*

**Step 3d.** Obtain the standard errors.

Obtain the standard errors taking into account the sampling stratification (for narrow confidence intervals).

*se2.mod <- moe(prob=prob.mod, rs=rs, wt=wt, psu=fies$fcluster, sampl.strata=fies$strata)*

*se2.sev <- moe(prob=prob.sev, rs=rs, wt=wt, psu=fies$fcluster, sampl.strata=fies$strata)*

*se2 <- c("FI\_mod+"=se2.mod, "FI\_sev"=se2.sev)*

*se2\*100*

**Step 4**. Background characteristics and disaggregates. This guide also describes the background variables (region, urban/rural) and diaggregates required for Feed the Future reporting—gendered household type and severity (medium, severe) of food insecurity. For the purposes of baseline report, and endline-baseline report, the analyst can add other disaggregates to the codes in Steps 2e and 4f.

Note: The gendered household type variable identifies each household’s gendered type—that is, adult male and female, adult female-only, adult male-only, or child-only. If the number of cases is too small for any level of diaggregate, the analysis will not generate prevalence estimates for the diaggregate (gendered household type).

**Step 4a.** Use the following function to loop through all levels of background characteristics. The function returns a list of results, within an entry in the list for each individual category of the background characteristic.

*FIES\_char <- function (df.char, fies. = fies, rr. = rr, ee. = ee)*

*{*

*list.x <- list()*

*for (i in levels(df.char) ) {*

*print(i)*

*# selection for this individual category*

*select <- !is.na(df.char) & df.char==i*

**Step 4b.** Sub-set the background characteristics into the questionnaire data, the weight variable, and the de jure members, and run the Rasch model for the category.

*xx <- fies[select,1:8]*

*wt <- fies$wt[select]*

*mem <- fies$mem[select]*

*list.x[[i]]$rr <- RM.w(xx, wt, write.file = TRUE, country = paste("COUNTRY",i))*

Note: In the code above, ‘*i’* is the *ith* position for the background characteristics (e.g., urban position, rural position).

**Step 4c.** Calculate the prevalence of food insecurity, assuming the same model as for the full sample, applied to the subsample, and display the prevalence estimates. The below code performs a probabilistic assignment of households to levels of severity along the latent trait and estimates the weighted probability of being beyond the threshold for moderate and severe food insecurity.

*list.x[[i]]$prev <- prob.assign(sthres = ee$adj.thres, flex=list(a=rr$a,se.a=rr$se.a,XX=xx,wt=(wt\*mem)))$sprob*

*list.x[[i]]$prev\*100*

**Step 4d.** Extract the raw scores and the probabilities for each score for moderate and severe food insecurity for each case.

*rs = rowSums(xx)*

*prob.mod=prob.rs[rs+1,1]*

*prob.sev=prob.rs[rs+1,2]*

**Step 4e.** Calculate standard errors, taking into account the sampling stratification.

*se2.mod <- moe(prob=prob.mod, rs=rs, wt=wt\*mem, psu=fies$fcluster[select], sampl.strata=fies$strata[select])*

*se2.sev <- moe(prob=prob.sev, rs=rs, wt=wt\*mem, psu=fies$fcluster[select], sampl.strata=fies$strata[select])*

*list.x[[i]]$se2 <- c("FI\_mod+"=se2.mod, "FI\_sev"=se2.sev)*

*list.x[[i]]$se2\*100*

The loop for the background characteristics within the function ends with the closing brace, and then returns a list of outputs for the background characteristic.

*}*

*return(list.x)*

*}*

**Step 4f.** Run the function to produce the results for each background characteristic.

*list.reg <- FIES\_char(fies$region)*

*list.res <- FIES\_char(fies$urbanrural)*

*list.ghh <- FIES\_char(fies$genhhtype\_dj)*

Additional background characteristics can be added. However, note that these should first be included under Step 2e.

**Step 5.** Output the results.

**Step 5a** The following function creates a data frame of the prevalence estimates for output for a background characteristic.

*FIES\_df <- function( df.char, list.x ) {*

*df.x <- data.frame()*

*for ( i in levels(df.char) ) {*

*df.x <- rbind(df.x,cbind(*

*"Little or no" = 100\*(1-list.x[[i]]$prev[1]),*

*"Moderate" = 100\*(list.x[[i]]$prev[1]-list.x[[i]]$prev[2]),*

*"Severe" = 100\*(list.x[[i]]$prev[2]),*

*"Moderate or severe" = 100\*(list.x[[i]]$prev[1]),*

*"Number" = sum(list.x[[i]]$rr$wt),*

*"CI Severe low" = 100\*(list.x[[i]]$prev[2]-list.x[[i]]$se2[2]),*

*"CI Severe high" = 100\*(list.x[[i]]$prev[2]+list.x[[i]]$se2[2]),*

*"CI Moderate+ low" = 100\*(list.x[[i]]$prev[1]-list.x[[i]]$se2[1]),*

*"CI Moderate+ high" = 100\*(list.x[[i]]$prev[1]+list.x[[i]]$se2[1])*

*))*

*}*

*row.names(df.x) <- levels(df.char)*

*return(df.x)*

*}*

**Step 5b.** Construct the data frame of results for all background characteristics

*df <- data.frame()*

*df <- rbind(df,FIES\_df(fies$region, list.reg))*

*df <- rbind(df,FIES\_df(fies$urbanrural, list.res))*

*df <- rbind(df,FIES\_df(fies$genhhtype\_dj, list.ghh))*

*# Add total to the bottom of the data frame*

*df.t <- data.frame()*

*prev <- ee$prevs*

*df.t <- cbind(*

*"Little or no" = 100\*(1-prev[1]),*

*"Moderate" = 100\*(prev[1]-prev[2]),*

*"Severe" = 100\*(prev[2]),*

*"Moderate or severe" =100\*(prev[1]),*

*"Number" = sum(rr$wt),*

*"CI Severe low" = 100\*(prev[2]-se2[2]),*

*"CI Severe high" = 100\*(prev[2]+se2[2]),*

*"CI Moderate+ low" = 100\*(prev[1]-se2[1]),*

*"CI Moderate+ high" = 100\*(prev[1]+se2[1])*

*)*

*row.names(df.t) <- c("Total")*

*df <- rbind(df,df.t)*

*# Write the resuts to a CSV file*

*write.csv(df, file="FTF ZOI SURVEY COUNTRY FIES table.csv")*

The above code will provide prevalence estimates for four levels of food insecurity: little to no food insecurity, moderate food insecurity, severe food insecurity, and moderate or severe food insecurity. The calculations will also generate outputs for each background characteristic and the required disaggregates.

#### References

Ballard, T., Kepple, A. W., and Cafiero, C. 2013. *The food insecurity experience scale: Development of a global standard for monitoring hunger worldwide*. Technical Paper. Rome: FAO. Available at: <http://www.fao.org/fileadmin/templates/ess/voh/FIES_Technical_Paper_v1.1.pdf>

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: FAO. Available at: <http://www.fao.org/3/a-i4830e.pdf>

FAO. 2019. *Voices of the hungry: Applying the FIES*. Available at: <http://www.fao.org/in-action/voices-of-the-hungry/using-fies/en/>

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: FAO. Available at: <http://www.fao.org/3/a-i3946e.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: FAO. Available at: <http://www.sesric.org/imgs/news/1752_Manual_on_RM_Weights_Package_EN.pdf>

# 15. Children’s and women’s nutrition indicators

This chapter describes the children’s and women’s nutrition indicators related to diet and anthropometry. 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.

## Guidelines to construct the indicators

This section provides the guidelines to construct the Feed the Future children’s and women’s nutrition indicators.

### 15.1.1 Children’s nutritional status indicators

Underweight, stunted, wasted, and healthy weight children under 5 years of age (0-59 months) are the major children’s nutritional status indicators:

* Prevalence of underweight children under 5 years of age
* Prevalence of stunted children under 5 years of age
* Prevalence of wasted children under 5 years of age
* Prevalence of healthy weight among children under 5 years of age

Stunting is an indicator of linear growth retardation, most often due to prolonged exposure to an inadequate diet and poor health. Reducing the prevalence of stunting among children, particularly children 0-23 months of age, is important because linear growth deficits accrued early in life are associated with cognitive impairments, poor educational performance, and decreased work productivity among adults. Better nutrition leads to increased cognitive and physical abilities, thus improving individual productivity in general, including improved agricultural productivity. Stunting reflects the effects of a systematic lack of adequate nutrition over a number of years and recurrent and chronic illness. It is, therefore, a measure of the long-term effects of malnutrition and does not vary significantly according to the season of data collection, unlike wasting.

Wasting is an indicator of acute undernutrition among children under 5 years of age, resulting from acute and dire food shortage or disease. It is a robust predictor of under-5 mortality.

To obtain nutritional status indicators from height or length[[90]](#footnote-89), 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, weight-for-age z-score (WAZ), relating to underweight, and weight-for-height z-score (WHZ), relating to wasting. For the Feed the Future phase one endline/phase two baseline ZOI Surveys, z-scores are calculated after data collection is completed but before data analysis begins—during the post-processing stage.[[91]](#footnote-90) The methodology for calculating z-scores is described below.

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 SD of the reference population, as shown in the following formula:

*Z-score=(Individual value of the child−Median value of children in the reference population)÷(Standard deviation of the reference population)*

Using the formula, the z-scores are calculated as follows:

*HAZ=(Height of child in the sample−Median value of height of children in the reference population having the same age and sex)÷(Standard deviation of height of children in the reference population having the same age and sex)*

*WAZ=(Weight of child in the sample−Median value of weight of children in the reference population having the same age and sex)÷(Standard deviation of weight of children in the reference population having the same age and sex)*

*WHZ=(Weight of child in the sample−Median value of weight of children in the reference population having the same height and sex)÷(Standard deviation 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 16.[[92]](#footnote-91) 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 16: Z-score Cutoffs for Children’s Nutritional Status Indicators**

| Nutritional Status Indicator | Z-score | Minimum Cutoff | Maximum Cutoff |
| --- | --- | --- | --- |
| Stunting | HAZ | -6 SD | +6 SD |
| Underweight | WAZ | -6 SD | +5 SD |
| Wasting | WHZ | -5 SD | +5 SD |

Source: WHO. 2006. *The WHO Child Growth Standards*

### 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 nutrition 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.

### 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, and feeding frequency, which is a proxy for energy density, by breastfeeding status, 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 feeding frequency and minimum dietary diversity for their respective age groups and breastfeeding status, then they are considered to receive 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 indicator is calculated using the following two proportions:

* (Breastfed children 6-23 months of age who had at least the minimum dietary diversity and the minimum meal frequency during the day and night preceding the survey) ÷ (Breastfed children 6-23 months of age)
* (Non-breastfed children 6-23 months of age who received at least two milk feedings and had at least the minimum dietary diversity not including milk feeds and the minimum meal frequency during the day and night preceding the survey) ÷ (Non-breastfed children 6-23 months of age)

#### Breastfed children

Minimum dietary diversity is based on the following seven food groups for breastfed children:

* Grains, roots, and tubers
* Legumes and nuts
* Dairy products (milk, yogurt, cheese)
* Flesh foods (meat, fish, poultry, and liver/organ meats)
* Eggs
* Vitamin-A rich fruits and vegetables
* Other fruits and vegetables

Minimum dietary diversity for breastfed children 6-23 months of age is defined as having consumed foods from at least four of the seven food groups.

Minimum meal frequency for breastfed children is defined as two or more feedings of solid, semi-solid, or soft food for children 6-8 months of age, and three or more feedings of solid, semi-solid, or soft food for children 9-23 months of age.

#### Non-breastfed children

Minimum dietary diversity is based on the following six food groups for non-breastfed children:

* Grains, roots, and tubers
* Legumes and nuts
* Flesh foods (meat, fish, poultry, and liver/organ meats)
* Eggs
* Vitamin-A rich fruits and vegetables
* Other fruits and vegetables

Minimum dietary diversity for non-breastfed children 6-23 months of age is defined using only six of the seven food groups; the dairy products group is not included. It is defined as having consumed foods from at least four of the six applicable food groups.

Minimum meal frequency for non-breastfed children is defined as four or more feedings of solid, semi-solid, soft food, or milk feeds for children 6-23 months of age, and at least two of these feedings must be milk feeds.

### 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-generation 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=weight (in kg)]÷[height (in meters) squared]. Criteria for categorizing women’s nutritional status using BMI are shown in Table 17. Women who have a BMI less than 12 or greater than 60 are excluded from the indicator calculation.[[93]](#footnote-92)

**Table 17: Women’s Nutritional Status Category by BMI**

| Women’s Nutritional Status | BMI |
| --- | --- |
| Moderately and severely underweight | <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 |

### Women’s dietary diversity: Mean number of food groups consumed by women of reproductive age

Women of reproductive age are at risk for multiple micronutrient deficiencies, which can jeopardize their health and ability to care for their children and participate in income-generating activities. Maternal micronutrient deficiencies during lactation can directly impact child growth and development, but the potential consequences of maternal micronutrient deficiencies are especially severe during pregnancy, when there is the greatest opportunity for nutrient deficiencies to cause long-term, irreversible development consequences for the child in utero. Dietary diversity—assessed here as the number of food groups consumed—is a key dimension of a high-quality diet with adequate micronutrient content, and thus is important to ensure the health and nutrition of both women and their children.

This validated indicator of the probability of micronutrient adequacy of the diet reports the mean number of food groups consumed in the previous day and night by women of reproductive age (15-49 years). To calculate this indicator, the following nine food groups are used:

* Grains, roots, and tubers
* Legumes and nuts
* Dairy products (milk, yogurt, cheese)
* Organ meat
* Eggs
* Flesh foods and other miscellaneous small animal protein
* Vitamin A-rich dark green leafy vegetables
* Other vitamin A-rich vegetables and fruits
* Other fruits and vegetables

The food groups are based on the micronutrient content that is necessary to maintain the health of women of reproductive age and provide necessary nutrients to a child in utero. Adequate consumption of these foods may also help reduce maternal micronutrient deficiencies during the lactation period and provide necessary micronutrients to lactating children.

The mean number of food groups consumed by women of reproductive age indicator is tabulated by averaging the number of food groups consumed during the day and night preceding the survey out of the nine food groups above across all women of reproductive age in the sample with data on dietary diversity.

### 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
* Other vegetables
* Other 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.

This indicator is a new version of the Women’s Dietary Diversity Score (WDDS) indicator. There are two main differences between this indicator and the WDDS. First, this indicator is a dichotomous indicator, whereas the WDDS is a quasi-continuous score. Dichotomous indicators, which reflect the proportion of a population of interest that is above or below a defined threshold (in this case, women who are consuming a diet of minimum diversity), are more intuitive and understandable to a broad audience of stakeholders. This indicator will be more useful for reporting and describing progress toward improved nutrition for women than the WDDS, which reports the mean number of food groups consumed by women. Second, the food groups used to calculate this indicator are slightly different from those used to calculate WDDS. This indicator uses 10 food groups, whereas WDDS uses 9.

## Step-by-step procedures to calculate nutrition indicators

This section describes the detailed step-by-step procedures to calculate the nutrition indicators.

### Prevalence of underweight 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 underweight in the ZOI population. It is based on information collected in the core ZOI Survey Module 5A, *Children’s Anthropometry.* The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_CHN.do* file*.*

#### Definitions

| Numerator | Number of de facto children 0-59 months of age whose WAZ is more than 2 SD below the 2006 WHO Child Growth standards population median |
| --- | --- |
| Denominator | Total number of de facto children 0-59 months of age |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Children under 5 years of age |
| Disaggregation levels | Child sex\*  Child age category (0-11, 12-23, 24-35, 36-47, 48-59 months)  Caregiver education  Gendered household type (de jure household member)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P1-ZOI |
| Treatment of missing data | Children who were not weighed or are missing weight 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 | *hhea, hhnum, samp\_stratum, wgt\_c5, waz* |
| Analytic variables used | *c0\_59m, hhmem\_df, cnut\_age, sex* |
| Analytic variables created | *wazd, chn\_uw* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create a new variable (*wazd*) by dividing the WAZ variable (*waz*) by 100. The WAZ variable was created using a CSPro data post-processing program and is included in the person-level data file exported from CSPro.[[94]](#footnote-93)

*Replace wazd=waz÷100*

*Label variable “Weight-for-age z-score, decimal”*

**Step 2.** Create a binary variable that flags children under 5 years of age who are moderately to severely underweight (*chn\_uw*), meaning that their WAZ score is less than -2. WAZ values of 99.96 and above correspond to missing or invalid z-scores and should be omitted from the variable.

*Set chn\_uw=missing*

*Replace chn\_uw=0 if wazd≥-2 and waz<9996 and c0\_59m=1*

*Replace chn\_uw=1 if wazd<-2 and c0\_59m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is moderately-to-severely underweight (WAZ<-2)”*

**Step 3.** After applying the children under 5 weight (*wgt\_c5*), calculate the percentage of de facto children under 5 years of age who are underweight using the *chn\_uw* analytic variable. Repeat using the child sex and child age disaggregates. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_c5], strata(samp\_stratum)*

*Svy, subpop(hhmem\_df): prop chn\_uw*

*Svy, subpop(hhmem\_df): prop chn\_uw, over(cnut\_age)*

*Svy, subpop(hhmem\_df): prop chn\_uw, over(sex)*

#### 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

World Health Organization. 2019. *Global database on child growth and malnutrition.* Available at: <https://www.who.int/nutgrowthdb/about/introduction/en/>

World Health Organization. 2006. *The WHO child growth standards.* Available at: <http://www.who.int/childgrowth/publications/technical_report_pub/en/>

World Health Organization and 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

### 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. It is based on information collected in the core ZOI Survey Module 5A, *Children’s Anthropometry.* The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_CHN.do* file*.*

#### Definitions

| Numerator | Number of de facto 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 de facto children 0-59 months of age in surveyed households |
| 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)  Caregiver education  Gendered household type (de jure household member)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P1-ZOI and P2-ZOI |
| 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 | *hhea, hhnum, samp\_stratum, wgt\_c5, haz* |
| Analytic variables used | *c0\_59m, hhmem\_df, sex, cnut\_age* |
| Analytic variables created | *hazd, chn\_stunted* |

\*Standard Feed the Future disaggregate

\*Standard Feed the Future disaggregate Feed the Future phase two only

#### Calculations

**Step 1.** Create a new variable (*hazd*) by dividing the HAZ variable (h*az*) by 100. The HAZ variable was created using a CSPro data post-processing program and is included in the person-level data file exported from CSPro.[[95]](#footnote-94)

*Replace hazd=haz÷100*

*Label variable “Height-for-age z-score, decimal”*

**Step 2.** Create a binary variable that flags children under 5 years of age who are stunted (*chn\_stunted*), meaning that their HAZ score is less than -2. HAZ values of 99.96 and above correspond to missing or invalid z-scores and should not be included.

*Set chn\_stunted=missing*

*Replace chn\_stunted=0 if hazd≥ -2 and haz<9996 and c0\_59m=1*

*Replace chn\_stunted=1 if hazd<-2 and c0\_59m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is moderately-to-severely stunted (HAZ<-2)”*

**Step 3.** After applying the children under 5 weight (*wgt\_c5*), calculate the percentage of de facto children under 5 years who are stunted using the *chn\_stunted* analytic variable. Repeat using the child sex and child age category disaggregates. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_c5], strata(samp\_stratum)*

*Svy, subpop(hhmem\_df): prop chn\_stunted*

*Svy, subpop(hhmem\_df): prop chn\_stunted, over(cnut\_age)*

*Svy, subpop(hhmem\_df): prop chn\_stunted, over(sex)*

#### 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

World Health Organization. 2019. *Global database on child growth and malnutrition.* Available at: <https://www.who.int/nutgrowthdb/about/introduction/en/>

World Health Organization. 2006. *The WHO child growth standards.* Available at: <http://www.who.int/childgrowth/publications/technical_report_pub/en/>

World Health Organization and 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

### 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. It is based on information collected in the core ZOI Survey Module 5A, *Children’s Anthropometry.* The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_CHN.do* file*.*

#### Definitions

| Numerator | Number of de facto 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 de facto children 0-59 months of age in surveyed households |
| 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)  Caregiver education  Gendered household type (de jure household member)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P1-ZOI and P2-ZOI |
| 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 | *hhea, hhnum, samp\_stratum, wgt\_c5, whz* |
| Analytic variables used | *c0\_59m, hhmem\_df, sex, cnut\_age* |
| Analytic variables created | *whzd, chn\_wasted* |

\*Standard Feed the Future disaggregate

\*\*Standard Feed the Future disaggregate Feed the Future phase two only

#### Calculations

**Step 1.** Divide the WHZ variable (*whz*) by 100. The WHZ variable was created using a CSPro data post-processing program and is included in the person-level data file exported from CSPro.[[96]](#footnote-95)

*Replace whz=whz÷100*

*Label variable “Weight-for-height z-score, decimal”*

**Step 2.** Create a binary variable that flags children under 5 years of age who are wasted (*chn\_wasted*), meaning that their WHZ score is less than -2. WHZ values of 99.96 and above correspond to missing or invalid z-scores and should be omitted from the variable.

*Set chn\_wasted=missing*

*Replace chn\_wasted=0 if whz≥ -2 and whz<9996 c0\_59m=1*

*Replace chn\_wasted=1 if whz<-2 and c0\_59m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child is moderately-to-severely wasted (WHZ<-2)”*

**Step 3.** After applying the children under 5 weight (*wgt\_c5*), calculate the percentage of de facto children under 5 years of age who are wasted using the *chn\_wasted* analytic variable. Repeat using the child sex and child age category disaggregates. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_c5], strata(samp\_stratum)*

*Svy, subpop(hhmem\_df): prop chn\_wasted*

*Svy, subpop(hhmem\_df): prop chn\_wasted, over(cnut\_age)*

*Svy, subpop(hhmem\_df): prop chn\_wasted, over(sex)*

#### 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

World Health Organization. 2019. *Global database on child growth and malnutrition.* Available at: <https://www.who.int/nutgrowthdb/about/introduction/en/>

World Health Organization. 2006. *The WHO child growth standards.* Available at: <http://www.who.int/childgrowth/publications/technical_report_pub/en/>

World Health Organization and 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

### 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. It is based on information collected in the core ZOI Survey Module 5A, *Children’s Anthropometry.* The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_CHN.do* file*.*

#### Definitions

| Numerator | Number of de facto 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 de facto children 0-59 months of age in surveyed households |
| 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)  Caregiver education  Gendered household type (de jure household member)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P2-ZOI |
| 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 | *hhea, hhnum, samp\_stratum, wgt\_c5, whz* |
| Analytic variables used | *whzd, c0\_59m, hhmem\_df, sex, cnut\_age* |
| Analytic variables created | *chn\_hw* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create a binary variable that flags children under 5 years of age who have a healthy weight (*chn\_hw*), meaning that their WHZ score is between -2 and 2. WHZ values of 99.96 and above correspond to missing or invalid z-scores and should be omitted from the variable.[[97]](#footnote-96)

*Set chn\_hw=missing*

*Replace chn\_hw=0 if (whz<-2 or whz >2) and whz<996 and c0\_59m=1*

*Replace chn\_hw=1 if (whz≥-2 and whz≤2) and c0\_59m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child has a healthy weight (WHZ≥-2 and WHZ≤2)”*

**Step 2.** After applying the children under 5 weight (*wgt\_c5*), calculate the percentage of de facto children under 5 years of age who have a healthy weight using the *chn\_hw* analytic variable. Repeat using the child sex and child age category disaggregates. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_c5], strata(samp\_stratum)*

*Svy, subpop(hhmem\_df): prop chn\_hw*

*Svy, subpop(hhmem\_df): prop chn\_hw, over(cnut\_age)*

*Svy, subpop(hhmem\_df): prop chn\_hw, over(sex)*

#### 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

World Health Organization. 2019. *Global database on child growth and malnutrition.* Available at: <https://www.who.int/nutgrowthdb/about/introduction/en/>

World Health Organization. 2006. *The WHO child growth standards.* Available at: <http://www.who.int/childgrowth/publications/technical_report_pub/en/>

World Health Organization and 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

### 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 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.do* file*.*

#### Definitions

| Numerator | Number of de facto children 0-5 months of age who received only breastmilk and no other liquid, solid food, or plain water in the day and night preceding the survey |
| --- | --- |
| Denominator | Number of de facto children 0-5 months of age |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Child under 2 years of age |
| Disaggregation levels | Child sex\*  Caregiver education  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P1-ZOI and P2-ZOI |
| 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 regarding feedings in the day and night preceding the survey will be omitted from the numerator and denominator. |
| Survey variables used | *hhea, hhnum, samp\_stratum, wgt\_c2, v521, v522, v526, v527, v529, v531, v532, v533, v535, v536, v537, v539-v559* |
| Analytic variables used | *c0\_5m, hhmem\_df* |
| Analytic variables created | *bf, water, othermilk, nonmilk, food, chn\_fmiss, chn\_ebf* |

\*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 (*bf*). This includes being breastfed by the mother; receiving breastmilk in a spoon, cup, or bottle; or being breastfed by another woman.

*Set bf=missing*

*Replace bf=0 if c0\_5m=1*

*Replace bf=1 if (v521=1 or v522=1) and c0\_5m=1Label values 0 “No” 1 “Yes”*

*Label variable “Child consumed breastmilk”*

**Step 2.** Create a binary variable that flags children 0-5 months of age who received any milk other than breastmilk (*othermilk*). This includes formula; milk, such as tinned, powdered, or fresh animal milk; or yogurt.

*Set othermilk=missing*

*Replace othermilk=0 if c0\_5m=1*

*Replace othermilk=1 if (v527=1 or v529=1 or v533=1) and c0\_5m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child consumed other milk (non-breastmilk)”*

**Step 3.** Create a binary variable that flags children 0-5 months of age who received plain water.

*Set water=missing*

*Replace water=0 if c0\_5m=1*

*Replace water=1 v526=1 and c0\_5m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child consumed plain water”*

**Step 4.** Create a binary variable that flags children 0-5 months of age who received non-milk liquid (*nonmilk*).Non-milk liquids include juice, juice drinks, porridge, and other liquids, such as glucose water or sugar water.

*Set nonmilk=missing*

*Replace nonmilk=0 if c0\_5m=1*

*Replace nonmilk=1 if (v531=1 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 consumed non-milk liquids”*

**Step 5.** Create a binary variable that flags children 0-5 months of age who received any food (*food*).

*Set food=missing*

*Replace food=0 if c0\_5m=1*

*Replace 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=1or v558=1 or v559=1) and c0\_5m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child consumed any food (non-liquid)”*

**Step 6.** Create a binary variable that flags children 0-5 months of age who are missing all data about feedings in the day and night preceding the survey (*chn\_fmiss*).

**Step 6a.** Recode all variables that indicate whether a child a particular foods so that ‘no’ responses have a value of ‘0’ rather than ‘2’ and so that all ‘don’t know’ (8) and missing (9) responses are set to blank (missing).

*For each variable (var) of variable list* *v526 v527 v529 v531 v532 v533 v535 v536 v537 v539-v560:*

*Set `var’x=`var’*

*Replae `var’x=0 if `var’=2*

*Replace `var’x=missing if `var’=8 or `var’=9*

**Step 6b.** Recreate the *chn\_fmiss* variable by counting the number of recoded food variables created in Step 6a that are missing. If not all food variables are missing—that is, if the value of *chn\_fmiss* is less than 32, recode the *chn\_fmiss* variable to be 0. However, if the value of *chn\_fmiss* is 32, set its value to 1.

*Set chn\_fmiss=missing*

*Replace chn\_fmiss=0 if c0\_23m=1*

*Replace chn\_fmiss=count number of variables in parentheses=missing(v526x v527x v529x v531x v532x v533x v535x v536x v537x v539x-v560x)*

*Replace chn\_fmiss=missing if c6\_23≠1*

*Replace chn\_fmiss=0 if chn\_fmiss<32*

*Replace chn\_fmiss=1 if chn\_fmiss=32*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child 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 (bf=1 and water=0 and othermilk=0 and nonmilk=0 and food=0) and c0\_5m=1*

*Replace chn\_ebf=missing if chn\_fmiss=1 and c0\_5m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child was exclusively breastfed”*

**Step 8.** Create a variable that captures the subpopulation 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 under 2 weight (*wgt\_c2*), calculate the percentage of de facto children 0-5 months of age who were exclusively breastfed using the *chn\_ebf* analytic variable. Repeat using the child sex disaggregate. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_c2], strata(samp\_stratum)*

*Svy, subpop(hhmem\_ c05m\_df): prop chn\_ebf*

*Svy, subpop(hhmem\_ c05m\_df): prop chn\_ebf, over(sex)*

#### References

Debes, A.K., Kohli, A., Walker, K., Edmond, K., and Mullany, L.C. 2013. “Time to initiation of breastfeeding and neonatal mortality and morbidity: a systematic review.” *BMC Public Health*. 13(3):S19. doi: 10.1186/1471-2458-13-S3-S19.

Khan, J., Vesel, L., Bahl, R., and Martines, J.C. 2015. “Timing of breastfeeding initiation and exclusivity of breastfeeding during the first month of life: Effects on neonatal mortality and morbidity—A systematic review and meta-analysis.” *Maternal and Child Health Journal*. 19(3):468-79. doi: 10.1007/s10995-014-1526-8.

Kramer, M.S., and Kakuma, R. 2012. Optimal duration of exclusive breastfeeding. *Cochrane database of systematic reviews*, Issue 8. Art. No.: CD003517. Available at: <https://doi.org/10.1002/14651858.CD003517.pub2>

UNICEF. *Nutrition: Breastfeeding*. Available at: <https://www.unicef.org/nutrition/index_24824.html>

World Health Organization. 2019. *e-Library of Evidence for Nutrition Actions (eLENA).* *Exclusive breastfeeding for optimal growth, development and health of infants.* Available at: who.int/elena/titles/exclusive\_breastfeeding/en/

World Health Organization. 2017. *Global nutrition monitoring framework: Operational guidance for tracking progress in meeting targets for 2025*. Geneva: World Health Organization; 2017. 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/>

### Prevalence 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 age, breastfeeding status, dietary diversity, and frequency of receiving solids and milk or milk products. It is based on information collected in the core ZOI Survey 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.do* file*.*

#### Definitions

| Numerator | Number of de facto children 6-23 months of age who achieve the minimum dietary diversity and meal frequency thresholds established for their age group and breastfeeding status |
| --- | --- |
| Denominator | All de facto children 6-23 months of age |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Child under 2 years of age |
| Disaggregation levels | Child sex\*  Child age category (6-11, 12-17, 18-23 months)  Caregiver education  Gendered household type (de facto household member)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index |
| Coverage | P1-ZOI and P2-ZOI |
| 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. They are included in the denominator but not in the numerator. Children missing **all** data regarding feedings in the past 24 hours will be 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, samp\_stratum, wgt\_c5, v521, v522, v527, v529, v533, v535, v539-v559* |
| Analytic variables used | *c6\_23m, chn\_fmiss* |
| Analytic variables created | *bf\_gr1, bf\_grp2, bf\_grp3, foodgrp1-foodgrp7, mdd\_nbf, mdd\_bf, mindietdiv, mfreq\_milk, mfreq\_milkplus, minmfreq, chn\_mad* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create variables that categorize children 6-23 months of age by age and breastfeeding status.

**Step 1a.** Create a variable that flags breastfed children 6-8 months of age (bf\_grp1).

*Set bf\_grp1=0 if c6\_8m=1*

*Replace bf\_grp1=1 if (v521=1 or v522=1) and c6\_8m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child 6-8 months, breastfeeds”*

**Step 1b.** Create a variable that flags breastfed children 9-23 months of age (bf\_grp2).

*Set bg\_grp2=0 if c9\_23m=1*

*Replace bf\_grp2=1 if (v521=1 or v522=1) and c9\_23m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child 9-23 months, breastfeeds”*

**Step 1c.** Create a variable that flags non-breastfed children 6-23 months of age (bf\_grp3).

*Set bf\_grp3=0 if c6\_23=1*

*Replace bf\_grp3=1 if ((v521=2 and v522=2) or (v521=2 and v522=2)) and c6\_23m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child 6-23 months, does not breastfeed”*

**Step 2.** Create variables for the different food groups.

Identify the variables associated with the seven food groups used to calculate children’s dietary diversity score. 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 18 lists the food groups and their corresponding variables in the ZOI Survey core questionnaire. Local adaptation may also have resulted in changes in the numbering of the questions in the questionnaire, so the variable numbering in Table 18 should also be reviewed for any needed adjustment.

**Table 18: Seven Food Groups Used to Construct Dietary Diversity for Breastfed Children**

| **Food Group** | **Variables** |
| --- | --- |
| Grains, roots, and tubers[[98]](#footnote-97) | *v535*=porridge  *v539*=grains  *v541*=roots or tubers  *v541a*=plantain or green banana |
| Legumes and nuts | *v552*=legumes  *v553*=nuts |
| Dairy products | *v527*=formula  *v529*=milk  *v533*=yogurt  *v554*=dairy products (milk, yogurt, cheese) |
| 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  *v558*=grubs, snails, or insects |
| Eggs | *v550*=eggs |
| 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 |
| Other fruits and vegetables | *v543*=any other vegetables  *v545*=any other fruits |

Create a binary variable for each of the seven food groups (*foodgrp1-foodgrp7)* used to calculate the minimum dietary diversity score.

*Set foodgrp1=0*

*Replace foodgrp1=1 if v535=1 or v539=1 or v541=1 or v541a=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate grains, roots, tubers”*

*Set foodgrp2=0*

*Replace foodgrp2=1 if v552=1 or v553=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate legumes, nuts”*

*Set foodgrp3=0*

*Replace foodgrp3=1 if v527=1 or v529=1 or v533=1 or v554=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate dairy products”*

*Set foodgrp4=0*

*Replace foodgrp4=1 if v546=1 or v547=1 or v548=1 or v549=1 or v551=1or v558=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate flesh foods”*

*Set foodgrp5=0*

*Replace foodgrp5=1 if v550=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate eggs”*

*Set foodgrp6=0*

*Replace foodgrp6=1 if v540=1 or v542=1 or v544=1 or v559=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate vitamin A-rich fruits and vegetables”*

*Set foodgrp7=0*

*Replace foodgrp7=1 if v543=1 or v545=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate other fruits and vegetables”*

**Step 3.** Create variables that flag children 6-23 months of age that meet the minimum dietary diversity threshold based on their breastfeeding status.

**Step 3a.** Create a binary variable that flags breastfed children who meet the minimum threshold for dietary diversity (*mdd\_bf*)—that is, they consumed foods from at least four of the seven specified food groups.

*Set mdd\_bf=0 if c6\_23m=1*

*Replace mdd\_bf=1 if (foodgrp1+foodgrp2+foodgrp3+foodgrp4+foodgrp5+foodgrp6+foodgrp7)≥4 and c6\_23m=1*

*Replace mdd\_bf=missing if (bf\_grp1≠1 and bf\_grp2≠1) or c6\_23m≠1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate foods from 4+ food groups, breastfed”*

**Step 3b.** Create a binary variable that flags non-breastfed children who meet the minimum threshold for dietary diversity (*mdd\_nbf*)—that is, they consumed foods from at least four of the six specified food groups.(For non-breastfed children, do not include dairy products, *foodgrp3*, in this count.)

*Set mdd\_nbf=0 if c6\_23m=1*

*Replace mdd\_nbf=1 if (foodgrp1+foodgrp2+ foodgrp4+foodgrp5+foodgrp6+foodgrp7)≥4 and c6\_23m=1*

*Replace mdd\_nbf=missing if bf\_grp3≠1 or c6\_23m≠1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child ate foods from 4+ food groups, non-breastfed”*

**Step 3c.** Create a binary variable that flags children 6-23 months of age who meet the criteria for minimum dietary diversity based on their breastfeeding status (*mindietdiv*).[[99]](#footnote-98)

*Set mindietdiv=0 if c6-23m=1*

*Replace mindietdiv=1 if (mdd\_bf=1 or mdd\_nbf=1) and c6-23m=1*

*Replace mindietdiv=missing if chn\_fmiss=1 and c6-23m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child meets minimum dietary diversity criteria”*

**Step 4.** Create variables for meal frequency among children 6-23 months of age.

**Step 4a.** Prepare the variables. Because values will be summed across variables to get a total number of feedings, reassign the numeric values assigned to missing and “don’t know” responses so that they are not included in the total. For *v528*, *v530*, and *v534*, recode “don’t know” responses (98) and missing responses (99) to zero (0) under the assumption that children had no dairy feedings. 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 v524x=0 if v534=98 or v534=99*

*Set v563x=0 if v562=2 or v562=8 or v562=9 or v563=98 or v653=99*

**Step 4b.** Create a variable that counts the total number of milk feeds (*mfreq\_milk*).

*Set mfreq\_milk=0*

*Replace mfreq\_milk=v528x+v530x+v534x*

*Label variable “Number of milk feeds day and night before survey”*

**Step 4c.** Create a variable that counts the total number of feeds, including milk-feeds plus soft, solid, or semi-solid food feeds (*mfreq\_milkplus*)

*Set mfreq\_milkplus=0*

*Replace mfreq\_milkplus=v528x+v530x+v534x+v563x*

*Label variable “Number of feeds (milk+food) day and night before survey”*

**Step 4d.** Create a binary variable that flags children 6-23 months of age who received the minimum meal frequency or more based on their age and breastfeeding status (*minmfreq*).

*Set minmfreq=0 if c6\_23m=1*

*Replace minmfreq=1 if bf\_grp1=1 and (v563≥2 and v563≠missing) [breastfed children ages 6-8 months]*

*Replace minmfreq=1 if bf\_grp2=1 and (v563≥3 and v563≠missing) [breastfed children ages 9-23 months]*

*Replace minmfreq=1 if bf\_grp3=1 and (mfreq\_milk≥2 and mfreq\_milk≠missing) and mfreq\_milkplus≥4 and mfreq\_milkplus≠missing) [non-breastfed children]*

*Replace minmfreq=missing if chn\_fmiss=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child meets minimum meal frequency criteria”*

**Step 5.** 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 mindietdiv=1 and minmfreq=1 and c6\_23=1*

*Replace chn\_mad=missing if chn\_fmiss=1 and c6\_23m=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Child meets minimum acceptable diet criteria”*

**Step 6.** Create a variable that captures the breastfeeding status of children 6-23 months of age (*bf\_stat*).

*Set bf\_stat=missing*

*Replace bf\_stat=0 if c6\_23=1*

*Replace bf\_stat=1 if c6\_23=1 and (bf\_grp1=1 or bf\_grp2=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable "Child is currently breastfeeding"*

**Step 7.** Create a variable that captures the subpopulation 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 defacto household member”*

**Step 8.** After applying the children under 2 weight (*wgt\_c2*), calculate the percentage of de facto children 6-23 months of age who received the minimum acceptable diet using the *chn\_mad* analytic variable. Repeat using the child sex disaggregate. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_c2], strata(samp\_stratum)*

*Svy, subpop(hhmem\_ c623m\_df): prop chn\_mad*

*Svy, subpop(hhmem\_ c623m\_df): prop chn\_mad, over(sex)*

#### References

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. 2017. *Global nutrition monitoring framework: Operational guidance for tracking progress in meeting targets for 2025*. Geneva: World Health Organization. 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>

### 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. It is based on anthropometry measurements collected in the core ZOI Survey Module 4A, *Women’s Anthropometry.* The step-by-step procedures to calculate the indicator follow the Stata syntax in the *FTF ZOI Survey [COUNTRY] [YEAR] syntax nut\_WHN.do* file*.*

#### Definitions

| Numerator | Number of de facto non-pregnant women of reproductive age (15-49 years) who have a BMI less than 18.5 |
| --- | --- |
| Denominator | Number of de facto non-pregnant women of reproductive age |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Women of reproductive age |
| Disaggregation levels | Age category (15-18, 19-49)\*  5-year age categories  Education  Gendered household type (de facto household member)  Household food insecurity  Wealth quintile (P2-ZOI only)  Poverty status  Shock exposure index |
| Coverage | P1-ZOI and P2-ZOI |
| 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 | *hhea, hhnum, samp\_stratum, wgt\_w, v402, an405,* *an406, an407* |
| Analytic variables used | *wra, wra\_cage* |
| Analytic variables created\* | *height\_m, weight\_m, whn\_bmi, whn\_bmi2, whn\_uw* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Create a continuous variable that indicates women’s height in centimeters converted to meters (*height\_m*).

*Set height\_m=missing*

*Replace height\_m=an406÷100*

*Replace height\_m=missing if height\_m=an406≥999.4*

*Label variable “Woman’s height (meters)”*

**Step 2.** Create a continuous variable that indicates women’s weight in kilograms (w*eight\_m*).

*Set weight\_m=missing*

*Replace weight\_m=an407*

*Replace weight\_m=missing if weight\_m=an407≥999.4*

*Label variable “Woman’s weight (kg)”*

**Step 3.** Create a continuous variable that indicates the BMI of non-pregnant women of reproductive age (*whn\_bmi*) by dividing their weight in kilograms by the square of their height in meters.

*Set whn\_bmi=missing*

*Replace whn\_bmi=weight\_m÷(height\_m)2 if wra=1 and (an405=2 or an405=8)*

*Label variable “Woman’s BMI (kg/m^2)”*

**Step 4.** Create a variable that excludes women with implausible BMI values—that is, women whose BMI is less than 12 or greater than 60 (*whn\_bmi2*).

*Set whn\_bmi2=missing*

*Replace whn\_bmi2=whn\_bmi if wra=1 and (an405=2 or an405=8) and whn\_bmi≥12 and whn\_bmi≤60*

*Label variable “Women’s BMI (kg/m^2), excluding values <12 & >60”*

**Step 5.** Create a variable that indicates whether women are underweight according to their BMI (*whn\_uw*).

*Set whn\_uw=0 if whn\_bmi≠missing*

*Replace whn\_uw=1 if whn\_bmi2<18.5*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman is underweight (BMI<18.5)”*

**Step 6.** After applying the women of reproductive age weight (*wgt\_w*), calculate the percentage of de facto women of reproductive age who are underweight using the *whn\_uw* analytic variable. Repeat using the age category disaggregate. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_w], strata(samp\_stratum)*

*Svy, subpop(hhmem\_df): prop whn\_uw*

*Svy, subpop(hhmem\_df): prop whn\_uw, over(wra\_cage)*

#### 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.htm#t=Guide_to_DHS_Statistics_DHS-7.htm>

World Health Organization. 2016. *Factsheet: Obesity and overweight*. Available at: <http://www.who.int/mediacentre/factsheets/fs311/en/>

World Health Organization Regional Office for Europe. 2019. *Body mass index—BMI*. Available at:<http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi>

### Women’s Dietary Diversity: Mean number of food groups consumed by women of reproductive age

This indicator estimates the average number of food groups, out of nine specified food groups, consumed by women of reproductive age (15-49 years) during the day and night preceding the survey in the ZOI population. 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.do* file*.*

#### Definitions

| Numerator | This is a quasi-continuous score indicator, so there is no numerator or denominator. The estimate is the mean number of food groups, out of nine specified food groups, consumed by de facto women of reproductive age. |
| --- | --- |
| Denominator |
| Unit of measure | Mean |
| Level of data | Individual |
| Sampling weight | Women of reproductive age |
| Disaggregation levels | None\* |
| Coverage | P1-ZOI |
| 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 **all** 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, samp\_stratum, wgt\_w, v409-v429* |
| Analytic variables used | *hhmem\_df, wra\_cage* |
| Analytic variables created | *wfoodgrp1-wfoodgrp9, whn\_wdds, whn\_fmiss* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Identify variables associated with the nine food groups used to calculate women’s dietary diversity score. Pay particular attention if the questionnaire was adapted to include local foods to ensure that they are assigned to the appropriate food group. Table 19 lists the food groups and their corresponding variables in the ZOI Survey core questionnaire. Local adaptation may also have resulted in changes in the numbering of the questions in the questionnaire, so the variable numbering in Table 19 should also be reviewed for any needed adjustment.

**Table 19: Nine Food Groups Used to Generate Women’s Dietary Diversity Score**

| **Food Group** | **Variables** |
| --- | --- |
| Grains, roots, and tubers | *v409*=grains  *v411*=roots and tubers  *v411a*=plantain or banana |
| Legumes and nuts | *v422*=legumes  *v423*=nuts or seeds |
| Dairy products | *v424*=dairy products (milk, yogurt, cheese) |
| Organ meat | *v416*=organ meats from domesticated animals  *v418*=organ meats from wild animals |
| Eggs | *v420*=eggs |
| Flesh foods, fish, and other miscellaneous small animal protein | *v417*=any meat from domesticated animals  *v419*=any flesh meat from wild animals  *v421*=fresh or dried fish, shellfish, or seafood  *v428*=grubs, snails, or insects |
| Dark green leafy vegetables | *v412*=any dark green leafy vegetables |
| Other vitamin A-rich fruits and vegetables | *v410*=pumpkin, carrots, squash, or sweet potatoes that are yellow or orange inside  *v414*=ripe mangoes or papayas  *v429*=foods made with red palm oil, red palm nuts, or red palm nut pulp sauce |
| Other fruits and vegetables | *v413*=any other vegetables  *v415*=any other fruits |

**Step 2.** Create nine binary food group variables—one for each food group (w*foodgrp1-wfoodgrp9*).

*Set wfoodgrp1=0*

*Replace wfoodgrp1=1 if (v409=1 or v411=1 or v411a=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate grains, roots, tubers”*

*Set wfoodgrp2=0*

*Replace wfoodgrp2=1 if (v422=1 or v423=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate legumes, nuts”*

*Set wfoodgrp3=0*

*Replace wfoodgrp3=1 if (v424=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate dairy products”*

*Set wfoodgrp4=0*

*Replace wfoodgrp4=1 if (v416=1 or v418=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate organ meat”*

*Set wfoodgrp5=0*

*Replace wfoodgrp5=1 if (v420=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate eggs”*

*Set wfoodgrp6=0*

*Replace wfoodgrp6=1 if (v417=1 or v419=1 or v421=1 or v428=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate flesh foods”*

*Set wfoodgrp7=0*

*Replace wfoodgrp7=1 if (v412=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate dark green leafy vegetables”*

*Set wfoodgrp8=0*

*Replace wfoodgrp8=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 and vegetables”*

*Set wfoodgrp9=0*

*Replace wfoodgrp9=1 if (v413=1 or v415=1)*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate other fruits and vegetables”*

**Step 3.** Create a binary variable that flags women of reproductive age who are missing all data about food that they consumed the day and night preceding the survey (*whn\_fmiss*).

*Set whn\_fmiss=missing*

*Replace whn\_fmiss=1 if wra=1 and v409=missing and v410=missing and v411a=missing and v412=missing and v414=missing and v415=missing and v416=missing and v417=missing and v418=missing abd v419=missing and v420=missing and v421=missing and v422=missing and v423=missing and v424=missing and v428=missing and v429=missing*

*Label variable “Woman is missing all food data”*

**Step 4.** Calculate women’s dietary diversity score (*wdds*) by summing the number of food groups consumed during the day and night preceding the survey.

*Set whn\_wdds=missing*

*Replace whn\_wdds=0 if wra=1*

*Replace whn\_wdds=(wfoodgrp1+wfoodgrp2+wfoodgrp3+………..+wfoodgrp9) and wra=1*

*Replace whn\_wdds=missing if whn\_fmiss=1 and wra=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman’s dietary diversity score”*

**Step 5.** After applying the women of reproductive age weight (*wgt\_w*), calculate the mean dietary diversity score for de facto women of reproductive age using the *when\_wdds* analytic variable. Repeat using the age category disaggregate. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_w], strata(samp\_stratum)*

*Svy, subpop(hhmem\_df): mean whn\_wdds*

*Svy, subpop(hhmem\_df): mean whn\_wdds, over(wra\_cage)*

#### 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>

Kennedy, G., Ballard, T., and 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

### 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.do* file*.*

#### Definitions

| Numerator | Number of de facto women 15-49 years of age who consumed at least 5 of 10 specified food groups during the day and night preceding the survey |
| --- | --- |
| Denominator | Number of de facto women 15-49 years of age |
| Unit of measure | Percentage |
| Level of data | Individual |
| Sampling weight | Women of reproductive age |
| Disaggregation levels | Age category (15-18, 19-49)\*  5-year age category  Education  Pregnancy status  Gendered household type (de facto household member)  Household food insecurity  Wealth quintile  Poverty status  Shock exposure index  Household ownership of agricultural land  Household ownership of livestock  Household production of targeted VCCs |
| Coverage | P2-ZOI |
| 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 **all** 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, samp\_stratum, wgt\_w,* *v409-v429* |
| Analytic variables used | *mddfgrp1-mddfgrp10, hhmem\_df, wra\_cage, whn\_fmiss* |
| Analytic variables created | *whn\_mdd\_w* |

\*Standard Feed the Future disaggregate

#### Calculations

**Step 1.** Identify the variables associated with the 10 food groups used to generate women’s food score. Pay particular attention if the questionnaire was adapted to include local foods. Table 20 lists the food groups and their corresponding variables in the ZOI Survey core questionnaire.

**Table 20: Ten Food Groups Used to Generate Women’s Food Score**

| **Food Group** | **Variables** |
| --- | --- |
| Grains, white roots, tubers[[100]](#footnote-99) | *v409*=grains  *v411*=roots/tubers  *v411a*=plantain or banana |
| Pulses | *v422*=beans, peas, lentils, legumes |
| Nuts and seeds (including groundnuts) | *v423*=nuts or seeds |
| Dairy products | *v424*=dairy products (milk, cheese, yogurt) |
| 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  *v428*=grubs, snails, or insects |
| Eggs | *v420*=eggs |
| Dark green leafy vegetables | *v412*=any dark green leafy vegetables |
| 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 |
| Other vegetables | *v413*=any other vegetables |
| Other fruits | *v415*=any other fruits |

**Step 2.** Create 10 binary variables, one for each food group variable (*mddfgrp1-mddfgrp10#*).

*Set mddfgrp1=0*

*Replace mddfgrp1=1 if v409=1 or v411=1 or v411a=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate grains, white roots, tubers”*

*Set mddfgrp2=0*

*Replace mddfgrp2=1 if v422=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate pulses”*

*Set mddfgrp3=0*

*Replace mddfgrp3=1 if v423=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate nuts, seeds”*

*Set mddfgrp4=0 if v400e=1*

*Replace mddfgrp4=1 if v424=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate dairy products”*

*Set mddfgrp5=0*

*Replace mddfgrp5=1 if v416=1 or v417=1 or v418=1 or v419=1 or v421=1 or v428=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate meat, poultry, fish”*

*Set mddfgrp6=0*

*Replace mddfgrp6=1 if v420=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate eggs”*

*Set mddfgrp7=0*

*Replace mddfgrp7=1 if v412=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate dark green leafy vegetables”*

*Set mddfgrp8=0*

*Replace mddfgrp8=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 and vegetables”*

*Set mddfgrp9=0*

*Replace mddfgrp9=1 if v413=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate other vegetables”*

*Set mddfgrp10=0*

*Replace mddfgrp10=1 if v415=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman ate other fruits”*

**Step 3.** Calculate the minimum dietary diversity food score (*mdd\_wfscore*) by summing the number of food groups consumed during the day and night preceding the survey.

*Set mdd\_wfscore=0*

*Replace mdd\_wfscore=(mddfgrp1+ mddfgrp2+ mddfgrp3+………..+mddfgrp9+mddfgrp10)*

*Label variable “Woman’s minimum dietary diversity food score”*

**Step 4.** Create a binary variable that flags a woman who achieved a minimum dietary diversity (whn\_*mdd*).[[101]](#footnote-100)

*Set whn\_mdd\_w=missing*

*Replace whn\_mdd\_w=0 if wra=1*

*Replace whn\_mdd\_w=1 if (mdd\_wfscore≥5 and mdd\_wfscore≠missing) and wra=1*

*Replace whn\_wdd\_w=missing if whn\_fmiss=1 and wra=1*

*Label values 0 “No” 1 “Yes”*

*Label variable “Woman achieved minimum dietary diversity”*

**Step 5.** After applying the women of reproductive age weight (*wgt\_w*), calculate the percentage of de facto women of reproductive age who achieved the minimum diet diversity score using the *whn\_mdd* analytic variable. Repeat using the age category disaggregate. (Sample code uses Stata syntax.)

*Svyset hhea [pweight=wgt\_w], strata(samp\_stratum)*

*Svy, subpop(hhmem\_df): prop whn\_mdd\_w*

*Svy, subpop(hhmem\_df): mean whn\_mdd\_w, over(wra\_cage)*

#### 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 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>

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., and 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/>

# Appendix A. A-WEAI summary information

**Table A1: Summary of A-WEAI Domains, Indicators, Survey Questions, Variables, Definitions, and Weights**

| Domain | Indicator Name | Survey Questions | ZOI Survey (2018-2019)  Questions | ZOI Survey (2018-2019)  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?”    “How much input did you have in making decisions about food crop farming, cash crop farming, livestock raising, and fishing or fishpond culture?”  “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?” | Q.6202 (a, b, c, f)  Q.6203 (a, b, c, f)  Q.6204 (a, b, c, f) | v6202\_1-v6202\_3,  v6202\_6  v6202\_1-v6202\_3,  v6202\_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 amedium or high extent (93 ‘no decision made’ coded as missing) | 1/5 |
| Access to resources | Indicator 2.1: Ownership of assets | “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”  “Do you own any of the item either by yourself or jointly with someone else?” | Q.6301a–Q.6301n  Q.6303a–Q.6303n | v6301\_01-v6301\_15  v6303\_01-v6303\_15 | 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 | “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.”  “Who made the decision to borrow from [SOURCE]?”  “Who makes the decision about what to do with the money/item borrowed from [SOURCE]?” | Q.6308a–Q.6308f  Q.6309a–Q.6309f  Q.6310a–Q.6310f | v6308\_1-v6303\_6  v6309\_1-v6309\_6  v6310\_1-v6310\_6 | Can alone or jointly make at least one decision regarding at least one source of credit | Household has no credit; OR household hascredit 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?”  “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” | Q.6205a–Q.6205f  Q.6204d,  Q.6204e,  Q.6204g,  Q.6204h | v6205\_1- v6206\_6,  v6204\_4,  v6204\_5, v6204\_7,  v6204\_8 | Has input into some; or most or all decisions on use of income for at least one productive/ 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 | Participates in activity but has no input in decisions about income, OR feels she or he has no 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 | 1/5 |
| 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?” | Q.6405a–Q.6405k | 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. | Q.6601 | 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. A-WEAI confidence interval and design effect calculations

## B1. Background and guidelines

This section provides guidelines for calculating the 95 percent confidence interval (CI) and design effect (DEFF) for the abbreviated Women’s Empowerment in Agriculture Index (A-WEAI) and A-WEAI context indicators. Due to the complicated nature of the non-linear A-WEAI and A-WEAI context indicators, non-standard methods must be used to compute the CI and DEFF. Section B1.1 lists the steps for CI estimation, and Section B1.2 lists the steps for DEFF estimation.

### B1.1. The confidence interval estimation

To compute the CI for the A-WEAI and A-WEAI context indicators, 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 (A-WEAI or A-WEAI context) can be calculated as follows:

,



in which



where is the estimate computed from the full sample of *C* clusters and

is the estimate computed from the reduced sample of (*C* – 1) clusters (*c*th cluster excluded).

To estimate CI for both A-WEAI and A-WEAI context indicators, the algorithm for the JRR method follows a sequence of six steps:

1. The indicator of interest, A-WEAI or A-WEAI context, is estimated based on the full sample as .
2. A total of *C* sample replicates are defined, where replicate *c* is a reduced sample of (*C* – 1) clusters where *c*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 , where is the summation of the survey weight for all units within stratum *h* and is the same summation but after cluster *c* is excluded.
4. From each sample replicate *c*, the indicator of interest, A-WEAI or A-WEAI context, is calculated as 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% Confidence intervals are estimated based on the estimated indicators and standard error as .

### B1.2. The Design effect estimation

Due to the reasons mentioned earlier, there’s not a constructed method to calculate the DEFF for the A-WEAI and A-WEAI context indicators. Because in both indicators, sets of linear variables are used in the computation, the average of DEFFs for these variables can be used as approximates for DEFFs as follows:

1. DEFF for the A-WEAI: calculated as the average of DEFFs for the four variables that are involved in the calculation of the A-WEAI main components (5DE and GPI). The four variables are ch\_20p, a\_20p, H\_GPI and ci\_average.
2. DEFF for the A-WEAI context: calculated as the average of DEFFs for the six A-WEAI indicators of achievement of adequacy.

### B1.3. Differences of indicators and test of significance

The A-WEAI and A-WEAI context indicators can be disaggregated, for instance, by women’s age category, comparing the estimated indicator for women under 30 years of age and women 30 years of age or older. Hypothesis testing can be used to determine whether the difference between the estimated indicators for women under 30 years of age and women 30 years of age or older is statistically significant or not.

Assuming that both the A-WEAI and A-WEAI context indicators follow the standard normal distribution, let denote the estimate computed for group 1, and denote the estimate computed for group 2. To test the null hypothesis, *No*: *1*=*2*, versus the alternative hypothesis, *No*: *1*≠*2*, the test statistic (Z-score) can be calculated as follows:

where SE1 is the standard error for

SE2 is the standard error for

Using the standard normal tables, the associated two-tailed p-value of the Z-score can be found as 2\**P(Z>)* where is the critical Z value for the chosen α level significance (e.g., 1.96 for the 95 percent confidence level and 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.

### 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>

## B2. Step-by-step procedures

### B2.1 A-WEAI indicator with standard error estimation

The step-by-step instructions to calculate the confidence intervals and design effect (DEFF) for the Feed the Future A-WEAI context indicator follow the steps in the “FTF ZOI Survey [COUNTRY] [YEAR] aweai\_calc\_JK” do file. The jackknife variance estimation depends on a second do file, “GPI\_fun.do,” which is called by the first do file. “GPI\_fun.do” defines the JKGPI function that calculates the Gender Parity Index (GPI) indicator within the jackknife iterations.

#### Five domains of empowerment (5DE)

**Step 1.** Load the Abbreviated Women’s Empowerment in Agriculture Index (AWEAI) individual indices data file, “aweai\_individual\_indices.dta,” created in the A-WEAI calculation do file (the most recent version of “FTF ZOI Survey [COUNTRY] [YEAR] syntax AWEAI\_2 calc.do”).

*Load “aweai\_individual\_indices.dta”*

**Step 2.** If not already created, create a variable (*age\_cat*) that indicates whether the primary adult decisionmaker is under 30 years of age or 30 years of age or older. If an individual is missing a self-reported age in the A-WEAI module, use the age information available in the household roster.

*Set age\_cat=missing*

*Replace age\_cat=1 if v6102<30*

*Replace age\_cat=2 if v6102≥30 and v6102<98*

*Replace age\_cat=1 if youth\_fdm\_dj=1 and age\_cat=missing and v6\_sex=2*

*Replace age\_cat=2 if youth\_fdm\_dj=0 and age\_cat=missing and v6\_sex=2*

*Replace age\_cat=1 if youth\_mdm\_dj=1 and age\_cat=missing and v6\_sex=1*

*Replace age\_cat=2 if youth\_mdm\_dj=0 and age\_cat=missing and v6\_sex=1*

*Label values 1 “< 30yo” 2 “30+ yo”*

*Label variable “Age category: <30 or 30+”*

**Step 3.** Create a sequential cluster variable (*id*) using the original cluster variable (*hhea*). The variable will range from 1 to the number of clusters in the analysis (N). First, sort the data by cluster number, lowest to highest. Then assign *id* for the first cluster to be 1, the second cluster to be 2, the third cluster to be 3, etc., until all clusters have a value.

*Sort by hhea*

*Set id=1 for first hhea*

*Replace id=2 for second hhea*

*Replace id=3 for third hhea*

*[Repeat for all clusters]*

*Replace id=N for Nth (final) hhea*

*Label variable “Sequential cluster ID variable”*

**Step 4.** Create total weight variables for sub-populations for which the indicator is being calculated.

**Step 4a.** Create a total weight variable (*alloc*) with strata and sex-specific values.

*Sort by samp\_stratum v6\_sex: set alloc=total(weight)*

**Step 4b.** Create a total weight variable (*alloc\_AGE*) with strata, sex, and age-specific values.

*Sort by samp\_stratum v6\_sex age\_cat: set alloc\_AGE=total(weight)*

**Step 5.** Drop records for primary adult male decisionmakers.

*Drop if v6\_sex=1*

**Step 6.** Save the data to a temporary data file, “AWEAI\_DEPR.”

*Save “Temp\AWEAI\_DEPR”*

**Step 7.** Specify the survey design—that is, the cluster, weight, and strata variables.

*Svyset hhea [pw=weight], strata(samp\_stratum)*

**Step 8.** Create variables for the ch\_20p and a\_20p sample-weighted means and their corresponding DEFFs for all women not missing values for any of the six A-WEAI indicators. The variables are created for three groups: (0) all women, (1) women under 30 years of age, and (2) women 30 years of age or older. Save the sample-weighted mean estimates in matrix ‘a’ and the DEFF estimates in matrix ‘d’ and then assign the relevant matrix components to the variables. First, create variables for the sample-weighted mean and DEFF for all women, and then create variables for the sample-weighted mean and DEFF for women under 30 years of age and women 30 years of age or older.

*For each value (x) of varlist ch\_20p a\_20p:*

*Svy: mean `x' if miss\_any=0*

*Estimate deff*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set mean\_0\_`x'=a[1,1]*

*Set deff\_0\_`x'=d[1,1]*

*Svy: mean `x' if miss\_any=0, over(age\_cat)*

*Estat effects*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set mean\_1\_`x'=a[1,1]*

*Set mean\_2\_`x'=a[1,2]*

*Set deff\_1\_`x'=d[1,1]*

*Set deff\_2\_`x'=d[1,2]*

**Step 9.** Because the values of the mean and DEFF variables are the same for all women, collapse the data so that the values for each of the three groups are contained in a dataset with a single row.

*Collapse mean mean\_0\_ch\_20p mean\_1\_ch\_20p mean\_2\_ch\_20p*

*mean\_0\_a\_20p mean\_1\_a\_20p mean\_2\_a\_20p*

*deff\_0\_ch20p deff\_1\_ch20p deff\_2\_ch20p*

*deff\_0\_a20p deff\_1\_a20p deff\_1\_a20p*

**Step 10.** Create variables that capture the 5DE score for each group (*EA\_20p0*, *EA\_20p1*, and *EA\_20p2*) by multiplying the mean ch\_20p and the mean a\_20p and subtracting the result from 1.

*Set EA\_20p0=1-(mean\_0\_ch\_20p\*mean\_0\_a\_20p)*

*Set EA\_20p1=1-(mean\_1\_ch\_20p\*mean\_1\_a\_20p)*

*Set EA\_20p2=1-(mean\_2\_ch\_20p\*mean\_2\_a\_20p)*

**Step 11.** Create variables that capture the 5DE score DEFF for each group by averaging the ch\_20p and a\_20p DEFFs.

*Set deff0=(deff\_0\_ch\_20p+deff\_0\_a\_20p)÷2*

*Set deff1=(deff\_1\_ch\_20p+deff\_1\_a\_20p)÷2*

*Set deff2=(deff\_2\_ch\_20p+deff\_2\_a\_20p)÷2*

**Step 12.** Create a binary variable (*ID*) equal to one and then change the data format from household level to item level—that is, from flat to rectangular format—so that there are three records, one for each category: all women, women under 30 years of age, and women 30 years of age or older. (See procedures outlined in Section 9.2.1, Step 1.)

*Set ID=1*

*Reshape long EA\_20p deff, i(ID) j(age\_cat)*

**Step 13.** Keep only the *EA\_20p*, *deff*, and *age\_cat* variables.

*Keep EA\_20p deff age\_cat*

**Step 14.** Initialize a counter variable (*replicate*) and initialize it to 0.

*Set replicate=0*

**Step 15.** Save the data to a temporary data file, “5DE\_JK.”

*Save “Temp\5DE\_JK”*

**Step 16.** Create a scalar variable (*PS2*) equal to the number of sampled clusters in the indicator calculation.

*Scalar PS2=[number of clusters]*

**Step 17.** Perform the jackknife computation for the 5DE using the loop in this step. Starting with a loop iteration variable (*i*) equal to 1, the loop:

* Loads in the “AWEAI\_DEPR” temporary data file created in Step 6
* Drops all records in the cluster that have an *id* value equal to the value of the loop iteration variable
* Creates a new total weight variable (*allocj)* with strata and sex-specific values and a new total weight variable (*allocj\_AGE)* with strata, sex, and age-specific valuesthat exclude records in the dropped cluster
* Creates adjusted weight variables by sex (*wtadj*) and by sex and age (*wtadj\_AGE*)
* Saves the data to a temporary data file, “AWEAI\_DEPR\_`i',” where i is the loop iteration
* Creates a dataset with one row in which the adjusted sample weight (*wtadj*) is used to generate ch\_20p and a\_20p values for all women
* Creates an age\_cat variable with a value equal to 0 and saves the data to a temporary data file, “NATIONAL\_`i',” where `i' is the loop iteration
* Loads the temporary data file, “AWEAI\_DEPR\_`i'
* Creates a dataset with one row in which the adjusted sample weight (*wtadj\_AGE*) is used to generate ch\_20p and a\_20p values for women by age category, and appends the data to a temporary data file, “NATIONAL\_`i',” where `i' is the loop iteration
* Creates a variable for the 5DE (*EA\_20p*), creates a variable (replicate) equal to the loop iterations, and drops the ch\_20p and a\_20p variables
* Appends the data to the temporary data file, “5DE\_JK,” saves the data file, erases the temporary data files created earlier in the loop iteration
* Adds 1 to the loop iteration variable

*Local i=1*

*While `i'≤PS2:*

*Load “Temp\AWEAI\_DEPR”*

*Drop if id=`i'*

*Sort by samp\_stratum: set allocj=total(weight)*

*Sort by samp\_stratum age\_cat: set allocj\_AGE=total(weight)*

*Set wtadj=weight\*(alloc÷allocj)*

*Set wtadj\_AGE=weight\*(alloc\_AGE÷allocj\_AGE)*

*Save “Temp\AWEAI\_DEPR\_`i'”*

*Collapse ch\_20p a\_20p [aw=wtadj]*

*Set age\_cat=0*

*Save “Temp\NATIONAL\_`i'”*

*Load “Temp\AWEAI\_DEPR\_`i'”*

*Collapse ch\_20p a\_20p [aw=wtadj\_AGE],by(age\_cat)*

*Append using “Temp\NATIONAL\_`i'”, force*

*Set EA\_20p=1-(ch\_20p\*a\_20p)*

*Keep age\_cat EA\_20p*

*Set replicate=`i'*

*Append using “Temp\5DE\_JK,” force*

*Save “Temp\5DE\_JK”*

*Erase “Temp\AWEAI\_DEPR\_`i'”*

*Erase “Temp\NATIONAL\_`i'”*

*Local i= `i' + 1*

**Step 18.** Erase the temporary “AWEAI\_DEPR” data file.

*Erase “Temp\AWEAI\_DEPR”*

**Step 19.** Change the data format from item level to primary decisionmaker level—that is, from rectangular to flat format.

*Reshape wide EA\_20p deff, i(replicate) j(age\_cat)*

**Step 20.** Save the data to a temporary file, “5DE\_JK.”

*Save “Temp\5DE\_JK”*

#### Gender Parity Index

**Step 1.** Load the data file “awei\_depr\_indicators” created in the A-WEAI calc do file, the most recent version of “FTF ZOI Survey [COUNTRY] [YEAR] syntax AWEAI\_2 calc.do.”

*Load “aweai\_depr\_indicators”*

**Step 2.** If not already created, create a variable (*age\_cat*) that indicates whether the primary adult decisionmaker is under 30 years of age or 30 years of age or older. If an individual is missing a self-reported age in the A-WEAI module, use the age information available in the household roster.

*Set age\_cat=missing*

*Replace age\_cat=1 if v6102<30*

*Replace age\_cat=2 if v6102≥30 and v6102<98*

*Replace age\_cat=1 if youth\_fdm\_dj=1 and age\_cat=missing and v6\_sex=2*

*Replace age\_cat=2 if youth\_fdm\_dj=0 and age\_cat=missing and v6\_sex=2*

*Replace age\_cat=1 if youth\_mdm\_dj=1 and age\_cat=missing and v6\_sex=1*

*Replace age\_cat=2 if youth\_mdm\_dj=0 and age\_cat=missing and v6\_sex=1*

*Label values 1 “< 30yo” 2 “30+ yo”*

*Label variable “Age category: <30 or 30+”*

**Step 3.** Create a variable to define the sub-sample that will be used to compute the GPI (*gpisub*), that is—only paired women and men in the same household with complete A-WEAI indicator information.

*Sort by hhea hhnum: set temp=\_N*

*Set gpisub=1 if temp=2*

*Drop temp*

*Keep if gpisub=1*

**Step 4.** Create a sequential cluster variable (*id*) using the cluster variable (*hhea*). The variable will range from 1 to the number of clusters in the analysis. First, sort the data by cluster number, lowest to highest. Then assign *id* for the first cluster to be 1, the second cluster to be 2, the third cluster to be 3, etc., until all clusters have a value.

*Sort by hhea*

*Set id=1 for first hhea*

*Replace id=2 for second hhea*

*Replace id=3 for third hhea*

*[Repeat for all clusters]*

*Replace id=N for Nth (final) hhea*

*Label variable “Sequential cluster ID variable”*

**Step 5.** Create total weight variables for specific sub-populations.

**Step 5a.** Create a total weight variable (*alloc*) with strata and sex-specific values.

*Sort by samp\_stratum: set alloc=total(weight)*

**Step 5b.** Create a total weight variable with (*alloc\_AGE*) strata, sex, and age category-specific values.

*Sort by samp\_stratum age\_cat: set alloc\_AGE=total(weight)*

**Step 6.** Save the data to a temporary data file, “AWEAI\_DEPR.”

*Save “Temp\AWEAI\_DEPR”*

**Step 7.** Create a variable that captures records that have complete information for all six A-WEAI indicators (*sample5do*), and set the variable *ci* to missing if *sample5do* is 0.

*Set sample5do=0*

*Replace sample5do=1 if feelinputdecagr≠missing and jown\_count≠missing and credjanydec\_any≠missing and incdec\_count≠missing and groupmember\_any≠missing and npoor\_z105≠missing*

*Replace ci=missing if sample5do=0*

*Label variable ci “Inadequacy count without parity”*

**Step 8.** Compute the censored inadequacy scores for men (*M\_cen\_ci)* and women (*W\_cen\_ci*) by household.

*Sort by hhea hhnum: set w\_ci\_id=ci if v6\_sex=2*

*Sort by hhea hhnum: set m\_ci\_id=ci if v6\_sex=1*

*Sort by hhea hhnum: set w\_ci=max(w\_ci\_id)*

*Sort by hhea hhnum: set m\_ci=max(m\_ci\_id)*

*Drop w\_ci\_id m\_ci\_id*

*Sort by hhea hhnum: set float W\_cen\_ci=W\_ci*

*Sort by hhea hhnum: replace W\_cen\_ci=0.20 if W\_cen\_ci≤0.20 and W\_cen\_ci≠missing*

*Sort by hhea hhnum: set M\_cen\_ci=M\_ci*

*Sort by hhea hhnum: replace M\_cen\_ci=0.20 if M\_cen\_ci≤0.20 and M\_cen\_ci≠missing*

**Step 9.** Create a variable (*ci\_above*) to identify households without gender parity.

*Sort by hhea hhnum: set ci\_above=(W\_cen\_ci>M\_cen\_ci)*

*Sort by hhea hhnum: replace ci\_above=missing if W\_cen\_ci=missing or M\_cen\_ci≠missing*

*Label var ci\_above “Equals 1 if woman is more disempowered than man in the same HH”*

**Step 10.** Create a binary variable (*female*) equal to 1 for women who can be identified as having a value for *ci\_above*.

*Set female=1 if v6\_sex=2 and ci\_above≠missing*

**Step 11.** Create a binary variable (*inadequate*) equal to 1 for women who are more disempowered than men in the same household.

*Set inadequate=missing*

*Replace inadequate=0 if ci\_above≠missing and v6\_sex=2*

*Replace inadequate=1 if ci\_above=1 and v6\_sex=2*

**Step 12.** Specify the survey design—that is, the cluster, weight, and strata variables—and calculate the weighted headcount ratio of inadequate women and the associated DEFF for all women and by age category. Save the sample-weighted mean estimates in matrix ‘a’ and the DEFF estimates in matrix ‘d’ and then assign the relevant matrix components to the variables. First, create variables for the sample-weighted headcount ratio of inadequate women and DEFF for all women, and then create variables for the sample-weighted headcount ratio of inadequate women and DEFF for women under 30 years of age and women 30 years of age or older.

*Svyset hhea [pw=weight], strata(samp\_stratum)*

**Step 12a.** Calculate the sample-weighted headcount ratio of inadequate women (*H\_GPI0*) and the associated DEFF (*deff\_H\_GPI0*) for all women.

*Svy: ratio inadequate÷female*

*Estimate deff*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set H\_GPI0=a[1,1]*

*Set deff\_H\_GPI0=d[1,1]*

**Step 12b.** Calculate the sample-weighted headcount ratio of inadequate women (*H\_GPI1* and *HGPI2*) and the associated DEFF (*deff\_H\_GPI1* and *deff\_H\_GPI2*) for women by age category.

*Svy: ratio inadequate÷female, over(age\_cat)*

*Estimate deff*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set H\_GPI1=a[1,1]*

*Set deff\_H\_GPI1=d[1,1]*

*Set H\_GPI2=a[1,2]*

*Set deff\_H\_GPI2=d[1,2]*

**Step 13.** Create a variable (*ci\_gap*) equal to the normalized, or average, gender parity gap.

*Set ci\_gap=missing*

*Replace ci\_gap=(W\_cen\_ci-M\_cen\_ci)÷(1-M\_cen\_ci) if ci\_above=1 and v6\_sex=2*

**Step 14.** Calculate the sample-weighted average gender parity gap and the associated DEFF for all women and by age category. Save the sample-weighted mean estimates in matrix ‘a’ and the DEFF estimates in matrix ‘d’ and then assign the relevant matrix components to the variables. First, create variables for the sample-weighted mean and DEFF for all women, and then create variables for the sample-weighted mean and DEFF for women under 30 years of age and women 30 years of age or older.

**Step 14a.** Calculate the sample-weighted average gender parity gap (*ci\_average*0) and the associated DEFF (*deff\_ci\_average0*) for all women.

*Svy: ratio ci\_gap÷inadequate*

*Estimate deff*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set ci\_average0=a[1,1]*

*Set deff\_ci\_average0=d[1,1]*

**Step 14b.** Calculate the weighted average gender parity gap (*ci\_average1* and *ci\_average2*) and the associated DEFF (*deff\_ci\_average1* and *deff\_ci\_average2*)for women by age category.

*Svy: ratio ci\_gap÷inadequate, over(age\_cat)*

*Estimate deff*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set ci\_average1=a[1,1]*

*Set deff\_ci\_average1=d[1,1]*

*Set ci\_average2=a[1,2]*

*Set deff\_ci\_average2=d[1,2]*

**Step 15.** Create a dataset with one row that contains variables for the mean sample-weighted headcount ratio of inadequate women and mean sample-weighted gender parity gap and their associated DEFFs for all women and by age category.

*Collapse (mean) H\_GPI0 H\_GPI1 H\_GPI2 deff\_H\_GPI0 deff\_H\_GPI1 deff\_H\_GPI2 ci\_average0 ci\_average1 ci\_average2 deff\_ci\_average0 deff\_ci\_average1 deff\_ci\_average2*

**Step 16.** Create GPI variables, overall (*GPI0*) and by age category (*GPI1* and *GPI2*), by multiplying the sample-weighted censored headcount ratio of inadequate women by the mean weighted average gender parity gap and subtracting the result from 1.

*Set GPI0=1-(H\_GPI0\*ci\_average0)*

*Set GPI1=1-(H\_GPI1\*ci\_average1)*

*Set GPI2=1-(H\_GPI2\*ci\_average2)*

**Step 17.** Create DEFF variables, overall (*deff0\_*) and by age (*deff1\_* and *deff2\_*), by averaging the DEFFs for the weighted censored headcount ratio of inadequate women and the mean weighted average gender parity gap.

*Set deff0\_=(deff\_H\_GPI0+deff\_ci\_average0)÷2*

*Set deff1\_=(deff\_H\_GPI1+deff\_ci\_average1)÷2*

*Set deff2\_=(deff\_H\_GPI2+deff\_ci\_average2)÷2*

**Step 18.** Initialize a counter (*replicate*) equal to 0.

*Set replicate=0*

**Step 19.** Keep only the *GPI*, *deff*, and *replicate* variables.

*Keep GPI0 GPI1 GPI2 deff0\_ deff1\_ deff2\_ replicate*

**Step 20.** Save the data to a temporary data file, “GPI\_JK.”

*Save “Temp\GPI\_JK”*

**Step 21.** Load the GPI function (JKGPI) to be used in the loop below.

*Load “GPI\_fun.do”*

**Step 22.** Create a scalar variable equal to the number of sample clusters included in the analysis.

*Scalar PS2=[number of sampled clusters in analysis]*

**Step 23.** Perform the jackknife computation for the GPI using the loop in this step. Starting with a loop iteration variable (*i*) equal to 1, the loop:

* Loads the “AWEAI\_DEPR” temporary data file created in Step 6
* Drops all records in the cluster that have an *id* value equal to the value of the loop iteration variable
* Creates a new total weight variable (*allocj)* with strata and sex-specific values and a new total weight variable (*allocj\_AGE)* with strata, sex, and age-specific valuesthat exclude records in the dropped cluster
* Creates adjusted weight variables by sex (*wtadj*) and by sex and age (*wtadj\_AGE*)
* Calls the JKGPI function, which performs Step 7 to Step 16 (excluding the DEFF variables in Steps 12, 14, and 15) using the adjusted weight variables and yields values for the following variables: *GPI0, GPIi, and GPI2* for the dataset, excluding records for the dropped cluster
* Appends the values for the variables to the temporary data file “Temp\GPI\_JK”
* Adds 1 to the loop iteration variable

*Local i = 1*

*While `i'≤PS2:*

*Use “Temp\AWEAI\_DEPR”*

*Drop if id=`i'*

*Sort by strata: set allocj=total(weight)*

*Sort by strata age\_cat: set allocj\_AGE=total(weight)*

*Set wtadj=weight\*(alloc÷allocj)*

*Set wtadj\_AGE=weight\*(alloc\_AGE÷allocj\_AGE)*

*Call the function (JKGPI) in the “GPI\_fun.do” program*

*Set replicate=`i'*

*Keep GPI0 GPI1 GPI2 replicate*

*Append using “Temp\GPI\_JK,” force*

*Save “Temp\GPI\_JK,”*

*Local i=`i'+1*

**Step 24.** Save the data to a temporary data file, “GPI\_JK,” and erase the “AWEAI\_DEPR” temporary data file.

*Save “Temp\GPI\_JK”*

*Erase “Temp\AWEAI\_DEPR”*

#### A-WEAI

**Step 1.** Load the “GPI\_JK” temporary data file and merge it with the “5DE\_JK” temporary data file using the replicate variable (loop iteration variable) to match the records in the two data files. Now the 5DE and GPI score variables for each jackknife iteration are contained in one data file for all women and for women by age category, along with the corresponding DEFF variables.

*Use “Temp\GPI\_JK”*

*Merge using “Temp\5DE\_JK” using replicate as the key matching variable*

**Step 2.** Create A-WEAI variables for all women (*AWEAI0*) and by age category (*AWEAI1* and *AWEAI2*) by multiplying the 5DE variable by 0.9, multiplying the GPI variable by 0.1, and summing the two products.

*Set AWEAI0=(0.9\*EA\_20p0)+(0.1\*GPI0)*

*Set AWEAI1=(0.9\*EA\_20p1)+(0.1\*GPI1)*

*Set AWEAI2=(0.9\*EA\_20p2)+(0.1\*GPI2)*

**Step 3.** Create A-WEAI DEFFs for all women (*deff0*) and by age category (*deff1* and *deff2*) by averaging the 5DE and GPI DEFFs.

*Replace deff0=(deff0+deff0\_)÷2*

*Replace deff1=(deff1+deff1\_)÷2*

*Replace deff2=(deff2+deff2\_)÷2*

**Step 4.** Drop the unneeded DEFF variables and save the data to a temporary data file “GPI\_5DE\_JK.”

*Drop deff0\_ deff1\_ deff2\_*

*Save “Temp\GPI\_5DE\_JK”*

**Step 5.** Create standard error variables for the A-WEAI.

**Step 5a.** Create variables that capture the A-WEAI indicator standard error (SE) for all women (*B\_WEAI0*)and by age category (*B\_WEAI1* and *B\_WEAI2*)for each iteration of the loop, and drop the record for iteration 0.

*For each variable (x) of variable list AWEAI0 AWEAI1 AWEAI2:*

*Set A\_`x' =(PS2\*`x'[1] - (PS2-1)\*`x')*

*Set B\_`x' = (A\_`x' - `x'[1])^2*

*Drop if replicate=0*

**Step 5b.** Create a dataset with only one row that contains the sum of the SEs for all loop iterations and rename the variables so that the “B” in the variable names is replaced by “JKSE.”

*Collapse (sum) JKSE\_WEAI0=B\_AWEAI0*

*JKSE\_WEAI1=B\_AWEAI1*

*JKSE\_WEAI2=B\_AWEAI2*

**Step 5c.** Create variables that capture the overall A-WEAI context indicator SE for all women (*JKSE\_AWEAI0*) and by age category (*JKSE\_AWEAI1* and *JKSE\_AWEAI2*).

*For each variable (x) of variable list JKSE\_AWEAI0 JKSE\_AWEAI1 JKSE\_AWEAI2:*

*Set SE\_`x' = sqrt(`x'÷(PS2\*(PS2-1)))*

*Drop `x'*

**Step 6.** Create a binary variable (*ID*) equal to 1 and save the variables created in Step 5c to a temporary data file, “GPI\_5DE\_SE.”

*Set ID=1*

*Save “Temp\GPI\_5DE\_SE”*

**Step 7.** Load the “GPI\_5DE\_JK” temporary data file and drop all observations in which the replicate variable is not equal to 0.

*Load “Temp\GPI\_5DE\_JK”*

*Drop if replicate≠0*

**Step 8.** Create a binary variable (*ID*) equal to 1 and merge the data file with the GPI\_5DE\_SE data file. Now all 5DE, GPI, A-WEAI, DEFF, and SE variables are contained in one row in one data file.

*Set ID=1*

*Merge using “Temp\GPI\_5DE\_SE” using ID as the key matching variable*

*Drop \_merge replicate*

**Step 9.** Rename the standard error variables *SE\_JKSE\_AWEAI0* to be *SE0*, *SE\_JKSE\_AWEAI1* to be *SE1,* and *SE\_JKSE\_AWEAI2* to be *SE2*.

*Rename SE\_JKSE\_AWEAI0 to be SE0*

*Rename SE\_JKSE\_AWEAI1 to be SE1*

*Rename SE\_JKSE\_AWEAI2 to be SE2*

**Step 10.** Create a variable (*Z0*) that captures the z-score for the test of difference comparing the A-WEAI score for women under 30 years of age and women 30 years of age or older.

*gen Z0=abs(AWEAI1-AWEAI2)/sqrt(SE1^2+SE2^2)*

**Step 11.** Change the data format from household level to item level—that is, from flat to rectangular format—keeping the 5DE, GPI, A-WEAI, DEFF, SE, and Z variables so that the values for all women and for each age category are contained in a separate row. (See procedures outlined in Section 9.2.1, Step 1.)

*Reshape long GPI EA\_20p AWEAI SE deff Z, i(ID)*

*Drop ID \_j*

**Step 12.** Create row labels.

*Label row 1 “All women”*

*Label row 2 “Young women”*

*Label row 3 “Old women”*

**Step 13.** Create variables equal to the upper (*UCI*) and lower (*LCI*) bounds of the 95 percent confidence intervals for the A-WEAI.

*Set LCI=AWEAI-(1.96\*SE)*

*Set UCI=AWEAI+(1.96\*SE)*

**Step 14.** Create a variable (*P*) equal to the p-values corresponding to the z-scores calculated in step 10.

*gen P = 2\*(1-normal(abs(Z)))*

**Step 15.** Save the data to a temporary data file, “AWEAI\_FINAL,” and erase the two temporary GPI data files.

*Order Label GPI EA\_20p AWEAI SE LCI UCI deff Z P, first*

*Save “Temp\AWEAI\_FINAL”*

*Erase “Temp\GPI\_5DE\_SE”*

*Erase “Temp\GPI\_5DE\_JK”*

**Step 16.** Export the A-WEAI results (GPI, 5DE, A-WEAI, SE, LCI, and UCI) to Excel.

*Export excel using “Results\AWEAI CI.xlsx”*

### B2.2 A-WEAI context indicator with standard error estimation

The step-by-step instructions to calculate the confidence intervals and design effect (DEFF) for the Feed the Future Abbreviated Women’s Empowerment in Agriculture Index (A-WEAI) context indicator follow the steps in the “FTF ZOI Survey [COUNTRY] [YEAR] aweai\_context\_JK” do file. The jackknife variance estimation for the indicator depends on a second do file, “CONT\_fun.do,” which is called by the first do file. “CONT\_fun.do” defines the JKCONT function that calculates the A-WEAI context indicator within the jackknife iterations.

**Step 1.** Load the A-WEAI individual indices data file, “aweai\_individual\_indices.dta,” created in the A-WEAI calc do file (the most recent version of “FTF ZO1 Survey [COUNTRY] [YEAR] syntax AWEAI\_2 calc.do”).

*Load “aweai\_individual\_indices”*

**Step 2.** Rename the censored headcount variables generated for the six A-WEAI empowerment indicators in the A-WEAI calc do file.

*Rename feelinputdecagr\_ndepr\_CH\_20p to be ind1\_ndepr\_CH\_20p*

*Rename jown\_count\_ndepr\_CH\_20p to be ind2\_ndepr\_CH\_20p*

*Rename credjanydec\_any\_ndepr\_CH\_20p to be ind3\_ndepr\_CH\_20p*

*Rename incdec\_count\_ndepr\_CH\_20p to be ind4\_ndepr\_CH\_20p*

*Rename groupmember\_any\_ndepr\_CH\_20p to be ind5\_ndepr\_CH\_20p*

*Rename npoor\_z105\_ndepr\_CH\_20p to be ind6\_ndepr\_CH\_20p*

**Step 3.** Define the weights to be used for the calculation, if not already done. For example, combine the sampling weight for the primary adult male decisionmaker if the individual is male, and the sampling weight for the primary adult female decisionmaker if the individual is female. Be sure to first divide the weight by the number of decimal places if the weight does not include decimals.

*Set pdm\_wgt=wgt\_fpdm if v6\_sex=2*

*Replace pdm\_wgt=wgt\_mpdm if v6\_sex=1*

*Label variable “Primary adult decisionmaker weight”*

**Step 4.** If not already created, create a variable (*age\_cat*) that indicates whether the primary adult decisionmaker is under 30 years of age or 30 years of age or older. If an individual is missing a self-reported age in the A-WEAI module, use the age information available in the household roster.

*Set age\_cat=missing*

*Replace age\_cat=1 if v6102<30*

*Replace age\_cat=2 if v6102≥30 and v6102<98*

*Replace age\_cat=1 if youth\_fdm\_dj=1 and age\_cat=missing and v6\_sex=2*

*Replace age\_cat=2 if youth\_fdm\_dj=0 and age\_cat=missing and v6\_sex=2*

*Replace age\_cat=1 if youth\_mdm\_dj=1 and age\_cat=missing and v6\_sex=1*

*Replace age\_cat=2 if youth\_mdm\_dj=0 and age\_cat=missing and v6\_sex=1*

*Label values 1 “< 30yo” 2 “30+ yo”*

*Label variable “Age category: <30 or 30+”*

**Step 5.** Create a sequential cluster variable (*id*) using the original cluster variable (*hhea*). The variable will range from 1 to the number of clusters in the analysis (N). First, sort the data by cluster number, lowest to highest. Then assign *id* for the first cluster to be 1, the second cluster to be 2, the third cluster to be 3, etc., until all clusters have a value.

*Sort by hhea*

*Set id=1 for first hhea*

*Replace id=2 for second hhea*

*Replace id=3 for third hhea*

*[Repeat for all clusters]*

*Replace id=N for Nth (final) hhea*

*Label variable “Sequential cluster ID variable”*

**Step 6.** Create total weight variables for sub-populations for which the indicator is being calculated.

**Step 6a.** Create a total weight variable (*alloc)* with strata and sex-specific values*.*

*By samp\_stratum v6\_sex: set alloc=total(pdm\_wgt)*

**Step 6b.** Create a total weight variable (*alloc\_AGE)* with strata, sex, and age-specific values*.*

*By samp\_stratum v6\_sex age\_cat: set alloc\_AGE=total(pdm\_wgt)*

**Step 7.** Save the data to a temporary data file, “AWEAI\_INDIV.”

*Save “Temp\AWEAI\_INDIV”*

**Step 8.** Specify the survey design—that is, the cluster, weight, and strata variables.

*Svyset hhea [pw=weight], strata(samp\_stratum)*

**Step 9.** Create variables for each of the six A-WEAI indicator sample-weighted means and their corresponding DEFFs by sex for all individuals not missing values for any of the six A-WEAI indicators. For each A-WEAI indicator, save the sample-weighted mean estimates in matrix ‘a’ and the DEFF estimates in matrix ‘d’ and then assign the relevant matrix components to the variables for men and women. Variables for men are indicated by “\_m\_” in their names, and variables for women are indicated by “\_f\_” in their names.

*For each value (x) of variables ind1\_ndepr\_CH\_20p ind2\_ndepr\_CH\_20p ind3\_ndepr\_CH\_20p ind4\_ndepr\_CH\_20p ind5\_ndepr\_CH\_20p ind6\_ndepr\_CH\_20p:*

*Svy: mean `x' if miss\_any=0, over(v6\_sex)*

*Estimate deff*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set mean\_m\_`x'=a[1,1]*

*Set mean\_f\_`x'=a[1,2]*

*Set deff\_m\_`x'=d[1,1]*

*Set deff\_f\_`x'=d[1,2]*

**Step 10.** Sum the variables capturing the six A-WEAI indicator sample-weighted means created for women in Step 9 and divide by six to create a variable (*mean\_f\_overall*) for the sample-weighted mean percentage of women achieving adequacy across the six A-WEAI indicators.

*Set mean\_f\_overall=(mean\_f\_ind1\_ndepr\_CH\_20p+mean\_f\_ind2\_ndepr\_CH\_20p+ mean\_f\_ind3\_ndepr\_CH\_20p+mean\_f\_ind4\_ndepr\_CH\_20p+ mean\_f\_ind5\_ndepr\_CH\_20p+mean\_f\_ind6\_ndepr\_CH\_20p)÷6 if v6\_sex=2*

*Label variable “Average % of women achieving adequacy across the 6 A-WEAI indicators”*

**Step 11.** Sum the variables capturing the DEFFs for the six A-WEAI indicator means created for women in Step 9 and divide by six to create a variable (*deff\_f\_overall*) for the DEFF of the sample-weighted mean percentage of women achieving adequacy across the six A-WEAI indicators.

*Set deff\_f\_overall=(deff\_f\_ind1\_ndepr\_CH\_20p+deff\_f\_ind2\_ndepr\_CH\_20p+ deff\_f\_ind3\_ndepr\_CH\_20p+deff\_f\_ind4\_ndepr\_CH\_20p+ deff\_f\_ind5\_ndepr\_CH\_20p+deff\_f\_ind6\_ndepr\_CH\_20p)÷6 if v6\_sex=2*

*Label variable “DEFF of average % of women achieving adequacy across the 6 A-WEAI indicators”*

**Step 12.** Sum the variables capturing the six A-WEAI indicator sample-weighted means created for men in Step 9 and divide by six to create a variable (*mean\_m\_overall*) for the sample-weighted mean percentage of men achieving adequacy across the six A-WEAI indicators.

*Set mean\_m\_overall=(mean\_m\_ind1\_ndepr\_CH\_20p+mean\_m\_ind2\_ndepr\_CH\_20p+ mean\_m\_ind3\_ndepr\_CH\_20p+mean\_m\_ind4\_ndepr\_CH\_20p+ mean\_m\_ind5\_ndepr\_CH\_20p+mean\_m\_ind6\_ndepr\_CH\_20p)÷6 if v6\_sex=1*

*Label variable “Average % of men achieving adequacy across the 6 A-WEAI indicators”*

**Step 13.** Sum the variables capturing the DEFFs for the six A-WEAI indicator means created for men in Step 9 and divide by six to create a variable (*deff\_m\_overall*)for the DEFF of the sample-weighted mean percentage of men achieving adequacy across the six A-WEAI indicators.

*Set deff\_m\_overall=(deff\_m\_ind1\_ndepr\_CH\_20p+deff\_m\_ind2\_ndepr\_CH\_20p+ deff\_m\_ind3\_ndepr\_CH\_20p+deff\_m\_ind4\_ndepr\_CH\_20p+ deff\_m\_ind5\_ndepr\_CH\_20p+deff\_m\_ind6\_ndepr\_CH\_20p)÷6 if v6\_sex=1*

*Label variable “DEFF of average % of men achieving adequacy across the 6 A-WEAI indicators”*

**Step 14.** Create variables that capture the sample-weighted mean percentages of individuals achieving adequacy in each of the six A-WEAI indicators and their corresponding DEFFs by sex and age for all individuals not missing values for any of the six A-WEAI indicators. For each A-WEAI indicator, save the sample-weighted mean estimates in matrix ‘a’ and the DEFF estimates in matrix ‘d’ and then assign the relevant matrix components to the variables for men and women by age. Variables for men under 30 years of age are indicated by “\_my\_” in their names, and variables for men 30 years of age or older are indicated by “\_mo\_” in their names. Variables for women under 30 years of age are indicated by “\_fy\_” in their names, and variables for women 30 years of age or older are indicated by “\_fo\_” in their names.

*For each value (x) of variables ind1\_ndepr\_CH\_20p ind2\_ndepr\_CH\_20p ind3\_ndepr\_CH\_20p ind4\_ndepr\_CH\_20p ind5\_ndepr\_CH\_20p ind6\_ndepr\_CH\_20p:*

*Svy: mean `x' if miss\_any=0, over(v6\_sex age\_cat)*

*Estimate deff*

*Matrix a=e(estimates)*

*Matrix d=e(deff)*

*Set mean\_my\_`var'=a[1,1]*

*Set mean\_mo\_`var'=a[1,2]*

*Set mean\_fy\_`var'=a[1,3]*

*Set mean\_fo\_`var'=a[1,4]*

*Set deff\_my\_`var'=d[1,1]*

*Set deff\_mo\_`var'=d[1,2]*

*Set deff\_fy\_`var'=d[1,3]*

*Set deff\_fo\_`var'=d[1,4]*

**Step 15.** Sum the variables capturing the six A-WEAI indicator sample-weighted means created for individuals in Step 14 and divide by six to create variables (*mean\_my, mean\_mo, mean\_fy, mean\_fo*) for the sample-weighted mean percentage of individuals achieving adequacy across the six A-WEAI indicators by sex and age.

*Set mean\_my=(mean\_my\_ind1\_ndepr\_CH\_20p+mean\_ my\_ind2\_ndepr\_CH\_20p+ mean\_my\_ind3\_ndepr\_CH\_20p+mean\_my\_ind4\_ndepr\_CH\_20p+ mean\_my\_ind5\_ndepr\_CH\_20p+mean\_ my\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “Average % of men<30yo achieving adequacy across the 6 A-WEAI indicators”*

*Set mean\_mo=(mean\_mo\_ind1\_ndepr\_CH\_20p+mean\_ mo\_ind2\_ndepr\_CH\_20p+ mean\_mo\_ind3\_ndepr\_CH\_20p+mean\_mo\_ind4\_ndepr\_CH\_20p+ mean\_mo\_ind5\_ndepr\_CH\_20p+mean\_ mo\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “Average % of men 30+yo achieving adequacy across the 6 A-WEAI indicators”*

*Set mean\_fy=(mean\_fy\_ind1\_ndepr\_CH\_20p+mean\_ fy\_ind2\_ndepr\_CH\_20p+ mean\_fy\_ind3\_ndepr\_CH\_20p+mean\_fy\_ind4\_ndepr\_CH\_20p+ mean\_fy\_ind5\_ndepr\_CH\_20p+mean\_ fy\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “Average % of women<30yo achieving adequacy across the 6 A-WEAI indicators”*

*Set mean\_fo=(mean\_fo\_ind1\_ndepr\_CH\_20p+mean\_ fo\_ind2\_ndepr\_CH\_20p+ mean\_fo\_ind3\_ndepr\_CH\_20p+mean\_fo\_ind4\_ndepr\_CH\_20p+ mean\_fo\_ind5\_ndepr\_CH\_20p+mean\_ fo\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “Average % of women 30yo+ achieving adequacy across the 6 A-WEAI indicators”*

**Step 16.** Sum the variables capturing the DEFFs for the six A-WEAI indicator means created for individuals in Step 14 and divide by six to create variables (*deff\_my, deff\_mo, deff\_fy, deff\_fo*) for the DEFFs of the sample-weighted mean percentage of individuals achieving adequacy across the six A-WEAI indicators by sex and age.

*Set deff\_my=(deff\_my\_ind1\_ndepr\_CH\_20p+deff\_ my\_ind2\_ndepr\_CH\_20p+ deff\_my\_ind3\_ndepr\_CH\_20p+deff\_my\_ind4\_ndepr\_CH\_20p+ deff\_my\_ind5\_ndepr\_CH\_20p+deff\_ my\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “DEFF of average % of men<30yo achieving adequacy across the 6 A-WEAI indicators”*

*Set deff\_mo=(deff\_mo\_ind1\_ndepr\_CH\_20p+deff\_ mo\_ind2\_ndepr\_CH\_20p+ deff\_mo\_ind3\_ndepr\_CH\_20p+deff\_mo\_ind4\_ndepr\_CH\_20p+ deff\_mo\_ind5\_ndepr\_CH\_20p+deff\_ mo\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “DEFF of average % of men 30+yo achieving adequacy across the 6 A-WEAI indicators”*

*Set deff\_fy=(deff\_fy\_ind1\_ndepr\_CH\_20p+deff\_ fy\_ind2\_ndepr\_CH\_20p+ deff\_fy\_ind3\_ndepr\_CH\_20p+deff\_fy\_ind4\_ndepr\_CH\_20p+ deff\_fy\_ind5\_ndepr\_CH\_20p+deff\_ fy\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “DEFF of average % of women <30yo achieving adequacy across the 6 A-WEAI indicators”*

*Set deff\_fo=(deff\_fo\_ind1\_ndepr\_CH\_20p+deff\_ fo\_ind2\_ndepr\_CH\_20p+ deff\_fo\_ind3\_ndepr\_CH\_20p+deff\_fo\_ind4\_ndepr\_CH\_20p+ deff\_fo\_ind5\_ndepr\_CH\_20p+deff\_ fo\_ind6\_ndepr\_CH\_20p)÷6*

*Label variable “DEFF of average % of women 30+yo achieving adequacy across the 6 A-WEAI indicators”*

**Step 17.** Create a dataset with one row containing the mean and DEFF variables created in Steps 10–13, 14, and 15.

*Collapse mean mean\_m\_overall deff\_m\_overall mean\_f\_overall deff\_f\_overall*

*mean\_my deff\_my mean\_mo deff\_mo*

*mean\_fy deff\_fy mean\_fo deff\_fo*

**Step 18.** Create a counter variable (*replicate*) and initialize it to 0.

*Set replicate=0*

**Step 19.** Save the data in a temporary data file, “CONT\_JK0.”

*Save “Temp\CONT\_JK0”*

**Step 20.** Drop all DEFF variables and save the data in a temporary data file, “CONT\_JK.”

*Drop deff\_m\_overall deff\_f\_overall deff\_my deff\_mo deff\_fy deff\_fo*

*Save “Temp\CONT\_JK”*

**Step 21.** Load the do file that contains the jackknife program (JKCONT) to be used in the loop in Step 23.

*Load “CONT\_fun.do”*

**Step 22.** Create a scalar variable (*PS2*) equal to the number of sampled clusters in the indicator calculation.

*scalar PS2=[number of clusters]*

**Step 23.** Perform the jackknife computation for the A-WEAI context indicator using the loop in this step. Starting with a loop iteration variable (*i*) equal to 1, the loop:

* Loads in the “AWEAI\_INDIV” temporary data file created in Step 7
* Drops all records in the cluster that have an *id* value equal to the value of the loop iteration variable
* Creates a new total weight variable (*allocj)* with strata and sex-specific values and a new total weight variable (*allocj\_AGE)* with strata, sex, and age-specific valuesthat exclude records in the dropped cluster
* Creates adjusted weight variables by sex (*wtadj*) and by sex and age (*wtadj\_AGE*)
* Calls the JKCONT function, which performs Steps 8, 9, 10, 12, 14, 15, and 17 using the adjusted weight variables and yields values for the following variables: *mean\_m\_overall, mean\_f\_overall, mean\_my, mean\_mo, mean\_fy, and mean\_fo*, for the dataset excluding records for the dropped cluster
* Appends the values for the variables to the temporary data file “Temp\CONT\_JK”
* Adds 1 to the loop iteration variable

The loop is repeated for all sampled clusters included in the analysis, so that the “Temp\CONT\_JK” data file contains one record for loop.

*Local i=1*

*while `i'≤PS2:*

*Load “Temp\AWEAI\_INDIV”*

*Drop if i=`i'*

*Sort by samp\_stratum v6\_sex: set allocj=total(pdm\_wgt)*

*Sort by samp\_stratum v6\_sex age\_cat: set allocj\_AGE=total(pdm\_wgt)*

*Set wtadj=pdm\_wgt\*(alloc÷allocj)*

*Set wtadj\_AGE=pdm\_wgt\*(alloc\_AGE÷allocj\_AGE)*

*Call the function (JKCONT) in the “CONT\_fun.do” program*

*Set replicate=`i'*

*Append using “Temp\CONT\_JK,” force*

*Save “Temp\CONT\_JK”*

*local i=`i'+1*

**Step 24.** Load the temporary data file, “CONT\_JK.”

*Load “Temp\CONT\_JK”*

**Step 25a.** Create variables that capture the A-WEAI context indicator standard error (SE) for each sex and age category (*B\_mean\_m\_overall, B\_mean\_f\_overall, B\_mean\_my, B\_mean\_mo, B\_mean\_fy, B\_mean\_fo*) for each iteration of the loop, and drop the record for iteration 0.

*For each value (x) of variable list mean\_m\_overall mean\_f\_overall mean\_my mean\_mo mean\_fy mean\_fo:*

*Set A\_`x'=(PS2\*`x'[PS2+1] - (PS2-1)\*`x')*

*Set B\_`x'=(A\_`x' - `x'[PS2+1])^2*

*Drop if \_n=PS2+1*

**Step 25b.** Create a dataset with only one row that contains the sum of the SEs for all loop iterations and rename the variables so that the “B” in the variable names is replaced by “JKSE.”

*Collapse (sum) JKSE\_mean\_m\_overall=B\_mean\_m\_overall*

*JKSE\_mean\_f\_overall=B\_mean\_f\_overall*

*JKSE\_mean\_my=B\_mean\_my*

*JKSE\_mean\_mo=B\_mean\_mo*

*JKSE\_mean\_fy=B\_mean\_fy*

*JKSE\_mean\_fo=B\_mean\_fo*

**Step 25c.** Create variables that capture the overall A-WEAI context indicator SE for each sex and age category (SE\_JKSE\_mean\_m\_overall, SE\_JKSE\_mean\_f\_overall, SE\_JKSE\_mean\_my, SE\_JKSE\_mean\_mo, SE\_JKSE\_mean\_fy, SE\_JKSE\_mean\_fo).

*For each variable (x) of variables JKSE\_mean\_m\_overall JKSE\_mean\_f\_overall JKSE\_mean\_my JKSE\_mean\_mo JKSE\_mean\_fy JKSE\_mean\_fo:*

*Set SE\_`x'= sqrt(`x'÷(PS2\*(PS2-1)))*

*Drop `x'*

**Step 26.** Create a dummy variable (*ID*) equal to 1 and save the variables created in Step 25c to a temporary data file, “CONT\_SE.”

*Set ID=1*

*Save “Temp\CONT\_SE”*

**Step 27.** Load the “CONT\_JK0” temporary data file, which contains one record with the mean and DEFF variables (see Step 19), and create a dummy variable (*ID*) equal to 1 to merge the file with the “CONT\_SE” temporary data file, and then drop the unneeded variables (*\_merge* and *replicate*). Now all mean, DEFF, and SE variables are contained in one row in one data file.

*Use “Temp\CONT\_JK0”*

*Set ID=1*

*Merge using “Temp\CONT\_SE,” using ID as the key matching variable*

*Drop \_merge replicate*

**Step 28.** In preparation for transforming the data in Step 29, rename the mean, DEFF, and SE variables so that the names are simplified and the overall female estimates are designated with a 1, the overall male estimates are designated with a 2, the estimates for women under 30 years of age are designated by a 3, the estimates for men under 30 years of age are designated by a 4, the estimates for women 30 years of age or older are designated by a 5, and the estimates for men 30 years of age or older are designated by a 6.

*Rename mean\_f\_overall to be R1*

*Rename mean\_m\_overall to be R2*

*Rename mean\_fy to be R3*

*Rename mean\_my to be R4*

*Rename mean\_fo to be R5*

*Rename mean\_mo to be R6*

*Rename deff\_f\_overall to be deff1*

*Rename deff\_m\_overall to be deff2*

*Rename deff\_fy to be deff3*

*Rename deff\_my to be deff4*

*Rename deff\_fo to be deff5*

*Rename deff\_mo to be deff6*

*Rename SE\_JKSE\_mean\_f\_overall to be SE1*

*Rename SE\_JKSE\_mean\_m\_overall to be SE2*

*Rename SE\_JKSE\_mean\_fy to be SE3*

*Rename SE\_JKSE\_mean\_my to be SE4*

*Rename SE\_JKSE\_mean\_fo to be SE5*

*Rename SE\_JKSE\_mean\_mo to be SE6*

**Step 29.** Create a variable (*Z1*) that captures the z-score for the test of difference comparing women under 30 years of age and women 30 years of age or older.

*gen Z1=abs(R3-R5)/sqrt(SE3^2+SE5^2)*

**Step 30.** Create a variable (*Z2*) that captures the z-score for the test of difference comparing men under 30 years of age and men 30 years of age or older.

*gen Z2=abs(R4-R6)/sqrt(SE4^2+SE6^2)*

**Step 31.** Change the data format from household level to item level—that is, from flat to rectangular format, keeping the means, DEFFs, SEs, and Z so that the values for each age and sex category are contained in a separate row. (See procedures outlined in Section 9.2.1, Step 1.)

*Reshape long R deff SE Z, i(ID) j(DOMAIN)*

**Step 32.** Create row labels.

*Label row 1 “All women”*

*Label row 2 “All men”*

*Label row 3 “Young women”*

*Label row 4 “Young men”*

*Label row 5 “Old women”*

*Label row 6 “Old men”*

**Step 33.** Create variables equal to the upper (*UCI*) and lower (*LCI*) bounds of the 95 percent confidence intervals for the A-WEAI context indicator by age and sex.

*Set LCI=R-(1.96\*SE)*

*Set UCI=R+(1.96\*SE)*

**Step 34.** Create a variable (*P*) equal to the p-values corresponding to the z-scores calculated in steps 29 and 30.

*Set P = 2\*(1-normal(abs(Z)))*

**Step 35.** Save the data to a temporary data file, “CONT\_FINAL,” and erase the temporary data files.

*Save “Temp\CONT\_FINAL”*

*Erase “Temp\AWEAI\_INDIV”*

*Erase “Temp\CONT\_JK”*

*Erase “Temp\CONT\_JK0”*

*Erase “Temp\CONT\_SE”*

**Step 33.** Export the A-WEAI context indicator results (mean, SE, DEFF, LCI, UCI, Z, and P) to Excel.

*Export excel using “Results\AWEAI\_CONTEXT CI.xlsx”*

1. Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>. [↑](#footnote-ref-0)
2. Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>. [↑](#footnote-ref-1)
3. Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>. [↑](#footnote-ref-2)
4. Feed the Future Indicator Handbook March 2018. Available at: <https://www.agrilinks.org/sites/default/files/ftf-indicator-handbook-march-2018-508.pdf>. [↑](#footnote-ref-3)
5. Information on the universe of Feed the Future indicators is available at: https://www.agrilinks.org/sites/default/files/ftf-indicator-handbook-march-2018-508.pdf. [↑](#footnote-ref-4)
6. The P1-ZOI may have changed between the phase one baseline and endline surveys, and the P2-ZOI may not be the same as the P1-ZOI. Consult the country-specific study protocol for the ZOI Survey 2018–2019 for more information about the ZOI. [↑](#footnote-ref-5)
7. Two additional indicators in phase one, “Prevalence of anemia among children 6-59 months” and “Prevalence of anemia among women of reproductive age,” are not required even if they were collected as baseline. They are, therefore, omitted from this list. [↑](#footnote-ref-6)
8. Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>. [↑](#footnote-ref-7)
9. Available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods>. [↑](#footnote-ref-8)
10. 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. [↑](#footnote-ref-9)
11. 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. [↑](#footnote-ref-10)
12. The A-WEAI is a special case, and instructions for computing confidence intervals, the design effects, and tests of difference for the A-WEAI are available in Appendix B. [↑](#footnote-ref-11)
13. Standard disaggregates are those that are entered into the FTFMS. Other disaggregates are those that are not entered into the FTFMS but are included in the ZOI Survey reports. [↑](#footnote-ref-12)
14. As noted in Section 3.1, sample weights should not be created so that they scale to the ZOI population, but they should be normalized so that the weighted numbers of observation are on the same scale as the sample (i.e., the summation of the weight will equal the total sample size and the average of the weight will be 1). If the population size (i.e., weighted n) that Stata outputs looks to be on a population scale, please revisit the weight variable and ensure it it normalized. [↑](#footnote-ref-13)
15. 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.” [↑](#footnote-ref-14)
16. 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 ‘*citype(logit)*’ with *svy: mean*, but the option does not work with all versions of Stata. [↑](#footnote-ref-15)
17. 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. [↑](#footnote-ref-16)
18. Survey design-adjusted t-tests are not required for Feed the Future descriptive reports. [↑](#footnote-ref-17)
19. Note, however, that if a disaggregate category has fewer than 30 observations in it, that category should not be analyzed. [↑](#footnote-ref-18)
20. StataCorp. (2019). Stata: Release 16. Statistical Software. College Station, TX: StataCorp LLC. [↑](#footnote-ref-19)
21. UC REGENTS. (2019). UCLA Institute for Digital Research and Education. *How can I do a t-test with survey data? | Stata FAQ.* <https://stats.idre.ucla.edu/stata/faq/how-can-i-do-a-t-test-with-survey-data/> Accessed on: August 27, 2019. [↑](#footnote-ref-20)
22. 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. [↑](#footnote-ref-21)
23. This is also what Stata and other statistical packages assume when conducting the statistical tests described in the previous sections. [↑](#footnote-ref-22)
24. Heeringa, S., West, B., Berglund, P. (2010). Applied Survey Data Analysis. New York: Chapman and Hall/CRC, <https://doi.org/10.1201/9781420080674> [↑](#footnote-ref-23)
25. An example of one such calculator is available at: https://www.socscistatistics.com/pvalues/normaldistribution.aspx [↑](#footnote-ref-24)
26. If female primary adult decisionmakers can be entered in line 1 of the data collection program household roster, use the following syntax instead: *Set mdm=1 if v101a=1 and m1\_line=1* [↑](#footnote-ref-25)
27. If female primary adult decisionmakers can be entered in line 1 of the data collection program household roster, use the following syntax instead: *Set fdm=1 if v101b=1 and (vtype=1 and m1\_line=2) | (v101b=1 and vtype=3 and m1\_line=1)* [↑](#footnote-ref-26)
28. The steps to create the variables that indicate who made selected key production decisions are not included in the core aquaculture module. If the questions are added to the aquaculture module during country adaption, users should include and adapt these steps. [↑](#footnote-ref-27)
29. The definition of what constitutes a household may vary by country, especially where polygamous relationships exist. [↑](#footnote-ref-28)
30. Note that in the template Stata do files, a generic variable, *pov\_stat*, is used as the poverty status disaggregate. Please replace the generic variable with the poverty status variable appropriate for the analysis being conducted. [↑](#footnote-ref-29)
31. Deaton, A., and S. Zaidi. 2002, [*A Guide to aggregating consumption expenditures, Living Standards Measurement Study*](http://siteresources.worldbank.org/INTPA/Resources/429966-1092778639630/deatonZaidi.pdf)*, Working Paper 135*. *Working Paper 135*. Available at: <http://siteresources.worldbank.org/INTPA/Resources/429966-1092778639630/deatonZaidi.pdf> [↑](#footnote-ref-30)
32. Grosh, M., and Muñoz, J. 1996. *A manual for planning and implementing the Living Standards Measurement Study Surveys. LSMS Working Paper #126*. *The World Bank*. Available at: <http://documents.worldbank.org/curated/en/1996/05/438573/manual-planning-implementing-living-standards-measurement-study-survey> [↑](#footnote-ref-31)
33. 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 household members in the household for which the reported value is found to be an outlier or is missing. [↑](#footnote-ref-32)
34. The cover page of the questionnaire will identify the administrative units that are used to identify the location of the households. [↑](#footnote-ref-33)
35. 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 this during data analysis. See the step-by-step process in Section 9.2.1. [↑](#footnote-ref-34)
36. For the food consumption sub-module, all households should report consumption of at least one food item. No interview should be marked as complete if all questions in this sub-module are coded as “NO” or “Don’t know.” If computer-assisted personal interviewing is used, the program will automatically reject this situation and will not allow closing the module if there is no food item recorded as consumed. [↑](#footnote-ref-35)
37. The cover page of the questionnaire will identify the administrative units that are used to identify the location of the households. [↑](#footnote-ref-36)
38. It is important to note that food consumption can be skewed toward the right with a long tail up to 5 SD from the mean, because of a few households with high level of per capita consumption. Further examination of these households and patterns of consumption should help determine whether they are true outliers. However, even if these observations are legitimate, it may be necessary to remove these households, because they will strongly impact sample statistics. [↑](#footnote-ref-37)
39. Available at: <http://databank.worldbank.org/data/home.aspx>, under the World Development Indicators database and Real Interest Rate (%) series. [↑](#footnote-ref-38)
40. Bangladesh, Ethiopia, Ghana, Guatemala, Honduras, Kenya, Mali, Nepal, Niger, Nigeria, Senegal, and Uganda [↑](#footnote-ref-39)
41. At least 20 years of data are available for Bangladesh, Guatemala, Honduras, Kenya, Nigeria, and Uganda. For Mali, Niger, and Senegal, the series starts in 2005, providing more than 10 years of data. There are no data for Ghana, Ethiopia, or Nepal. [↑](#footnote-ref-40)
42. 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. [↑](#footnote-ref-41)
43. If the household owns more than one item of the durable goods, the average value and average age should be used. [↑](#footnote-ref-42)
44. For a review of approaches in measuring the value of housing for the welfare aggregate, see: Balcazar, C.F., et al. 2014. *Rent imputation for welfare measurement. A review of methodologies and empirical findings.* Policy Research Working Paper 7103. The World Bank. [↑](#footnote-ref-43)
45. In regions where there is an active rental market, respondents may have a reasonable idea of the rent 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. [↑](#footnote-ref-44)
46. If there is insufficient rental data to use hedonic regression, please contact USAID to discuss and an agree on an alternate approach. [↑](#footnote-ref-45)
47. Other functional forms can be tested, such as a log-log function. [↑](#footnote-ref-46)
48. Note that question 204 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. [↑](#footnote-ref-47)
49. The age of the house (8.603) is asked only of homeowners, so this variable can be obtained from the dataset only for these cases in which a homeowner can also provide a rental value of the house (8.604). [↑](#footnote-ref-48)
50. This is not included in survey questionnaire; it is added during the post-processing stage using the country-specific definition of urban and rural. [↑](#footnote-ref-49)
51. These observations are not available from the standard dataset, but could potentially be obtained from secondary geographic information systems sources. [↑](#footnote-ref-50)
52. In countries in which the ZOI has changed in phase two, one combined survey could still be conducted but over two samples (albeit with possible overlap) with two resulting datasets. [↑](#footnote-ref-51)
53. The International Comparison Program conducts comprehensive market surveys that are used to compute global PPP and real expenditures (<http://siteresources.worldbank.org/ICPEXT/Resources/ICP_2011.html>). A comprehensive market survey was conducted in 2005 (2005 PPP) and more recently in 2010 (2011 PPP). The new 2011 PPP is required to compute the poverty indicator estimates based on the new international poverty line of USD $1.90 per capita per day, and 2005 PPP is required to compute the estimates based on international poverty line of USD $1.25 per capita per day. [↑](#footnote-ref-52)
54. PPP conversion factor, private consumption (LCU per international $), 2005 International Comparison Program. Available at: <http://data.worldbank.org/indicator>. They are also provided for all former Feed the Future countries in Table 5. [↑](#footnote-ref-53)
55. Average U.S. CPI for year 2005 (base 2005 prices) was 195.30 (<http://www.bls.gov/cpi/cpid05av.pdf>), and for year 2010 was 218.06 (<http://www.bls.gov/cpi/cpid10av.pdf>). Hence, the CPI ratio is= 218.06/195.30=1.1165. [↑](#footnote-ref-54)
56. The World Bank Group. 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. [↑](#footnote-ref-55)
57. The World Bank Group. World Bank Country and Lending Groups. <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>. Accessed on July 17, 2019. [↑](#footnote-ref-56)
58. Rutstein S.O. 2014. Steps to Constructing the New DHS Wealth Index. Rockville, MD: ICF International. [↑](#footnote-ref-57)
59. 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 recreated 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 affects the wealth scores and quintiles calculated. [↑](#footnote-ref-58)
60. 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. [↑](#footnote-ref-59)
61. 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 comparative wealth index indicator. [↑](#footnote-ref-60)
62. Rutstein, S.O. 2014. *Steps to constructing the new DHS wealth index.* Rockville, MD: ICF International. [↑](#footnote-ref-61)
63. 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. [↑](#footnote-ref-62)
64. 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. [↑](#footnote-ref-63)
65. 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) in 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. [↑](#footnote-ref-64)
66. 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. [↑](#footnote-ref-65)
67. 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. [↑](#footnote-ref-66)
68. 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. [↑](#footnote-ref-67)
69. 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. [↑](#footnote-ref-68)
70. 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. [↑](#footnote-ref-69)
71. 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. [↑](#footnote-ref-70)
72. Note that for the wealth quintiles generated for the CWI reference survey, household weights were used rather than household member weights. [↑](#footnote-ref-71)
73. 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 1a of the Part1. AWI step-by step procedures. [↑](#footnote-ref-72)
74. In Stata, this can be done with the “reshape long” command. [↑](#footnote-ref-73)
75. Hillesland, M. “Causal Mapping of the Gender Integration Framework,” accessed March 13, 2018. Available at: https://agrilinks.org/sites/default/files/resource/files/CausalPathwaysGender\_fullpaper\_29Jan2016.pdf. [↑](#footnote-ref-74)
76. Appendix A Table 1 is adapted from Malapit et al., 2015; Alkire et al., 2012; and Alkire et al., 2013. [↑](#footnote-ref-75)
77. These asset categories 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. [↑](#footnote-ref-76)
78. The sample for the calculation of the A-WEAI index excludes cases with missing values for any of the indicators of adequacy in the 5DE. [↑](#footnote-ref-77)
79. Appendix A, Table A1 is adapted from Malapit et al., 2015; Alkire et al., 2012; and Alkire et al., 2013. [↑](#footnote-ref-78)
80. The Feed the Future indicator that measures application of improved management practices and technologies at the implementing mechanism level, *EG.3.2-24 Number of individuals in the agriculture system who have applied improved management practices or technologies with USG assistance [IM-level]*, includes an additional category for climate mitigation, which captures practices and technologies “selected because they minimize emission intensities relative to other alternatives.” improved management practice. This category is not included in the ZOI-level indicator because the ZOI Survey does not collect information on alternative practices or technologies that producers might have used and whether they would result in a higher or lower amount of emissions—that is, whether a practice or technology produces less greenhouse gasses than the alternative. [↑](#footnote-ref-79)
81. Feed the Future définition of a smallholder extends to livestock producers, although for the purpose of the indicator, the definition is relevant for only crops. [↑](#footnote-ref-80)
82. There will be a separate weight for producers of each targeted agricultural commodity included in the survey. [↑](#footnote-ref-81)
83. Create an eighth category, *imp\_oth*, to capture improved management practices and technologies that should be classified into the other category. This category is not included in the step-by-step procedures because the core ZOI Survey questionnaire does not include any questions related to this category. [↑](#footnote-ref-82)
84. Aquaculture producers could also apply practices or technologies in the “Other” 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. [↑](#footnote-ref-83)
85. Create a ninth category, *imp\_oth*, to capture improved management practices and technologies that should classified into the other category. This category is not included in the step-by-step procedures because the core ZOI Survey questionnaire does not include any questions related to this category. [↑](#footnote-ref-84)
86. 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. [↑](#footnote-ref-85)
87. Before moving to this step, repeat the calculations for each additional targeted VCC that the Feed the Future team included in the indicator and for which data on identified promoted practices and technologies were collected. [↑](#footnote-ref-86)
88. Note that data in Stata version 13.0 or above should be saved in Stata versions 11.0 or 12.0 for compatibility before reading. [↑](#footnote-ref-87)
89. The link for the spreadsheet is available at: <http://www.fao.org/fileadmin/user_upload/voices_of_the_hungry/docs/EPE_Example_05.xlsx> [↑](#footnote-ref-88)
90. 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, unless they are unable to stand still for the measurement; then they are also measured lying down. [↑](#footnote-ref-89)
91. For more information, please see the Feed the Future ZOI Survey Data Processing Manual, available at: <https://www.agrilinks.org/post/feed-future-zoi-survey-methods> [↑](#footnote-ref-90)
92. WHO. 2006. *The WHO Child Growth Standards:* <http://www.who.int/childgrowth/en/> and <http://www.who.int/childgrowth/publications/technical_report_pub/en/> [↑](#footnote-ref-91)
93. 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>. [↑](#footnote-ref-92)
94. The WAZ variable created in CSPro does not include decimals to avoid issues with floating numbers that can occur when exporting and importing data into different analysis programs. [↑](#footnote-ref-93)
95. The HAZ variable created in CSPro does not include decimals to avoid issues with floating numbers that can occur when exporting and importing data into different analysis programs. [↑](#footnote-ref-94)
96. The WHZ variable created in CSPro does not include decimals to avoid issues with floating numbers that can occur when exporting and importing data into different analysis programs. If the prevalence of healthy weight children indicator has already been calculated, this step has already been performed. [↑](#footnote-ref-95)
97. The *whzd* variable was created in Step 1 of the Prevalence of wasted children under age 5 indicator step-by step procedures. See this step if the *whzd* variable has not yet been created. [↑](#footnote-ref-96)
98. Also includes plantains and green bananas, which can be starchy staple foods in some areas [↑](#footnote-ref-97)
99. The *chn\_fmiss* variable is created in Step 6 of the Prevalence of exclusive breastfeeding indicator step-by step procedures. See this step if the *chn\_fmiss* variable has not yet been created. [↑](#footnote-ref-98)
100. Also includes plantains and green bananas, which can be starchy staple foods in some areas [↑](#footnote-ref-99)
101. The *whn\_fmiss* variable is created in Step 3 of the Women’s Dietary Diversity indicator step-by step procedures. See this step if the *whn\_fmiss* variable has not yet been created. [↑](#footnote-ref-100)