INFANT AND YOUNG CHILD FEEDING PRACTICES

Collecting and Using Data: A Step-by-Step Guide

CARE USA
January 2010

CARE is a humanitarian organization leading the fight against global poverty. CARE places special focus on working alongside poor women because, equipped with the proper resources, women have the power to help whole families and entire communities escape poverty. As part of its mission, CARE works with survivors of war and natural disasters to help people rebuild their lives.

Funding for this Guide comes from a private family foundation donation to the CARE USA Infant and Young Child Feeding in Emergencies project and the Window of Opportunity program.

We acknowledge logistical support, help in carrying out data collection and technical input from CARE Kenya, the Dadaab IYCF Team (including UNHCR and GTZ), CARE Sierra Leone, and CARE Indonesia. We also thank the following individuals who reviewed and commented on this Guide: Kristin Helz, Dama Soekarjo, Jan Stucki Wilson, Mark Myatt, Ali Maclaine and Victoria Sibson.

CARE is an Equal Opportunity Employer and Affirmative Action Employer (AA/M/F/D/V) dedicated to workplace diversity.

CARE© and CARE Package© are registered marks of CARE. Copyright ©2010 by Cooperative for Assistance and Relief Everywhere, Inc. (CARE). All rights reserved. Unless otherwise indicated, all photos ©CARE. Photo credits: front cover: Erin Lubin/CARE; Mary Lung'aho/CARE; Jane Dempster/CARE.

Published January 2010.

Recommended Citation: Infant and Young Child Feeding Practices: Collecting and Using Data: A Step-by-Step Guide. Cooperative for Assistance and Relief Everywhere, Inc. (CARE). 2010.

For more information or to obtain copies, please contact Bethann Cottrell.
Dr. Bethann Cottrell
Director of Child Health and Nutrition
CARE USA
Phone: 404-979-9413
Email: bcottrell@care.org

Dr. Kirk Dearden, Consultant
Boston University, Center for Global Health and International Development
and Department of International Health
801 Massachusetts Avenue, Crosstown Center, 3rd Floor
Boston, MA 02118
Phone: 617-414-1451
Fax: 617-414-1261
Email: kdearden@bu.edu

Dr. Mary S. Lung'aho
IYCF Special Advisor
CARE USA
114 Ridgewood Road
Glastonbury, CT 06033-3639
Phone: 860-633-3437
Fax: 860-659-4200
Email: mary@nutritionpolicypractice.org

TABLE OF CONTENTS

Abbreviations i

Introduction to the Guide ii

Chapter 1: About This Guide 1
1.1 What the Guide is 1
1.2 Who should use the Guide 2
1.3 How to use the Guide 3
1.4 Additional Resources 4

Chapter 2: Modifying Your Survey to Collect Data on Infant and Young Child Feeding . . . 9
2.1 Use the standard IYCF questions 10
2.2 Thinking about categories of liquids and foods 16

Chapter 3: Sampling 19
3.1 What is sampling? 20
3.2 Answering questions about the sampling strategy you will use 21
3.2.1 Do you prefer a simple random sample, a systematic random sample or cluster sampling? 21
3.2.2 Are you conducting a one-time assessment of IYCF practices or will you collect data at several points in time? 23
3.3 Sample size 24
3.3.1 The five steps to calculating sample size 24
3.4 Finding your sample 37
3.4.1 Choosing your households with simple random sampling 37
3.4.2 Choosing your households with systematic random sampling 38
3.4.3 Choosing your households with cluster sampling 39
Annex: Instructions about how to sample if you are collecting data on IYCF behaviors and anthropometry at the same time 41

Chapter 4: Getting Started with Data Entry 43
4.1 Why we use Excel 43
4.2 How to develop a data screen 44

Chapter 5: Checking and Cleaning Data in Excel 51
5.1 Finding missing data or individuals who shouldn't be in your dataset 52
5.1.1 Cleaning your own data 53
5.1.2 What to do with individuals who should not be in your dataset 56
5.2 Visually scanning your data 57
5.3 Range checks 59
5.4 Consistency checks 62
5.5 Cleaning data using Excel 64

Chapter 6: Analyzing Infant and Young Child Feeding Data 65
6.1 The importance of age measurement 65
6.2 Calculating the age of the infant 66
Table 6.2.1 Converting child's age from days to months 68
6.2.2 Checking age calculations 70
6.3 Ten indicators to measure infant and young child feeding 71
6.4 Calculating indicators for IYCF 72
6.4.1 Timely initiation of breastfeeding (children 0-23 months) 74
6.4.2 Exclusive breastfeeding under 6 months 79
6.4.3 Timely complementary feeding 84
6.4.4 Introduction of solid, semi-solid or soft foods 88
6.4.5 Continued breastfeeding at 1 year 90
6.4.6 Minimum dietary diversity 92
6.4.7 Minimum meal frequency 96
6.4.8 Minimum acceptable diet 99
6.4.9 Consumption of iron-rich or iron-fortified foods 103
6.4.10 Bottle feeding 105

Chapter 7: Producing and Interpreting Tables and Bar Graphs 109
7.1 Using the report card to present results 109
7.2 Interpreting results 111
7.3 Creating and interpreting figures in Excel 114
7.4 Creating and interpreting tables in Microsoft Word 114
7.5 Using statistics to compare groups 116

Glossary 119

ABBREVIATIONS

BF	Breastfeeding or breastfed
CATCH	The *Rapid* CATCH (Core Assessment Tool on Child Health)
DHS	Demographic and Health Survey
ENA	Emergency Nutrition Assessment
FAO	Food and Agriculture Organization
IYCF	Infant and Young Child Feeding
KPC	Knowledge, Practices and Coverage survey
LQAS	Lot Quality Assurance Sampling
M&E	Monitoring and Evaluation
MICS	UNICEF Multiple Indicator Cluster Survey
SMART	Standardized Monitoring and Assessment of Relief and Transitions
WHO	World Health Organization
UNICEF	United Nations Children's Fund

INTRODUCTION TO THE GUIDE

Worldwide, more than 9 million children under 5 years of age die each year. Malnutrition underlies a majority of these U5 deaths, 70% of which occur in the first year of life. Infant and young child feeding practices directly impact the nutritional status and, ultimately, the child survival of children under 2 years of age. Therefore, improving infant and young child feeding is critical to ensuring their optimal health, nutrition and development.

Extensive research has contributed to global feeding recommendations to maximize health and minimize morbidity and mortality for infants and young children. In 1991, WHO released a set of indicators designed to be used in population-based surveys to measure adherence to recommended feeding practices.[1] However, most of these indicators focused on breastfeeding practices. Recently, the indicators have been updated to include a greater focus on appropriate feeding practices for children of complementary feeding age (6-23 months). This 5-year effort by an interagency working group has resulted in a set of simple, valid and reliable indicators that measure food-related aspects of complementary feeding (including dietary variety and frequency of eating episodes), as well as current guidance on the feeding of non-breastfed infants and young children up to 24 months of age.

The newly-developed WHO indicators are described in two companion documents:

- Indicators for assessing infant and young child feeding practices
  - Part 1: Definitions (WHO, UNICEF, IRPRI, UC Davis, and USAID/AED/FANTA
    - Part 1 describes the background, justification, uses, limitations, and definitions of a set of eight core and seven optional indicators.
- Indicator for assessing infant and young child feeding practices
  - Part 2: Measurement (DRAFT July 7, 2009)
    - Part 2 is the operational guide, which provides tool for collection and calculation of the indicators. It is intended primarily for use by large-scale surveys. It includes a short discussion of sampling, an example questionnaire with interviewer instructions, instructions for calculating indicator values, and suggestions for adapting the questionnaire to the survey context, as well as tools in 4 annexes.

The Step-by-Step Guide provides additional instruction to help CARE staff adapt the WHO operational guidance for use with smaller-scale surveys where data are obtained from mothers or primary caregivers of living children 0-23 months of age. The Guide provides assistance with the selection of indicators for a small-scale survey, choice of sampling strategy, entering and cleaning IYCF data, and analyzing the data and reporting the results.

Through use of the Guide, staff will be better able to:

- Assess trends in infant and young child feeding practices
- Target populations at risk
- Make decisions about resource allocation
- Monitor and evaluate progress toward achieving goals, and
- Determine the impact of programs

[1] Indicators for assessing breastfeeding practices. Geneva, World Health Organization, 1991 (WHO/CDD/SER/91.14).

CHAPTER 1: ABOUT THIS GUIDE

Goal: To familiarize you with this Guide and its use

By the end of this chapter, you will be able to:

- Understand what this Guide can help you do
- Understand who should use the Guide
- Know where to find additional materials to help you collect and use data

1.1 What the Guide is

This Guide will help you collect, manage, and use data on IYCF behaviors. Specifically, it will teach you to:

- Create or modify a survey to collect data for calculation of standard WHO indicators on breastfeeding and complementary feeding (Chapter 2)
- Identify and use an appropriate sampling strategy (Chapter 3)
- Enter and clean IYCF data using Excel (Chapters 4 and 5)
- Analyze data and report results (Chapters 6 and 7)

This Guide will not:

- Tell you how to develop questionnaires or come up with indicators for infant and young child feeding that are different from WHO recommended indicators. There are other excellent guides available to help you develop questionnaires. We recommend some of them in the Additional Resources section at the end of this chapter.
- Help you collect data on anthropometry. Anthropometry surveys measure nutritional status using the height (or length), weight and mid-upper arm circumference of children. For more information on collecting anthropometry data, see the Additional Resources at the end of this chapter.

1.2 Who should use the Guide

This Guide is designed for:

- Staff from implementing agencies who collect data on IYCF behaviors
- People who collect data on other topics. While the examples used in this Guide focus on surveys conducted on infant and young child feeding practices, the Guide can be used by staff working in other areas as well

This Guide is especially helpful for technical staff that has at least some experience:

- Entering and cleaning data
- Analyzing and interpreting data
- Creating tables and figures on infant and young child feeding behaviors or have the desire to learn how to do so.

If you have done some or all of these things, this Guide can be a helpful review.

To use this Guide most effectively, you should have some experience with:

- Collecting, entering, analyzing and/or interpreting infant and young child feeding data
- Computer program Microsoft Excel

If you do not have experience with Excel, do not worry. This Guide will show you the steps for entering, managing, and analyzing data. Experience just makes this process easier.

Note: This Guide was written using Microsoft Excel 2003. If you use a more recent version of Excel, not all of the commands and syntax will work. One option is to click on the 'Alt' key to get a drop down menu of commands in Excel 2003.

Why another Guide? Hasn't someone already explained all of this? No. There are many sources of guidance on collecting information on infant and young child feeding and the recent WHO guidance provides information on standard indicators. However, there are no Guides that help you enter, clean, and analyze IYCF data. The intent of this Guide is to address such gaps.

1.3 How to use the Guide

This Guide has two parts:

1. The Guide itself, which you can read on a computer or on paper.
2. A set of Excel files and other files (which we call "tools") that you will use to practice the skills this Guide teaches.

Each chapter of the Guide teaches a specific skill. To learn the skill, you will be given:

- An example of typical work done in programs focused on malnutrition
- A set of data with which to practice
- One or more tools. The beginning of each chapter will tell you what tools you need to complete the assigned task
- A series of steps to show you how to carry out a given task. You will be able to practice small tasks, one at a time, to help improve learning and retention

Be sure to practice all the steps. As you complete each of the steps in the Guide you will feel more confident. Once you feel comfortable using the practice dataset, you can follow the same steps with datasets you create for your own projects. You will also have a resource you can use the next time you conduct a survey.

We recommend that you **take several days to review the entire Guide**. There is a lot of new information that will take time to absorb. **You should do no more than ONE chapter at a time.** For Chapters 5 and 6, we recommend you do no more than one section of each chapter in one day. When you get to Chapters 5 and 6, you will see where we recommend you take a break.

Here is a list of files that come with the Guide. At the beginning of each chapter, we name the files you need to complete the tasks for that chapter.

Table 1.1 Tools (files) that come with the Guide:

Chapters Where You Will Use the Tool	Tool	Name of File
1	Guide (an electronic version of the Step-by-Step Guide for IYCF)	Guide.doc
2	Questionnaire	Questionnaire.doc
3	Sample size calculator	Sample Size IYCF.xls
6	Indicators for assessing infant and young child feeding practices	WHO IYCF Indicators.pdf (note: this file has 2 parts)
4-6	Practice data file	Practice data file.xls
7	Report card template	Report card template.xls
7	Calculator for chi-square tests	Chi-square tests.xls

Please note that the Practice data file does not include all of the variables specified in the questionnaire. A large data file may make it difficult to learn how to carry out basic tasks. We use a small dataset to help you learn more effectively.

1.4 Additional Resources

There are many steps to measuring infant and young child feeding behaviors. However, this Guide only talks about the things you need to do to manage, analyze, and use IYCF data. Below you will find resources to help you with the other steps. Because recent WHO guidance is critical to calculating each of the 10 indicators, here we describe the 2009 WHO guidance in detail. Also, throughout the Guide, we refer extensively to these documents.

WHO's *Indicators for assessing infant and young child feeding practices. Part 1: Definitions* (WHO, 2009) reflects current guidance on breastfeeding, complementary feeding, and feeding of non-breastfed infants and young children up to 24 months of age. *Part 1* describes the background, justification, uses, limitations, and definition of core and optional indicators. Part 2 (Measurement) provides tools for collection and calculation of the indicators.

We also refer extensively to the SMART methodology for assessing nutritional status. To learn more about SMART, see the following:

SMART (Standardized Monitoring and Assessment of Relief and Transitions) is widely used by NGOs. It helps staff collect, manage and report on anthropometry and mortality data in emergency situations.

SMART and ENA are different from what you do in this Guide. SMART helps you with anthropometry and mortality data. The Guide helps you with IYCF behaviors. You can get more information about SMART, along with ENA software and manuals at: www.smartmethodology.org.

From the SMART website: SMART is an inter-agency initiative, which was launched in 2002 by a network of organizations and humanitarian practitioners including: donors, policymakers, and leading experts in emergency epidemiology and nutrition, food security, early warning systems, and demography. The SMART methodology is an improved survey method based on the two most vital, basic public health indicators to assess the severity of a humanitarian crisis: nutritional status of children under-five and mortality rate of the population. These indicators are useful for assessing needs and prioritizing resources, as well as for monitoring the extent to which the relief system is meeting the needs of the population and thus, the overall impact of the relief response.

SMART comes with ENA software and a manual. The ENA software helps you collect and handle data on anthropometry and mortality. In particular, ENA:

- Addresses issues of survey design
- Assesses the quality of training and indicates which individuals require additional training or replacement
- Helps with identification of errors during data entry
- Examines the internal structure of the data to see where the quality of the survey is inadequate (reports the SD, skewness and kurtosis of the sample and for each team, and the Poisson distribution of the malnourished cases, etc.) is inadequate
- Provides a more accurate method for estimating death rates taking into account a variable recall period method for dealing with mass migration
- Produces a report in a standard format that contains all the information needed to judge the quality of the survey as well as presenting the data in a standard format

To learn more about modifying survey instruments so that you can collect IYCF data, refer to:

Child Survival Technical Support Project, CORE Group, US Agency for International Development. KPC 2000+ Knowledge, practices and coverage survey. Module 2: Breastfeeding and Infant and Young Child Feeding. Interviewer Instructions. Unpublished manuscript, June, 2006. Calverton, MD: Child Survival Technical Support Project.

Access this for free at: http://www.childsurvival.com/kpc2000/kpc2000.cfm#FieldGuide.This Guide has questionnaires NGOs use for evaluating child survival programs, including infant and young child feeding practices. If you wish, use the questionnaire on infant and young child feeding. You can access this updated module through the childsurvival website (as noted above: www.childsurvival.com) or through the CORE Group website (www.coregroup.org).

A simpler questionnaire on IYCF is available through the CORE Group and the Child Survival Technical Support program. It is called Rapid Catch. The full reference is: Child Survival Technical Support Project, CORE Group, US Agency for International Development. KPC 2000 Rapid Core Assessment Tool on Child Health). Interviewer Instructions. Unpublished manuscript, October, 2000. Calverton, MD: Child Survival Technical Support Project.

To learn more about sampling, refer to:

Sarriot E, Winch P, Weiss WM, Wagman J. 1999. Methodology and sampling issues for KPC surveys. Unpublished manuscript. Baltimore: Bloomberg School of Public Health, Johns Hopkins University, Department of International Health.

This Guide is free: http://www.childsurvival.com/kpc2000/method.pdf It helps you understand monitoring and evaluation and sampling. It may be hard to understand some of the sections on sampling but if you want details about how to sample, this is a good Guide.

The Save the Children Fund. 2004. Emergency Nutrition Assessment: guidelines for field workers. London: Save the Children Fund.

You need to pay for this Guide (£20.00). You can buy it at: http://www.savethechildren.org.uk/scuk/jsp/resources/details.jsp?id=2313&group=resources&section=publication&subsection=details&pagelang=en.

This Guide explains how to 1) Identify what causes malnutrition 2) Determine the prevalence of malnutrition in your area 3) Carry out anthropometric measurements 4) Assess mortality rates 5) Interpret findings, and 6) Evaluate program impact and coverage. It contains a description of sampling methodologies used for nutritional or anthropometric surveys.

A good guide on how to select households for interview once a sample size has been chosen:

Magnini R. 1997. Sampling guide. Food and Nutrition Technical Assistance (FANTA): Washington DC. You can get this guide for free at: http://www.fantaproject.org/downloads/pdfs/sampling.pdf

Chapter 4 is especially helpful for determining how to select respondents in the field.

To learn more about selecting indicators on IYCF and analyzing results, refer to:

World Health Organization. 2007. Indicators for assessing infant and young child feeding practices. Part I: conclusions of a consensus meeting held 6-8 November 2007. Washington DC: Author.

This WHO guidance provides the latest thinking on standard indicators for IYCF. We use this material in the calculation of indicators specified in our Guide. The Consensus statement referenced above is included electronically with our Guide.

World Health Organization. 2009. Indicators for assessing infant and young child feeding practices. Part II: measurement.

This WHO guidance helps the user calculate core and optional indicators for IYCF.

Arimond M, Ruel MT. 2003. Generating indicators of appropriate feeding of children 6 through 23 months from the KPC 2000+. Unpublished manuscript. Washington DC: Food and Nutrition Technical Assistance Project.

Download this document for free at: http://www.fantaproject.org/publications/kpc.shtml

This document tells you how to sample, develop questionnaires, adapt the Knowledge, Practices and Coverage (KPC) 2000+ survey to your needs, interview, create data entry screens in EpiInfo, tabulate indicators and present results on infant and young child feeding. *This document offers ideas for infant and young child feeding indicators but the indicators are now out of date and should not be used.*

Lung'aho M, Huffman SL, Labbok MH, Sommerfelt E, Baker J. 1996. Tool kit for monitoring and evaluating breastfeeding practices and programs. Unpublished manuscript. Washington DC: US Agency for International Development and Wellstart International.

The Wellstart toolkit is included electronically with our Guide. *This document gives you ideas for infant and young child feeding indicators but the indicators are now out of date and should not be used.* The Wellstart toolkit (an older document) provides useful information to help you select and use appropriate indicators for breastfeeding, and avoid problems when you collect and analyze data.

To learn more about how to determine child's date of birth when the caregiver does not remember it, refer to:

Food and Agriculture Organization. 2008. Guidelines for estimating the month and year of birth for infants and young children. Rome: Author.

This is an excellent resource that tells you how to construct a calendar of local events (religious, political, climatic, etc.) that can be used to help caregivers remember when their children were born. This resource also provides suggestions for how to train interviewers in the use of the calendar. It is an indispensable resource and should be used every time information on IYCF is collected.

For a basic textbook on biostatistics, refer to:

Sullivan L. Essentials of Biostatistics in Public Health. 2008. Boston: Jones and Bartlett.

For a workbook on how to use Excel to produce tables and graphs, summarize data and conduct data analysis, refer to:

Sullivan L. Essentials of Biostatistics Workbook: Statistical Computing Using Excel. 2008. Boston: Jones and Bartlett.

CHAPTER 2: MODIFYING YOUR SURVEY TO COLLECT DATA ON INFANT AND YOUNG CHILD FEEDING

Goals: To understand the importance of using a standardized questionnaire and how to modify the questionnaire to reflect locally available foods and liquids.

By the end of this chapter, you will:

- Understand the importance of using a standardized questionnaire to measure IYCF behaviors
- Name the major categories of liquids and foods you will ask about when you collect data
- Describe how to modify the questionnaire so that the questions in your data collection instrument reflect the specific liquids and foods available in the area where you work

Tool	Purpose of tool	Name of file
Questionnaire	Provides guidance on the questions that need to be asked to calculate the standard infant and young child feeding indicators and helps to organize them into a questionnaire.	Questionnaire.doc

First, an important note about age:

You should make every effort to get accurate information on the infant's age. The information on the IYCF indicators will not be meaningful without accurate information on infant age. All infant and young child feeding recommendations are age-dependent, and analysis of all current status indicators depends on the child's exact age. We will identify the 10 core IYCF indicators in a minute.

The best source of information on age is an identification card such as a government health or vaccination card, a birth or baptismal certificate. You may also want to consult a card used by a local implementing agency (your agency or another agency) to see if date of birth is recorded there.

In cases where no date of birth is recorded, the child's age should be estimated using a local calendar of events. This calendar helps the mother remember date of birth with greater precision. Estimated date of birth will only be accurate if determined properly. Here is an excellent resource:

> Food and Agriculture Organization. 2008. Guidelines for estimating the month and year of birth for infants and young children. Rome: Author.

This resource can be downloaded for free at: www.foodsec.org. This resource will tell you how to construct a calendar specific to the area where you work. It also tells you how to train interviewers to use the calendar to collect date of birth in situations where caregivers do not know when the child was born. In some instances, it is simply not possible to determine date of birth, but this is likely a rare event.

In the questionnaire that accompanies this guide, you will see that Question 3a asks about the child's age in completed years and question 3b asks about the child's age in completed months. This means that a child who is 1 year and 5 months old (i.e., 17 months) should be recorded as being 1 year of age in question 3a and 5 months in question 3b. A child who is 2 weeks old would be recorded as being 0 years of age in question 3a and 0 months of age in question 3b. We'll talk more about age later on (and in particular, in Chapter 6).

As noted later in this Guide, we will use the following language to describe the ages of children:

- 0-5 months means 0-5.9 months (a period of 6 completed months)
- 6-8 months means 6-8.9 months (a 3 month period)
- 6-9 months means 6-9.9 months (a 4 month period)
- 0-23 months means birth through 23.9 months (or birth up to 24 months; it includes all children under 2 years of age)

2.1 Use the standard IYCF questions

When surveying IYCF, use the questions from the questionnaire included with this Guide.

- These questions have been carefully written and tested over many years. They are based on the WHO Indicators for Assessing Infant and Young Child Feeding Practices (2009), and the KPC 2000+ and Rapid CATCH 2007 survey guidance.
- These questions provide the information you need to calculate the 10 key variables on IYCF. If you change the questions, you will measure breastfeeding and complementary feeding appropriately.
- If you use these exact questions you will be able to compare your results to other surveys. If you change the questions, your results cannot be compared to MICS (Multiple Intervention Cluster Surveys) and DHS (Demographic and Health Surveys) data.

The complete questionnaire is included with the Guide. It is called questionnaire.doc. Below, find each of the 10 indicators and the questions you will use to calculate each indicator. The table demonstrates the importance of collecting data on each of the questions contained in the questionnaire. We cite the question numbers that appear on questionnaire.doc. However, question numbers for your data collection instrument may be somewhat different.

Note: All 10 of the indicators are restricted to living children and require accurate information about the child's age to be calculated correctly. For this reason, the portion of the questionnaire that collects information on both the date of birth (the most accurate method of calculating child age) and the reported age or the child in completed months is essential.

Summary table of indicator definitions, calculation and questions used to construct indicators

Indicator Name	Definition	Calculation
1. Timely initiation of breastfeeding (children 0-23 months)	Proportion of children 0-23 months who were put to the breast within one hour.	Numerator: Number of children 0-23 months who were put to the breast within the first hour of birth (Q6=00) Divided by Denominator: Number of children 0-23 months
2. Exclusive breastfeeding under 6 months	Proportion of infants 0-5 months of age who were fed exclusively with breastmilk in the past 24 hours.	Numerator: Number of infants 0-5 months who received breastmilk in the past 24 hours (Q8=1) and did not receive any other foods or liquids in the past 24 hours (all of Q9c-9j and all of Q10aa-qq=0) Divided by Denominator: Number of infants 0-5 months old
3. Timely complementary feeding	Percent of infants 6-9 months of age who receive breastmilk and a solid or semi-solid in the previous 24 hours. Solid, semi-solid and soft foods are defined as mushy or solid foods, not fluids.	Numerator: Number of infants 6-9 months who breastfed in the past 24 hours (Q8=1) and who also received at least one food in the past 24 hours (Q10aa-10qq=at least one food listed). Divided by Denominator: Number of breastfed infants 6-9 months
4. Introduction of solid, semi-solid or soft foods	Proportion of infants 6-8 months who receive solid, semi-solid or soft foods.	Numerator: Number of infants 6-8 months who received at least one food in the past 24 hours (Q11>0). Divided by Denominator: Number of infants 6-8 months

Indicator Name	Definition	Calculation
5. Continued breastfeeding at 1 year	Proportion of children 12-15 months old who are fed breastmilk.	Numerator: Number of children 12-15 months who received breastmilk in the past 24 hours (Q8=1) Divided by Denominator: Number of children 12-15 months An alternative indicator – suggested by WHO – is continued breastfeeding at 2 years of age (when children are 20-23 months old)
6. Minimum dietary diversity	Proportion of children 6-23 months who received food from 4 or more food groups in the past 24 hours. The 7 food groups used to calculate this indicator are: 1. Grain, roots tubers 2. Legumes and nuts 3. Dairy products (milk, yogurt, cheese) 4. Flesh foods (meat, fish, poultry, liver/organ meats) 5. Eggs 6. Vitamin A rich fruits and vegetables 7. Other fruits and vegetables	Numerator: Children from 6-23 months who received food from 4 or more of the 7 food groups in the past 24 hours as follows: Grains: Q10aa, Q10bb Legumes: Q10dd Dairy: Q9d, Q9e, Q9i, Q10mm Flesh foods: Q10hh-kk Eggs: Q10ll Vitamin A rich fruits and vegetables: Q10cc, ee, ff, oo Other fruits and vegetables: Q10gg Divided by Denominator: Number of children 6-23 months

Indicator Name	Definition	Calculation
7. Minimum meal frequency	Proportion of breastfed and non-breastfed children 6-23 months of age who receive solid, semi-solid or soft foods the minimum number of times or more.	There are two separate calculations for this indicator, one for breastfeed children and one for non-breastfed children. Number of breastfed children 6-23 months of age who received solid, semi-solid or soft foods the minimum number of times or more during the previous day Divided by Number of breastfed children 6-23 months old And Non-breastfed children 6-23 months of age who received solid, semi-solid or soft foods the minimum number of times or more during the previous day Divided by Non-breastfed children 6-23 months old Meals and snacks are listed in Q9c-9j and Q10aa-qq=0. See the definition in Chapter 6 for 'minimum' number of meals and snacks.

Indicator Name	Definition	Calculation
8. Minimum acceptable diet	Proportion of children 6-23 months of age who receive a minimum acceptable diet (apart from breastmilk) Calculate twice: For breastfed and non-breastfed children	**Breastfed children** Numerator: Breastfed children 6-23 months who had a least the minimum dietary diversity and minimum meal frequency in the past 24 hours Divided by Denominator: Breastfed children 6-23 months **Non-breastfed children** Numerator: Non-breastfed children 6-23 months who had a least the minimum dietary diversity and minimum meal frequency in the past 24 hours Divided by Denominator: Non-breastfed children 6-23 months For breastfed and non-breastfed children, use indicators 6 and 7 (above) to calculate minimum acceptable diet.
9. Consumption of iron-rich or iron-fortified foods	Proportion of children 6-23 months old who receive an iron-rich food or iron-fortified food that is specially designed for infants and young children or that is fortified in the home.	Numerator: Number of children 6-23 months who received at least one iron-rich or iron-fortified food (from the types of foods listed in Q10hh-10kk and Q12-Q15) Divided by Denominator: Number of children 6-23 months
10. Bottle feeding	Proportion of children 0-23 months who were fed with a bottle during the previous day.	Numerator: Children 0-23 months who were fed with a bottle during the previous 24 hours (Q16=1) Divided by Denominator: Number of children 0-23 months

Please note that per WHO guidance, we strongly recommend that you collect information about the child's diet in the previous 24 hours. This approach has been used in many settings around the world.

One of the questions (Question 6) refers to liquids that were given to the child in the first three days after delivery. Generally, these include plain water, sugar water, fruit juice, teas, honey, and infant formula. In the area where you work there may be other liquids that mothers give their newborns. We do not provide input here on calculating the proportion of children who receive prelacteals. However, the generic questionnaire we have included with this Guide has information on whether prelacteals were given. You can use this information to calculate this indicator – especially if the provision of prelacteals is common in the area where you work. The WHO 1991 guidance contains information about how to calculate this. To find out about other prelacteals that are given, you can:

- Ask staff and nutritionists from your organization
- Ask staff from other organizations
- Visit homes and talk with mothers or other primary caregivers

Questions 9 and 10 ask about the liquid and solid foods infants have received in the previous 24 hours. You need to modify this list to include foods that are available in the area where you are working. As noted by WHO guidance from 2009, we know that what people eat varies by geography, wealth, and custom. Therefore, IYCF Questions 9 and 10 must be reviewed and modified before each survey to reflect commonly consumed liquids and foods. These questions must be adapted with caution to ensure that all commonly consumed liquids and foods are identified and categorized correctly into the liquid/food groups listed on the generic questionnaire. If it is known that certain foods/groups are not eaten by anyone in the survey area, they can be deleted from the questionnaire. Please see Annex 4 (pages 66-81) of the WHO guidance, part 2, for more information about how to do this.

When modifying the questionnaire, it is important to ***make it easy for the mother to understand***. It is also important to make the questionnaire easy for the interviewer and data analysts to use, but making it understandable for mothers is most important.

This questionnaire is an example of an adapted questionnaire we developed for a specific region. In that region, the fruits and vegetables that were available included guava, bananas, oranges/lemons, watermelon, pineapple, tomato, onion, green pepper, cabbage and avocado. These are only examples. Your questionnaire may include different fruits and vegetables. **Thus, while you should use the questionnaire we have provided you, it will be necessary to modify the liquid and food categories to reflect liquids and foods available in your area.**

2.2 Thinking about categories of liquids and foods

As you adapt the questionnaire and modify the list of foods on your questionnaire, you must think about each category of food that is available and accessible to the population. The major categories are:

- Food from plants
  - Grains, roots and tubers
  - Legumes and nuts
  - Vitamin-A-rich fruits and vegetables
  - Other fruits and vegetables
- Food from animals
  - Dairy products (milk, yogurt, cheese)
  - Flesh foods (meat, fish, poultry and liver/organ meats)
  - Eggs

As you ask about foods from each category,

- **Name specific foods that your target population is familiar with.** For example, in the sample questionnaire, we ask about foods in the category of grains, roots and tubers. They appear on the questionnaire as very specific lists of grains, roots, and tubers available in the area: e.g. pasta, rice, maize, sorghum, bulgher, ugali, bread biscuits, chappati, anjera, mandazi, sweet potatoes, carrots, potatoes. Your questionnaire should replace these specific foods with the ones found in your area.
- **When asking about commercially available foods, use local brand names.** For example, this questionnaire lists brand names of milk such as COAST, HILAW, HAYAT, Black Cow and White Cow. Your questionnaire should replace these brand names with the brands in your area.

Experience with previous surveys has shown that classification into the vitamin A-rich fruit and vegetables groups can be challenging. The food list provided in Annex 4 (pages 66-81) of the WHO guidance, part 2 should be very helpful for this task.

As guidance from WHO makes clear, there is no food group on the generic questionnaire for distributed commodities. However, these foods are often listed separately, so that coverage can be assessed. If foods are iron-fortified to meet the needs of infants and young children, these foods should be listed twice: under Question 10, so as to be categorized (e.g. as grain-based food) and counted in the minimum dietary diversity indicator and also in Questions 12-15, as iron-fortified foods.

A copy of this portion of the questionnaire appears below. You should use this exact questionnaire format.

The purpose of the survey is to gain a greater understanding of infant feeding practices in the area where you work. The objective with respect to 24 hour dietary recall is to obtain a detailed understanding of what children eat. In some cases, implementing agencies may choose to shorten the list of foods described in Question 10. If you want, you can abbreviate Question 10 so that mothers are asked only about *any* solid/semi-solid food, without specifying which ones. In this case it is very important to include a set of standard probes, to be used with all respondents, to ensure that interviewers get full information on semi-solid and solid food. Also, the question on frequency of feeding of solids/semi-solids does not need to be asked for infants 0-5 months of age. See the Rapid CATCH 2000 for examples of how to do this.

You've done well! **Take a break and proceed to Chapter 3 on another day!**

Please tell me everything that [NAME] ate yesterday during the day or night (whether at home or outside the home). Think about when [NAME] first woke up yesterday. Did [NAME] eat anything at that time?

Keep probing **'Anything else?'** until the respondent says 'nothing else.' If nothing else was given when the child first got up, ask:

What did [NAME] do after that? Did [NAME] eat anything at that time?

If yes, ask: **Please tell me everything [NAME] ate at that time.** Probe: **'Anything else?'** until respondent says 'nothing else.'

Repeat questions above until respondent says the child went to sleep until the next day.

As the respondent recalls each food, underline the food in the food group below. If any one or more foods is mentioned in a food group, circle 'Y' in the column to the right.

If respondent mentions mixed dishes like a sauce or stew, probe: **What ingredients were in that [MIXED DISH]?** Probe: **'Anything else?'** Until respondent says 'nothing else.'

If a food recalled by the respondent is not listed in any of the food groups below, write the food in the box labeled 'other foods'.

If foods are used in small amounts for seasoning or as a condiment, include them under the condiments food group.

Once the respondent has finished recalling foods eaten, read each food group where '1' was NOT circled. ASK the question **'Yesterday, during the day or night, did [NAME] drink/eat any (FOOD GROUP ITEMS)?'** Circle '1' if respondent says YES, '2' if NO, and '8' if DON'T KNOW. **Every** line must have a code.

Other foods: please write down other foods in this box that the respondent mentioned but are not in the list below. When data are entered into the computer, the other foods will be assigned to one of the 7 categories of foods:

If at least one food from the food group has been given in the past 24 hours, circle 'Y' in the column below. If no food groups have been given, circle 'N.' If the respondent doesn't know, circle 'DK.'

Question			Last 24 hours		
	a	Bread, rice, noodles, or other foods made from grains, including thick grain-based porridge?	Y	N	DK
	b	White potatoes, white yams, manioc, cassava, or any other foods made from roots?	Y	N	DK
	c	Pumpkin, carrots, squash, or sweet potatoes that are yellow or orange inside?	Y	N	DK
	d	Any foods made from beans, peas, lentils or nuts, including Plumpy Nut?	Y	N	DK
	e	Any dark green leafy vegetables?	Y	N	DK

Question			Last 24 hours		
	f	Ripe mangoes, ripe papayas or (insert other local vitamin A-rich fruits)?	Y	N	DK
	g	Any other fruits or vegetables?	Y	N	DK
	h	Liver, kidney, heart or other organ meats?	Y	N	DK
	i	Any meat such as beef, pork, lamb, goat, chicken or duck?	Y	N	DK
	j	Fresh or dried fish, shellfish or seafood?	Y	N	DK
	k	Grubs, snails or insects?	Y	N	DK
	l	Eggs?	Y	N	DK
	m	Cheese, yogurt or other milk products?	Y	N	DK
	n	Any oil, fats or butter, or foods made with any of these?	Y	N	DK
	o	Foods made with red palm oil, red palm nut and red palm nut pulp sauce?	Y	N	DK
	p	Any sugary foods such as chocolates, sweets, candies, pastries, cakes or biscuits?	Y	N	DK
	q	Condiments for flavor such as chilies, spices, herbs or fish powder?	Y	N	DK

CHAPTER 3: SAMPLING

Goal: Calculate sample size and find the right number of people for your sample

By the end of this chapter, you will:

- Decide upon an appropriate sampling strategy based on your needs and circumstances
- Calculate sample size
- Determine how you will find the right number of individuals for your sample

Tool	Purpose of tool	Name of file
Sample size calculator	Help calculate sample size if you carry out a survey to obtain information on IYCF practices	Sample Size.doc

In this chapter, we talk about how to sample. We recommend that for one-time prevalence surveys and for baseline and endline surveys, data should be collected on each of the 10 IYCF indicators using sample sizes that you calculate. You will calculate your own sample sizes to establish the prevalence of IYCF behaviors with sufficient precision (if you are doing a one-time survey) or to detect significant differences in behaviors (if you are collecting baseline and endline data). You may choose to collect data on topics in addition to IYCF. However, for our purposes, we assume that one of your primary objectives is to find out about IYCF behaviors. Consequently, we tell you how to calculate sample size for infant and young child feeding assessments and surveys.

If resources are limited or there is some other compelling reason to use a different sampling strategy, you might consider alternatives to the sampling approach we propose. At the end of this chapter, we discuss ONE of the alternatives to sampling: 'piggybacking' or adding IYCF questions to nutrition (anthropometry) surveys. These are often done as part of a 30 x 30 cluster sample. While such an approach offers some advantages, we feel that calculating your own sample size is a better approach to sampling.

In other instances – for example, when you collect data on an annual basis – you may want to use a shorter questionnaire. This allows you to gauge progress toward achieving program targets and can be helpful to program staff as they think about issues that may need special attention.

You will need to answer a couple of questions before calculating sample size and find the right number of people for your sample. First, do you want a simple random sample, a systematic random sample, or cluster sampling? Second, are you conducting a one-time assessment of IYCF practices or will you collect data at several points in time (e.g., baseline and endline)?

Once these questions are answered, you can calculate sample size. Calculating sample size requires a number of assumptions about precision, power, design effect, subject refusals and so on. We will define these terms later. Once you know more about these concepts, you will need to either accept our assumptions about precision, power, design effect, and refusals or pick these numbers yourself.

Because knowing a little about sampling before making the decisions we have described above, we will give you background information. Then we will help you make decisions about what works best for you. Your decisions will affect your overall sample size.

3.1 What is sampling?

When you collect data – on infant and young child feeding or on any other topic – you have two options: interview all households or interview a sample. Very few implementing agencies interview everyone in a given population. It is too expensive and time-consuming. Most of the time, staff interview a *sample* of individuals. Using correct sampling procedures, you can collect data from *some* of the children (a sample) and make inferences about what is happening in the larger population. However, in order for a sample to be representative, you must follow correct procedures. Here we assume that you will interview a *sample* of individuals.

A note about what this chapter does not cover: This chapter *does not* provide a detailed summary of sampling. It also does not tell you how to randomly choose respondents once you are in the field. There are several good manuals you can consult for help with these tasks. These include:

> Magnini R. 1997. "Sampling Guide." Food and Nutrition Technical Assistance (FANTA): Washington DC.

Chapter 4 is helpful in understanding how to sample respondents.

Another good guide is:

> The Save the Children Fund. Emergency nutrition assessment: guidelines for field workers. Save the Children Fund, London, 2004.

Some of what we tell you below comes from the SMART (Standardized Monitoring and Assessment of Relief and Transitions) manual. If you are interested in more detail about sampling, the SMART manual is a very good place to start.

3.2 Answering questions about the sampling strategy you will use

As noted above, you will need to answer a couple of questions before you can calculate sample size.

3.2.1 Do you prefer a simple random sample, a systematic random sample, or cluster sampling?

There are many approaches to sampling. The three used most commonly by implementing agencies are simple random sampling, systematic random sampling, and cluster sampling. Cluster sampling is most commonly used but has some distinct disadvantages. We never recommend convenience sampling to collect data to describe IYCF practices at the population level.

Simple Random Sampling:

If you have a complete list of households in your area, you can conduct a simple random sample. A simple random sample requires random selection of households in a given site regardless of where individuals live. This is the "gold standard" (i.e., best approach) to sampling but you must have *complete*, up-to-date records on *every* household with children less than 2 years of age. Often, program staff does not have these data.

Systematic Random Sampling:

Another approach is to conduct systematic sampling. As noted in the SMART manual (page 42):

"Systematic random sampling is used when there is either a list of households or where the population is geographically concentrated and all the dwellings are arranged in a regular geometric pattern. Such a situation may occur in a camp where tents are pitched row after row, in blocks of flats, where streets are laid out in a grid, or where the houses are all along the edge of a river, coast, road, or other major feature. The first household is chosen at random. The subsequent households are visited systematically using a 'sampling interval.' You may want to consider systematic random sampling if the houses in the population of interest are arranged in a systematic way."

One of the advantages of systematic random sampling is that you do need accurate population data.

Cluster Sampling:

Cluster sampling is the approach to sampling that we recommend the least but has been used frequently. You use it when there is no particular structure to how households are distributed. Because many implementing agencies use cluster sampling a lot (for example, when they collect data on anthropometry) and because there is a great deal of information about how to do cluster sampling (see references in Chapter 1), we will not say more about it here but will help you decide whether it is right for you. The table below compares each of the sampling options.

Cluster sampling affects overall sample size as well as how children are found for interview and measurement. Below we tell you how you can adjust your sample size to account for the effects of clustering. We will also tell you how to find children using cluster sampling.

Type of sampling	Use this strategy if:
Simple random sampling	You used this approach for a previous survey of this same project (see Section 3.2.2 for a discussion of one-time data collection versus gathering data at multiple points in time) In addition to assessing IYCF behaviors (either at one point in time or at multiple points), you also intend to publish your findings or explore some topics in detail You are not yet certain how may infants and young children should be in your sample; you want to determine this on your own You have *complete*, up-to-date records on *every* household You want to minimize sampling error You want to use the 'gold standard' in sampling
Systematic random sampling	You used this approach for a previous survey of this same project In addition to assessing IYCF behaviors (either at one point in time or at multiple points), you also intend to publish your findings or explore some topics in greater detail You are not yet certain how may infants and young children should be in your sample; you want to determine this on your own You have 1) A list of households or 2) A population that is geographically concentrated and all the dwellings are arranged in a regular geometric pattern You want to minimize sampling error You want to use a powerful approach to sampling
Cluster sampling	You used this approach for a previous survey of this same project Your primary purpose is to understand infant feeding patterns at one point in time or at multiple points (for example, baseline and endline), to see if your program has had an impact

3.2.2 Are you conducting a one-time assessment of IYCF practices or will you collect data at several points in time?

Sometimes implementing agencies conduct a one-time assessment of IYCF practices. They do this because they want to understand what is happening *now*. They want to know about the prevalence of each of the 10 recommended IYCF practices. They use one-time surveys to more accurately target at-risk individuals for intervention. We use the term "assessment" to describe these activities.

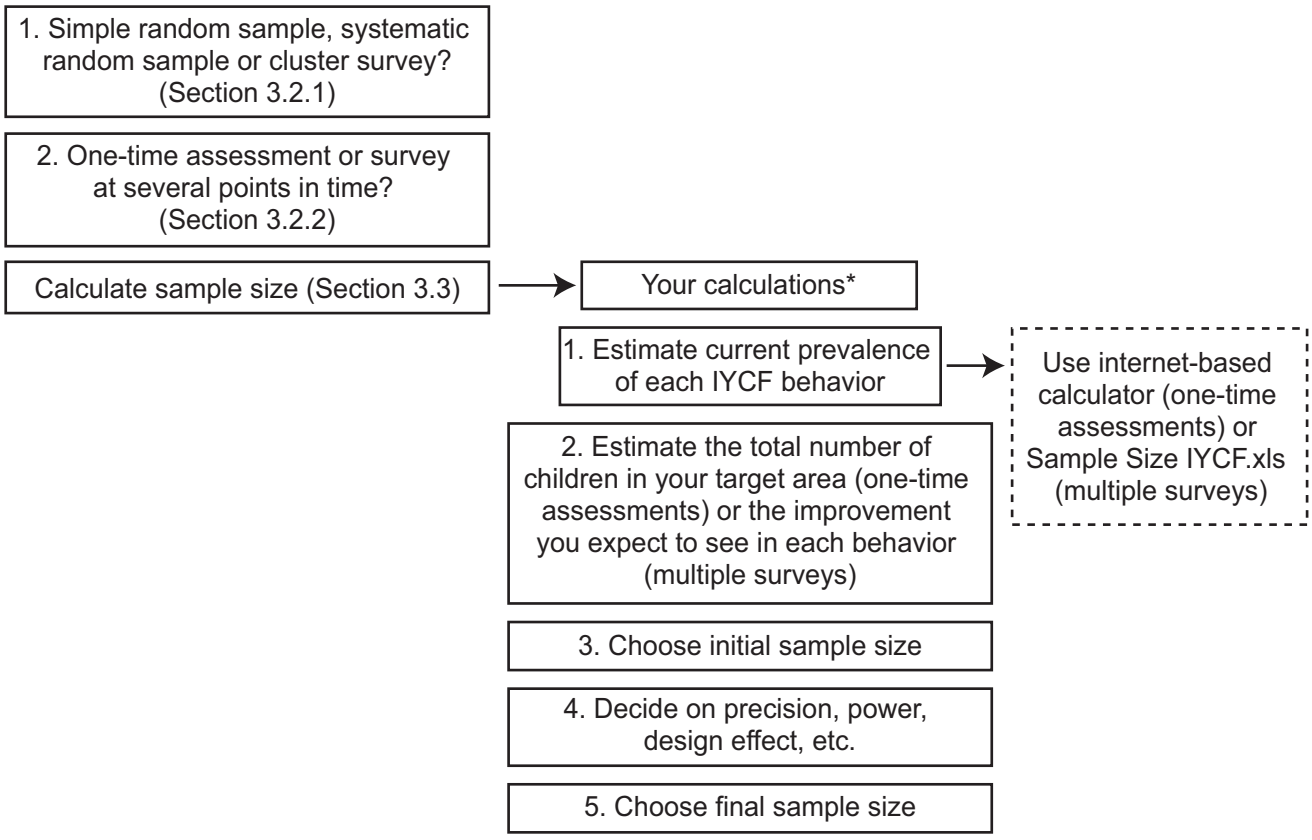
At other times implementing agencies conduct surveys at several different points in time. Usually this occurs at baseline (before an intervention begins) and at endline (when the intervention is over) but surveys can take place in the middle of a program as well. Often, the major purpose of collecting IYCF data at multiple points is to evaluate the impact of a project or program.

If you conduct a one-time assessment of IYCF behaviors, you will use one type of sample size calculator to determine sample size. If you carry out multiple surveys, you will use another sample size calculator. Because the purpose of your assessment and surveys is different and because the sample size calculators themselves are different, the number of respondents required will not be the same for one-time assessments and surveys at multiple points.

The diagram below helps you decide which strategy you'll use and where you can find information about next steps. Since most of what we discuss in this chapter is relevant to everyone, all should read this chapter in its entirety.

Please note that if you are collecting data at several points, you should use the same strategy now that you have used in the past. Choosing the same sampling strategy helps ensure comparability between surveys across time.

Steps for determining sample size:



If you need additional information before making a decision, please consult the manuals we recommend above. Be sure to consider other factors that influence which sampling strategy you use. These issues include cost, staff availability, time availability, desire for precision, and so on.

3.3 Sample size

Regardless of how you answered the two questions above, you need to calculate sample size before collecting data.

3.3.1 The five steps to calculating sample size

Everyone needs to follow similar steps to calculate the size of their sample. The five steps are:

1. **Estimate** the percent of mothers currently practicing each of the IYCF behaviors. If you are conducting a *one-time assessment*, you will also need to estimate the total number of children in your target area. If you are conducting *multiple surveys*, you will also need to estimate the improvement you expect to see in each of these practices by the end of your project.
2. **Enter** these estimates into a sample size calculator.
3. **Choose** your initial sample size.
4. **Decide** how other factors such as precision and power will affect sample size.
5. **Choose** your final sample size.

Here is how you carry out each of the steps:

Step 1: Estimate the percent of mothers practicing each of the IYCF behaviors

Before you can calculate your sample size, you need to estimate the current prevalence for each of the 10 IYCF indicators in your target group. These indicators are:

- Timely initiation of breastfeeding (children 0-23 months)
- Exclusive breastfeeding under 6 months
- Timely complementary feeding
- Introduction of solid, semi-solid or soft foods
- Continued breastfeeding (at 1 year)
- Minimum dietary diversity
- Minimum meal frequency
- Minimum acceptable diet
- Consumption of iron-rich or iron-fortified foods
- Bottle feeding

WHO has published two documents that give guidance on IYCF indicators. These documents were published in 1991 and 2009. In this Guide, we use the new guidance (2009). For two indicators – timely initiation of breastfeeding (children 0-23 months) and timely complementary feeding – we use the 1991 guidance. You will find more details about all of these indicators in Chapter 6 and the following document:

> Indicators for assessing infant and young child feeding practices, Part 1 Definitions: Conclusions of a consensus meeting held 6-8 November 2007 in Washington DC. This is a joint publication of USAID, AED, UC Davis, UNICEF, IFRPI and WHO. This is the 2009 manual referenced above.

Here it is worth mentioning that *timely* initiation is calculated on *living* children. It is different from *early* initiation (the indicator WHO recommends) which measures the number of *living and deceased* children who received breastmilk in the first hour of life.

Data sources for estimating current prevalence:

In most programs there will be at least 2 sources of useful data for the 10 indicators:

- Existing project records. These include any monitoring you have carried out already, including previous surveys that have data on similar indicators.
- Other surveys. These include surveys carried out by the government, by NGOs, by UN agencies (including the UNICEF MICS), as well as DHS surveys (described below).

You may choose to use these surveys if:

- You do not have your own data to estimate the prevalence of IYCF behaviors in your target population.
- You are not confident about the data you have.

Be aware of the quality of data from other sources. Sometimes these data do not reflect the population's diversity or provide enough information to allow you to calculate some of the other indicators (e.g., indicators for children 6-9 months of age).

Where possible, you should use survey data from your program population. Please note that IYCF behaviors among refugee populations may more closely resemble the IYCF behaviors from their country of origin rather than those of their country of residence.

Here is a description of different types of national-level data, should you choose to use them:

The DHS (Demographic and Health Survey) is one of the best sources of national IYCF data. Most DHS reports have information on timely initiation, exclusive breastfeeding, and continued breastfeeding. Some may report on timely and appropriate complementary feeding as well as other indicators. To review DHS data on IYCF indicators in your country, go to www.measuredhs.com and click on "StatCompiler." Follow the instructions or use available hard copies of DHS reports you have access to (these reports are also available at www.measuredhs.com). You may need to use DHS data from the country of origin of the refugee population, not the country where refugees now live.

You can also use the UNICEF MICS (Multiple Indicator Cluster Survey) data. To use MICS data, go to: www.childinfo.org/nutrition.html. This website contains information on exclusive breastfeeding, timely complementary feeding and continued breastfeeding but not timely initiation of breastfeeding. You can also access a complete report, by country at: www.childinfo.org/mics3_surveys.html

Using data from another organization

The biggest problem with using data from another organization (for example, an NGO or a UN agency) is that you may not know how staff collected data. Sampling correctly and collecting, managing, and analyzing data can be difficult. Do not assume the other organizations collected data in a satisfactory manner. Examine the report for details of sampling, questions used, data cleaning criteria, and indicator definitions that we use in this Guide.

In the end, take a look at all of your numbers (data from NGOs and UN agencies and MICS and DHS surveys, described below). In some cases, you may need to make an educated guess about IYCF behaviors. If you are really uncertain about the prevalence of a behavior, choose 50% prevalence. By choosing 50%, you maximize your sample size, thus increasing the cost of data collection and entry. Percents lower or higher than 50% yield smaller sample sizes.

Please note that MICS and DHS data represent the entire country, NOT the specific region where you work. IYCF practices in the area where you work may differ from IYCF practices in other areas of the country. Even so, you should use MICS and DHS data as a check for your estimates. Are MICS and DHS percents for each of the indicators a lot higher or lower than the percents for your project area? Sometimes this is okay if you have reason to believe they should be different. However, very large differences could mean that your data are not reliable.

When considering whether or not to consult DHS or MICS data, you may experience these problems:

- There may not be a DHS or MICS survey for your country of interest.
- The DHS or MICS survey may be old.
- Survey data may not represent IYCF behaviors in the area where you work.

Write down current rates for each of the 10 IYCF behaviors in ***your*** area.

Indicator	Current rate
Timely initiation of breastfeeding (children 0-23 months)	
Exclusive breastfeeding under 6 months	
Timely complementary feeding	
Introduction of solid, semi-solid or soft foods	
Continued breastfeeding at 1 year	
Minimum dietary diversity	
Minimum meal frequency	
Minimum acceptable diet	
Consumption of iron-rich or iron-fortified foods	
Bottle feeding	

The second part of Step 1 is as follows:

- **For one-time assessments:** Estimate the total number of children in your target area.
- **For multiple surveys:** Estimate the improvement you expect to see in each behavior by the end of your project.

Because instructions for one-time assessments and multiple surveys differ, please read them carefully!

Step 1.1: For those of you who are conducting a one-time assessment

In order to calculate sample size, estimations are needed of the prevalence of each of the 10 indicators *and* of the number of infants and children in your study area. If you followed the steps above, you already have estimates for each of the 10 indicators. Now estimate the number of individuals in the *entire* population where you work (your program's target area).

Because all users of this manual are collecting IYCF data on children 0-23 months of age, we will describe how to come up with the total number of children 0-23 months old. In many populations in Southern ('developing') countries, the percent of children in each 6-month age interval is about 2%. You can multiply 2% by the number of 6 month intervals in the age group you will assess. Based on this estimate, the percentage of children 0-23 months of age is 8% of the entire population (2% x four 6-month age intervals).

Now multiply 8% by your total population. Let us assume you have a population of 60,000. You estimate that there are 4,800 children 0-23 months of age (8% x 60,000). Remember that when you use percents, you are using decimal points. Thus, you enter 8% into a calculator like this: .08.

Step 1.2: For those of you who are collecting data at multiple periods

You MUST estimate changes in proportions for each of the 10 indicators. There is no rule about how much change you should expect to see by the end of your project. If you have results showing the changes in behavior that another agency has achieved, use that information to see how much improvement there was between baseline and endline. If you do not have this information, assume improvements in behaviors by *at least* 5% per year. If you assume a change of 5% per year, you will have:

- 5% in 1 year
- 10% in 2 years
- 15% in 3 years
- 20% in 4 years, and
- 25% in 5 years

This means for example that if your exclusive breastfeeding rate is 27% at baseline, you should estimate that it will be 32% after 1 year, 37% after 2 years, 42% after 3 years and so on.

In most cases, it is unreasonable to expect changes greater than this. However, some projects have been quite successful. See for example: "Experience LINKAGES: Results" at: http://www.linkagesproject.org/publications/index.php?detail=23

It is often easier to change some behaviors – for example, timely initiation of breastfeeding (children 0-23 months) – than others (exclusive breastfeeding). The expected increase you anticipate for one behavior does not have to be the same as for the other behaviors.

Write down current estimated and expected rates for each of the IYCF behaviors for ***your*** project here:

Indicator	Current prevalence	New proportion you expect to see as a result of your project
Timely initiation of breastfeeding (children 0-23 months)		
Exclusive breastfeeding under 6 months		
Timely complementary feeding		
Introduction of solid, semi-solid or soft foods		
Continued breastfeeding at 1 year		
Minimum dietary diversity		
Minimum meal frequency		
Minimum acceptable diet		
Consumption of iron-rich or iron-fortified foods		
Bottle feeding		

Remember that programs try to *reduce* bottle use so your estimated change for this indicator should be *lower* rather than higher (for example, dropping from 37% at baseline to 22% at follow-up). For all of the other behaviors, anticipate an *increase*.

Step 2: Enter these estimates in a sample size calculator

Now that you have your estimates, you will need to enter your numbers into a sample size calculator. We will discuss 2 sample size calculators: 1) Calculators for one-time assessments, and 2) Calculators for multiple surveys (e.g., baseline and endline).

One-time assessments

There are many good calculators for determining sample size for *one-time assessment*s. One that is excellent and free of charge is GNU sampsize. You can access this calculator at the following website: sampsize.sourceforge.net/iface/index.html#prev

Access the calculator by pointing your mouse on the address above and while holding down the control button, click on the link itself. Or, if you prefer, you can simply copy and paste the link into your web browser. Once you have gone to the link, you will see that the calculator requires 4 pieces of information, including:

1. Precision
2. Prevalence
3. Population, and
4. Level

Unless you want to make assumptions that are different from our own, assign a 'Precision' of 5% and a 'Level' of 95%. The remaining two pieces of information you will need (prevalence and population) come from step 1, above. Let us take timely initiation of breastfeeding (children 0-23 months) as an example. In GNU sampsize, type in the following numbers:

Precision=5

Prevalence=45 (this is our estimated prevalence of timely initiation of breastfeeding (children 0-23 months))

Population=4800

Level=95

When you enter these numbers and click on 'calculate,' you get a sample size of 353. Your output will look something like this:

Sample size results:

Assumptions:

Precision=5.00%

Prevalence=45.00%

Population size=4800

95% Confidence Interval specified limits [40%-50%]
(these limits equal prevalence plus or minus precision)

Estimated sample size: n=353

Now that you've calculated sample size for timely initiation, determine sample sizes for the remaining 9 indicators. For this exercise, assume the following levels of prevalence:

Indicator	Current prevalence
Timely initiation of breastfeeding (children 0-23 months)	45%
Exclusive breastfeeding under 6 months	23%
Timely complementary feeding	57%
Introduction of solid, semi-solid or soft foods	60%
Continued breastfeeding at 1 year	40%
Minimum dietary diversity	21%
Minimum meal frequency	33%
Minimum acceptable diet	17%
Consumption of iron-rich or iron-fortified foods	10%
Bottle feeding	37%

You should get the following results:

Indicator	Current prevalence	Sample size needed
Timely initiation of breastfeeding (children 0-23 months)	45%	353
Exclusive breastfeeding under 6 months	23%	258
Timely complementary feeding	57%	350
Introduction of solid, semi-solid or soft foods	60%	343
Continued breastfeeding at 1 year	40%	343
Minimum dietary diversity	21%	243
Minimum meal frequency	33%	318
Minimum acceptable diet	17%	208
Consumption of iron-rich or iron-fortified foods	10%	135
Bottle feeding	37%	334

According to the Save the Children Fund's manual, 'Emergency nutrition assessment: guidelines for field workers,' "Note that a common misunderstanding about sample sizes is that the size of the sample needed for a 'good' nutrition survey is dependent on the population size … Population size only affects sample size in populations less than 5000" (page 97).

The Sampsize calculator asks about the total size of the population so that it can make an adjustment in the case of a small population size. For most population sizes, the only factors that will affect the sample size are the expected precision, the expected prevalence, and the design effect.

If you are not doing multiple surveys, skip to step 3. Otherwise, read on!

Multiple surveys

We have developed a sample size calculator for IYCF *surveys that occur more than once*. It is called Sample Size IYCF.xls and it comes with this Guide.

We will use an example with fictitious data for a project lasting 3 years. In some cases, we do not follow the 5% per year rule. We do this to demonstrate that sometimes you will change IYCF behaviors by more or less than 5% per year.

Try out the calculator by entering the current and expected prevalence for timely initiation of breastfeeding (children 0-23 months). Using fictitious data, it is as follows:

Indicator	Current prevalence	Estimated change in prevalence after 3 years
Timely initiation of breastfeeding (children 0-23 months)	45%	65%

You will enter 45 and 65 in cells D8 and E8, respectively. This yields a sample size of 146. Now try it with the remaining percentages:

Indicator	Current prevalence	Estimated change in prevalence after 3 years
Timely initiation of breastfeeding (children 0-23 months)	45%	65%
Exclusive breastfeeding under 6 months	23%	38%
Timely complementary feeding	57%	72%
Introduction of solid, semi-solid or soft foods	60%	70%
Continued breastfeeding at 1 year	40%	55%
Minimum dietary diversity	21%	36%
Minimum meal frequency	33%	48%
Minimum acceptable diet	17%	32%
Consumption of iron-rich or iron-fortified foods	10%	25%
Bottle feeding	37%	22%

For the time being, do not worry about the blue columns labeled precision, power, design effect, and individual response rate. You should get these results:

Indicator	Number of children required per indicator
Timely initiation of breastfeeding (children 0-23 months)	97
Exclusive breastfeeding under 6 months	148
Timely complementary feeding	160
Introduction of solid, semi-solid or soft foods	357
Continued breastfeeding at 1 year	174
Minimum dietary diversity	142
Minimum meal frequency	168
Minimum acceptable diet	129
Consumption of iron-rich or iron-fortified foods	101
Bottle feeding	145

Notice that as you enter numbers, some of the other cells not shaded in blue change their values. Do not worry about this. Excel is performing calculations.

Step 3: Choose your initial sample size

Everyone should read this section.

The age ranges of some indicators are very narrow. For example, the indicator, 'introduction of solid, semi-solid and soft foods' includes only children who are 6-8 months of age. There may not be enough children 6-8 months of age living in your target area to reach that sample size. Even if there are, when you calculate sample size on a narrow age range, you need lots of children of that age before you can have confidence in the results. To keep overall sample sizes reasonable, we recommend that you make final decisions about sample size based on indicators with *wide* age ranges. There are seven of them:

1. Timely initiation of breastfeeding (children 0-23 months)
2. Exclusive breastfeeding under 6 months
3. Minimum dietary diversity
4. Minimum meal frequency
5. Minimum acceptable diet
6. Consumption of iron-rich or iron-fortified foods
7. Bottle feeding

Choosing your initial sample size if you are conducting a one-time assessment

Look at sample size calculations for these 7 indicators and choose the largest one. If you are doing a *one-time assessment*, the largest of the following numbers:

Indicator	Current prevalence	Sample size needed
Timely initiation of breastfeeding (children 0-23 months)	45%	353
Exclusive breastfeeding under 6 months	23%	258
Minimum dietary diversity	21%	243
Minimum meal frequency	33%	318
Minimum acceptable diet	17%	208
Consumption of iron-rich or iron-fortified foods	10%	135
Bottle feeding	37%	334

is 353. If you were carrying out a survey based on our fictitious data, you would select 1412 (353 x 4) children 0-23 months of age.

Why multiply sample size calculations by 4?

There are 4 age categories for children 0-23 months old: 0-5, 6-11, 12-17 and 18-23. The smallest age range for any of the 7 indicators above is 6 months. While some age ranges are larger, we have made a conservative estimate that you will need the calculated number of children *per 6-month age group*. If you do not multiply your estimated sample size by 4, you will be able to say much statistically about key indicators such as exclusive breastfeeding that covers a 6 month time span. Using the example above, you will need 353 children 0-5 months of age, 353 children 6-11 months of age, 353 children 12-17 months of age and 353 children 18-23 months of age.

Choosing your initial sample size if you are conducting multiple surveys

Let us move on to multiple assessments. If you are doing *multiple surveys*, the largest of the following numbers:

Indicator	Number of children required per indicator
Timely initiation of breastfeeding (children 0-23 months)	97
Exclusive breastfeeding under 6 months	148
\Minimum dietary diversity	142
Minimum meal frequency	168
Minimum acceptable diet	129
Consumption of iron-rich or iron-fortified foods	101
Bottle feeding	145

is 168. Notice that in the sample size calculator we gave you for multiple surveys, we have created purple boxes where you can look to determine the number on which your final sample size will be based (i.e., the highest number in the purple boxes in column Px4). If you were carrying out a survey based on our fictitious data, you would select 672 children 0-23 months of age (168 children in each of the age categories: 0-5, 6-11, 12-17 and 18-23 months).

If you are carrying out multiple assessments, see the comment immediately above which gives an explanation for: ***Why multiply sample size calculations by 4?***

Step 4: Decide how other factors such as precision and power will affect sample size

Precision (also known as level of significance) refers to the possibility of rejecting a "correct" null hypothesis (called alpha; often we use the symbol α to represent alpha). As a reminder: the null hypothesis states there is no difference in feeding practices between two groups. The alternative hypothesis (the one you hope to see) is that there IS a difference in feeding practices. In other words, α is the chance that we conclude, based on our data, that there IS a difference, when in fact there is NO difference between the groups.

Power is the possibility of NOT rejecting a "false" null hypothesis (called beta; often we use the symbol β to represent beta). In other words, β is the chance that we conclude that there is NO difference, when in fact there IS a difference (the power to detect a small difference).

To reduce the possibility of rejecting a "correct" null hypothesis, we want to increase **precision**. To reduce the possibility of NOT rejecting a "false" null hypothesis, we want to increase **power**. In short, the more precision and power you have, the more confident you are of your results. But this also means a larger sample – in some cases, a much, much larger sample! You will need to weigh the need for power and precision against your budget and the feasibility of carrying out data collection on a large sample.

Instructions for modifying assumptions about precision and power if you are conducting a one-time assessment

If you want to use your assumptions for power and confidence (and not ours), type in the number you feel most appropriate for power (box 1 in GNU sampsize located at: sampsize.sourceforge.net/iface/index.html#prev) and level (box 4 in GNU sampsize). A good starting point is 5 for power and 95 for level. However, if the resulting sample is simply too big for you, you will need to reduce power, level, or both. You can play around with the numbers to see how changing one or both of them affects sample size. Be careful to not reduce power and level too much; otherwise, you won't be able to say anything meaningful about your results. We recommend you not enter a number greater than 10 for precision (i.e., no less than 90%) and a number lower than 10 (i.e., no less than 80%) for power.

Two other factors that affect your overall sample are design effect, and "refusals/ incomplete surveys." We will discuss those factors in a minute.

Instructions for modifying assumptions about precision and power if you are conducting an assessment multiple times

Default settings for precision and power (95% and 80%, respectively) are already included in our sample size calculator. Be careful to not reduce power and level too much; otherwise, you will not be able to say anything meaningful about your results. We recommend you not enter a number greater than 10 for precision (i.e., no less than 90%) and a number lower than 10 (i.e., no less than 80%) for power.

As we have already mentioned, 2 other factors you'll need to think about are design effect and refusals/incompletes. We will define these terms in a minute. As noted below, if you are conducting a simple random sample or a systematic random sample, you do not need to worry about design effect but you *do* need to account for refusals and incompletes. On the other hand, if you are using cluster sampling, you *do* need to worry about design effect but *don't* need to worry about refusals and incompletes: the cluster sampling framework already accounts for this.

If you use this sampling strategy	**Do this:**	**Refusals and incompletes** (individual response rate)	**Design effect.** Multiply your sample size calculation by this much using a regular calculator.
Simple random	→	Increase your sample size by *at least* 10%	Do nothing; there is no design effect
Systematic random	→	Increase your sample size by *at least* 10%	Do nothing; there is no design effect
Cluster	→	Do nothing; cluster sampling accounts for attrition	Determine design effect. Usually this number is between 1.5 and 3.0

Design effect is the correction you make when you use cluster sampling. As noted in the SMART manual (page 44), "Cluster sampling requires a larger sample size than for simple or systematic random sampling. This is because subjects within the same cluster are generally more similar to each other than to members of different clusters, which results in a decrease in precision." You compensate for this flaw by increasing your sample size through the design effect.

In simple and systematic random samples, there is no design effect. In cluster samples, you'll need to account for it. There are lots of opinions about how large this effect should be. Many people think that the design effect for cluster sampling should be between 1.5 and 3. Let us assume for a moment that we estimate a design effect of 1.5. This means that whatever our sample size calculations, the final sample must be 50% larger than what we originally estimated.

Instructions for modifying assumptions about design effect

For *one-time assessments*, the GNU sampsize calculator does not allow you to account for design effect. This is not a problem if you are conducting a simple random sample or a systematic sample. However, if you are carrying out a cluster survey that is different from a 30 x 30 cluster design (described at the end of this chapter), you will need to account for design effect. This is easily remedied. Once you have calculated your sample size using GNU sampsize, multiply it by your design effect using a regular calculator. So, if your calculation of sample size yielded 1412 and you assume a design effect of 1.5, simply multiply 1412x1.5=2118. This is obviously a large sample. You may need to go back and modify some of your other assumptions so that you have a more reasonable sample size.

For *assessments occurring at multiple time periods*, using our sample size calculator, simply type your design effect in column N. This should be a number between 1.5 and 3. The design effect must be the same for all 10 indicators.

Instructions for modifying assumptions about refusals and incompletes

Not all mothers/caregivers will agree to participate in your survey. These are called "refusals." Some mothers will start the survey but not complete it, so you cannot use information from those women either. "Incompletes" refer to respondents who stop part way through the survey.

You should increase your sample size by at least 10% to account for refusals and incompletes.

If you are conducting a *one-time assessment*, simply multiply your sample size by the expected level of refusals and incompletes. You can perform this calculation using a regular calculator. If, for example, you expect that your sample size needs to be 10% higher than originally projected because of refusals and incompletes, your final sample size will be: 388 (353x1.10), *per 6-month age category* (i.e., you'll need to multiply this sample size by 4).

If you are carrying out an evaluation at multiple time periods, simply change your individual response rate in column O of the sample size calculator that comes with this Guide. If you assume a 10% rate for refusals and incompletes, type 110% in all of column O. Excel will give you your final sample size, automatically accounting for refusals and incompletes.

A note about response rates: not all mothers/caregivers will agree to participate in the survey. These are called "refusals." Some mothers will start the survey but not complete it so you can't use information from those women either. "Incompletes" refer to respondents who stop part way through the survey. For example, the interviewer is only able to ask a third or half the questions. These mothers/caregivers should be eliminated from your dataset. Later, we will tell you how to do this. Sometimes, mothers complete the entire survey but leave a few questions unanswered. For the purposes of sampling, you do not need to worry about mothers who complete almost all of the questions. They will be included in your final dataset.

Step 5: Choose your final sample size

Whether you conduct a one-time assessment or a survey at multiple points, this is the highest number among the 7 IYCF indicators we talked about in step 3, above. You increased this number as part of step 4. Look at your modified sample sizes for all 7 indicators and choose the largest number.

Below, we tell you how to choose households for your survey whether you've chosen simple random sampling, systematic random sampling or a cluster survey.

If you are conducting a baseline survey for a program that operates in a number of different sites (villages, refugee camps, etc.), you now need to decide whether you want results that give the prevalence for your entire program or results for individual sites. You have three options, outlined below.

Option	Assumptions and limitations	Sample size
1. Collect data from one site	IYCF behaviors in one site are similar to IYCF behaviors in other sites (this is a *big* assumption)	You will collect information in one site using the sample size you've already calculated
2. Collect data from multiple sites	You will not be able to say anything about statistical differences between sites	You will collect information from 2 or more sites using the sample size you've already calculated. In other words, you will collect data from multiple sites to reach your total sample size
3. Collect data from multiple sites with a full sample size in each site	This option is expensive	You will collect information from 2 or more sites using the sample size you've already calculated. Each site has the full sample size

In brief, adequate sample sizes from each area allow you to make comparisons between sites and arrive at statistically meaningful conclusions about differences. But conducting large surveys in multiple sites can be expensive and time consuming.

3.4 Finding your sample

Now that you have calculated sample size, the next step is to find the individuals in your sample. Go to the appropriate section below:

Simple random sampling	3.4.1 Choosing your households with simple random sampling
Systematic random sampling	3.4.2 Choosing your households with systematic random sampling
Cluster sampling	3.4.3 Choosing your households with cluster sampling

3.4.1 Choosing your households with simple random sampling

Simple random sampling works best among small populations where you have detailed information that allows you to locate randomly chosen households. To find households for your sample, carry out the following 3 steps (see page 50 of the SMART manual for more detailed instructions):

1. Determine the number of households that should be visited. If you already have detailed information about how many children less than 2 years of age live in each household, you do not need to carry out this task or task 2, below:
   - Calculate the average number of children per household. You can get information on the total number of children per household from your census or from another complete listing of children. Divide the total number of children by the number of households in your population.
   - Divide your calculated sample size (step 5 under Section 3.3. Sample size, above) by the average number of children 0-23 months old per household. In our example above for surveys at multiple points, there are 1008 children 0-23 months old in our entire sample. Assuming there are 0.5 children 0-23 months old per household, the number of households we need to visit is: 1008/0.5=2016.

2. Use the Nutrisurvey software included in the SMART manual to determine which households will be visited (Nutrisurvey which is free of charge and can be found at www.smartmethodology.org).
   - In Nutrisurvey, enter the total number of households in the population (i.e., the total number of households in your *entire* target area) as well as the number required for the sample. Nutrisurvey automatically selects the required house numbers at random. It then sorts your list in ascending order so that visiting each house is easier.

3. Collect IYCF data on each of the selected households.

Important: Do not worry about collecting data from a fixed number of children 0-5 months of age, 6-11 months of age, 12-17 months old and 18-23 months old. Given that you are sampling randomly, on average, your sample should have similar numbers of children in each age group. If not, do not worry about it unless there are far more children of certain age groups than others. This may suggest that systematic random sampling was done incorrectly.

Also, you will collect IYCF data on *every* child 0-23 months old. When you encounter multiple births (for example, twins and triplets) or multiple children 0-23 months old in a given household, interview mothers/primary caregivers of all age-eligible children in the family (a new questionnaire for each child).

3.4.2 Choosing your households with systematic random sampling

Once you have calculated your sample size, you are ready to choose your households using the following 8 steps (also see page 50 of the SMART manual). If you have complete census information on every house with a child less than 2 years of age, only completion of steps 1, 2, 6 and 7 will be necessary.

1. Pull out a map of the site where you will collect data. In order for you to use systematic random sampling, your map has to include *all* households in your target area. *Incomplete maps cannot be used for systematic random sampling.*

2. On your map draw a continuous route that passes in front of *every* house. See Figure 2 on page 51 of the SMART manual for an example.

3. Determine the number of inhabitants and the number of households in the population. Let us assume there are 60,000 people and 20,000 households in our target area.

4. Estimate the number of children 0-23 months old in the population (e.g., 4800 from the example above). You can estimate this based on local knowledge of the average number of children per household (check multiple households to make sure you get a good estimate). Alternatively, you can use the proportion of children in the population. In some countries, about 8% of the total population is children 0-23 months of age.

5. Determine the number of households that should be visited for your IYCF survey:
   - If you know the average number of children per household, there is no need to calculate it. If you do not know the average number, you can divide the number of children 0-23 months old by the number of households. Here is the number for our fictitious data above: 20,000/60,000=.3. In other words, we estimate that there are 0.3 children per household. If we need a sample size of 1008 children, we will need to go to 3360 households (1008/0.17=3360).

6. Now you are ready to determine your sampling interval. The sampling interval is usually a small number of houses, one of which is randomly selected as a starting point, followed by systematic sampling of households in each subsequent interval. To calculate *your* sampling interval, divide the total number of households by the number that should be visited. Again, using our example, there are 20,000 households and we must visit 3360 of them. Our sampling interval is (20,000/3360=5.9). The SMART manual suggests you add 10% to this number to make sure you have an adequate sample size. In this case, an additional 10% moves the interval from 5.9 to 6.5. The SMART manual also recommends you round your number down to the nearest whole number. In this case, the nearest whole number rounded down is 6. By rounding down, you are certain you get enough children. Thus, using fictitious data, we calculate that we need to visit 1 in every 6 houses to reach our sample size of 1008.

7. Now select the first household to be visited. You can do this using the Nutrisurvey software or by writing numbers representing your sampling interval on separate pieces of paper, folding them, and drawing one randomly. In this case, we need to randomly select a number between 1 and 6. If you randomly select 3, the third house is your starting point. Once you've interviewed in your first randomly selected house, add the sampling interval to the first house you visited. In this example, your next house will be number 9 (household 3 was randomly selected and our sampling interval is 6; in other words, 3+6=9). The third house in our survey is house number 15 (9+6=15). The fourth house is number 21 (15+6=21) and so on.

Important: Do not worry about collecting data from a fixed number of children 0-5 months of age, 6-11 months of age, 12-17 months old and 18-23 months old. Given that you are sampling randomly, on average, you should have similar numbers of children in each age group. If you do not have similar numbers in each age group, do not worry, unless there are far more children of certain age groups than others. This may suggest that systematic random sampling was done incorrectly.

Also, you will collect IYCF data on *every* child 0-23 months old. When you encounter multiple births (for example, twins and triplets) or multiple children 0-23 months old in a given household, interview mothers/primary caregivers of all age-eligible children in the family (a new questionnaire for each child).

3.4.3 Choosing your households with cluster sampling

Because cluster designs are so widely used, we will spend very little time here talking about how to select clusters and households within clusters. In this section, we will talk about:

- Deciding how many clusters to choose
- Determining how many children to choose per cluster
- Locating children for interview

3.4.3.1 Deciding how many clusters to choose

The more clusters you choose (and the fewer respondents per cluster) the more you reduce design effect. Here, the important thing to remember is that *you should choose no less than 30 clusters*. Choosing less than 30 clusters increases the design effect and, more importantly, the chances that your sampling strategy will be flawed. If you use cluster sampling, it is better to have more clusters with fewer individuals per cluster than fewer clusters with more individuals per cluster. *If you want to increase your sample size, add clusters, not individuals within clusters.* For more information about deciding how many clusters to choose, consult the following sources:

The CDC Manual: Indicators and methods for cross-sectional surveys of vitamin and mineral status of populations.

The Save the Children Fund. 2004. Emergency nutrition assessment: guidelines for field workers. London: Save the Children Fund.

SMART (Standardized Monitoring and Assessment of Relief and Transitions) available at www.smartmethodology.org.

3.4.3.2 Determining how many children to choose per cluster

In this section, we talk about how to determine how many children 0-23 months of age you should select per cluster. If your calculated sample size was 1008 and you were collecting IYCF data from 40 clusters, you would have about 25 children per cluster (1008 divided by 40). If you were collecting data from 45 clusters, you would have about 23 children per cluster (1008 divided by 45).

3.4.3.3 How do I locate these children?

- When you get to the first household, ask whether there is at least one child 0-23 months of age.
- Go to the next household and do the same, asking for *all* children 0-23 months old.
- Do not worry about getting the exact same number of children for each 6-month age category in a given cluster. By the time you finish your survey, you should have similar numbers of children in each 6-month age category.

Some general reminders regardless of your sampling strategy:

- Use your IYCF survey questionnaire (the one you modified in Chapter 2) to ask the mother or primary caregiver of *every* child 0-23 months old in the household about her child's feeding and food intake patterns.
- 23 months old means the child has not completed his or her 2-year-old birthday. If the child's 2-year-old birthday is on the day of the survey or if the child already had his or her 2-year-old birthday, the child SHOULD NOT be included in the IYCF survey. But if the mother has one or more other children who are 0-23 months old, be sure to use the IYCF survey to determine feeding practices for each child. When you encounter multiple births (for example, twins and triplets), interview all age-eligible children in the family.
- Sometimes you will collect data on a day when feeding practices might be different – for example, a Friday, Saturday or Sunday, a holiday, or a religious day. Feeding patterns may be different for other reasons as well. For example, food may be less available, the child may be sick, and so on. Regardless of these differences, you should still collect data on the actual feeding practices for the previous day for all children.
- By collecting data on all children, you get information that reflects what is happening in the *population*, which is what you want.
- **Document anything unusual about data collection.** For example, were data collected during fast or feast days? Days when children were sick? Unfortunately, childhood illness is a "regular" part of life for many. If you see large differences when you compare results at 2 or more points in time – for example, baseline and endline – discrepancies may be due to collection of data on an usual day for the entire population (e.g., many ill children), fasting, feasting and so on. While there is not a great deal you can do about this, you should document it nonetheless. Alternatively, if you anticipate that data collection will occur during an "unusual" period, you could add questions to your data collection instrument about whether the child was sick, whether the community was celebrating a religious event that might influence feeding practices, etc.

Annex: Instructions about how to sample if you are collecting data on IYCF behaviors *and* anthropometry at the same time

In some situations, especially if you lack sufficient funds to carry out an independent survey but want to track trends in program impact, it is useful to collect IYCF data at the same time you administer an anthropometry survey. An anthropometry survey measures nutritional status using age, height (or length for children less than 2 years old), weight, and mid-upper arm circumference. It is possible to collect information on IYCF practices at the same time you obtain anthropometry data.

We would like to share with you one strategy that we have used to obtain data on IYCF practices in situations where we have lacked the resources to carry out a large scale IYCF survey. We have done this by 'piggybacking' or adding IYCF questions to nutrition (anthropometry) surveys that are frequently conducted by agencies operating in emergency situations. They frequently employ a 30x30 cluster survey methodology based on the following assumptions:

- The prevalence of malnutrition is 50%
- The design effect is 2
- The error is 5%
- The precision is 5%

The sample size is also based on the assumption that 15% of mothers will refuse to participate in the survey or provide incomplete information. When these assumptions are made, sample size is about 900 (30 individuals each in 30 clusters).

Combining data collection on IYCF and anthropometry has implications for sampling. For example, staff who conduct an evaluation of anthropometry *and* IYCF behaviors at the same time often rely on cluster sampling. However, if you combine IYCF and anthropometry, *you do not have to use cluster sampling*. You might consider the other approaches. Previously, we talked about simple random sampling, systematic random sampling, and cluster sampling in the context of an IYCF survey.

In this annex, we assume you are using cluster sampling. We'll talk about:

- Deciding how many clusters to choose
- Determining how many children to choose per cluster
- Locating children for interview

Deciding how many clusters to choose:

When implementing agencies carry out data collection for anthropometry surveys, they often choose 30 clusters. You should always choose *at least* 30 clusters. The more clusters you choose (and the fewer respondents per cluster) the more you reduce design effect. Choosing less than 30 clusters increases the design effect and, more importantly, the chances that your sampling strategy will be flawed. If you use cluster sampling, it is better to have more clusters with fewer individuals per cluster than fewer clusters with more individuals per cluster. *If you want to increase your sample size, add clusters, not individuals within clusters.* For more information about 30x30 cluster sampling, see the references listed in section 3.4.3.1.

Determining how many children to choose per cluster:

In this section, we talk mostly about how to determine how many children 0-23 months of age you should select per cluster. If you were using a 30 x 30 cluster design to collect data on anthropometry only, you would have 100 children per 6-month age groups (900 children 6-59 months of age divided by 9 six-month age groups). If you were to also include children 0-5 months of age among your 900 hundred children, each 6-month group would have only 90 children (900 children 0-59 months of age divided by 10 six-month age groups). In fact, you will want to have a similar sample size; therefore, adding additional children to your sample may be necessary. If you add 3 more clusters of 30 children 0-59 months of age, you will have about 1000 (990) children 0-59 months old which is about 100 children for each 6-month age category. Therefore, collect data on about 1000 children (33 clusters x 30 children per cluster). Make sure you take anthropometry measurements on all age-eligible children as you normally would *and* information on IYCF practices among children 0-23 months old.

Locating children for interview:

- Conduct a 30 cluster sample as you normally would but add an additional 3 clusters of 30 children 0-59 months.
- When you get to the first household, ask whether there is at least one child in the age range you have specified (0-59 or 0-23 months of age since by definition all children 0-23 months of age are included in an evaluation of IYCF behaviors).
- Conduct the appropriate anthropometry measurements and interviews as specified above.
- Go to the next household and do the same, asking for *all* children in your age range.
- Do not worry about getting the exact same number of children for each 6-month age category in a given cluster. By the time you finish your survey in all 33 clusters, you should have similar numbers of children in each 6-month age category.

IMPORTANT: You will collect IYCF data on *every* child 0-23 months old. When you encounter multiple births (for example, twins and triplets) or multiple children 0-23 months old in a given household, interview mothers/primary caregivers of all age-eligible children in the family (a new questionnaire for each child).

You've done well! **Take a break and proceed to Chapter 4 on another day!**

CHAPTER 4: GETTING STARTED WITH DATA ENTRY

Goal: develop the data entry screen

By the end of this chapter, you will be able to:

- Explain why Excel should be used to enter data
- Describe different types of data fields
- Develop a data entry screen
- Put checks into your data entry screen to reduce data entry errors
- Lock (or freeze) variable names to improve data entry

Tool	Purpose of tool	Name of file
Data file with 20 mothers	Helps you understand data entry	Practice data file.xls

4.1 Why we use Excel

Before you enter data from your survey, you need to develop a data entry screen. The data entry screen allows you to enter and check data before you analyze it. By creating a good data entry screen ahead of time, you will minimize errors in data entry, reduce the amount of time needed to manage and clean data and improve confidence in your results.

There are several software packages that help you create data entry screens including EpiInfo, CSPro and Excel. We recommend you use Excel because it is widely available and because staff is usually more familiar with it than other software packages. Here are some other advantages and disadvantages of using Excel:

Advantages:

- Easy to tell Excel which values are okay for each variable
- Excel can do data analysis
- Easy to produce good graphics
- Compatible with Mac computers

Disadvantages:

- Your data entry screen doesn't look like your questionnaire
- During data entry, you can't tell Excel when to skip questions for individuals who should not respond to them
- Data cleaning can be more difficult than in SPSS or other statistical software programs. This is especially true if you have a long questionnaire
- Data analysis is more difficult than in SPSS and other statistical software packages
- If you want to conduct statistical analysis, you need to load the **Analysis Toolpak**. (The Toolpak is available on the same CD as your installation copy of Excel so there is no extra charge for loading it onto your computer.)

In this Guide we assume you have access to Excel. If you ever need additional help with Excel, on the main menu, go to Help and click on Microsoft Excel help or Microsoft Office online. If several of you are using the Guide, you may want to have pairs of users together. Those with more experience using Excel can be in pairs with staff who have less experience.

4.2 How to develop a data screen

There are many ways to use Excel to develop data entry screens. We tell you the simplest way to do it. If you know Excel and want to create a more user-friendly screen, you can do so. However, building fancier screens can be difficult.

Before you start, print out hard copies of these files that came with this Guide:

- Questionnaire.doc
- Practice data file.xls

You will want to refer to these files as you are working.

When should you develop your data entry screen?

Develop your data entry screen *after* you finalize your questionnaire.

Screens should be developed after the questionnaire is field tested and interviewers have been trained. If you develop data entry screens after finalization of the instrument, you can test your screen with the field test data. You will still be ready for data entry before completed surveys come in.

Step 1: Assign variable names

Explanation

Look at the Questionnaire. Each space where you fill in information is called a **field**. A **variable** is a collection of all the values you have entered into the field. On your data entry screen, you will need to assign names to each variable.

When the questionnaire is finished, make an electronic copy of the file and insert variable names from the data entry screen (in another color and font) into the questionnaire. This will be a great help to data entry staff as well as to those who clean and analyze the data. For example, for question 4, sex of child, you could put something like this: 4. Sex of child [sex].

Open up a blank screen in Excel. You will use this screen for data entry. In this spreadsheet you will create 61 variables. The first variable (cell A1) will be called **serial** and the last one (cell BI) will be **bottle**.

Try it: assign variable names

1. Open Excel, and press Ctrl-N to open a new spreadsheet.

2. Click on cell A1. You will see a black box around the cell. You will also see a plus sign.

3. Enter a name for your first variable. The first variable is the information that goes in the first blank on the questionnaire. In this case, enter the word **serial**.

   Note: For many questions you may need to shorten the variable name. For example, with the question, "Did you ever breastfeed [name]?," you could assign the variable name **EverBF**. Do not leave a space in your variable name. **EverBF** is okay, but **Ever BF** is not.

4. In cell A1, type **serial** (if you have not already done so), then hit the right arrow key. Go to cell B1 and type **result**. Go to cell C1 and type **team**. In cell D1 type **leader**. In cell E1 type **interviewerid**. In cell F1 type **DOI**. All of your variables should be in row 1. Do not leave row 1 blank. Do this for ALL variables. **Bottle** should be the last variable in your Excel spreadsheet. If you are using the questionnaire that comes with the Guide, **bottle** will be in cell BI1. There should be 56 variables in your Excel data entry screen. While 56 is a lot of variables, it doesn't take long to enter their names.

5. Because some of the first few variables require special consideration, we will describe what you have to do once you finish entering date of interview (**DOI**), cluster number (**cluster**), household number (**hhno**) and child number (**childno**). Remember, you are creating a NEW data entry screen, not modifying the "Practice data file.xls" document.

   Sometimes staff members working in emergency situations will use the questionnaire. To account for this, in the Excel spreadsheet, we've added the variables **camp** and **block**. Note that these variables are not on the generic questionnaire. If you don't work in an emergency situation, you don't need to add these variables. However, if you don't add them, be aware that your total number of variables and their location may be different from ours. If you do decide to add **camp** and **block** even though you aren't in an emergency situation, you can simply not enter data for these variables.

   Enter variable names for name of child (**name**), child's date of birth (**DOB**) and source (**source**). For child's age, responses to question 3 will include years, months or both, depending on the age of the child. Interviewers will fill in one or both blanks but the data entry specialist should create a single variable (**age**) to reflect the child's age. We strongly recommend reporting the child's age in months. For all children who are at least 1 year old, the data entry specialist will need to add 12 to the age in months. Thus, if a child is 1 year and 5 months old. **Age** in the Excel spreadsheet should be entered as 17 (12 months in a year + 5 additional months).

   Create a variable for sex of child (**sex**) and whether the child was ever breastfed (**everbf**).

   Recording when the child was first put to the breast is tricky. The interviewer will circle 000, 1, 2 or 98 then fill in the hours or days after birth that the child was put to the breast. However, you only need to enter information for hours or days after birth that the child was put to the breast. Create 2 variables after **everbf**: **initiatehours** and **initiatedays**. If the child initiated breastfeeding in the first day (0-23 hours after birth), when you get to the point of entering data, you will enter the number of hours under **initiatehours** and leave **initiatedays** blank. If the child began breastfeeding 24 hours or later after birth, enter number of whole days under **initiatedays** but leave **initiatehours** blank.

   Create a variable to add information about whether the mother offered the first fluids that came from the breasts. Call it **firstfluids**.

   Question 7 asks about what liquids were given to the infant in the first 3 days of life. Multiple responses are possible. For example, the mother (or someone else) may have given sugar water and formula. You want to know all of the liquids given. Instead of creating a single variable, you will create

5 variables: **plainwater**, **sugarwater**, **animal**, **formula**, **otherliq**. The responses for each of these variables can be 1 (yes), 0 (no) and 8 (don't know.) On the questionnaire, there is space to specify what other liquids were given. We haven't created a new variable to record this, but you could if you'd like. The variable should be to the right of **otherliq**. You could call it **otherliqspecify**. For mothers who indicate other liquids were given in the first 3 days of life, you should type the names of the liquids given.

Now enter the remaining variables.

If you are using the questionnaire that comes with the Guide, bottle will be in cell BI1. There should be 61 variables in your Excel data entry screen. While 61 is a lot of variables, it doesn't take long to enter their names.

Note: Please do not give 2 different variables the same name. Doing so creates difficulties when you calculate your indicators.

Step 2: Tell Excel what type of variable you want to create

Explanation: This step helps prevent errors with data entry.

There are many different types of variables:

- 'Whole number' fields allow you to enter numbers such as 1, 2 and 3 but not numbers with decimals (for example, 1.1, 2.1 and 3.1)
- 'Date' fields allow you to enter dates using a variety of formats (for example, day/month/year or month/day/year)
- 'Text length' fields allow you to enter words (for example, the name of each team leader)

Try it: Assign variable type for whole numbers

- Highlight column C that you have labeled with the variable name "team". Click on **Data**, then on **Validation**. When the dialogue box appears, click on the **Settings** tab.
  - In the Allow drop down box, click on **Whole number**
  - In the Data drop down box, click **between**
  - For the purposes of this example, we assume you have 3 teams
  - In the Minimum drop down box, enter the number **1**
  - In the Maximum drop down box, enter the number **3**

- Click on the **Input Message**
  - Place a check in the box **Show input message when cell is selected**
  - In the Title box, type **Team leader**
  - Under Input message, type:
    - **1=team 1** then <return>
    - **2=team 2** then <return>
    - **3=team 3** then <return>
  - Click OK

Now Excel knows that your variable **team** is a whole number and that it can have only 3 values (1, 2 and 3 corresponding to the number of interviewer teams in our example). Just now you gave both a short name

(team) and longer title (team leader) to your first variable. You also told Excel what each value means.

Note: You do not need to use the input message feature to create longer names/descriptors for variables. In fact, in some instances, it may be advantageous to NOT use them because sometimes when you copy variables to other columns, all formatting (including long variable names) is copied as well. In short, your long and short names for variables may be inconsistent.

When assigning variables for yes or no answers (as in the "result" column), always use **1** for yes and **0** for no. You can follow the pattern we used in the file 'data entry screen.' It is important that the numbers you enter are *consistent*. You *could* enter 'no' as '2' but as you will see later, this creates problems when you calculate each of the four indicators. '0' is best.

Now click on any cell in column C. You will notice that Excel gives you the longer title of the variable (in case you forgot what **team** really means) and the value for each variable. This comes in handy during data entry if you forget what each variable name and value mean.

Try it: Assign variable type for dates

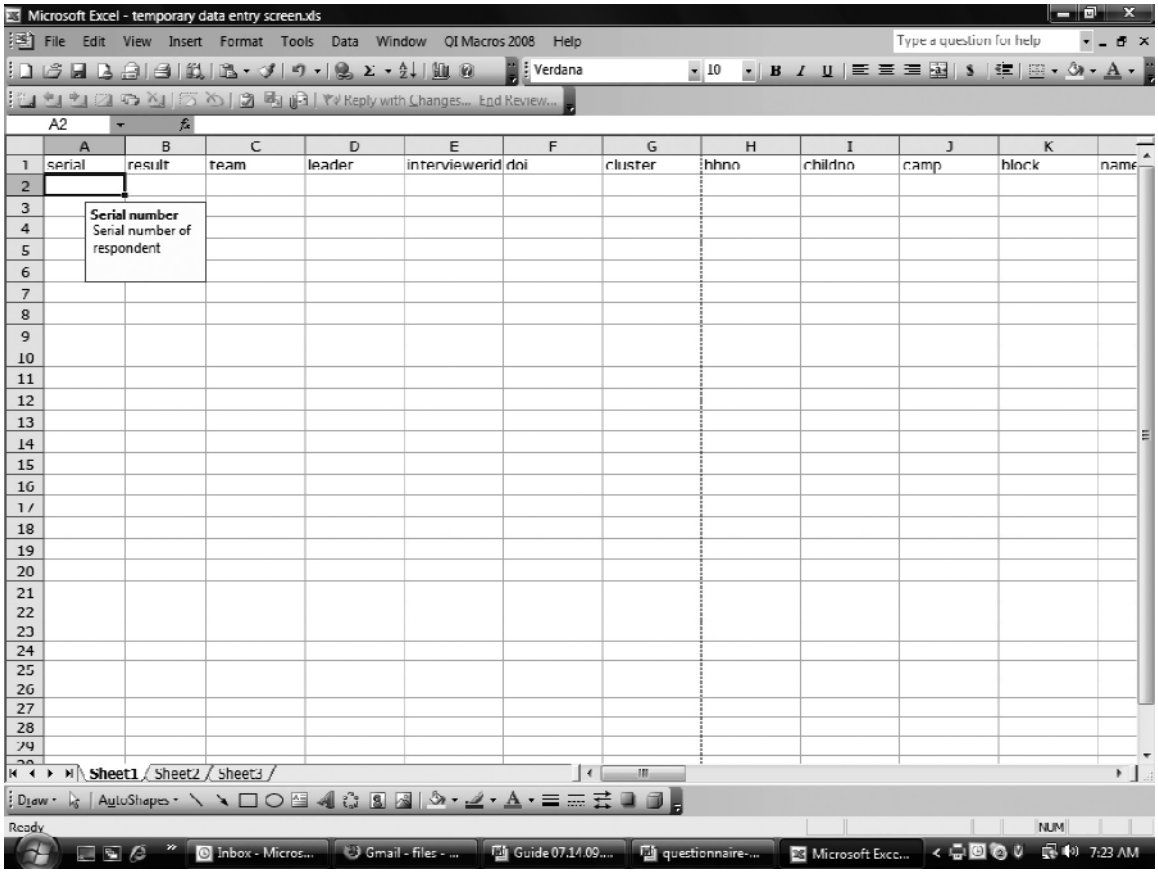
As you will see later on, it is very important that your dates are formatted correctly. Otherwise, you cannot calculate the age of the child and you may not even be able to enter dates either. Some parts of the world use an American system for writing dates: month-day-year (for example 12-21-2008). Some parts of the world use a European system (for example 21-12-2008). When you use this Guide, you should enter dates the way you normally would in your country. Your computer should be set up to accept dates the way you usually write them. However, if your computer is not set up properly, you can change the date format through your control panel (click on the **Start** button at the bottom of your computer screen then **Control panel; Date, time, language and regional options** then **Date and time properties** and **Time zone**). From the drop down menu, select your country. In some cases, your country will not appear in the list. If so, select your time zone relative to Greenwich Meridian Time (the time in London, England, also known as GMT). For example, Sierra Leone is the same time zone as GMT so you can select GMT). If your time zone is before or after GMT, select a country that is in the same time zone as yours.

- Highlight column F labeled with the variable name **DOI**. (We'll use column F in your practice worksheet because that is the column we used for date of interview in the file 'data entry screen.'
- Click on **Data, Validation** and **Settings**.
  - In the Allow drop down box, click on **Date**
  - In the **Date** drop down box, click **between**
  - In the Start date drop down box, enter the **earliest date when you began interviewing**
  - In the End date drop down box, **enter the date when all interviewing ended (or will end)**
- Click on the **Input Message** tab
  - Place a check in the box **Show input message when cell is selected**
  - In the Title box, type **Date of interview**
  - Click OK.

Note about data validation: There is no need to use Data validation for text variables, such as the names of your team leaders in column B. Even so, you should make sure you spell words in text fields the same way (for example, always 'Christy' not 'Cristy.')

Try it: Use your own data

Now go back through Step 1 and Step 2 using your own questionnaire. Below is a picture of a portion of a completed Excel data entry screen.



It is important that the numbers you enter are *consistent*. For example, 'yes' should always be entered as '1' and 'no' should be entered as '0.'

Step 3: Freeze rows

Excel allows you to 'lock' or 'freeze' rows and columns. This comes in handy when you are entering so much data that Row 1 scrolls off the screen. You can freeze Row 1 so it will remain visible and you can see all your variable names. This allows you to make sure you enter data in the right fields.

Try it: Freeze row 1

- Move the cursor to row 2 on the far left side of the screen. The cursor will turn into a right-pointing arrow. Click on it to highlight Row 2.
- From the drop-down menus at the top of the screen, click on **Window**.
- Click on **Freeze Panes**.

This will keep Row 1 visible at the top of the screen no matter how far down the page you go. If you want to unfreeze this row, go back to the Window menu and click **Unfreeze**.

Sometimes, after developing the data entry screen, you may decide you want to add additional variables. You can use steps 1 and 2 above to create a new variable at the end of your data file. But what if you want to insert a new variable somewhere between the first and last variable of your dataset? You can do this. Here is how:

- Find the place where you want to create a new variable. Click on the column to the right of it. The entire column should be highlighted.
- Now right click using the mouse.
- Select **Insert**.
- You now have a column for your new variable. Type the name of the new variable in row 1. You can give it a name and properties using steps 1 and 2 above.

Try it: Add a new variable

Let us say you want to create a new variable between **team** and **leader**. You may want this field to contain information about the number of interviewers per team. Highlight column D (leader) and right click on **Insert**. In row 1, give the variable the new name **size**. Note: if you have used Data Validation on the variable leader (e.g., Whole numbers if you enter an ID for the leader rather than his/her name), you will not be able to type the new variable name in the inserted column, as the Data Validation is copied. In that case, highlight your new column B, click Data, click Data Validation, choose Any Value and click OK. Type in your new variable name and continue with Data Validation for your new variable.

Adding new variables is very important. For example, you can insert variables for indicators of your choice. Let us say you want to add a variable to count the number of infants who receive only formula. You can do this using the steps above.

Important: At this point, do not add or delete any columns. If you have already added columns, delete them so that your data entry screen looks exactly like ours: **serial, result, team, leader, interviewerid, DOI,** etc. If your screen doesn't look like ours, you will have difficulty calculating indicators because your variables will be located in a place that is different from ours.

Now is a good time to enter the data from your survey. In Chapter 5, you will learn how to clean data.

Note: When data entry technicians enter information on diet (Question 9 for liquids and question 10 for solid, semi-solid and soft foods), you will enter numbers, not letters.

Enter:

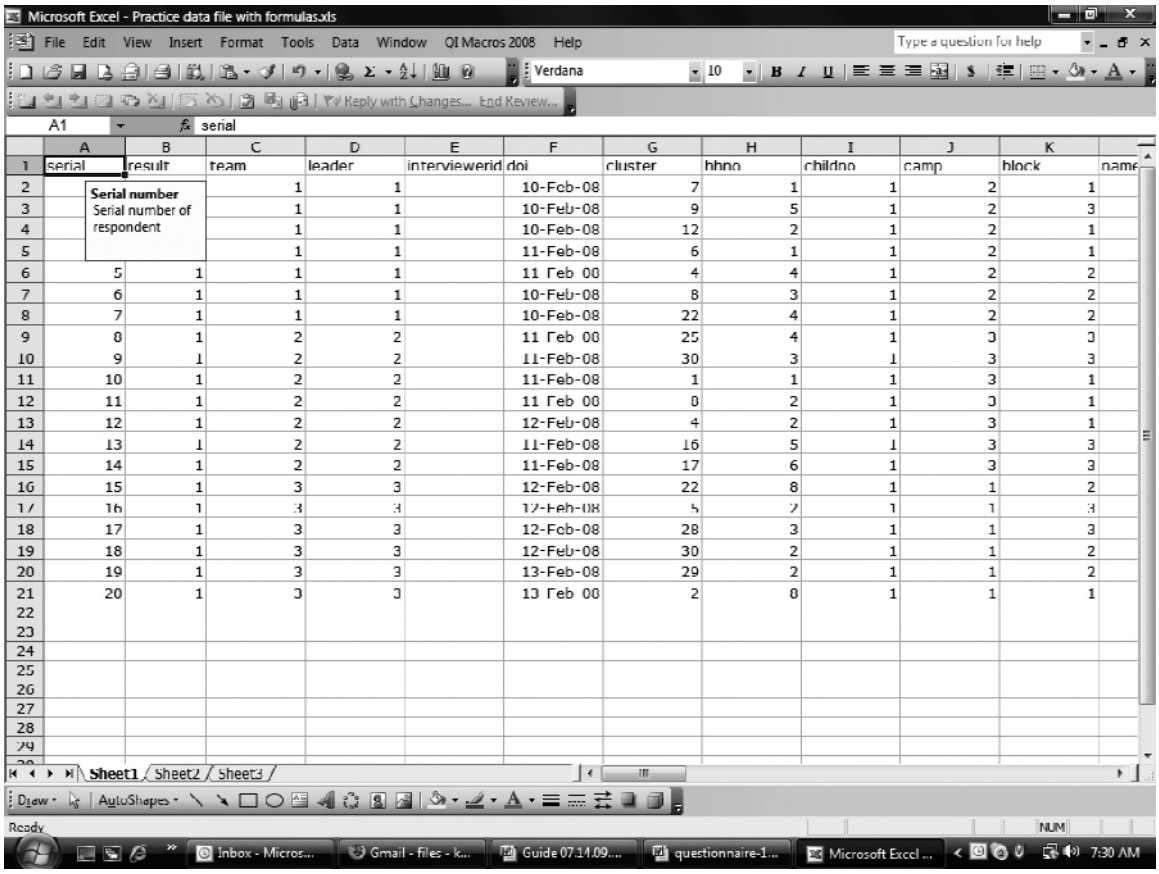
- **1** for **yes**, not **Y**
- **0** for **no**, not **N**
- **8** for **DK**, not **DK**

When we developed Questions 9 and 10, we put **Y, N** and **DK** in the questionnaire rather than **1, 0** and **8** to make it easier for the interviewer to correctly mark what the child had been given to drink and eat.

Now, enter all of your data. You will likely need to do this in several sessions. Entering all of your data in one sitting is likely to result in fatigue and data entry errors.

Note: Entering information into the computer about when the child began breastfeeding can be tricky. When you look at your questionnaire, you will see that the child can be breastfed immediately (coded 000), in the first hour of life (coded 00), in the first day of life (coded 01 to 23 depending on the hour when the child began breastfeeding) and after the first day of life (coded in days). In your practice data file, under the column labeled **initiatehours**, you will enter **0** if the child began breastfeeding immediately or in the first hour of life. Since we are only concerned with whether the child began breastfeeding in the first hour of life, we do not need to worry about the distinction between 000 and 00. Under **initiatehours**, you will enter number of hours (1-23) if the child began breastfeeding between 1 and 23 hours of age. Under **initiatedays**, enter number of days if the child began breastfeeding after the first day of life. Note that hours are entered under **initiatehours** and days are entered under **initiatedays**. Each individual in your dataset will have data for **initiatehours** OR **initiatedays** but not both!

The practice data file.xls gives you an example of a completed data entry screen for 20 observations. Below, you will see a screen shot of *part* of the data entry file.



You've done well! **Take a break and proceed to Chapter 5 on another day!**

CHAPTER 5: CHECKING AND CLEANING DATA IN EXCEL

Goal: learn how to check and clean your database

By the end of this chapter, you will be able to:

- Find cases with missing data as well as individuals who should not be in your dataset
- Name and use three ways of checking your data
- Clean data using Excel

Tool	Purpose of tool	Name of file
Data file with 20 mothers	Helps you understand data entry then check your data	Practice data file.xls

Always save multiple copies of your original dataset. It's a good idea to save them in different locations (hard drive, jump/flash/thumb drive, CD-ROM, etc.) in case some get lost or corrupt.

When working with your own data, you might call this file **IYCF database original**. This dataset includes all individuals who completed the entire survey, as well as those who finished only a portion of the survey. You may have some basic information about those who did not complete the survey. If some individuals refused to give any information, they will not be in your dataset.

Then save a copy of the file using a different name. This is the file you will work on. If you make a mistake with this copy, you can always go back to the original file.

In this chapter, you will first practice cleaning data using the data we provide in the file **practice data file.xls**. Based on what you learn, you will be able to clean your own data in the future. The checks you've developed (described in Chapter 4) will reduce the amount of cleaning you need to do, but you still need to examine your data carefully before calculating IYCF indicators.

If you are working in pairs and you have been at the keyboard, give your partner a chance to be at the computer.

There are lots of ways to check your data including:

1. Finding cases with missing data or individuals who should not be in your dataset
2. Visually scanning your data to make sure they are clean
3. Range checks, and
4. Consistency checks

Later (in Chapter 6), we will tell you how to check for measurement and recall bias, including age heaping.

We will talk about all 4 ways of checking data.

5.1 Finding missing data or individuals who should not be in your dataset

In the next exercise, we will use the **Practice data file.xls**. It is a file that you can use to try out your new skills cleaning and managing data. It looks a lot like the data entry screen you developed in Chapter 4. But there are a few small differences. We have added some variables that you will need to make sure your data are 'clean' for example **serial**, described in the next paragraph.

Sometimes we collect information on people who did and did not complete the survey. Look at **Practice data file.xls** and you will see a variable called **serial**. It is the first variable in **Practice data file.xls**. In your data file, you probably used a number of variables to give each child a unique number, for example, cluster, household number, child number, camp (if you are working in an emergency situation) and/or block. To keep things simple, in this chapter we use a single variable (**serial**) to refer to the unique identification number of each child. The serial number will help you identify children we talk about in this chapter (Chapter 5).

Note: Our serial numbers are in order. They do not have to be but if you decide to sort your dataset by serial number, make sure you sort the *entire* dataset, not just the column marked **serial**. To do so, highlight serial, then **data, sort**. When you sort **serial**, Excel will give you the opportunity to **expand the selection**. Click on **expand the selection** to make sure that all of the information that corresponds with **serial** 1 stays with **serial** 1. Information for **serial** 2 should stay with **serial** 2, and so on.

We have also included a second variable (**result**). For the practice data file, participants are coded as: **result**=1 (completed the entire survey), 2 (completed only part of the survey), and 3 (refused to take the survey). For all indicators, you want to limit your analyses to those mothers who completed the entire interview (**result**=1).

Note: Even if you did not collect information on whether individuals completed the survey, please read through this section. It contains important information about filtering that you will need to clean your data.

5.1.1 Cleaning your own data

1. Save a copy of your original (practice) dataset. Call it “Practice data file.”
2. Save another copy of your original dataset using a different name, for example “Practice data file version 2.”
3. In the *copy* of the original dataset, look for individuals who have completed only part of the survey. These are individuals who have a 2 or 3 for **result**. As you will see, there are no individuals with a code of 2 and 3.

There are at least three ways of looking for incomplete data. One is to visually scan your data to look for individuals who are coded as having an incomplete record (see Section 5.2 below for how to scan data). Another way is to use the **Autofilter** command. This is a command we will use a lot in Excel so we will explain it in detail in a minute. A third way is to sort data by result. Above, we described how to sort data. Please note that when you sort individuals by **result**, those coded 2 or 3 should be at the bottom of the list.

Try it: Use Autofilter to find incomplete data

1. In the copy you made of the practice data file, click on the top of the column that contains the variable **result** (column B). The whole column should be highlighted.
2. Now click on **Data, Filter** and **Autofilter**. A down arrow (↓) will appear to the right of your variable name (**result**).
3. Click on the down arrow. A list will appear showing all the unique answers in this column.
4. Click on **1**. This will show you all individuals in your dataset for whom you have complete information on **result**. If there were individuals for whom **result** was 2 or 3, you could click on the down arrow, then on 2 or 3 and see records for all individuals who completed only part of the survey (2) or none of the survey at all (3).
5. If you are using **Autofilter**, don’t worry if your data screen is sometimes empty. The data are still there. You are simply asking Excel to show you all the data that meet the criterion you specified.
6. To restore the original view with all the data, click on the down arrow (next to **result**) then (**All**).

Sometimes you may want to use Autofilter to sort for more than one variable. Here is how you do it:

Try it: Use Autofilter to sort for more than one variable

1. To sort for more than one variable, you need to switch on **Autofilter** for the entire page. That way you are sure you have a complete dataset before filtering some observations. First turn off **Autofilter** for the entire worksheet. To do so, click on the box in the upper left hand corner of your worksheet. It is located above row 1 and to the left of column A. Your worksheet should be completely highlighted. Now click on **Data** and **Filter**. If there is a check mark next to **Autofilter**, click it to turn **Autofilter off**.
2. Then click **Data, Filter, Autofilter** again to turn it on for the whole page. Down arrows will appear to the right of each variable name.
3. Click on the down arrow to the right of the variable **sex** (column P) then click on **1**. Excel selects all observations that are boys. The down arrow to the right of the **sex** changes to blue. This is a reminder that this variable is sorted.

4. Now click on the arrow to the right of **porridge** (column AL) and click **1**. Excel shows only those observations that are male AND receiving porridge. In other words, Excel has filtered for sex and porridge. (Also, note that the arrow next to both of these variable names is now blue.)
   - Only one child (serial #2) was male and got porridge

5. Now suppose you want to look at all observations (whether the child is male or female) to see whether they got porridge. You can go back to the variable **sex**, click the down arrow, and select (**All**). Now you will see all children who got porridge, whether they were male or female.
   - Five children got porridge. Serial numbers for these children are 2, 4, 9, 14 and 18.

An alternative to **Autofilter** is the **List** command. It is quicker than **Autofilter** but you can't select all variables at once. Rather, you need to highlight each variable one by one then create a **List**. If you prefer using List over **Autofilter**, here is how you use it:

- Highlight the column you want to List. Click on Data, List, Create List. Then click on **My List Has Headers** and **OK.**

- Your column now has a down arrow (↓) just like when you use **Autofilter**. You can use **Autofilter** in the same fashion as you did previously (see steps 3-6 above).

Let us say you want to look at all individuals who had a **result** value other than 1. Here is how you would check this.

Try it: Use Autofilter to exclude an individual from the list

Only exclude an individual if there is truly insufficient information to calculate the 10 IYCF indicators. In the example below, a mother provided incomplete information making it impossible to calculate each indicator. Consequently, we will exclude this individual from data analyses.

1. Click on the down arrow next to **result**
2. Click on (**Custom...**)
3. In the first drop down menu box, click on **does not equal**
4. Go to the drop down menu box immediately to the right and type in **1**
5. Excel will now filter out any observation that does not equal **1**.
6. Since all of our **result** values are **1** (no observations are not equal to 1), our screen will be blank.
7. To restore the original view with all the data, click on the down arrow (next to **result**) again and then on (**All**).
8. We are now certain that everyone in our practice dataset completed the full interview.
   Note: in the Practice data file, the person whose serial number is 6 is missing information on liquids given. It may be that in the data file for your survey some individuals may be missing additional information.

Important: Autofilter is additive. That means that Excel will continue to filter variables if you select more than 1 variable. Here's an example:

Now let us take another look at your dataset. How many individuals have only a few missing pieces of information?

- If respondents are missing only a little information and the information isn't critical to the calculation of the 10 indicators (listed in Section 6.1), you should keep these individuals in your dataset.

- If respondents are missing a lot of information or if they are missing information that is critical to the calculation of the indicators (for example, age of child, without which the data cannot be analyzed), delete them from the copy of your dataset.

Here is how to delete an individual from the data set:

1. Make sure you have saved a copy of your data file using a name that is different from the original dataset (for example, Practice data version 2 if you are saving the practice file that comes with this Guide)

2. Use **Autofilter** to detect who is missing data. Highlight variables that you think are particularly important, for example, **age** and **everbf**. Click (**Blanks**).

3. Highlight the row corresponding to the respondent who has incomplete data. To do this, click on the row number at the far left of the row. Move your cursor to that row then click the mouse once using the left mouse key. The row should be highlighted.

4. Click the right cursor. On the pop-up menu screen, click **Delete**.

5. Do this for every individual who has a lot of incomplete data for key variables such as age and breastfeeding. You should only do this if the individual is missing information that is critical for calculating your variables. For example, an individual would need to be missing information on most of the following variables before deleting him or her from the dataset:
   a. Age (a child must have date of interview and date of birth or reported **age** if dates of interview and birth are missing) so any child with missing age should be deleted
   b. Everbf
   c. Initiatehours
   d. Lots of missing liquids (columns T through AL in the practice data file)
   e. Lots of missing foods (columns AM through BC)
   f. Solids
   g. Information on iron-rich or iron-fortified foods (columns BE through BH)
   h. Bottle

6. How do you decide when to delete an individual? You will need to decide this for each of the 10 IYCF indicators. Step 5 above gives general guidance but you can find specific instructions about information each individual must have for the calculation of each of the 10 IYCF indicators by referring to the table in Chapter 2, section 2.1 (Summary table of indicator definitions, calculation and questions used to construct indicators).

7. Now save a copy of your dataset using a different name for your file.

Usually, very few individuals refuse to participate in the survey (and they will not be entered anyway). The number of people who have incomplete data should be small. It is the interviewer's job to ensure the data are complete – or as complete as possible.

5.1.2 What to do with individuals who should not be in your dataset

Some individuals should not be in your dataset. For example, if you are using a cluster sample and the household number is greater than what you originally projected, interviewers did not follow instructions about sampling. Check with interviewers and the field supervisor to make sure the cluster number was not recorded wrong. If it was, correct the cluster number. Here's how you do it:

1. Highlight the cell that has the incorrect cluster number.
2. After speaking with the interviewer and/or field supervisor to determine the correct cluster number, type in the new cluster number and hit enter.

If, in fact, a particular house should not have been approached for interview, it is best to exclude these individuals from the dataset. Here is how you do it:

1. Highlight the row containing the household that should not be included in the dataset.
2. Click the right cursor. On the pop-up menu screen, click **Delete**.

Undoing changes:

Sometimes you need to undo changes or deletions you have made. To do so, you can:

- Click on the back arrow symbol at the top of the page.
- Press Ctrl-Z, or
- If you are using a version of Excel prior to the 2007 edition, click on **Edit** at the top of the screen and then **Undo**.

You've done well! **Take a break and proceed to section 5.2 on another day!**

5.2 Visually scanning your data

Now that we have checked our data for incomplete observations, we are ready to scan our data to see if they make sense. It is a good idea to do this before doing more formal range and consistency checks.

Practice Dataset: Let us look at the practice dataset (**'Practice data file version 2.'**). It has only 20 observations. We made the dataset small so you can understand information for each child. Your own data will include more than 20 children.

Tip: As you scan the data, you want to see the unique identification for each child. The best way to do this is to freeze panes. We mentioned how to freeze rows in Chapter 4. This time you will freeze the *column* next to the unique identification number for the child.

- Move the cursor to the top of the first column, for the variable **serial**. The cursor will turn into an arrow that points down. Click to highlight the column.
- From the drop-down menus at the top of the screen, click on **Window**, then **Freeze Panes**.

The identification number for each child will now remain visible at the left of the screen, no matter how far to the right you scroll.

There are many variables we could check. We will look at just a few of them. After opening the file, look at the following variables:

1. **Source:** this variable gives the source of information about the child's date of birth. It is located in column N. Thirteen of the 20 mothers interviewed were able to give the interviewer the card with the date of birth (source=1). This is your most reliable source of information for date of birth because the interviewers actually saw the card showing the child's date of birth. Five mothers (observations 11, 12, 13, 16 and 17) were not able to show their child's card so they estimated the child's birth date (source=2). Two mothers (mothers 1 and 7) simply did not know the date of birth for their child (source=3).

2. **EverBF:** all infants in this example were breastfed at some point (EverBF=1).

3. **Initiatehours:** mothers 2, 6-9 and 16-17 began breastfeeding their infant in the first hour of life (initiate=0). The remaining mothers began breastfeeding after the first hour (initiate=1). No one responded 'don't know'.

4. **Bflastnight:** infants 6, 14, 18 and 19 weren't breastfed in the previous 24 hours (Bflastnight was missing or Bflastnight=0). Because they did not breastfeed last night, we cannot consider these infants exclusively breastfed. In fact, we know nothing about what infant 6 ate in the last 24 hours. We know that he started breastfeeding in the first hour of life, but that is all we know for certain. If the hard copy of the questionnaire does not give information about what the child had eaten and if there is no other way of determining what the child consumed, it might be good to delete this child from the database. We have not deleted the child from our practice database but when you use your own data, you may want to do this for infants who are missing a lot of information.

5. Now look at infant with serial number 2. He breastfed in the last 24 hours (**bflastnight**=1). This child also got porridge (**porridge**=1). Interestingly, the child appears to have received solids, semisolids or soft foods once (**solids**=1) in the last 24 hours. Look at the liquids and foods child 2 received. There are no foods recorded (porridge is considered a liquid if it is thin and given to younger children. 'Stiff' or 'thick' porridge eaten by older children or adults is classified as a 'grain.' Consequently, either **solids** (the number of times the child received a solid, semi-solid or soft food) should equal 0 or at least one of the foods should be yes (1).

6. Look at these variables:
    a. Bflastnight
    b. Vitamins
    c. ORS
    d. Water
    e. Infantformula
    f. Milk
    g. Juice
    h. Broth
    i. Otherliquids
    j. Sourmilk
    k. Commercial
    l. Porridge
    m. Grains
    n. Roots
    o. Beans
    p. Orange
    q. Greenveg
    r. Vitafoods
    s. Otherfruitveg
    t. Organmeats
    u. Meat
    v. Fish
    w. Grubs
    x. Eggs
    y. Cheese
    z. Oil
    aa. Palm oil
    bb. Sugar
    cc. Condiments

Your Dataset: The 20 infants in your dataset got different foods and liquids. Visually scan your data to see which infants you think are exclusively breastfeeding. Later, in Chapter 6, when you calculate the prevalence of exclusive breastfeeding for the practice dataset, we will see whether you got the same prevalence by scanning your data.

Note: We have highlighted some of the variable names in the first row of **Practice data file.xls** with different colors. We highlighted variables that you will refer to again and again. The colors should help you keep track of which variables we are talking about when we give instructions about cleaning data and using each variable to construct indicators. Throughout the rest of this document we refer to these colors. You can get the exact name of the color we refer to by clicking on the variable itself then clicking on the down arrow of the icon **Fill color** at the top right hand corner of your screen. As you pass the mouse over each color, Excel provides you the 'official' name for that color. For example, we colored the variables mentioned above as follows:

bflastnight – pink

initiatehours and **initiatedays** – light yellow

water and **porridge** – pale blue

Not all variables have colors. Also, Excel has not given every color a name (for example, the color for **bflastnight** is called **color scheme**. If you don't like the colors we have chosen, you are free to select your own colors.

5.3 Range checks

Range checks are one way of checking the quality of your data. You look at the values for one variable at a time to make sure they are within the range you specified. Because you have set up your database to accept variables in a certain range, you should be okay for most variables. But you should do range checks regardless. For example, the variable **initiatehours** allows you to specify how many hours after birth the baby was put to the breast. The only legitimate answers are 0-23 and 98. If the mother responded 25, **initiatehours** should not be completed but **initiatedays** should be (in this case, **initiatedays**=1). Consequently, when you look at **initiatehours**, check for any values that are different from 0-23 and 98. Check **initiatedays**. A high number of days (for example, 5 or more) might mean incorrect data.

Range checks are easy if you use **Autofilter**. We will use the same variables we examined in Section 5.2 to try out the **Autofilter** feature.

PRACTICE DATASET

Try it: Check one variable

1. Click on **Data, Filter**, and **Autofilter**. Make sure there is no check mark next to **Autofilter**. If there is a check mark, click next to Autofilter to turn it off.
2. Highlight the entire dataset. To do so, click on the box in the very upper left corner of the page to the left of column A and above row 1.
3. Now click on **Data, Filter, Autofilter**. All variables will have a down (↓) next to them
4. Go to the variable **source** and click on the down arrow.
5. Click on (**Custom...**) then in the drop down box click on does not equal.
6. Choose **1** in the box to the right then click **OK**.
7. Excel shows all the observations where the value for **source** doesn't equal **1**. It should show you the exact same values you discovered scanning your data in Section 5.2. The serial numbers for observations where **source** doesn't equal **1** are 1, 7, 11, 12, 13, 16 and 17.

When doing range checks (as described in Section 5.3), always make sure **Autofilter** is turned off and all observations are visible before turning **Autofilter** on and looking at individual variables.

Note: it is perfectly acceptable to have **source** equal 2 or 3, but be more cautious in your interpretation of these data because you cannot verify with certainty the child's age.

Try the same with the variables **everBF, initiatehours**, and **bflastnight**.

1. Click on the column for **everBF** (column Q).
2. Click on **Data, Filter, Autofilter**.
3. Click on **Custom**.
4. In the first drop down box, click on **does not equal**.
5. In the second drop down box type 0.
6. Now click on the radio button **AND** (the radio button is a circle; once you click on it, Excel fills the circle with black).
7. In the third drop down box, click on **does not equal**.
8. In the fourth drop down box type 1.

Alternatively, you can use the **List** feature described above.

In the Practice data file, Excel will show you no values because in our dataset, **everBF** is always 0 or 1. Perhaps in your dataset some individuals were incorrectly given a number other than 0 or 1. Try the same thing with **initiatehours** and **bflastnight**.

YOUR DATASET: Now that you have tried using **Autofilter** for the 4 variables above, you are ready to check your entire database. You can do so using a shortcut.

Shortcut:

1. Highlight your entire database by clicking on the box in the upper left cell (between column A and row 1).
2. Activate Autoformat (**Data, Format, Autofilter**).
3. Go to the first variable in your database (column A).
4. Click the down arrow then (**Blanks**). This shows you if there is missing information for this variable. If there is missing information, double check the hard copy of the questionnaires for individuals who are missing information. In some cases, there is a legitimate reason that individuals are missing information. If you click on the down arrow for a given variable and (**Blanks**) does not appear, this means that there is no missing information. Look at the variable **serial** (column A). It is an example of a variable with no missing information.
5. Now that you've checked for missing information (blanks), click on the down arrow for your first variable then click (**All**). This restores all individuals to your worksheet.
6. When you click (**All**), notice that in the drop down menu, Excel gives you every value contained in your database for that variable. For example, in our Practice data file, under our **serial** column, all of the values from 1 to 20 appear in the drop down menu when we use **Autofilter**. If we see the value 21, we know something is wrong with data entry because there should only be 20 individuals in our practice dataset.

Try it: Range check shortcut

1. For each variable in your dataset, check to see if there is any missing information. Do this by following steps 1-4 above (under 'Shortcut')

2. Now go back to your first variable and, with **Autofilter** activated, click on the drop down arrow for each variable. Are the values listed for each variable in the range specified on the questionnaire? If not, check your hard copies for interviewing errors or data entry problems.

If you are using the sample dataset, there are lots of things you can check. Here are some of them:

- Check date of interview. Are all the dates during the time period when you collected data? In our example, all of our interviews were conducted between 10 and 13 February of 2008. This corresponds to what we have in the database.
  - Question: Are there lots of interviews on one date but not on the others? If so, date of interview may have been recorded incorrectly.
- Check child number. Most mothers do not have many children less than 2 years of age. If you have a lot of child numbers that are 2 or higher, interviewers may have recorded child number incorrectly or data entry technicians may have entered the wrong information.
- Check date of birth. Are there a lot of missing birthdays? If so, see what is on the completed questionnaires. You may need to gather this information again or make sure that you have recorded date of birth correctly. Check that all children in your database have acceptable birthdays. Since this part of the questionnaire only includes children less than 24 months of age, date of birth should not be two years or more before the date of interview.
- In most cases, the percent of mothers who have ever breastfed should be high. If possible, look at other studies in the area where you conducted your survey. Are the percent of women who ever breastfeed similar to what you found?
- Make sure you have a variable that tells you how many hours after birth the child began breastfeeding. In the practice dataset, this variable is called **initiatehours**.
  - Your data should tell you *hours* after birth, not *days* after birth. Sometimes, interviewers record hours when they mean days or write down days when they mean hours. It is difficult to correct this mistake after interviewing has taken place. During training sessions for interviewers, trainers should carefully review how to record and distinguish between hours and days.
  - Sometimes the problem is with data entry. If interviewers collected and recorded the information correctly, this problem is easy to fix. You must go back to the hard copies of the completed questionnaires to see if data are entered correctly.
    - It is easy to confuse zeros with ones. Normally, 0 means 'no' and 1 means 'yes.' But for the variable **initiate**, 0 means the child began breastfeeding in the first hour of life, 1 means the child did not.

You've done well! **Take a break and proceed to Section 5.4 on another day!**

5.4 Consistency checks

Now let us see if your variables are consistent with each other. When you look at the consistency of your data, in most cases, you are looking at two variables at the same time. For example, if a mother says she never breastfed her child but told the interviewer how soon after birth she began breastfeeding, there is an inconsistency you will need to correct. Three things may have contributed to this error:

- The mother may have answered incorrectly
- The interviewer may have coded her response incorrectly, and
- The person entering data may have put the wrong information into the computer.

It is also possible that all of the above may have occurred for the same case. Hopefully, these discrepancies will have been discussed during the training of supervisors and such issues identified and resolved while the interview team is still in the field. If not, the following section tells you how to deal with these issues.

Let us look at an example of an inconsistency. Some children did not receive solid, semi-solid or soft foods in the previous 24 hours but have information indicating they received foods 1 or more times:

1. In the Practice data file, make sure **Autofilter** is turned on. Click the down arrow to the right of **solids** (column BD) and choose (**Custom**)
2. In the first drop-down menu, choose **greater than or equal** to and enter the number 1. Click on OK. The dataset now shows only those children with a 1 or higher in the solids column. This includes children with the following serial numbers: 2, 4, 8-12, 14, 15, and 18-20.
3. Go to **serial** (column A) and find child 2. Click on the row number to highlight the entire row. Notice that child number 2 has a 1 in the **solids** column. This indicates that child 2 had solid, semi-solid, or soft foods one time in the previous 24 hours.
4. Now examine the list of solid and semi-solid foods in columns AM though BC. The data indicate that this child did not receive any foods. This child has inconsistent data.

Clearing Autofilter before a new sort:

- When using **Autofilter**, make sure you're not accidentally searching with a list that has already been sorted for a different purpose.
- Before you start on a new sort, go to **Data, Filter, Autofilter** and make sure there is not a check mark next to **Autofilter**. If there is, click it to turn Autofilter off. This restores all your data to the screen.
- Then you can turn on **Autofilter** again to start with your entire dataset.

Here is another example: a child must have breastfed at least once to have initiated breastfeeding in the first hour of life. That is, if **initiatehours** equals 0-23, **everBF** must also equal 1. Here is how you check for this:

1. Make sure **Autofilter** is turned on and click on the down arrow to the right of **everBF** (column Q).
2. Click on (**Custom**) and in the drop down boxes, click on **does not equal** and 1
3. Now activate Autofilter for **initiatehours** (column R)
4. Click on (**Custom**) then **is greater than or equal to 0 AND is less than or equal to 1**

Steps 1 and 2 tell Excel to show us all of the children who never breastfed. Steps 3 and 4 tell Excel to show us all of the individuals who never breastfed AND who initiated breastfeeding. Since every child in our database had ever breastfed, the number of children who are 'inconsistent' is 0.

Let us suppose the second child in our dataset had the following codes:

everBF=0

initiatehours=0

In other words, this child never began breastfeeding (**everBF**=0) but, according to our data, began breastfeeding in the first hour of life (**initiatehours**=0). This is an inconsistency. In this example, Excel would have shown us the data for this child using the 4 steps above.

Try it: Look for inconsistencies

Look at a hard copy of your questionnaire. What are some inconsistencies in your data? Since your data will be different from the practice dataset, we can not tell you what the inconsistencies are, but here are a few you can check for in the practice dataset.

- No 'date of birth for child' is recorded but 'source of date of birth' is specified. Note: if the interviewer follows the skip pattern correctly, 'source of data' should be completed if date of birth is specified.
- Child not breastfeeding but the questions on initiation and whether the mother is still breastfeeding are answered.

5.5 Cleaning data using Excel

Cleaning data means correcting whatever mistakes you can.

To clean the data:

1. First, before you make changes, save your file. Save your new file using a name that is different from your original dataset.
2. Once you have identified an incorrect value, you can change it. Make sure you have figured out the correct value before making the change. You can often get the correct value by checking the hard copy of the questionnaire or visiting the mother again. Note: hard copies should be checked the same day as the interview so that if necessary the mother can be re-visited. Questionnaires given to data entry staff should be complete and already checked for consistency by the field supervisor.
3. To change the value, simply go to the cell with the incorrect information, type in the new value, hit Enter, then save your file.
4. You do not have to save your file after every change, but you should save it every 5 to 10 minutes.

Important Tip: Before you start cleaning your data, make a list of potential inconsistencies, how you will check them, and what you will do when you are in front of the computer so that you apply the same approach every time to dealing with inconsistencies. Let us suppose a mother initiates breastfeeding after the first day of life. The variable **Initiatedays** should have a number that is 1 or greater. The variable **Initiatehours** should be left blank. You could have a separate piece of paper that says something like this: 'If **Initiatedays** is greater than or equal to 1, **Initiatehours** should be blank. That way, you always apply the same rule for cleaning **Initiatedays** and **Initiatehours**.

In summary, when you check and clean your data, be sure to:

- Have a good look at your data by visually scanning them.
- Use **Autofilter** to find inconsistencies in your data.
- Note which individuals have inconsistencies.
- Check the hard copies of your questionnaire to see if data were entered incorrectly.
- Re-interview if necessary and make changes to the data file.

Important: Clean the data for *all* inconsistencies you have identified for your database. It is important that you do this. Otherwise, you will analyze a 'dirty' database that can give you misleading information.

You've done well! **Take a break and proceed to Chapter 6 on another day!**

CHAPTER 6: ANALYZING INFANT AND YOUNG CHILD FEEDING DATA

Goal: calculate indicators for infant and young child feeding

By the end of this chapter, you will be able to:

- Understand the importance of accurate age measurement to determining infant and young child feeding practices
- Calculate the age of infants using a standardized age chart and Excel
- Define each of the 10 indicators you should collect as part of an assessment on IYCF, including definitions for the numerator and denominator
- Calculate each of the 10 indicators by hand and using Excel

Tool	Purpose of tool	Name of file
Practice data file	Helps you calculate age of child and each of the 10 indicators	Practice data file.xls

In this chapter, we show you how to calculate the 10 standard indicators that we analyze with data from living children 0-23 months of age. However, you may want to calculate other indicators as well. For example, the use of prelacteals. We do not describe how to calculate all of the indicators you may be interested in but you can use the principles we outline in this chapter to carry out calculations for indicators in addition to the 10 we mention.

6.1 The importance of age measurement

It is very important that you determine the age of the child with accuracy. Recommendations for infant and young child feeding are based on the child's age; consequently, if information used to calculate age is incorrectly collected during the interview, misrecorded, entered incorrectly into the database and/or miscalculated in Excel, the indicators below will be incorrect.

To calculate the recommended indicators, you will need data on infant age in **completed months**. An infant aged 0-30 days has not yet completed one month, so he or she is classified as less than one month, or "0 completed months." This can be written as: "0 months" or "0-<1 month," or "0-0.9 months."

While there are many ways of referring to age, we will follow the WHO recommendation to include the first and last month of the age interval. Thus, the age of the child who has not yet had his 6-month birthday is written as 0-5 [completed] months, NOT 0-<6 months or 0-5.9 months.

6.2 Calculating the age of the infant

Before you calculate each of the 10 indicators – either by hand or using Excel – you need to calculate the age of the infant.

The best way to calculate the exact age of the child is by subtracting the child's **date of birth** from the **date of interview**. When you set up your Excel data entry screen, you told the software that date of interview (**DOI**) and date of birth (**DOB**) are **date** fields. If you did not set up these fields as date fields, please go back and do so now (**Format, Cells, Date**). Otherwise, you won't be able to calculate the age of the infant.

If the child's day of birth is missing, you will need to substitute a number for the missing data. Per WHO guidance, you should substitute **15** for the day of the month when it is missing. If you have not entered **15** for children with missing days, do so now. While **15** is an assumption, it is one that you will use consistently. It allows you to recover data on date of birth for those children who are missing this information.

Here is how to get Excel to automatically calculate the age of the infant in days:

1. In **Practice data file version 2.xls**, scroll to the right until you find your date of birth variable, **DOB**. Make a note of the column letter (in our file, it is column M). Remember that **Practice data file version 2.xls** is the new name you assigned to Practice data file.xls. You received **Practice data file.xls** with the Guide but not **Practice data file version 2.xls**.

2. Go to the end of your file. The first blank column at the end of the practice file is BJ.

3. In row 1 of the first blank column at the end of your file, type the name of the variable where you will store information on calculated age. **CalcAgeDays** (Calculated Age in Days) is a good name. Hit enter.

4. Now find the column where date of interview (**DOI**). In our example, it is column F. Make a note of the 'address' of the column.

5. Return to the column that contains the calculated age variable you are creating. In our example, this is column BJ.

6. In row 2 of that column, type '=' then the cell address of your **DOI** variable then the minus sign then the cell address for **DOB**. In our example, row 2 of that column looks like this: **=F2-M2**

7. Once you press enter, Excel will calculate the number of days between **DOB** and **DOI**. Do not worry if your number is high. We will explain high numbers in a minute.

8. If Excel gives you a date rather than age of child in days, click on the very top of the column (**CalcAgeDays**) to highlight the entire column. Now go to the main menu and click **Format, Cells, Number**, and reduce the number of decimal places to **0**.

9. Now that you have the formula correctly entered for row 2, you can copy it to other observations. To do so:
    a. Click on the cell where you calculated age for the child in row 2. Move the cursor to the lower right corner of that cell. In our case, this is cell BJ2. Do not click on the corner, simply move the cursor to that spot. When you do, you will see a + sign

    b. Now left click and drag the formula you entered in row 2 (under **CalcAgeDays**) straight down the column to the last observation in your file. Excel will automatically copy the formula down the column.

    c. If you have problems dragging formulas, you can do the following instead:
        - Right click on the cell you want copied
        - Click on 'copy'
        - Go to the cell(s) where you want to copy the information.
        - Right click on those cells
        - Click on 'paste'

10. You will notice that two observations (serial #1 and serial #7) have a calculated age in days that is very large (38,026)! These ages are high because Excel could not find a date of birth for these children. We cannot calculate ages for these individuals because we do not know dates of birth for these children.

Now we have the age of the child in days, but we need to know the age of the child in *months* before we can calculate each of the 10 indicators. **It is very important that you use Excel to calculate the *exact* age of the child.** For the purposes of this Guide, other ways of determining age are inaccurate, including mother's report of child's age (though you sometimes need to use mother's report if the child has no recorded birthday).

To convert child's age in days to the child's age in months, look at the chart below. If a child is between the ages of 0 (born on the date of interview) and 30 days old (inclusive), he or she is 0 months old. If the child is 31 days old, he or she is considered to be 1 month old. The chart below is different from what you see on a regular calendar. The chart contains standards about the number of days in a completed month. These are standards that have been agreed upon by experts who measure IYCF.

Question: How many months old is a child who has completed 583 days? How old is a child who is 91 days old?

The correct answers are 19 months old and 2 months old.

Table 6.2.1 Converting child's age from days to months

Age of Child in Days		Completed Months
If the child is at least this many days old:	**And no older than this many days:**	**The child is this number of months old:**
0	30	0
31	61	1
62	91	2
92	122	3
123	152	4
153	182	5
183	213	6
214	243	7
244	274	8
275	304	9
305	335	10
336	365	11
366	395	12
396	426	13
427	456	14
457	487	15
488	517	16
518	547	17
548	578	18
579	608	19
609	639	20
640	669	21
670	700	22
701	730	23

Now you need to convert child's age in days to age in months. Here's how you do it:

1. Go to the first blank column at the end of your file. It should be after the variable that contains calculated age in days (**CalcAgeDays**). In row 1, name this column **CalcAgeMonths**. In our example, this is cell BK1. Highlight **CalcAgeMonths** in the color rose.

2. Now go to the bottom left corner of your Excel file. You will see tabs called 'sheet 1,' 'sheet 2' and 'sheet 3.' Click on 'sheet 2.' Now print out and look at Table 6.2.1 above.

3. You will now type in the first day of each month in column A and the month number in column B (you find this information in column 1 and 3 of the table). To do this, in cell A1 of sheet 2, type 0. Now go to cell A2. Type 31. Go to cell A3. Type 62. Keep on typing the numbers in the first column of Table 6.2.1. Your numbers should start with 0 in cell A1 and end with 731 in A25. Technically, you should end with the number 701 in cell A24. However, adding 701 to cell A25 makes sure you don't include any child in your dataset older than 23 months of age.

4. In column B of sheet 2, you will type in numbers that correspond to the months 0 through 24. (Again, you only need to type up to 23 but ending at 24 helps ensure you have not added any children who are too old for your calculations. Thus, in cell B1, type 0. In cell B2, type 1. In cell B3, type 2, and so on until you get to cell B25 where you will type 24.

5. Now go back to sheet 1. In cell BK2, type: '**=lookup**' then click on your calculated age in days for the individual in row 2 of the practice data file. That information is located in cell BJ2 of our practice data file. Now type ',' (the comma sign) and go to sheet 2. Highlight cells A1 through B25. Hit enter. You have now told Excel to look up each calculated age in days, compare it to the values in column A of sheet 2 and assign it the value in column B of sheet 2. While this exercise might seem complicated, it is the easiest way to calculate age in months and will save you a great deal of time later.

6. If for some reason you have difficulty with the preceding steps, you can simply type the following formula into cell BK2 (assuming your calculated age in days is located in cell BJ2: **=LOOKUP(BJ2,Sheet2!A1:B25)**.

7. Before you copy your formula down the rest of the column, make sure the cursor is located in cell BK2. On the toolbar at the top of your Excel sheet, you will see a window that contains the information in that cell. It is called 'fx' for 'function. Click just before 'A1:B26.' Type the dollar sign ($) just before the A, then just before the 1, then just before the B, then just before the 25. Your formula will now look like this: **=LOOKUP(BJ2,Sheet2!$A$1:$B$25)**. You use the dollar sign to tell Excel to always go to the same spot (Sheet 2, cells A1 through B25) to look up age in months. If you do not use dollar signs, your calculations will be incorrect.

8. Copy your formula in BK2 down the remainder of your data file.

9. You should now have a list of values in **CalcAgeMonths**. Unfortunately, **CalcAgeMonths** will be wrong for individuals who are missing **DOI** or **DOB**. In theory, **DOI** for any given individual should not be missing because **DOI** should have been checked by the interviewer and the field supervisor. The same is true for **DOB**. If the interviewer or field supervisor knows that a given child is missing **DOI, DOB** or **Age**, they should reject the form. If the form is rejected, it should not be entered in Excel. Since our practice data file represents fictitious data and we want to illustrate what to do if you are missing **DOI, DOB** or **Age**, we have left this individual in the dataset.

10. Remember that children with serial numbers 1 and 7 had a very high age calculated in days? If you look at **CalcAgeMonths** (should be highlighted in rose), you will see that their age according to our calculations is 24 in both cases. If you look at the variable **Age**, you will see that their actual age is 5 (for child with serial number 1) and 2 (for child with serial number 7). What to do? **This is very important:**
    a. You will need to type in the correct age for children who are missing information on date of interview and/or date of birth. In our practice data file, you will substitute the formula in cell BK2 with the number 5 and BK8 with the number 2. If you do not do this for each of the children who are missing information on date of interview and/or date of birth, your IYCF indicators will be wrong!

    b. Remember, you only need to type in the age of the child if you are missing date of interview and/or date of birth. Otherwise, leave the formula you created above.

6.2.2 Checking age calculations

It is very important to check your age calculations before you start calculating the 10 indicators. We have mentioned one potential problem above: children who are missing date of interview and/or date of birth. Yet another problem is 'age heaping.'

Age heaping occurs when you have a lot more children of one age than another. While age heaping can occur at any age, it is likely that if age heaping exists, it will happen when children are *about* 6, 12 and 18 months of age. There is a natural tendency among parents to report that their children are a year old, for example, when in fact, the child may be 11 months old or 13 months old. If you have a lot of children who are 12 months old, you may have age heaping.

While age heaping for date of birth poses a smaller challenge than age heaping for reported age in months, it can still occur. For example, if there is no way to verify the child's exact date of birth, parents may be more likely to report that the child was 'born around' a given holiday (Ramadan, Christmas or another holiday) or another special event.

In Chapter 2, we talked about constructing a local events calendar to help the mother remember date of birth with greater precision. As mentioned in Chapter 2, you can use the following resource to create the local events calendar.

> Food and Agriculture Organization. 2008. Guidelines for estimating the month and year of birth for infants and young Children. Rome: Author.

This resource should help you avoid age heaping but it is always possible that you will encounter additional problems with age. It is very difficult to correct for age heaping at the data analysis stage. As we noted earlier, the best way to address this challenge is during training of interviewers and data quality checks. Sufficient training time should be devoted to helping interviewers probe on date of birth. In addition, when interviewers and supervisors check completed incoming questionnaires, they should make sure there is no age heaping. You can also check for age heaping when you calculate age as specified in Section 6.2, above. Look at your calculated age column. Are there lots of children of certain ages? More than you would expect to see? Is there more age heaping in one site over another? You may have age heaping. If you do, go back to your variables that measure date of interview, date of birth, and age of child. Check your database against hard copies of your questionnaires. Have data been entered correctly? If not, make sure data are properly entered.

It is difficult to adjust for age heaping in your analyses. If you were conducting more complicated analyses, the issue of age heaping could be at least partially addressed. But given that your analyses are simple, the best thing to do is to determine the extent to which age heaping occurs, check data, and make corrections. Also be aware that you may have small numbers of individuals in some categories because of age heaping. For example, you may have few children 0-5 months old because some mothers said their children were 6 months old when in fact they were only 5 months old.

Note: You can use SPSS or a similar program to calculate age and each of the indicators below. If you choose to use SPSS, we have included SPSS syntax in the annex. The syntax will need to be adapted to reflect the names you give your variables but it will provide guidance about calculation.

You've done well! **Take a break and proceed to Section 6.3 on another day!**

6.3 Ten indicators to measure infant and young child feeding

There are 10 indicators you will use to measure infant and young child feeding in the program areas where you work:

1. Timely initiation of breastfeeding (children 0-23 months)
2. Exclusive breastfeeding under 6 months
3. Timely complementary feeding
4. Introduction of solid, semi-solid or soft foods
5. Continued breastfeeding at 1 year
6. Minimum dietary diversity
7. Minimum meal frequency
8. Minimum acceptable diet
9. Consumption of iron-rich or iron-fortified foods
10. Bottle feeding

Remember that timely initiation is calculated on *living* children. As we talked about previously, it is different from *early* initiation (the indicator WHO recommends) which measures the number of *living and deceased* children who received breastmilk in the first hour of life. The sections below tell you how to calculate each of these indicators – both by hand and using Excel.

It is important to note that while indicators 3 and 4 (timely complementary feeding and introduction of solid, semi-solid or soft foods) appear to be similar, they measure different things. Timely complementary feeding requires that children 6-9 months of age receive breastmilk and consume a complementary food (either a solid, semi-solid or soft food) at least one time in the past 24 hours, which you determine by looking at ALL the foods the child got. Introduction of solid, semi-solid or soft foods simply measures whether children of a somewhat different age group (6-8 months old) get ANY solid, semi-solid or soft food based on a single question at the end of the section on 24-hour recall **(How many times did [NAME] eat solid, semi-solid or soft foods other than liquids yesterday during the day or at night?)**. Introduction of solid, semi-solid or soft foods does NOT require that the child receive breastmilk.

6.4 Calculating indicators for IYCF

You are now ready to calculate the 10 indicators.

Important note about who is and is not included in each indicator: As noted in the 2009 guidance from WHO, although your sampling universe covers a two–year age range (0-23 months), most indicators are calculated for age sub-groups. For example, the exclusive breastfeeding indicator is tabulated from questions included in the IYCF module that are asked of all caregivers of children 0-23 months. The tabulation of the indicator, however, uses only the data collected for children 0-5 months. In this case, the reference group, or denominator, used for tabulation of the indicator (children 0-5 months) is a subset of the age group defined for the sampling universe (children 0-23 months). Later, will explain 2 different methods for deciding which children go in your denominator for each indicator.

Please note that our calculations for each of the 10 IYCF indicators include ALL children 0-23 months – even when multiple children were interviewed per household. In contrast, WHO recommends the calculation of indicators based on ONE child per household (the last born child). While there are disadvantages to including all age-eligible children, we are recommending what we think is feasible for most implementing agencies.

Important note about how to handle 'don't know' responses: It is likely that some of your respondents indicated they did not know the answer to a given question. How should you handle 'don't know' responses? According to WHO guidance, one approach is to recode all 'don't know' responses as missing data and to not include the missing values in the numerator or denominator for any indicator. This is the approach used by many small-scale surveys. A second approach, used in Demographic and Health Surveys and other large-scale assessments, is to recode 'don't know' and missing data as 'no' (if the question has a yes/no response) or '0' (if a numeric response is required), and to include the recoded data in the numerator and denominator of indicators. Most implementing agencies using this Guide will likely carry out relatively small surveys (compared to the DHS, for example). Consequently, we recommend that you treat 'don't know' responses as missing data: they should not be included in the numerator or denominator for any indicator.

As you calculate indicators, highlight each one in tan. What you are calculating are *prevalence* levels because you are measuring the proportion of your target population that practices a given behavior. For more information about each of the indicators, see this document:

> World Health Organization. Indicators for assessing infant and young child feeding practices. Part 2: measurement" (DRAFT July 7, 2009). Washington, DC.

When you calculate indicators by hand, you will first use your variable for calculated age in months to determine whether the child is the right age for a particular variable. Then you will see if the child meets all of the criteria for inclusion in your indicator.

When you calculate indicators using Excel, you will tell Excel what it needs to do to see if the child (or caregiver) is practicing the behavior. You need to tell Excel 3 things:

1. Which children should be included in this indicator, based on the age of the child

2. What requirements parents and children need to fulfill before we can consider them practicing the behavior, and

3. How to code the child if he or she has (or has not) met the requirements. We usually assign '1' to individuals practicing the behavior and '0' to those who do not.

As an example, here is code we used to calculate timely initiation of breastfeeding (children 0-23 months): **=IF(AND(BK2<24,R2=0),'1','0')**

It has the 3 elements we mentioned above:

1. **BK2<24** tells Excel to look in cell BK2 which is where we calculated the age of the infant, in months. If BK2 is less than 24 (months old), this child will be included in our calculation. Otherwise, she will not be included in our calculation.

2. **R2=0** tells Excel to look in cell R2 (**initiatehours**) which is where we recorded information on when the child began breastfeeding. If R2 equals 0, the child started breastfeeding in the first hour of life.

3. The **1** and **0** tell Excel what to do when the statement is true and false. If the child is less than 24 months old AND the child began breastfeeding in the first hour of life, the child initiated breastfeeding in a timely fashion (1); otherwise, the child did not (0).

Remember to highlight this indicator and each of the other 9 indicators in tan. Below, we give you specific instructions for calculating each indicator. We will show you how to calculate all 10 indicators using Excel. We will also demonstrate how to hand calculate the first 3 indicators: timely initiation of breastfeeding (children 0-23 months), exclusive breastfeeding and timely complementary feeding by hand. Hand calculations do not work very well unless you have a small sample, for example, if you are using Lot Quality Assurance Sampling (LQAS). With large datasets (for example, 100 or more infants), it is very tedious and time-consuming to calculate the indicators by hand. Even so, doing hand calculations for a few indicators will be helpful as you think about which infants are and are not included in each calculation.

6.4.1 Timely initiation of breastfeeding (children 0-23 months)

Definition: Proportion of children 0-23 months who were put to the breast within one hour of birth

$$\frac{\text{Children 0-23 months who were put to the breast within one hour of birth}}{\text{Children 0-23 months}}$$

Calculation:

$$\frac{\text{Age in months<24 and Q6=0}}{\text{Age in months<24}}$$

Notes:

Throughout the Guide, we have referred to the upper age range as the number just below the cut-off age. For example, we call the age range 0-<24, 0-23 months. However, in Excel, you will need to create syntax that refers to the number below which you want to include all individuals. So, 0-23 will be written in Excel as 0-<24.

Q6 is coded 00 when the respondent reports the baby was put to the breast in the first hour. (Excel will convert 00 to 0).

Timely initiation is calculated on *living* children. It is different from *early* initiation (the indicator WHO recommends) which measures the number of *living and deceased* children who received breastmilk in the first hour of life. The "Timely initiation of breastfeeding (children 0-23 months)" indicator is recommended in this guide for use as it is appropriate for smaller scale surveys.

Disaggregation: It is recommended that this indicator be further disaggregated and reported for (i) live births of children 0-11 months; and (ii) live births of children 12-23 months, if sample size permits. Previously, the standard indicator on children put to the breast within one hour of birth measured only children 0-11 months; calculating the indicator for this specific age group will be useful in order to compare data over time.

This indicator uses the older (1991) WHO definition based on children 0-11 months rather than the newer one (2009). The new WHO indicator requires data on all children born in the previous 23 months, living and deceased, whereas the older indicator includes only living children. To collect data on all living and deceased children, you would need to use a household roster. It is often difficult to administer a household roster. It takes a long time and is something implementing agencies do not usually do. Because of differences between our definition of timely initiation of breastfeeding and the definition used by WHO, the name we use for this indicator – timely initiation of breastfeeding (children 0-23 months) – is slightly different from the term used by WHO (timely initiation of breastfeeding).

Here is how you calculate timely initiation of breastfeeding (children 0-23 months) by hand:

1. Using a separate piece of paper, write down all of the **serial** numbers of all children in your dataset. This is your first column and should be on the left-hand side of your paper. In our example, we have serial numbers 1-20 written on separate lines.

2. Now look at the variable you created in Excel that calculated the age of the infant in days. Previously, we called it **CalcAgeDays**. In the next column, write down the age of each infant in days.

3. Now using the same piece of paper, create a third column – to the right of **CalcAgeDays**. Look at 'Table 6.2.1. Converting child's age from days to months,' above. Based on the age of the child in days and the information in Table 6.2.1, write down the age of the child in months. Child's age in months goes in the third column on your piece of paper. Your chart might look something like this:

Serial number	Age of infant in days (from the Excel worksheet)	Age of infant in months (from Table 6.2.1)

4. Now create a new table on a new piece of paper. It should look like this:

Serial numbers of children 0-23 months old	Child began breastfeeding in the first hour of life

5. Now write down all of the serial numbers of children 0-23 months old. We include all children 0-23 months old because this is the age range for this particular indicator. In this case, all 20 children will be listed in the column for serial numbers. If your dataset is very large (for example, more than 300 individuals), you may want to do this for the first 50 individuals only. This gives you practice assigning ages by hand but avoids a lot of extra work.

6. Now look at the variable that contains information about when the child began breastfeeding. In our dataset, it is called **InitiateHours**.

Note: If the child began breastfeeding 24 hours after birth or later, you will need to convert days to hours. You can do this by multiplying the number of days by 24. So for example, if **InitiateDays** is 3 (and **InitiateHours** is blank), in the **InitiateHours** column, type 72 (3 days x 24 hours per day=72). If you have only a few infants who began breastfeeding 24 hours after birth or later, you can do the multiplication in your head and simply fill in the number under **InitiateHours**. However, if quite a few infants began breastfeeding 24 hours after birth or later, you can have Excel calculate this for you. Let us assume **InitiateDays** is in column S (highlighted in light yellow) and that the infant who began breastfeeding after the first day of life is located in S2. To calculate how many hours after birth this infant began breastfeeding, in cell R2 (assuming this is where you have **InitiateHours**), type: **=S2*24**. Copy this formula to all cells where the child began breastfeeding 24 hours after birth or later. *Do not copy this formula to all cells under **InitiateDays**. Doing so will wipe out all of the data you have for children who breastfed in the first 23 hours of life.*

7. For each infant listed under the serial number column, look to see if the child began breastfeeding in the first hour of life (**InitiateHours**=000 or 00). Place a checkmark in the boxes where the child was both age eligible (first column) and began breastfeeding in the first hour of life. Your table should look something like this:

Serial numbers of children 0-23 months old	Child began breastfeeding in the first hour of life
1	
2	√
3	
4	
5	
6	√
7	√
8	√
9	√
10	
11	
12	
13	
14	
15	
16	√
17	√
18	
19	
20	
TOTAL: 20	**7**

8. Add up the number of age-eligible children (20 in this case) and the number of checkmarks in the right column (7 here). 7 out of 20 infants 0-23 months of age began breastfeeding in the first hour of life; in other words, 35%.

Note: If you have a very large database (for example, 300 individuals or more), it is tedious to manually count all of the individuals who meet your criteria for timely initiation of breastfeeding (children 0-23 months). Assuming each of your 10 indicators is coded as 1 (practices the behavior) or 0 (does not practice the behavior), you can use the sum function in Excel to add up the number of mothers practicing a given behavior. If you are familiar with a statistical software package (for example, SPSS), another option is to export the data to that software package and calculate a frequency for each indicator.

Now you are ready to calculate timely initiation of breastfeeding (children 0-23 months) using Excel. Here is how you do it:

Try it: Calculate timely initiation of breastfeeding (children 0-23 months)

1. Open your dataset (if you used the name we suggested, it is called **Practice data file version 2**).

2. Go to the last column of your database (the one furthest to the right). Next to it you will see a blank column. Click on the first row of that column.

3. Type the name of the new variable (the indicator for timely initiation of breastfeeding). **TI** (which stands for timely initiation of breastfeeding (children 0-23 months)) is a good name.

4. Hit enter.

5. In row 2, type the formula for calculating timely initiation of breastfeeding. This is the formula but you will need to change the cell addresses to match your database:

   =IF(AND(BK2<24,R2=0),1,0)

6. Replace **BK2** with the address in your database for the age of the child in months. Right now, you are only doing this for the first child in your database: the child in row 2.

   a. Remember that if you are missing date of birth or date of children for any of your children, Excel will incorrectly calculate the age of the child. Follow step 7 at the end of Section 6.2 to make sure you have the correct age for every child.

7. Replace **R2** with the address for the variable that tells you whether the child began breastfeeding in the first hour of life.

Question: Let us suppose that in your dataset, you have the age of the first child stored in cell F2 and information about when the child began breastfeeding in cell K2. How would you tell Excel how to determine whether this child began breastfeeding in the first hour of life?

Answer: **=IF(AND(F2<24,K2=0),1,0)**

8. Now look at your own database and type in the correct cell addresses.

9. Once you have finished, copy your formula all the way down the column (to the last observation in the dataset). You can use the mouse to grab on to the lower right-hand corner of the cell, as explained in Section 6.2 step 12.

Note: You can also calculate each of the indicators by inserting functions. We do not explain how to do this in this Guide because we feel it does not save you any time. But if you're interested in having Excel insert functions for you, you can click on Insert at the top of your screen, then function. You can also look up insert function in Excel's help menu.

When we used Excel to calculate timely initiation of breastfeeding for the Practice data file we got these results:

There are 20 mothers. We were able to calculate age based on date of birth for 18 of the 20 mothers. For the remaining two mothers (children with serial numbers 1 and 7), we were able to get a good estimate of the age of the child in months so we typed the values found in our column for **Age** in **CalcAgeMonths**. We used all 20 mothers for this calculation: all had age of child and all were 0 to 23 months of age. Seven of the 20 mothers began breastfeeding in the first hour of life. Their serial numbers are 2, 6-9 and 16-17. Here is our calculation:

$$\frac{\textbf{7 children born in the last 23 months who were put to the breast within 1 hour of birth}}{\textbf{20 children born in the last 23 months}}$$

This equals 0.350. Multiply this number by 100 to get 35.0%. We will round all numbers to the nearest whole number.

Answer: 35% of mothers of infants less than 2 years of age began breastfeeding in the first hour of life.

Write down the percent of infants who initiated breastfeeding in the first hour of life based on your own data. You will need to refer back to this number in Chapter 7.

When you present your results to others, please use language that is this specific. Later, we will tell you how to produce a graphic with this information. We will also suggest ways you can interpret your results. Also, double-check your hand calculation against the calculation you did using Excel. The answers here are consistent.

Tip: It is easy to hand tally the number of infants who practice an optimal feeding behavior when your dataset is small. For example, our dataset has only 20 individuals. However, if you file is much larger, it will take a long time to count up all the infants who practice a given infant feeding behavior. One shortcut is to have Excel add up all the 1's in your indicator column. Thus, for timely initiation of breastfeeding (children 0-23 months), you can go to the bottom of the **TI** column and after the last observation, type =**sum** (then click on all of the observations in the column, from the first observation to the last. Your syntax will look something like this: **=SUM(BL2:BL21)**. If you have more than 20 observations, your range will be greater. Excel adds up all of the individuals who have a **1** for **TI**. A shortcut is to go to the first blank cell after your last observation (under **TI**), then click on the icon at the top of the screen that looks like this: Σ. Excel then asks you which cells you want included in your summation. Usually, Excel picks the cells automatically. *Make sure that Excel includes all of your observations for TI from the very first one to the last one. Then hit enter. Be sure to occasionally double-check how Excel is summing up. It is easy to make a mistake when you tell Excel what to add.* The above step gives you the number of infants in your numerator.

Now you need to determine how many age-eligible infants there are in your denominator. *Who* belongs in your denominator depends on the *age* of the infant or child and, in some cases, whether the child is *breastfeeding*. For timely initiation of breastfeeding for living children, you want all children in your dataset since the indicator covers children 0-23 months of age. Your denominator is the total number of children 0-23 months old in your dataset.

6.4.2 Exclusive breastfeeding under 6 months

Definition: Proportion of infants 0-5 months who are fed exclusively with breastmilk

$$\frac{\text{Infants 0-5 months who received only breastmilk during the previous day}}{\text{Infants 0-5 months}}$$

Calculation:

$$\frac{\begin{array}{c}\text{Age in months<6}\\ \text{AND}\\ \text{(Q8=1) AND (Q9c-j all=0) AND (Q10aa-qq all=0)}\end{array}}{\text{Age in months<6}} \times 100$$

Notes on calculation:

Q8 asks about feeding of breastmilk yesterday.

Q9 captures information about liquids that child had yesterday, and Q10 captures the information about foods the child ate yesterday.

Q9 includes two items that are allowed under the definition of exclusive breastfeeding (ORS and vitamin and/or mineral supplements).

The instructions above may be adapted if country-specific liquids or foods are added to Q9 or Q10 to incorporate new liquids and foods in the calculation.

This indicator includes breastfeeding by a wet nurse and feeding expressed breastmilk. The indicator is based on recall of the previous day for all living infants 0-5 months. This indicator does *not* represent the proportion of infants that are exclusively breastfed until just under 6 months and *should not* be interpreted as such. Rather, it represents the proportion of children of various ages who are less than 6 months old who breastfed exclusively in the previous 24 hours. Despite this limitation, the indicator is the best option for estimating exclusive breastfeeding. See Calculating Indicator Values, Exclusive Breastfeeding Indicator from WHO's Indicators for Assessing Infant and Young Child Feeding Practices.

Disaggregation: It is recommended that the indicator be further disaggregated and reported for the following age groups: 0-1 month, 2-3 months, 4-5 months and 0-3 months, if sample size permits.

Hand calculation:

Follow the instructions above for hand calculating timely initiation of breastfeeding (children 0-23 months). For exclusive breastfeeding, you will have to determine whether the child:

- Is age eligible (look at the **CalcAgeMonths** variable)
- Breastfed yesterday or last night (look at the **BFlastnight** variable)
- Received no other liquids nor foods (here you have to look at ALL of the variables between and including 'water' and 'condiments.'

Your table and answers should look something like this:

Serial numbers of children 0-5 months old	Children who breastfed yesterday or last night	Children who breastfed yesterday or last night AND who received no liquids or foods yesterday or last night
1	√	√
2	√	
3	√	√
4	√	
7	√	√
12	√	√
13	√	√
14		
17	√	√
TOTAL: 9		**6**

Add up the number of age-eligible children (9 in this case) and the number of checkmarks in the column furthest to the right (6). 6 out of 9 infants 0-5 months of age began breastfeeding in the first hour of life, in other words, 66.7%.

A couple of important reminders about hand calculations:

- *All criteria have to be met.* For example, a child must be 0-5 months of age AND breastfeeding in the previous 24 hours AND receiving no other foods or liquids.
- As you determine this, move from the left-hand column to the right-hand column. You only need to check boxes for children who are eligible. In the example above, when we check our dataset to determine whether the child had any liquids or foods yesterday or last night, we need to look at only the following children: 1, 2, 3, 4, 7, 12, 13 and 17. We do not need to look at child 14 because even though she is age eligible, she did not breastfeed in the previous 24 hours so she cannot be considered to be exclusively breastfed.

Calculating the indicator in Excel:

1. Insert a column to the right of TI (your timely initiation indicator; it can be tan). In our practice file, this is column BM. Call it **LIQFOODS** (it can be light blue). To the right of **LIQFOODS**, create another variable called **EBF** (it should be tan).

2. There are 3 parts to this indicator: age, whether the child was breastfed yesterday or last night and whether the child got any liquids or foods in the previous 24 hours.

3. Let us start with age. For timely initiation of breastfeeding (children 0-23 months), you chose infants 0-23 months of age. Now you need only those infants who are 0-5 months old. In a minute we will tell you how to select these individuals. But first let us determine whether the child was breastfed (at all) yesterday or last night.

4. Where is the information about whether the child was breastfed yesterday or last night? In our file, it is in column Y (**bflastnight**). In order for a child to have been breastfed in the previous 24 hours, Y must equal 1.

5. Where is the information about the liquids and foods given to the child in the previous 24 hours? In our file it is located in columns AB through BC. (Remember that a child can be given vitamins/minerals and ORS and still be considered exclusively breastfed so we don't need to include these variables in our calculation). The easiest way to determine whether the child was given *any* liquids and foods is to 'add' them up. Remember that when you entered data about whether a child received a given liquid or food, 1=yes and 0=no. When you add all of the liquids and foods, any number equal to or greater than 1 means the child got at least 1 substance. By definition, these children are *not* exclusively breastfed.

Note: Some mothers may not have known whether their child received a given liquid or food. When you entered this information into your database, you assigned an 8 for this particular child and liquid or food. This information is helpful in understanding how often children's parents did not know what their child drank or ate. However, if you leave these responses as 8's now, they will throw off your calculation of **LIQFOODS. For ALL liquids and foods (columns AB through BC), delete the 8's and leave these cells blank.** Doing so helps you avoid assigning unnecessarily high numbers to **LIQFOODS**.

Now is a good time to delete ALL of the 8's you have in your database (except for the variable InitiateHours).

6. We will add up the liquids and foods and put this information in the column we just created (**LIQFOODS**).

7. In row 2 under **LIQFOODS** type: =**sum** then go to the first cell in row 2 that contains a liquid. Our first liquid in the questionnaire is water. It is located in column AB. Click on the first observation in this column (AB2). Now click on the '+' sign to indicate you want to add other variables as well. You will add all of the variables between AB2 and BC2 *except* for **FORMULATIMES, MILKTIMES** and **OTHERTIMES**. You will not add these variables because they tell you how many times each of these substances was given, not whether they were given. Your syntax for this variable should look like this:

   =AB2+AC2+AE2+AG2+AH2+AI2+AJ2+AL2+AM2+AN2+AO2+AP2+AQ2+AR2+AS2+AT2+AU2+AV2+AW2+AX2+AY2+AZ2+BA2+BB2+BC2+BG2

8. Please note that if lipid-based nutrient supplements are made from peanuts (for example, Plumpy Nut), they should be counted as a nut (as well as an iron fortified food). We have accounted for this by including the cell BG2 in our syntax above.

9. Once you have highlighted the last food, hit enter. You have now added up all the liquids and foods the mother gave. You are ready to calculate exclusive breastfeeding!

10. In row 2 of your EBF variable, type =**if(and)** and click on the cell containing information about the age of the child. In our case, it is cell BK2. Once you have clicked on the cell, type **<6,** (<6 followed by the comma) to tell Excel that you want children less than 6 months of age (0-5 months old). Now click on the cell that contains information about whether the child was breastfed yesterday or last night. In our file, it is located in cell **Y2**. Once you have clicked on that cell, type **=1,** (=1 followed by the comma). Now find the cell that adds up all the liquids and foods. For us, it is **BM2**. Click on it and type **=0**), then type **1,0**) and enter. Your syntax will look like this: **=IF(AND(BK2<6,Y2=1,BM2=0),1,0)**.

11. You have just determined whether the first child in your database breastfed exclusively. You can find out which other infants breastfed exclusively by copying the formula in row 2 to the end of your file.

12. Do not forget to highlight this indicator and all other indicators using the color tan.

13. Now you need to have Excel tally the number of infants in your numerator and denominator. For the numerator, follow the same steps you used to calculate the numerator for timely initiation of breastfeeding (see Section 6.4.1) using **=sum** or the Σ function. For the denominator, you only want children who are 0-5 months old. Use **Autofilter** to select age-eligible infants, in this case, infants who are 0-5 months old. Per previous instructions, you do this by highlighting your **CalcAgeMonths** variable, clicking on **Data** and **Autofilter**, then clicking on the down arrow next to the variable name (**CalcAgeMonths**). Now click on **Custom** and **is greater than or equal to**. In the box on the right hand side, type 0. Now click on **is less than or equal to** and type 5 and **OK**. You have now limited your sample to children who are the appropriate age for the calculation of exclusive breastfeeding. Here you will need to manually count up the total number of individuals who remain on your screen after the filter. Summing them only adds the number of children who are exclusively breastfeed, not all children 0-5 months old.

14. For an alternative approach to determining how many children are in your denominator – for this indicator and the other 9 indicators – see the end of Chapter 6.

When we used Excel to calculate exclusive breastfeeding for the Practice data file, we got the following results:

There are 20 mothers. We used only the mothers whose children were 0 to 5 months of age (9 mothers). (**Note:** One infant was 6 months old but that is too old for this indicator; the child must be *less than* 6 months of age.) Six of the nine women breastfed their infant in the previous 24 hours and gave no other liquids or foods. Here is our calculation:

$$\frac{\textbf{6 infants 0-5 months of age who received only breastmilk during the previous day}}{\textbf{9 infants 0-5 months of age}}$$

This equals 0.666. Multiply this number by 100: 66.6%.

Answer: 67% of mothers of infants less than 6 months of age breastfed exclusively in the previous 24 hours.

Write down the percent of infants who were exclusively breastfeed based on your own data. You will need to refer back to this number in Chapter 7.

Please be careful when interpreting results from the practice dataset or from other small datasets. Because there are so few observations, it is difficult to arrive at meaningful conclusions.

When you use Excel to calculate indicators, remember that **Excel will assign a 1 to all individuals who meet the criteria you specify**. This is your denominator (the top half of your equation). This number is correct. **However, Excel assigns a 0 to ALL remaining individuals in the dataset, regardless of age.** So, for exclusive breastfeeding, Excel assigns a 0 for every child who does not meet the criteria for exclusive breastfeeding **(including children 6-23 months of age)**. Obviously, you only want infants 0-5 months of age in your indicator of exclusive breastfeeding. How do you correct this?

1. You already have the correct numerator (number of children 0-5 months old who exclusively breastfeed). It is all the **1**'s under **EBF**. You can count this number by hand, you can use **Autofilter** or you can use a **=sum** function at the bottom of your **EBF** column. Having Excel sum up the **1**'s is the easiest approach:
   a. Select all cells then click on **Data, Filter, Autofilter** (make sure **Autofilter** is on)
   b. For your column labeled **EBF**, click on the down arrow and select 1
   c. Count up the number of individuals with a 1 under **EBF**
2. Now reset **Autofilter** so that all children are displayed in your database
3. Use **Autofilter** to determine how many children there are in the denominator. For the exclusive breastfeeding indicator, **CalcAgeMonths** should be less than 6.
   a. Select all cells then click on **Data, Filter, Autofilter** (make sure **Autofilter** is on)
   b. For your column labeled **CalcAgeMonths**, click on the down arrow and select **Custom, is less than** and 6
   c. Count up the number of individuals once **Autofilter** is on. Remember, you are interested in the total number of infants and children less than 6 months of age, regardless of whether they exclusively breastfed or not.

You should follow the same procedure for all other indicators too. You will need to select the appropriate age range for each indicator, for example, infants 6-8 months of age for the introduction of solid, semi-solid or soft foods.

6.4.3 Timely complementary feeding

Definition: Proportion of infants 6-9 months who received breastmilk and a solid or semi-solid food (based on 24-hour dietary recall). Solid and semi-solid foods are defined as mushy or solid foods, not fluids.

$$\frac{\text{Number of infants 6–9 months who received breastmilk } \textit{and} \text{ a solid or semi-solid food}}{\text{Total number of infants 6–9 months}}$$

Calculation:

$$\frac{\begin{array}{c}\text{Age in months 6<10 months}\\ \text{AND}\\ \text{(Q8=1) AND (one or more of Q10aa-10qq)}\end{array}}{\text{Age in months 6<10 months}} \times 100$$

Notes:

Q8 asks about feeding of breastmilk yesterday. The infant can receive breastmilk either by breastfeeding or another means.

Q10 captures the information about foods the child ate yesterday.

The "timely complementary feeding" indicator is sometimes called "breastfed with complementary foods." "Timely complementary feeding" is historically a standard indicator for measuring IYCF practices. Because it requires that children have received both breastmilk and a solid or semi-solid food in the last 24 hours, it can be more difficult to interpret than the newer WHO 2009 complementary feeding indicator measuring "introduction of solid, semi-solid and soft foods" (see indicator below). However, we recommend calculating "timely complementary feeding" as it may be useful for making comparisons over time using historical data.

Also note that this indicator is calculated for a 4-month time period while the "introduction of solid, semi-solid or soft foods" indicator is calculated for a 3-month time period.

Disaggregation: Further disaggregation of this indicator is not recommended because of the narrow age range.

Hand calculation:

For timely complementary feeding, you will have to determine whether the child:

- Is age eligible (look at the **CalcAgeMonths** variable; it is rose)
- Breastfed or received breastmilk yesterday or last night (look at the **BFlastnight** variable; it is pink)
- Received solid, semi-solid or soft foods

Remember, it is tedious to calculate this indicator and any of the 9 other indicators by hand. If you have a large dataset, determine timely complementary feeding for the first 50 individuals only. That way you practice calculating the indicator by hand.

Your table and answers should look something like this:

Serial numbers of children 6-9 months old	Children who breastfed or received breastmilk yesterday or last night	Children who breastfed or received breastmilk yesterday or last night AND had solid, semi-solid or soft food yesterday or last night
8	√	√
12	√	
16	√	
19		
20	√	√
TOTAL: 5		**2**

Add up the number of age-eligible children (5 in this case) and the number of checkmarks in the column furthest to the right (2). Two out of 5 infants 6-9 months of age received breastmilk and a solid, semi-solid or soft food in the previous 24 hours; in other words, 40%.

Calculating the indicator in Excel:

1. Insert a column to the right of EBF (your exclusive breastfeeding feeding indicator). Call it **TCF**.

2. To calculate timely complementary feeding, we want to first look at age of the child then whether the child was breastfed last night, followed by whether the child got any other foods.

3. Only infants who are at least 6 months and less than 10 months old will be included in your calculation.
   a. The first part of your syntax will look something like this (your age variable may be somewhere other than column BK): **=IF(AND(BK2>=6,BK2<10**. This tells Excel that you want to include all infants who are at least 6 months old (that is, had their six-month birthday today) or are older than 6 months but less than 10 months old (that is, they had not yet turned 10 months of age).

4. Now you need to have Excel check to see:
   a. Whether the child is currently being breastfed, and

   b. Whether the child is getting solid foods, semi-solid foods or soft foods or some combination. It does not matter whether the child receives any other foods or liquids (including non-human milk and formula) as long as he or she gets breastmilk and at least one solid, semi-solid, or soft food.

5. In our sample file (Practice data file), we have information on whether the child is currently being breastfed in column Y (**Bflastnight**) so this part of our syntax is simply **Y2=1**.

6. Now we want to see if the child got any other foods. To do this, create a new variable. Do so by inserting a column to the right of your **EBF** column. The new variable should be located in column BO. Call the new variable **solidsemisolidsoft** by typing that name in cell BO1. If you want, you can use the color plum for cell BO2.

7. In row 2 of **solidsemisolidsoft**, tell Excel where you have information on solid, semi-solid and soft foods stored. In our example, these foods start with grains in column AM and end with condiments in column BC. Type: **=sum(AM2:BC2)**

   This tells Excel to add up all the 1's in row 2 from column AM to column BC. Your column addresses will probably be different. Be sure to change them to indicate the columns for your solid and semi-solid food variables. We have 17 variables measuring solid, semi-solid and soft foods (columns AM through BC). You may have more or fewer variables. Simply tell Excel the starting and ending columns for variables measuring foods.

8. Copy the formula in row 2 down to the end of the column.

9. Now that you have a single measure of how many solid, semi-solid and soft foods were given, creating the syntax is much easier. For example, we know that if the total in column BO (**solidsemisolidsoft**) is greater than 0, the child received complementary foods and has therefore met this part of the requirement for receiving complementary foods.

10. Putting the different parts of our syntax together, we get:

    IF(AND(BK2>=6,BK2<10,Y2=1,BO2>0),1,0)

    which is another way of saying 'if the child is at least 6 months old but less than 10 months of age, still breastfeeding and getting at least one solid, semi-solid or soft food, then the child has been introduced to solid, semi-solid and soft foods.

11. Using step 10 as an example, enter this formula with the correct cell addresses for your database in row 2 of the column you labeled TCF. You can type in cell addresses or simply click on them. For example, you could type 'BK2' or you can simply click on cell BK2 and Excel will put this 'address' into the formula.

12. Copy the formula to the rest of the column.

13. Use Excel to calculate the numerator (see previous instructions in Section 6.4.1 and 6.4.2). The denominator includes all children who are 6-9 months of age. You will need to use **Autofilter** to select individuals based on age. Remember, the denominator includes ALL children 6-9 months of age, not just children who were breastfed. To select infants 6-9 months of age, highlight the entire workbook (click on the upper left corner of your screen - the cell to the left of column A and above row 1). Click on **Data**, **Filter**, **Autofilter** and then click on the down arrow for **CalcAgeMonths** and use the **Custom** feature to select all children with **CalcAgeMonths** between 6 and 9. When you count up the total number of children, this is your denominator.

When we used Excel to calculate introduction of complementary foods for the Practice data file, we got these results:

There are 20 mothers. We used only the mothers whose children were 6 to 9 months old (5 mothers). Two of the 5 women breastfed their infant in the previous 24 hours AND gave a solid or semi-solid food. Here is our calculation:

$$\frac{\textbf{2 infants 6–9 months receiving breastmilk and solid or semi-solid foods}}{\textbf{5 infants 6–9 months}}$$

This equals 0.400. Multiply this number by 100: 40.0%.

Answer: 40% of mothers of infants 6-9 months of age introduced complementary foods.

Write down the percent of infants who were given complementary foods in a timely manner based on your own data. You will need to refer back to this number in Chapter 7.

These numbers are so small that we cannot say anything meaningful about our results. However, once you calculate this proportion for the study, you should be more confident about your findings because you will have a much larger sample size.

You've done well! **Take a break and proceed to Section 6.4.4 on another day!**

6.4.4 Introduction of solid, semi-solid or soft foods

Definition: Proportion of infants 6-8 months who receive solid, semi-solid or soft foods

$$\frac{\text{Number of infants 6-8 months old who received solid, semi-solid or soft foods during the previous day}}{\text{Infants 6-8 months old}}$$

Calculation:

$$\frac{\text{Age in months 6<9 AND (Q11>0)}}{\text{Age in months 6<9}} \times 100$$

Notes:

Q 11 asks about frequency of feeding of solid, semi-solid or soft foods yesterday.

According to the WHO Indicator Modules (2009), "Solid, semi-solid or soft foods include family foods, and also many special dishes prepared for infants and young children. Thick soups and stews should be included. Thick paps and porridges are also included. Very thin, watery soups and gruels should not be included because infants and young children do not get enough energy (calories) from very thin soups and gruels. Liquids do not count for this question. Also, very small snacks, such as a bite or two of someone else's food, should not be counted." (page 15)

Because the indicator has a relatively narrow age range of 3 months, estimates from surveys with small sample sizes are likely to have wide confidence intervals.

Disaggregation: Further disaggregation of this indicator is not recommended because of the narrow age range.

1. Insert a column to the right of **TCF** (your indicator for introducing complementary foods). This should be column BP. Call it **intro**.

2. To choose infants who are 6-8 months old, your syntax will begin:

   =IF(AND(BK2>=6,BK2<9,

3. Here is what your syntax should look like (substituting in the appropriate cell addresses). Cell BD2 is a plum color:

   =IF(AND(BK2>=6,BK2<9,BD2>0),1,0)

4. Using step 4 as an example, put in the correct cell addresses for your database.

5. Copy the formula to the rest of the column.

6. Use Excel to sum your numerator. Use **Autofilter** in Excel to help you determine your denominator.

When we used Excel to determine the percent of infants who received solid, semi-solid or soft foods using the Practice data file, we got these results:

There are 20 mothers. We used only the mothers whose children were 6 to 8 months old (4 mothers). Three of the 4 women gave their child a solid, semi-solid or soft food. Here is our calculation:

$$\frac{\textbf{3 infants 6–8 months old receiving solid, semi-solid or soft foods}}{\textbf{4 infants 6–9 months}}$$

This equals 0.750. Multiply this number by 100: 75.0%.

Answer: 75% of mothers of infants 6-8 months of age introduced solid, semi-solid or soft foods in a timely fashion.

Write down the percent of infants who were given solid, semi-solid or soft foods based on your own data. You'll need to refer back to this number in Chapter 7.

Again, these numbers are very small. We are only using them to show you how to calculate each indicator.

6.4.5 Continued breastfeeding at 1 year

Definition: Proportion of children 12-15 months old who are fed breastmilk

$$\frac{\textbf{Children 12-15 months old who received breastmilk during the previous day}}{\textbf{Children 12-15 months}}$$

Calculation:

$$\frac{\textbf{Age in months 12<16 AND Q8=1}}{\textbf{Age in months 12<16}} \textbf{ x 100}$$

Notes:

Q8 asks about feeding of breastmilk yesterday.

Because the indicator has a relatively narrow age range of 4 months, estimates from surveys with small sample sizes are likely to have wide confidence intervals.

This indicator includes breastfeeding by a wet nurse and feeding expressed milk. The infant can receive breastmilk either by breastfeeding or another means.

Disaggregation: Further disaggregation of this indicator is not recommended because of the narrow age range.

We recommend, however, that a similar indicator be calculated at around 2 years – i.e., continued breastfeeding at 2 years (20-23 months).

1. Insert a column to the right of **intro** (your indicator for solid, semi-solid or soft foods). Call it **continuedBF**
2. To choose infants who are 12-15 months old, your syntax will begin: **=IF(AND(BK2>=12,BK<16,**
3. Ask Excel whether the child is still breastfeeding: **Y2=1**. This is a pink cell.
4. Putting our syntax together, we get:

   =IF(AND(BK2>=12,BK2<16,Y2=1),1,0)

   which is another way of saying 'if the child is at least 12 months old, but less than 16 months old and still breastfeeding, the child is considered to have continued breastfeeding.'
5. Using step 4 as an example, put in the correct cell addresses for your database.
6. Copy the formula to the rest of the column.
7. Use Excel to sum your numerator. Use **Autofilter** in Excel to help you determine your denominator.

When we used Excel to determine the percent of infants who continue to breastfeed, we got these results:

There were two infants 12 to 15 months old. Both of them gave breastmilk the previous day. Here is our calculation:

$$\frac{\textbf{2 infants 12–15 months old receiving breastmilk}}{\textbf{2 infants 12–15 months}}$$

This equals 1.0. Multiply this number by 100: 100.0%.

Answer: 100% of mothers of infants 12 to 15 months of age continued to breastfeed their children.

Write down the percent of infants who continued to breastfeed at 1 year of age based on your own data. You will need to refer back to this number in Chapter 7.

6.4.6 Minimum dietary diversity

Definition: Proportion of children 6-23 months who receive foods from 4 or more food groups

$$\frac{\text{Children 6-23 months old who received foods from 4 or more food groups during the previous day}}{\text{Children 6-23 months}}$$

Calculation:

To calculate a value for this indicator, a 7 food group score variable needs to be created. The instructions below show how to calculate the 7 food group score. This is followed by instructions for calculating the minimum dietary diversity indicator.

Calculation of 7 food group score:

The 7 food groups used for calculation of this indicator are:

- Grains, roots and tubers
- Legumes and nuts
- Dairy products (milk, yogurt and cheese)
- Flesh foods (meat, fish, poultry and liver/organ meats)
- Eggs
- Vitamin-A rich fruits and vegetables
- Other fruits and vegetables

Construct the 7 food group score as follows:

- Begin with a score of 0.
- For each of the 7 food groups, add a point if any food in the group was consumed.
  - Food group 1 (Grains) — Add 1 point if: Q10aa=1 OR Q10bb=1
  - Food group 2 (Legumes) — Add 1 point if: Q10dd=1
  - Food group 3 (Dairy) — Add 1 point if: Q9d=1 OR Q9e=1 OR Q9i=1 OR Q10mm=1
  - Food group 4 (Flesh foods) — Add 1 point if: Q10hh=1 OR Q10ii=1 OR Q10jj=1 OR Q10kk=1
  - Food group 5 (Eggs) — Add 1 point if: Q10ll=1
  - Food group 6 (Vitamin-A rich fruits and vegetables) — Add 1 point if: Q10cc=1 OR Q10ee=1 OR Q10ff=1 OR Q10oo=1
  - Food group 7 (Other fruits and vegetables) — Add 1 point if: Q10gg=1

Calculation of minimum dietary diversity indicator:

$$\frac{\text{(Age in months 6<24) AND (7 food group score>4)}}{\text{Age in months 6<24}} \text{ x 100}$$

Notes:

Q9 and Q10 ask about liquids and foods the child consumed the previous day. The instructions may need to be adapted if country-specific liquids and foods are added to Q9 and Q10 as separate items.

Some foods are included in Q10 for completeness, but do not appear in any of the food groups above. They are included in the question because if they were omitted from the list, interviewers might assign them to one of the other groups that do count in the score. In addition, a complete list is also required for calculation of the exclusive breastfeeding indicator.

It does not matter how much food the child gets; any amount counts. There is no minimum quantity. However, if the food is only used as a condiment, don't count it.

Breastmilk is not counted in any of the above food groups because this indicator is meant to reflect the quality of the complementary food diet. See the timely complementary feeding indicator above for a measure of complementary feeding that includes breastmilk.

Disaggregation: Results may be reported separately for breastfed and non-breastfed children. However, diversity scores for breastfed and non-breastfed children should not be directly compared because breast milk is not "counted" in any of the above food groups. As a consequence, this indicator may show "better" results for children who are not breastfed than those who are breastfed in populations where formula and/or milk are commonly given to non-breastfed children.

It is recommended that the indicator be further disaggregated and reported for the following age groups: 6-11 months, 12-17 months and 18-23 months, if sample size permits.

1. You will insert 7 columns to the right of ContinuedBF (your continued breastfeeding indicator). Call them: **grainsgroup** (column BS), **legumesgroup** (column BT), **dairygroup** (column BU), **fleshfoodsgroup** (column BV), **eggsgroup** (column BW), **vitafoodsgroup** (column BX) and **otherfruitsveggroup** (column BY). You can use the same coloring scheme that you used above for each of these variables. So for example, **grainsgroup** can be brown, **legumesgroup** can be gold and so on. This will help you keep track of which foods belong to which groups. The colors will also help you calculate this indicator.

2. Go to cell BS2. Add up all of your grains, roots and tubers. In our dataset, these are located in cells AM2 and AN2. These locations might be different for your grains, roots and tubers. Your syntax in cell BS2 should look like this: **=AM2+AN2**. These are the brown cells.

3. Go to cell BT2. Add up all of your legumes and nuts. In this case, you have only one column representing legumes and nuts. The syntax in BT2 is: **=AP2+BG2**. This is a gold cell.

4. Go to cell BU2. Add up all of your dairy products. Your syntax should look like this: **=AC2+AE2+AJ2+AY2**. These are sky blue cells.

5. Go to cell BV2. Add up all of your flesh foods. Your syntax will look like this: **=AT2+AU2+AV2+AW2**. These are red cells.

6. Go to cell BW2. Your syntax for the eggs group should look like this: **=AX2**. This is a yellow cell.

7. Go to cell BX2. Your syntax for vitamin A rich foods will look like this: **=AO2+AQ2+AR2+BA2**. These are light orange cells.

8. Go to cell BY2. Your other fruits and vegetables should look like this: **=AS2**. This is a green cell.

9. The number of servings per food group may be helpful to you in your program as you think about what food groups (and how many food groups) children receive.

10. Now that you have categorized all of your food groups, you are ready to add them up. However, if you simply add them, your calculations may not be correct, especially if a given child got more than 1 food in a given group.

11. You will now insert 7 additional columns to the right of **othfruitsveggroup** (column BY). Call the first variable **grainsgroup2** (column BZ), **legumesgroup2** (column CA), **dairygroup2** (column CB), **fleshfoodsgroup2** (column CC), **eggsgroup2** (column CD), **vitafoodsgroup2** (column CE) and **otherfruitsveggroup2** (column CF). You can use the same coloring scheme that you used above for each of these variables. So for example, **grainsgroup2** can be brown, **legumesgroup2** can be gold and so on.

12. In **grainsgroup2** (column BZ), type in the following syntax in cell BZ2: **=IF(AND(BS2>=1),1,0)**. You will need to change your cell address if **grainsgroup** is not in column BS. Copy this formula to the end of the column.

13. Click on the handle in cell BZ2 and drag and copy this formula to **otherfruitsveggroup2** (cell CF2), moving *across* the columns (not down).

14. With cells BZ2 to CF2 highlighted (**grainsgroup2** through **otherfruitsveggroup2**), drag the handle *down* to the end of the columns. You've now copied all of the formulas such that Excel can determine whether the child got at least 1 item within each food group. If so, the child is assigned a 1. If not, the child is assigned a 0.

15. Now add a column to the right of **otherfruitsveggroup2**. Call it **numbergroups**. In our file, it is located in column CG. In cell CG2 under **numbergroups**, add up cells BZ2 through CF2. You can use the sum function (Σ) if you want. Your syntax will look like this: **=SUM(BZ2:CF2)**. Copy this formula to the end of the column.

16. Insert a new variable to the right of **numbergroups**. Call it **diversity**. Choose infants who are 6-23 months old: **=IF(AND(BK2>=6,BK2<24,**

17. Use the **numbergroups** variable to see if the child received foods from 4 or more groups: **CG>=4** in our example.

18. Putting the different parts of our syntax together, we get:

    =IF(AND(BK2>=6,BK2<24,CG>=4),1,0)

    which is another way of saying 'if the child is at least 6 months old but less than 24 months old and getting at least 4 of the 7 food groups, then the child is considered to have adequate dietary diversity.'

19. Copy this formula to the end of the column.

20. Now insert a variable to the right of **numbergroups** (cell CA1). Call it **diversity**. This is the indicator for minimum dietary diversity. In our file it is in column CH. You can highlight this indicator in tan. Type the syntax above (step 18) in column CH2. Copy the formula down column CH until you reach the last observation.

21. Using step 4 as an example, put in the correct cell addresses for your database.

22. Use Excel to sum your numerator. Use **Autofilter** in Excel to help you determine your denominator.

When we used Excel to determine minimum dietary diversity, we got these results:

There were 11 infants 6 to 23 months old. One of them received foods from 4 or more groups during the previous day. Here is our calculation:

$$\frac{\textbf{1 child 6-23 months old who received foods from 4+ food groups in the previous day}}{\textbf{11 children 6-23 months of age}}$$

This equals 0.090. Multiply this number by 100: 9.0%.

Answer: 9% of mothers of infants 6 to 23 months of age met the standards for minimum dietary diversity.

Write down the percent of infants who demonstrated minimum dietary diversity based on your own data. You'll need to refer back to this number in Chapter 7.

6.4.7 Minimum meal frequency

Definition: Proportion of breastfed and non-breastfed children 6-23 months who receive solid, semi-solid or soft foods (but also including milk feeds for non-breastfed children) the minimum number of times or more

$$\frac{\textbf{\textit{Breastfed} children 6-23 months who received solid, semi-solid or soft foods the minimum number of times or more during the previous day}}{\textbf{Breastfed children 6-23 months old}}$$

And

$$\frac{\textbf{\textit{Non-breastfed} children 6-23 months who received solid, semi-solid or soft foods or \textit{milk feeds} the minimum number of times or more during the previous day}}{\textbf{Non-breastfed children 6-23 months old}}$$

Minimum is defined as:

- 2 times for breastfed infants 6-8 months
- 3 times for breastfed children 9-23 months
- 4 times for non-breastfed children 6-23 months

'Meals' include both meals and snacks (other than trivial amounts), and frequency is based on caregiver report.

Calculation:

$$\frac{\textbf{(Age in months 6<9) AND (Q8=1) AND (Q11>=2) OR (Age in months 9<24) AND (Q8=1) AND (Q11>=3) OR (Age in months 6<24) AND (Q8=0) AND ((Q9d +Q9e +Q9i+Q11)>=4)}}{\textbf{Age in months 6<24}} \textbf{ x 100}$$

Notes:

Q8 asks about feeding of breastmilk yesterday.

Q11 asks about frequency of feeding of solid, semi-solid or soft foods yesterday.

Q9d asks about the number of time the child received infant formula yesterday, Q9e asks about the number of times the child received animal milk or powdered milk yesterday and Q9i asks about the number of times the child received yogurt or sour milk yesterday. Questions on "milk feeds" are calculated for non-breastfed children only.

Disaggregation: It is recommended that the indicator be further disaggregated and reported for the following age groups: 6-11 months, 12-17 months, and 18-23 months, if sample size permits. Results may also be reported separately for breastfed and non-breastfed children.

1. Create three new columns to the right of 'diversity.'
   - Call the first column **BF6-8** This column focuses only on breastfed infants 6-8 months old
   - Call the second one **BF9-23** This column focuses only on breastfed infants 9-23 months old
   - Call the third one **NBF6-23** This column focuses only on non-breastfed infants 6-23 months old

2. Highlight these three variables in lavender. In the first column (BF6-8), type:

   =IF(AND(BK2>=6,BK2<9,Y2=1,BD2>=2),1,0)

   As indicated earlier, if you want to reduce the amount of typing you do, you can click on cells (e.g., BK2) rather than typing them. Be sure to substitute in the appropriate cell addresses for your database. This syntax tells Excel to look for all infants 6-8 months old who are breastfed and have eaten at least 2 solid or semi-solid foods.

3. In the second column (BF9-23), type:

   =IF(AND(BK2>=9,BK2<24,Y2=1,BD2>=3),1,0)

   This syntax tells Excel to look for all infants 9-23 months old who are breastfed and have eaten at least 3 solid or semi-solid foods.

4. Now create a new variable between BF9-23 and BF6-23. In our spreadsheet, it is column CK. You will use this variable to add up the number of times infant formula, milk and sour milk or yoghurt plus solid, semi-solid or soft foods were given to children. Your syntax will look like this:

   =SUM(AD2+AF2+AK2+BD2)

   Copy this formula to the end of the column.

5. In the next column (NBF6-23), type:

   =IF(AND(BK2>=6,BK2<24,Y2=0,CK2>=4),1,0)

6. Now create a new column to the right of NBF6-23. Call it **MealFreq**. This variable represents whether infants met the minimum meal frequency, not how many meals children got. 1=yes, the child received the minimum meal frequency. 0=no, the child did not receive the minimum meal frequency. **MealFreq** can be colored tan.

7. In the first row, type:

   =SUM(CI2+CJ2+CL2)

8. Using step 7 as an example, put in the correct cell addresses for your database.

9. Copy the formula to the rest of the column.

10. Use Excel to sum your numerator. Use **Autofilter** in Excel to help you determine your denominator. Since all children were either breastfed yesterday or not, your denominator is all children 6-23 months of age.

When we used Excel to determine minimum meal frequency, we got these results:

There were 11 infants 6 to 23 months old. For one of them, we have no information on whether the child was breastfed so we excluded the child from our analyses leaving 10 children total.

- 2 of the 4 breastfed infants who were 6-8 months old met the requirement for minimum number of meals
- 2 of the 4 breastfed infants who were 9-23 months old met the requirement for minimum number of meals
- 1 of the 2 non-breastfed infants 6-23 months old met the requirement for minimum number of meals

Here is our calculation:

$$\frac{\textbf{5 children 6-23 months old received the minimum number of meals in the previous day}}{\textbf{10 children 6-23 months of age}}$$

This equals 0.5. Multiply this number by 100: 50.0%.

Answer: 50% of infants 6 to 23 months of age met the standards for minimum meal frequency.

Write down the percent of infants who demonstrated minimum meal frequency based on your own data. You will need to refer back to this number in Chapter 7.

6.4.8 Minimum acceptable diet

Definition: Proportion of children 6-23 months of age who receive a minimum acceptable diet (apart from breastmilk)

$$\frac{\textbf{\textit{Breastfed} children 6-23 months of age who had at least the minimum dietary diversity and the minimum meal frequency during the previous day}}{\textbf{Breastfed children 6-23 months old}}$$

And

$$\frac{\textbf{\textit{Non-breastfed} children 6-23 months of age who had at least the 2 milk feedings and had at least the minimum dietary diversity and the minimum meal frequency during the previous day}}{\textbf{Non-breastfed children 6-23 months old}}$$

Calculation:

For breastfed children, you have already calculated indicators for each of the practices incorporated into minimum acceptable diet (minimum dietary diversity and minimum meal frequency). Each child with a positive score for *both* of the composite indicators will score positively on this summary indicator.

However, for non-breastfed children, the dietary diversity component of this indicator is different from the individual practice indicator calculated earlier (which did not distinguish between breastfed and non-breastfed children). You will need to calculate dietary diversity again using a 6 food group score (instead of the 7 food group score above) in order to exclude the dairy group from the calculation. You will then need to count the number of milk feeds separately for non-breastfed children to determine if the child received at least 2 milk feeds, the minimum number required for this indicator.

The instructions below show how to calculate the 6 food group score for non-breastfed children.

Calculation of 6 food group score:

The 6 food groups used for calculation of this indicator are:

- 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

Construct the 6 food group score as follows:

- Begin with a score of 0.
- For each of the 6 food groups, add a point if any food in the group was consumed.

Food group 1 (Grains)	Add 1 point if: Q10aa=1 OR Q10bb=1
Food group 2 (Legumes)	Add 1 point if: Q10dd=1
Food group 3 (Flesh foods)	Add 1 point if: Q10hh=1 OR Q10ii=1 OR Q10jj=1 OR Q10kk=1
Food group 4 (Eggs)	Add 1 point if: Q10ll=1
Food group 5 (Vitamin-A rich fruits and vegetables)	Add 1 point if: Q10cc=1 OR Q10ee=1 OR Q10ff=1 OR Q10oo=1
Food group 6 (Other fruits and vegetables)	Add 1 point if: Q10gg=1

Calculation of Minimum acceptable diet indicator:

$$\frac{\begin{array}{c}\textbf{(Age in months 6<9) AND (Q8=1) AND (Q11}\geq\textbf{ 2) AND (7 food group score}\geq\textbf{ 4)}\\ \textbf{OR}\\ \textbf{(Age in months 9<24) AND (Q8=1) AND (Q11}\geq\textbf{ 3) AND (7 food group score}\geq\textbf{ 4)}\\ \textbf{OR}\\ \textbf{(Age in months 6<24) AND (Q8=0) AND ((Q9d +Q 9e +Q9i+Q11)}\geq\textbf{ 4) AND}\\ \textbf{((Q9d +Q 9e +Q9i)}\geq\textbf{ 2) AND (6 food group score}\geq\textbf{ 4)}\end{array}}{\textbf{Age in months 6<24}} \textbf{ x 100}$$

Notes:

Q8 asks about breastfeeding yesterday.

Calculation of food group scores are based on Q9 and Q10, which ask about liquids and foods the child consumed the previous day. The 7 food group score (see minimum dietary diversity indicator) is used for breastfed children. The 6 food group score (see above), which excludes dairy products, is used for non-breastfed children when calculating this indicator. This is because milk feeds are considered a separate and required element for non-breastfed children in this multi-dimensional indicator. Exclusion of milk feeds from the diversity score here avoids double-counting of this food group and allows use of this indicator in comparisons between populations with different rates of continued breastfeeding.

Q9d, Q9e and Q9i ask about the number of times the child received infant formula, milk, or yogurt yesterday.

Q11 asks about the frequency of feeding of solid, semi-solid and soft foods yesterday.

Disaggregation: It is recommended that the indicator be further disaggregated and reported for the following age groups: 6-11 months, 12-17 months, and 18-23 months, if sample size permits.

1. Insert a column to the right of **MealFreq**. Call it **AcceptDietBF6-23**

2. Choose infants who are 6-23 months old: **=IF(AND(BK2>=6,BK2<24,**

3. Determine which infants meet both the requirement for minimum dietary diversity (CH2; in tan) and the requirement for minimum meal frequency (CM2; in tan). For the Practice data file, the syntax is as follows:

   =IF(AND(BK2>=6,BK2<24,CH2=1,CM2=1),1,0)

4. Using step 3 as an example, put in the correct cell addresses for your database.

5. Copy the formula to the rest of the column.

6. Now you need to calculate acceptable diet for non-breastfeed children. To do so, insert a column to the right of **AcceptDietBF6-23**. Call it **liqtimesnosolids**. Here you are adding up the number of times liquids (but not solids) were given to the child. Your syntax should be as follows:

   =AD2+AF2+AK2

7. Now insert a column to the right of **liqtimesnosolids**. Call it **number6groups**. Here you will add up all the food groups you created before except the dairy food group. Your syntax will look like this (unless your cell addresses are different):

   =SUM(BZ2+CA2+CC2+CD2+CE2+CF2)

8. Now you are ready to calculate the number of non-breastfed children who have a minimal acceptable diet. Insert a column to the right of **number6groups**. Call it **AcceptDietNBF6-23**. Per the instructions above about how to calculate this indicator for non-breastfed children, your syntax will look like this:

   =IF(AND(BK2>=6,BK2<24,Y2=0,CK2>=4,CO2>=2,CP2>=4),1,0)

9. To determine how many children – breastfed or not – have an acceptable diet, insert a column to the right of **AcceptDietNBF6-23**. Call it **AcceptDiet**. Here you simply add up all the breastfed children who have an acceptable diet and all the non-breastfed children who have an acceptable diet. Your syntax will look like this:

   =CN2+CQ2

10. Use Excel to sum your numerator. Use **Autofilter** in Excel to help you determine your denominator. Since all children were either breastfed yesterday or not, your denominator is all children 6-23 months of age.

When we used Excel to determine minimum acceptable diet, we got these results:

There were 11 infants 6 to 23 months old. However, one infant was excluded so there were only 10 left. None of the breastfed children 6-23 months old and none of the non-breastfed children 6-23 months old met the requirements for this indicator. Here is our calculation:

$$\frac{\textbf{0 child 6-23 months old who met the requirements for minimal acceptable diet}}{\textbf{10 children 6-23 months of age}}$$

Answer: 0% infants 6 to 23 months of age met the standards for minimum acceptable diet.

Write down the percent of infants who demonstrated minimum acceptable diet based on your own data.

You've done well! **Take a break and proceed to Section 6.4.9 another day!**

6.4.9 Consumption of iron-rich or iron-fortified foods

Definition: Proportion of children 6-23 months old who receive an iron-rich food or iron-fortified food that is specially designed for infants and young children, or that is fortified in the home.

$$\frac{\text{Children 6-23 months old who received an iron-rich food or a food that was specially designed for infants and young children and was fortified with iron, or a food that was fortified in the home with a product that included iron during the previous day}}{\text{Children 6-23 months}}$$

Calculation:

$$\frac{\text{(Age in months 6<24) AND (Q10hh=1 OR Q10ii=1 OR Q10jj=1 OR Q12=1 OR Q13=1 OR Q14=1 OR Q15=1)}}{\text{(Age in months 6<24)}} \times 100$$

Notes:

Suitable iron-rich or iron-fortified foods include flesh foods, commercially fortified foods specially designed for infants and young children which contain iron, or foods fortified in the home with a micronutrient powder containing iron or a lipid-based nutrient supplement containing iron.

Disaggregation:

It is recommended that the indicator be further disaggregated and reported for the following age groups: 6-11 months, 12-17 months and 18-23 months, if sample size permits.

1. Insert a column at the end of your Excel file (to the right of **AcceptDiet**). Call it **IronCount**. Pick a color for **IronCount**. This variable will calculate how many iron-rich and iron-fortified foods the child consumed.

2. Now we want to see if the child got any iron-rich or iron-fortified foods. This information is located in columns AT (organmeats), AU (meat), AV (fish) and BE (fortified), BF (sprinkles), BG (lbns) and BH (fortformulas). For the Practice data file, we would type: **=SUM(AT2,AU2,AV2,BE2,BF2,BG2,BH2**). Remember, the flesh foods are in red. The iron-fortified foods can be any color you want. Grubs are not included in iron-rich foods.

3. Insert a column at the end of your Excel file (to the right of **IronCount**). Call it **IronRich**. This variable will calculate whether each child is of the appropriate age (6-23 months old) and getting at least 1 iron-rich or iron-fortified foods. In our database, **IronCount** is located in column CS and **IronRich** is in column CT.

4. Choose infants who are 6-23 months old: **=IF(AND(BK2>=6,BK2<24,**

5. Complete the syntax as follows:

   =IF(AND(BK2>=6,BK2<24,CS2>=1),1,0)

6. Using the steps above as an example, put in the correct cell addresses for your database.

7. Copy the formula to the rest of the column.

8. Use Excel to sum your numerator. Use **Autofilter** in Excel to help you determine your denominator.

When we used Excel to calculate consumption of iron-rich or iron-fortified foods, we got these results:

There were 11 infants 6 to 23 months old. Three of them consumed either an iron-rich or iron-fortified food in the previous day. Here is our calculation:

$$\frac{\textbf{3 children 6-23 months old who consumed iron-rich or iron-fortified foods during the previous day}}{\textbf{11 children 6-23 months of age}}$$

This equals 0.272. Multiply this number by 100: 27.2%.

Answer: 27% of infants 6 to 23 months of age consumed iron-rich or iron-fortified foods in the previous 24 hours.

Write down the percent of infants who consumed iron-rich or iron-fortified foods based on your own data. You will need to refer back to this number in Chapter 7.

6.4.10 Bottle feeding

Definition: Proportion of children 0-23 months old who were fed with a bottle during the previous day

$$\frac{\textbf{Children 0-23 months old who were fed with a bottle during the previous day}}{\textbf{Children 0-23 months}}$$

Calculation:

$$\frac{\textbf{(Age in months 6<24) AND (Q16=1)}}{\textbf{Age in months 6<24)}} \textbf{ x 100}$$

Notes:

Q16 asks about bottle feeding.

We include children less than 23 months who received any food or drink from a bottle with a nipple/teat during the previous day (including breast milk), regardless of whether or not the infant was breastfed.

Disaggregation: It is recommended that the indicator be further disaggregated and reported for the following age groups: 0-5 months, 6-11 months, and 12-23 months, if sample size permits.

Information on bottle feeding is helpful because bottle feeding can discourage optimal breastfeeding practices and because of the link between bottle feeding, sickness, and death. Bottles with a nipple are very difficult to clean and carry a high risk of contamination from bacteria and other substances.

1. Insert a column to the right of **IronRich**. Call it **BottleFeed**. It should be tan.

2. Choose infants who are 0-23 months old: **=IF(AND(BK2>=0,BK2<24,**

3. Complete the syntax as follows:

   =IF(AND(BK2>=0,BK2<24,BI2=1),1,0)

4. Using step 3 as an example, put in the correct cell addresses for your database.

5. Copy the formula to the rest of the column.

6. Use Excel to sum your numerator. Use **Autofilter** in Excel to help you determine your denominator.

When we used Excel to calculate bottle feeding, we got these results:

There were 20 infants 0 to 23 months old. Seven of them received a bottle the previous day. Here is our calculation:

$$\frac{\textbf{7 children 0-23 months old who were fed with a bottle during the previous day}}{\textbf{20 children 0-23 months of age}}$$

This equals 0.350. Multiply this number by 100: 35.0%.

Answer: 35% of mothers of infants 0 to 23 months of age gave their child a bottle in the previous 24 hours.

Note: Our calculations assume that the children who did not have date of birth are included in the calculation. However, as noted previously, children without **DOB** should be excluded from the database prior to analysis.

Write down the percent of infants who were fed using bottles based on your own data. You will need to refer back to this number in Chapter 7.

An alternative approach to determining which children should be included in your denominator: Rather than having to use **autofilter** for age then having to hand count all of the children in a given age range (for example, 0-5 months of age), you can have Excel count up the numbers for you. In order to calculate denominators for each of the 10 indicators, you need to know how many children are in each of the following age categories:

- 0-23 months old
- 0-5 months old
- 6-9 months old
- 6-8 months old
- 12-15 months old, and
- 6-23 months old

It is easy to calculate the number of children in your database who are 0-23 months old. If you have limited your database to children of this age, it should be the total number of children. This information is needed for the indicator on timely initiation of breastfeeding.

We will explain how to calculate the number of children 0-5 months of age. It is your denominator for exclusive breastfeeding. Once you know how to do this, it will be easy to calculate the other age ranges. We will provide the syntax for those calculations as well.

Instructions for determining the number of children 0-5 months old:

1. Create a new variable at the end of your Excel file. Call it **0-5m**

2. Let us assume your variable for calculated age in months is in cell BK

3. In cell BK2, type: **=IF(AND(BK2<6),1,0)**. This tells Excel to look at the calculated age of the child (in months) and determine whether the child is 0-5 months old. If the child is 0-5 months old, Excel assigns the child a '1.' If not, the child is assigned a value of '0.'

4. Now go all the way to the bottom of this column (after your last observation) and use the sum function (Σ) to add up all of the individuals who are 0-5 months old. With your cursor in the first cell after your last observation, click on the Σ. Now highlight all of the observations in the column **0-5m**. Your first observation should be in row 2.

5. Sometimes Excel suggests a range that it thinks you should add up. This range may very well be different from the one you want. Simply highlight the first cell you want (usually in row 2 under **0-5m**) and drag the cursor all the way to the end of the column.

6. Click enter.

7. Excel will add up the number of children 0-5 months old. For example, in one country, there were 896 children 0-23 months of age. When we added up the total number of children 0-5 months old, we found 243 of them. In this case, 243 was our denominator for exclusive breastfeeding.

Now do this for each of the other age ranges above. The other age ranges will be at the very end of your Excel file. Assuming your calculated age in months is in column BK, your syntax will look like this:

- 6-9 months old: **=IF(AND(BK2>=6,BN2<10),1,0)**
- 6-8 months old: **=IF(AND(BK2>=6,BN2<9),1,0)**
- 12-15 months old: **=IF(AND(BK2>=12,BN2<16),1,0)**
- 6-23 months old: **=IF(AND(BK2>=6,BN2<24),1,0)**

You might call these variables **6-9m, 6-8m, 12-15m and 6-23m**. It may be easier to do all of this if you freeze your first row so you can see the names of the variables.

What about stratified analyses?

In most cases, you'll want to know if percents for each of the 10 indicators are different for different sites, for example, intervention and control sites. There are a number of ways of stratifying in Excel, including sorting (the method we'll use) and pivot tables. If you are stratifying once or twice (e.g., separate calculations for intervention and control sites), sorting is the easiest way to go. If you have a lot of strata and/or you are very comfortable with Excel, you might try pivot tables. Excel will tell you how to do this. Here we give directions for using sorting for stratified analyses. For the purposes of this example, we'll assume you want to stratify by intervention and control communities.

What to do:

1. Make a copy of your dataset. Save it under a new name. This is important. When sorting, it's easy to change a database in ways that you don't want to! Always keep a couple of backups of your original data.

2. Go to the worksheet that has all of the children's records in it. Move your cursor to the top of the column that contains the information about whether the child lives in an intervention or control community. Make sure the entire column is highlighted.

3. Click on Data, Sort. Excel now wants to know whether you want to sort the current column (no) or all of the variables (yes). You want to sort all of the variables to make sure that all of the information you have on one child stays with that child. To do this, make sure the button "Expand the selection" is highlighted. Now click on Sort.

4. Another dialogue box now appears. Make sure the variable you are using to sort appears in the menu box then click on OK.

5. You have now sorted by site (intervention versus control).

6. Make TWO copies of the sorted worksheet: one for intervention sites and one for control sites. The best way to do this is to go to the bottom of the worksheet then click on the tab at the bottom where you've named your worksheet. If you haven't yet named your worksheet, do so now. Use a

name that indicates that the current worksheet contains ALL children. Once you've clicked on the tab, right click and click on "Move or copy." In the "Move or copy" dialogue box, click on "Copy" then click on "(Move to end)." You may have to scroll down to see the "(Move to end)."

7. Once you've done this twice, you have 3 worksheets with identical data. Label one of the worksheets "Intervention." Label the other one "Control." So you should have 3 worksheets with identical data: "All," "Intervention" and "Control" (preferably in that order). You can label each worksheet by double clicking on the tab at the bottom of the worksheet and giving it a new name.

8. What you need to do now is to delete the records of all children in the control communities from the "Intervention" worksheet and all records of children from the intervention communities from the "Control" worksheet. This is relatively straightforward since you've already sorted each worksheet by intervention versus control.

9. In the intervention worksheet, scroll up to the record of the child who is listed as the first individual in your CONTROL group. Place your cursor in the column furthest to the left for this individual. The whole row should be highlighted. Now, continue to keep the cursor clicked and scroll down to the last individual who is in the control site. Let's suppose that children living in intervention areas are coded as 1 for the variable "Intervention" (or whatever name you have given this variable) and 2 if they don't live in intervention areas (i.e, they are "controls"). Assuming you are in the worksheet for intervention sites, during this step, you will delete all of the children who are coded 2, while keeping all of the children who are coded 1.

10. Now do the same in the "Control" worksheet. Here you will keep all of the children who are coded as 2 and delete all children coded as 1.

11. Save your file.

12. You will now calculate your numerator, denominator and percent for each of the 10 indicators in the exact same way you did for ALL children. You can copy all the formulas at the bottom of the worksheet with all children to your "Intervention" and "Control" worksheets. Remember that you have fewer children in your intervention and control worksheets than you do in the worksheet containing all children. So when you copy formulas, check to make sure Excel is adding up the correct number of observations. Let's suppose you have 600 children overall. 350 of them live in intervention sites and 250 live in control sites. In your intervention worksheet, you are adding up indicator values for 350 children, not 600 children. In control sites, you add up values for 250 children. Please make sure that your formulas for each indicator reflect this change.

13. If you don't want to copy these formulas, you can simply recreate them at the bottom of each worksheet.

14. The very bottom line of each worksheet will give you percents for each of the 10 IYCF indicators for your overall sample and for children in intervention and control sites, calculated separately.

15. If you'd like, you can continue to stratify results using the same approach. For example, you may want to stratify by geographic region within intervention sites. Repeat the same process using only children living in intervention sites as your master worksheet. Be aware that sorting is not the best way to carry out stratified analyses if you want to look at lots of different strata. In this case, it is advisable to carry out these analyses using SPSS or a similar package.

You may want to proceed to Chapter 7 today.

CHAPTER 7: PRODUCING AND INTERPRETING TABLES AND BAR GRAPHS

Goal: Report and interpret indicators of infant and young child feeding

By the end of this chapter, you will be able to:

- Report results of your analyses in a clear fashion including the use of standard language for each indicator
- Produce a 'report card' on infant and young child feeding
- Interpret findings
- Create figures and tables to improve the presentation of results

Tool	Purpose of tool	Name of file
Excel file with report card	Use this file to help you report findings for each of the 10 indicators and create additional figures on IYCF	Report card template.xls
Calculator for chi-square tests	Use this tool to help you calculate chi-square and p-values	Chi-square tests.xls

7.1 Using the report card to present results

In many countries, students receive report cards telling how well they have done in their studies. We created a report card to help you share information on IYCF practices at different stages of program implementation (for example, at the beginning, the middle, and the end of your project). This information helps you know whether or not your program is on track. You can use this information to decide what changes might be needed to improve program performance. You can also share this information with community members, colleagues in other institutions and donors. In Section 7.5, we tell you how to calculate chi-square tests. These are among the most common tests which compare 2 or more groups to see if they are the same or different. You may want to use chi-square tests with the graphics in the report card mentioned below.

The file 'Report card template' helps you share results from surveys on IYCF. Let us take a look at it.

1. Open the file **Report card template.xls**

2. Notice the tabs at the bottom left of the screen. There are 10 separate sheets, one named for each of the 10 IYCF indicators. There are also other tabs for graphics and figures you may want to produce related to IYCF. You probably cannot see all of the names at once. Use the left and right scroll buttons at the bottom left corner of your screen to see the other tabs.

3. Let us look at the first sheet. It's called **Timely Initiation**. This sheet and the 9 other sheets have the following information: **indicator name, definition, program target, results for different time periods,** and **a figure** that helps you present your results.

4. The **definition** gives you the exact meaning of the indicator. **For Timely initiation of breastfeeding (children 0-23 months)**, the definition is: Percent of children born in the last 23 months who were put to the breast within one hour of birth. We will use this information below (Section 7.2) to help interpret our results.

5. The **program target** is the goal you wanted to achieve at the beginning of your project. You probably picked a program target when you made a proposal to a donor. In our example, we hoped that by the end of the project, 55% of mothers would initiate breastfeeding in the first hour of life. To see what this looks like, type **55** into cell E15 (worksheet is Timely Initiation). Your goal should be the same number you used when you calculated your sample size (step 3 in Section 3.2).

6. The **results** Section is broken into three periods, beginning with the oldest (least recent) and ending with the newest (most recent).

   You may have results from more or less than 3 time periods.

   If you have results from just 1 or 2 time periods: fill in the most recent data in the right most box (Period 3, most recent) and older findings (if you have them) in the Period 2 box.

   If you have results from more than 3 time periods: fill in the most recent 3 time periods or, if you prefer, the 3 most meaningful time periods, for example, baseline, mid-term and final evaluation.

7. The **figure** is called a 'bar chart.' It tells you visually how things have changed. In this case, it tells you how the prevalence of timely initiation has gone up or down over time.

8. You will notice that there is a line across each figure. It tells you what your target is. It helps you compare your results to the target you set at the beginning of your program.

Try it: Create a report card

1. Open up the file **Report card template.xls**. The chart is blank. Begin by adding your program target. Since we are using a made-up example now, if you have not already done so, you can put the number **55** in cell E15 (Program Target). For your own data, your target will probably be different. You can fill in your number later.

2. Let us assume we collected data at baseline (period 1), mid-term (period 2) and follow-up (period 3).

3. Under period 1 (cell F15), type the number **37**.

4. Under period 2 (cell G15), type the number **42**.

5. Under period 3 (cell H15), type the number **57**.

6. Notice that the bar chart figure changes as you put the numbers into the white boxes located in cells F15, G15 and H15.

7. When you enter the data for Timely Initiation for your program, your bar chart will reflect your own numbers.

7.2 Interpreting results

How should you interpret your findings?

First, remember the definition. For **Timely Initiation of Breastfeeding (children 0-23 months)**, it is: percent of children born in the last 23 months who were put to the breast within one hour of birth. Look back at the box in Section 6.3.1. You will notice that our indicator includes two parts:

$$\frac{\text{Children born in the last 23 months who were put to the breast within one hour of birth}}{\text{Children born in the last 23 months}}$$

In other words, for Timely Initiation of Breastfeeding (children 0-23 months), we are only considering children *less than* 24 months of age (from birth to 23 months old). What percent of all children 0 to 23 months of age began breastfeeding in the first hour of life? It is important to think about who is and is not included in this indicator. Only those infants who provided information on their age and information about when they started breastfeeding are eligible.

Try it: Interpreting results

Look at the information we entered into the **Report Card template.xls,** above. What percent of children initiated breastfeeding at baseline (time period 1)? Answer: 37%. You can find the answer in cell F15. This means that at program baseline, 37% of all children 0-23 months of age began breastfeeding in the first hour of life.

Did you meet your Program Target?

What percent of children initiated breastfeeding in a timely fashion at mid-term? At the final evaluation? Answer: 42% at mid-term; 57% at final evaluation. Did you meet your program target half way through your program? No. You hoped that 55% of all women would initiate breastfeeding in the first hour of life. In fact, only 42% did so. Did you meet your program target by the end of the project? Yes. 57% of all mothers of infants 0 to 23 months of age began breastfeeding in the first hour of life. Your goal was 55% so you did a little better than expected.

Compare your results to other studies:

First, check your results against the Demographic and Health Survey (DHS) for your country. (These statistics are available at www.measuredhs.com. Once there, click on StatCompiler and select the statistics you want.)

You may need to check your results against other studies. The DHS is not available for all countries. Sometimes, data for the DHS were collected a long time ago. If your country does not have a recent Demographic and Health Survey, you can check your results against another national study (for example, a MICS study; results are at www.childinfo.org) or a local study. If there are several good studies that have been done in your area or country, look at all of them.

Are your numbers high, low or what you would expect?

Let us say that the DHS for our country indicates that only one in five infants (20%) began breastfeeding in the first hour of life, a percent much lower than the ones we got (37% at baseline and 57% at final evaluation). Does this mean our findings are wrong? Not necessarily. It may be that our population is more urban or rural than the DHS population. Also, if you are working with a refugee population, it may be different from the population of the country where they now live. Also, it may be difficult to conduct a survey on a population that is 'on the move.' Alternatively, it may be that NGOs were able to improve breastfeeding practices.

Now let us look at the 9 remaining indicators.

1. In the Practice data file, find your results for Exclusive Breastfeeding, as we figured them in Chapter 6.
2. Enter your results in the Report Card, on the sheet named Exclusive Breastfeeding.
3. As you enter the results, Excel will complete the bar chart for you.
4. Use the remaining sheets to enter results for all of your other indicators. As you enter results, Excel will continue to complete the bar charts for you.

Here is how you interpret findings for the 9 remaining indicators:

Exclusive breastfeeding

The proportion of all children 0-5 months of age who receive only breastmilk in the previous 24 hours. These children were given nothing else – no other foods or liquids (not even water) except for breastmilk (ORS and vitamins are okay). By asking only about yesterday, you may overestimate the proportion of children exclusively breastfed because some children may have received liquids and foods on other days, but not yesterday. This risk of estimation is known and accepted.

It is helpful to determine rates of exclusive breastfeeding separately for infants 0-1 months of age, 2-3 months of age and 4-5 months of age. Often, rates of exclusive breastfeeding are high in the first 2 months of life and then drop off thereafter.

Timely complementary feeding

The proportion of all infants 6-9 months of age who receive breastmilk and a solid or semi-solid food in the previous 24 hours. Only those children who received breastmilk and a solid or semi-solid food (or both) are included in the numerator of this indicator. Remember that **timely complementary feeding** is an indicator that comes from the 1991 WHO guidance and is not found in the new (2009) guidance. Even so, we feel it is important to determine which children receive breastmilk and a solid or semi-solid food.

Introduction of solid, semi-solid or soft foods

The proportion of all infants 6-8 months of age who receive solid, semi-solid or soft foods.

Continued breastfeeding at 1 year

The proportion of all children 12-15 months old who are fed breastmilk. Children who in the previous day received any breastmilk whatsoever are included in the numerator.

Minimum dietary diversity

The proportion of all children 6-23 months of age who in the previous 24 hours receive foods from 4 or more food groups. There are 7 food groups. Children must have received food from at least 4 of these groups to have achieved minimum dietary diversity. You may want to look at the following age groups separately: 6-11 months, 12-17 months and 18-23 months. By asking only about yesterday, you may underestimate dietary because the child may have received foods from multiple groups on other days but not yesterday.

Minimum meal frequency

The proportion of breastfed and non-breastfed children 6-23 months who receive solid, semi-solid or soft foods (but also including milk feeds for non-breastfed children) the minimum number of times or more. There are two separate calculations for this indicator, one for breastfeed children and one for non-breastfed children.

Minimum acceptable diet

The proportion of all children 6-23 months of age who receive a minimum acceptable diet (apart from breastmilk). There are two separate calculations for this indicator, one for breastfeed children and one for non-breastfed children.

Consumption of iron-rich or iron-fortified foods

The proportion of all children 6-23 months old who receive an iron-rich food or iron-fortified food that is specially designed for infants and young children, or that is fortified in the home. Also, look at the following age groups separately: 6-11 months, 12-17 months and 18-23 months. Be aware that experts are still working on how to construct this variable. Even so, it is an important indicator of nutrient adequacy.

Bottle feeding

This means the proportion of all children 0-23 months old who were fed with a bottle during the previous 24 hours. We include children who received any food or drink from a bottle with a nipple/teat during the previous day including those who were given breastmilk from a bottle. You may want to look at the following age groups separately: 6-11 months and 12-23 months. By asking only about yesterday, you may underestimate bottle feeding because the child may have received a bottle on other days but not yesterday.

7.3 Creating and interpreting figures in Excel

Often, it is wise to present your findings using graphics. Because graphics display visual information that your audience is likely to remember, graphics may be particularly helpful in conveying your message about IYCF.

We developed additional figures you may want to use. The file **Report card template.xls** has figures but no data. As you add data, the report card will automatically be generated using the numbers you provide. Feel free to modify how the graphics look. You may want to develop other graphics on your own.

Try it: Create a graph

1. Use Excel to calculate the number of food groups consumed by children in your sample. Open the file **Report card template.xls** and go to the tab labeled **Number of food groups**.

2. Fill in cells N5 through N9 with the percent of infants in your sample who received 0, 1, 2, 3, and 4+ food groups in the previous 24 hours. As you fill in the numbers, the shape of the pie chart changes to match your data.

7.4 Creating and interpreting tables in Microsoft Word

If you want to present your findings using a table, it is often easiest to construct it using Microsoft Word. In general, it is best to use a table when you want to present a LOT of information and/or detailed information and a graph when you want to convey major findings that donors and others can understand best if they see a visual representation of results.

In this section, we describe how to construct a table. If you want to construct one or more tables for your indicators, read this section. Then you can practice creating tables on your own.

Below, we use made up data for our table. Here is what the table looks like:

Table 7.4.1. Percent of infants aged 0-5 months exclusively breastfed, by site, Country X

Exclusively Breastfed	Camp 1 (n=52)	Camp 2 (n=48)	Chi-square (p-value)
Yes	19.2	22.9	0.651
No	80.8	77.1	
TOTAL	100.0	100.0	

Note: The chi-square test compares the feeding practices of infants 0-5 months old in camp 1 and camp 2.

Note: If you did not design the survey in a way that allows you to compare different samples (i.e., if you did not calculate sample size for population 1 and separately for population 2), you should not perform chi-square tests since this was not the original purpose of your evaluation. While you always have the option of comparing the two (or more) groups, it is likely that statistically, your calculation will be underpowered.

What to look for in the table:

1. A table should be self-explanatory so all information about the table should be present in the title, the body of the table and in the footnotes. That said, the person reading the table (a colleague or donor) should be able to quickly grasp the main message of the table. Do not give too many numbers, lines, or other details.

2. The title clearly describes what is in the table (percent of infants exclusively breastfed, by site), your source of information (in this case, a survey on infant and young child feeding) and the area where you are getting information (country X).

3. There are only a few lines in the table: darker lines for the top and bottom of the table and lighter lines to separate the column headings from the results. You can refer to a manual on Word to find out more about formatting.

Tip: Highlight the boxes you want formatted, click on **Format, Borders and shading, Width**. Then choose the width you want. You can find width at the bottom (middle) of the dialogue box (borders and shading). We chose a width of 1½ for the top and bottom and ½ for the bottom of the header column. Also, you can copy and paste the table above into your report and change the title and the percents.

1. Each header is centered and describes what is in the column below it. If you need to explain 'units,' put the units in parentheses below the column header.

2. The number of individuals in each category appears below the column headings. For example, the number of infants on whom we collected data in site 1 is 52.

3. We will talk about chi-squares and p-values in Section 7.5 below. It is best to present the p-value rather than the actual chi-square statistic.

4. We have left space between each row to make it easier to read the numbers.

5. We have left a note (in smaller font) at the bottom of the table to make clear what groups we are comparing.

Here is how you interpret this table: About 1 in 5 infants in both sites breastfeed exclusively. There are no statistical differences between sites in the prevalence of exclusive breastfeeding as shown by the p-value (0.651). A p-value of <0.05 means that the groups are different.

7.5 Using statistics to compare groups

We use statistics – in this case, a chi-square test – to see if there are differences between groups of respondents (for example, infants in one site compared to infants in another site). See below for a more detailed description of chi-square tests and p-values.

When should I perform a statistical test?

Sometimes statistical tests are more helpful than others. Use a statistical test when:

- You want to communicate how big the difference is between groups (the two or more groups you want to compare, for example, infants in each of 3 refugee camps).
- You want to communicate how certain you are of your results.

Statistical tests may be less helpful when:

- You have a very large sample size. In this case the p-value for your test may suggest a very big difference between groups. Large sample sizes make even small differences look big.

Regardless of whether or not you use a statistical test, you should scan your findings to see if the differences between groups are large or small.

Note: If you visually scan your results, there is no hard and fast rule to help you decide whether differences between groups are large. In general, differences of only a few percentage points are not meaningful. When groups vary by only a few percentage points, you can often conclude that groups are similar.

What is a chi-square test and when do I use it?

The chi-square is a test statistic that is the sum of the squares of observed values minus expected values divided by the expected values. When you calculate a chi-square, it gives you a number. You can then compare this number against a chart with the chi-square distribution. This will tell you whether or not the chi-square test itself is significant. In short, the chi-square test is a statistical way of comparing groups. Chi-square tests help you compare categorical data, or data that can assume only a few specific values. For example, you might have *categories* for the age of the child (0-1, 2-3 and 4-5 months old). When data are grouped, we call these categorical data.

Another type of data is *continuous*. For example, age of infant in days can assume a lot of values (0 to 730 in our example). Usually, categorical data require the chi-square test; continuous data require the t-test.

Because the use of t-tests is relatively rare for IYCF data, we suggest you look at other sources for interpretation of those results. The Sullivan book on statistics that is referenced at the beginning of this Guide will help you interpret results from t-tests.

What is a p-value?

Whether you are using a chi-square test or a t-test, you can determine the significance of a comparison by the p-value. We often consider a p-value less than .05 significant. Using the example in Table 7.4.1 above, we see that our p-value is 0.651. An easy way to make sense of this p-value is as follows: If you were to repeat your study 1000 times, 349 times out of 1000 (1000-651), you would get the same or a similar result. Usually, any p-value above .05 (ours is 0.651) means that we are not certain there really is a difference between groups.

Suppose you have a p-value of .020. This means that if you were to repeat your study 1000 times, 980 times out of 1000 you would get the same or similar result. The smaller the p-value, the more certain you are of your results. In this case, we would conclude there is a difference between the two sites because our p-value is less than .05.

Using a calculator to determine chi-square and p-values

One of the tools that comes with this Guide is a calculator in Excel for determining chi-square and p-values (see the file Chi-square tests.xls). You can use the calculator to find out if there are differences between groups. Your 'groups' could be geographic areas (for example, refugee camp 1 compared to camp 2), time periods (baseline, mid-term and final evaluations), or ages of children. You may want to list the chi-square test (i.e., the actual statistic) and p-value in your report card, a report to a donor and so on.

1. Open the file **Chi-square tests.xls**
2. Decide what your groups are. Let us assume you want to compare 2 sites.
3. Click on the tab at the bottom of the screen to view the worksheet **Chi-square 2 groups**. In this example, you use the worksheet for 2 groups because you are comparing only 2 groups. If you were comparing 3 or 4 groups, you would use the corresponding worksheets.
4. Enter data for your sites in the white cells. Use the instructions given at the top of the worksheet for using the chi-square calculator.

Important: You are entering the ***NUMBER*** of individuals in each group, ***NOT*** the percent. If you use percents, your calculations will be wrong.

Here are the groups for which you are entering numbers:

1. Site 1 practicing the behavior
2. Site 1 not practicing the behavior
3. Site 2 practicing the behavior
4. Site 2 not practicing the behavior

Now that you have entered the data, you need to look at the p-value. Look at the 'chi-square 2 groups' worksheet. You can find the p-value in cell G30.

Making comparisons between sites

If you want to compare sites, you need to tell Excel the number of individuals in each site who do and do not practice a given behavior. For example, you need to let Excel know how many children in site 1 do and do not have a minimum dietary diversity, how many children in site 2 do and do not have a minimum dietary diversity, and so on.

How do you find this out? Use Autofilter to select for each of your sites. For example,

1. Select site 1 by clicking on the down arrow for the variable **site** (column F in our Practice data file). Then choose **Data, Filter, Autofilter** to make sure Autofilter is turned on, and choose **1**, for camp 1.

2. In the column where you have calculated minimum dietary diversity, count up (by hand) the number of individuals in camp 1 who had a minimum dietary diversity. Record the number. Do the same for children in camp 1 who did not meet the requirements for minimum dietary diversity.

3. Now Autofilter for camp 2, repeat the process for camp 2 and continue on to camp 3 and so on until you get all of the numbers for each camp.

Note: Carrying out the 3 steps above will be difficult if you have a large dataset. If you have a large dataset, you have 2 options: 1) You can use the Sum function in Excel (Σ) to add up all of the individuals meeting the criteria for this indicator or 2) You can export your data to SPSS or another statistical software and 'run a frequency' for your indicator.

How would you write this up in a report? We will use a fictitious example to demonstrate. Let us suppose our chi-square for minimum dietary diversity was 21.0 with p <0.0001. We might write something like this: "There was a statistically significant difference in minimum dietary diversity between camps 1 and 2 (chi-square 78.6, p<0.0001)." Please note that your percent (in this case, 21.0%) and chi-square statistic (in this case, 78.6) are almost always different. The first number is a percent and the second one is the actual chi-square statistic, followed by the p-value.

Making comparisons between time periods

Now let us assume your 'groups' are different time periods. Suppose you want to know whether minimum dietary diversity has gone up over time. How would you calculate chi-square?

Let us assume you want to compare changes in dietary diversity for 3 time periods: baseline, midterm and endline. In the worksheet **chi-square 3 groups** in the file 'chi-square tests,' type in the number of children at baseline who did and did not have a minimum level of dietary diversity. In all likelihood, you have information from baseline stored in one database, information for midterm in another and information for the endline survey in yet another database. Information for baseline goes in cells B13 and C13, respectively. In cells B14 and C14, put the number of children who had minimum dietary diversity at midterm. Put the number of children with minimum dietary diversity at endline in cells B15 and C15. Remember, you are entering *number* of children, not percent of children.

Where is the p-value? Look at the 'chi-square 2 groups' worksheet. You can find the p-value in cell G25.

How would you write this up in a report? We will use a fictitious example to demonstrate. Let us suppose our chi-square for minimum dietary diversity was 78.6 with p <0.0001. We might write something like this: "There was a statistically significant difference in minimum dietary diversity between baseline, mid-line and endline (chi-square 78.6, p<0.0001)."

GLOSSARY

Anthropometric survey: A survey that measures nutritional status using height (or length), weight and mid-upper arm circumference of children.

Assessment: A *one-time* evaluation of the current nutritional status of children and/or IYCF behaviors.

Categorical data: Data that can assume only a few values. For example, a child's sex is labeled as either male or female. Likewise, only 3 values are possible for 'source of information' for child's date of birth: card, mother, don't know.

Chi-square test: Chi-square tests allow you to see if there are differences between two ore more groups – for example, differences in IYCF behaviors in two refugee camps. Chi-square tests are used when you have categorical data. There are different types of chi-square tests including the Pearson chi-square test, which is the most common.

Cluster sample: A sampling strategy in which the overall population is subdivided into geographically-based groups called clusters and a random sample of clusters is selected for inclusion in the survey. Then, a sub-sample of respondents within the cluster are selected in a random fashion. All data from the selected clusters will make up your sample. A cluster sampling strategy is used when census data or other information that gives a complete picture of the population are lacking or when households are not arranged so that they can be visited systematically. Cluster sampling is frequently used by NGOs and other implementing agencies.

Consistency checks: Consistency checks are one way of checking the quality of your data. You use consistency checks to make sure that information from one variable does not contradict data found in a second variable. For example, if a mother indicates that her child received no foods whatsoever but reports the number of times the child got solid or semi-solid foods, there is an inconsistency in the data.

Continuous data: Data that can assume many different numerical values. For example, in our survey, a child can be anywhere from 0 to 23 months old.

Data analysis: Analyses helps you examine your data. For example, data analysis can show you the proportion of all infants 0-23 months old who began breastfeeding in the first hour of life. Your analyses also tell you whether your findings are statistically significant.

Data cleaning: Once you enter data, you need to clean it. Usually, this means making sure your data are within the limits you specify and that your data are consistent.

Data entry: Putting information from your survey into the computer using a data entry screen.

Data entry screen: A special screen designed in Excel where you can enter data about infant and young child feeding practices using specific fields you have set up.

Data set: An electronic file where you store all of the information you have collected on IYCF practices.

Excel: A Microsoft Office program used for data entry, management, and analysis.

Field: The place on a data entry screen where information is entered. For example, the field where you enter the sex of a specific child could be column M in an Excel spreadsheet. This term is generally used when referring to data entry.

Indicator: Indicators of infant and young child feeding practices give you a sense of how parents and others feed their infants. If you have data from 2 or more different time periods, you can determine how these practices have changed over time. Indicators are used to assess, target and monitor and evaluate. Indicators of infant and young child feeding practices are measures that help to quantify whether you have achieved a result (in the case of infant and young child feeding, the achievement of a particular breastfeeding or complementary feeding practice).

Range checks: Range checks are one way of checking the quality of your data. You look at the values for one variable at a time to make sure they fall within the minimum and maximum values.

Sampling: This involves the selection of a subset of individuals. If done well, sampling will let you confidently infer the IYCF practices of the broader population based on what you learn from some individuals in that population.

Survey: An activity you carry out that includes data collection and analysis for a given population. During the analysis portion of your survey, you examine a particular issue, for example, anthropometric status or infant and young child feeding practices.

Syntax: Language you use to tell computer programs what to do.

Variable: A variable is like a field but 'variable' refers to the field once you have entered all of the data. The field is the place where you enter information (for example, the child's sex in column M) and the variable is the place where the data are stored in your database. 'Field' refers to data entry and 'variable' refers to data cleaning and analysis.

Variable: A field that contains a particular element of data in the data set (for example, 'age', 'date of birth'). You will enter your data in the 'age' *field*, and then use the 'age' *variable* during data cleaning and analysis.