USAID
FROM THE AMERICAN PEOPLE

THE SENSITIVITY OF ANTHROPOMETRIC ESTIMATES TO ERRORS IN THE MEASUREMENT OF HEIGHT, WEIGHT, AND AGE FOR CHILDREN UNDER FIVE IN POPULATION-BASED SURVEYS

DHS METHODOLOGICAL REPORTS 28

AUGUST 2020

This publication was produced for review by the United States Agency for International Development. It was prepared by Thomas W. Pullum, Courtney Allen, Sorrel Namaste, and Trevor Croft.

DHS Methodological Reports No. 28

The Sensitivity of Anthropometric Estimates to Errors in the Measurement of Height, Weight, and Age for Children Under Five in Population-Based Surveys

Thomas W. Pullum
Courtney Allen
Sorrel Namaste
Trevor Croft

The DHS Program ICF
Rockville, Maryland, USA
August 2020

Corresponding author: Thomas W. Pullum, International Health and Development, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: +1 301-407-6500; fax: +1 301-407-6501; email: tom.pullum@icf.com

Acknowledgments: The authors thank Shea Rutstein for helpful comments on an earlier draft of this report.

Editor: Diane Stoy

Document Production: Chris Gramer

This study was conducted with support provided by the United States Agency for International Development (USAID) through The DHS Program (#720-OAA-18C-00083). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Additional information about The DHS Program can be obtained from ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850 USA; telephone: +1 301-572-0200, fax: +1 301-572-0999, email: info@DHSprogram.com, internet: www.DHSprogram.com.

Recommended citation: Pullum, Thomas W., Courtney Allen, Sorrel Namaste, and Trevor Croft. 2020. *The Sensitivity of Anthropometric Estimates to Errors in the Measurement of Height, Weight, and Age for Children Under Five in Population-Based Surveys*. DHS Methodological Reports No. 28. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES v
FIGURES vii
PREFACE ix
ABSTRACT xi
ACRONYMS AND ABBREVIATIONS xiii

1 INTRODUCTION 1

2 RELATIONSHIPS AMONG THE HAZ, WAZ, WHZ AND HEIGHT, WEIGHT, AGE 3
2.1 Anthropometric indicators: HAZ, WAZ, and WHZ 3
2.2 Sensitivity of the HAZ, WAZ, and WHZ formulas to differences in ht, wt, and days 6
2.3 Correspondences among height, weight, and age implied by the HAZ, WAZ, and WHZ 9
2.4 Relationship between stunting and the mean and standard deviation of the HAZ 11

3 ANTHROPOMETRY DATA IN DHS SURVEYS 17
3.1 Overview of surveys conducted between 2010 and 2018 17
3.2 The Peru 2012 and Nepal 2016 surveys 19

4 THE POTENTIAL IMPACT OF POPULATION HETEROGENEITY 29

5 THE POTENTIAL IMPACT OF ERRORS IN THE MEASUREMENT OF HEIGHT, WEIGHT, OR AGE 33
5.1 Systematic errors and random errors 33
5.2 Assumptions and model for random errors in the measurement of height, weight, and age 34
5.3 Simulating the impact of random errors of measurement on the dispersion of the Z scores 39
5.4 The impact of random measurement error on the standard deviations of the Z scores 41
5.5 The impact of random measurement error on the anthropometric outcomes 43

6 CONCLUSIONS 49

REFERENCES 53

APPENDIX 1 MEANS AND STANDARD DEVIATIONS OF Z SCORES IN STANDARD DHS SURVEYS CONDUCTED DURING 2010-18 55

APPENDIX 2 SIMULATING RANDOM ERRORS OF MEASUREMENT 57

APPENDIX 3 RESULTS OF THE MICRO SIMULATIONS IN CHAPTER 5 59

TABLES

Table 2.1 Percentage of children who are stunted for combinations of the mean (m) and standard deviation (s) of the HAZ under the assumption that the HAZ is normally distributed ... 14

Table 3.1 Weighted mean and standard deviation of the HAZ, WAZ, and WHZ for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys ... 20

Table 4.1 Observed means and standard deviations of the HAZ in the Nepal 2016 DHS, by region ... 30

Table A1 Weighted mean and standard deviation of the HAZ, WAZ, and WHZ for children under age 5 in DHS surveys conducted during 2010-18 that included measurements of height and weight ... 55

Table A3.1 The average effect of simulated random error in height, with mean zero and standard deviation s, on the standard deviation of the HAZ and WHZ, and on the estimates of stunting, overweight, and wasting. Children age 0-4 in the Peru 2012 DHS survey. ... 60

Table A3.2 The average effect of simulated random error in height, with mean zero and standard deviation s, on the standard deviation of the HAZ and WHZ, and on the estimates of stunting, overweight, and wasting. Children age 0-4 in the Nepal 20162 DHS survey. ... 61

Table A3.3 The average effect of simulated random error in weight, with mean zero and standard deviation s, on the standard deviation of the WAZ and WHZ, and on the estimates of underweight, overweight, and wasting. Children age 0-4 in the Peru 2012 DHS survey. ... 62

Table A3.4 The average effect of simulated random error in weight, with mean zero and standard deviation s, on the standard deviation of the WAZ and WHZ, and on the estimates of underweight, overweight, and wasting. Children age 0-4 in the Nepal 2016 DHS survey. ... 59

Table A3.5 The average effect of simulated random error in age, with mean zero and standard deviation s, on the standard deviation of the HAZ and WAZ, and on the estimates of stunting, and underweight. Children age 0-4 in the Peru 2012 DHS survey. ... 61

Table A3.6 The average effect of simulated random error in age, with mean zero and standard deviation s, on the standard deviation of the HAZ and WAZ, and on the estimates of stunting, and underweight. Children age 0-4 in the Nepal 20162 DHS survey. ... 61

FIGURES

Figure 2.1 The increase in the HAZ and the decrease in the WHZ due to an increase of 5 centimeters of height, restricted to boys and ignoring whether height is measured lying or standing 7

Figure 2.2 The increase in the WAZ and the WHZ due to an increase of 2 kilograms of weight, restricted to boys and ignoring whether height is measured lying or standing 8

Figure 2.3 The decrease in the HAZ and the WAZ due to an increase of 90 days of age, restricted to boys and ignoring whether height is measured lying or standing 8

Figure 2.4 Expected gain in height per day implied by a HAZ score of -3, 0, or +3 9

Figure 2.5 Expected gain in weight per day implied by a WAZ score of -3, 0, or +3 10

Figure 2.6 Expected gain in weight per unit of height implied by a WHZ score of -3, 0, or +3 10

Figure 2.7 Relationship between the level of stunting and the mean and standard deviation of the HAZ distribution, if it is exactly normal, for combinations of the mean HAZ (-1 and 0) and standard deviation of the HAZ (1.0 and 1.4) 12

Figure 2.8 Relationship between the percentage stunted and the mean and standard deviation of the HAZ distribution, if it is exactly normal, for any value of the mean HAZ between -2 and 0 and values of the standard deviation of the HAZ between 1.0 and 1.4 13

Figure 2.9 Relationship between the percentage stunted and the mean and standard deviation of the HAZ distribution, if it is exactly normal, for any value of the mean HAZ between -2 and 0, and values of the standard deviation of the HAZ from 1.0 and 1.4 in increments of 0.1 14

Figure 3.1 Scatterplot of the standard deviation and mean of the HAZ for children under age 5 in all DHS surveys since 2010 18

Figure 3.2 Scatterplot of the standard deviation and mean of the WAZ for children under age 5 in all DHS surveys since 2010 18

Figure 3.3 Scatterplot of the standard deviation and mean of the WHZ for children under age 5 in all DHS surveys since 2010 19

Figure 3.4 Estimates of stunted, underweight, overweight, and wasted for children under age 5 in the Peru 2012 DHS survey, calculated from both the Z scores and the best-fitting normal distributions 21

Figure 3.5 Estimates of stunted, underweight, overweight, and wasted for children under age 5 in the Nepal 2016 DHS survey, calculated from both the Z scores and the best-fitting normal distributions 22

Figure 3.6 The distribution of height, in cm, for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted 24

Figure 3.7 The distribution of weight, in kg, for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted 25

Figure 3.8 The distribution of age, in days, for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted 25
Figure 3.9 The distribution of the HAZ for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted 26
Figure 3.10 The distribution of the WAZ for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted 26
Figure 3.11 The distribution of the WHZ for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted 27
Figure 5.1 Values of the new or simulated TEM that are produced by combinations of the true TEM and the standard deviation (s) of additional random error 38
Figure 5.2 Change in the standard deviation of the HAZ and WHZ when height is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys 41
Figure 5.3 Change in the standard deviation of the WAZ and WHZ when weight is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys 42
Figure 5.4 Change in the standard deviation of the WAZ and HAZ when age is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys 42
Figure 5.5a Change in the estimates of stunting, overweight, and wasting when height is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 DHS survey 44
Figure 5.5b Change in the estimates of stunting, overweight, and wasting when height is disturbed by random error with standard deviation s for children under age 5 in the Nepal 2016 DHS survey 44
Figure 5.6a Change in the estimates of underweight, overweight, and wasting when weight is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 DHS survey 45
Figure 5.6b Change in the estimates of underweight, overweight, and wasting when weight is disturbed by random error with standard deviation s for children under age 5 in the Nepal 2016 DHS survey 45
Figure 5.7a Change in the estimates of stunting and underweight when age is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 DHS survey 46
Figure 5.7b Change in the estimates of stunting and underweight when age is disturbed by random error with standard deviation s for children under age 5 in the Nepal 2016 DHS survey 46

PREFACE

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

One of the objectives of The DHS Program is to continually assess and improve the methodology and procedures used to carry out national-level surveys as well as to offer additional tools for analysis. Improvements in methods used will enhance the accuracy and depth of information collected by The DHS Program and relied on by policymakers and program managers in low- and middle-income countries.

While data quality is a main topic of the DHS Methodological Reports series, the reports also examine issues of sampling, questionnaire comparability, survey procedures, and methodological approaches. The topics explored in this series are selected by The DHS Program in consultation with the U.S. Agency for International Development.

It is hoped that the DHS Methodological Reports will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low- and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor

Director, The DHS Program

ABSTRACT

Height-for-age (HAZ), weight-for-age (WAZ), and weight-for-height (WHZ) scores for under-5 children are based on a transformation of height, weight, and age into three pairs of bivariate relationships. The WHO 2006 transformation uses formulas to describe the pairwise relationships and coefficients estimated in a population of healthy, well-nourished children, from six settings around the world. The Z scores in that reference population are normally distributed with mean 0 and standard deviation 1. The tails of the Z distributions provide widely used estimates of the percentages of children in a population who are stunted, underweight, overweight, or wasted.

In order to provide estimates of the prevalence of these outcomes, many household surveys conducted by The Demographic and Health Surveys (DHS) Program and the Multiple Indicator Cluster Surveys (MICS) program include measurements of height and weight. A child's age is important for many indicators and is always obtained, even in surveys without anthropometry. Recent surveys have emphasized the training and supervision of the fieldworkers who measure height and weight. Nevertheless, measurement errors do occur, especially for height. In many countries the birthdates of children and their exact ages, measured in days, are often inaccurate.

Over-dispersion of the Z scores is an important indicator of measurement error. When the standard deviations are too large, it is challenging to identify the source or magnitude of the measurement error. The true values of the standard deviations are unknown and are unlikely to be exactly 1. Each measurement affects two Z scores. The impact of error depends on whether the input is in the numerator or the denominator of the Z score.

This report describes the potential impact of measurement error on the Z scores, under a larger perspective of examining the sensitivity of the means and standard deviations of the Z scores, and the estimates of the four problematic outcomes, to specified displacements in the inputs. For illustrative purposes, two DHS surveys with different nutritional profiles and evidence of good data quality are employed: the Peru 2012 and Nepal 2016 surveys. We used three strategies: analysis, macro-simulation, and micro-simulation. With the analytical approach, the formulas and coefficients are used to calculate the Z scores for pairs of hypothetical children, who differ by specified amounts of height, weight, or age. The amounts were arbitrarily set at maximum differences of 5 cm of height, 2 kg of weight, and 90 days of age, because these are approximately 5% of the full range of height, weight, and age, respectively.

A macro-simulation approach describes the sensitivity of the percentage of children who are stunted to changes in the mean and standard deviation of the HAZ. Those results extend to the sensitivity of the other outcomes to the mean and standard deviation of the relevant Z score. Micro-simulation is used to describe the effect of random and normally distributed bidirectional displacements in height, weight, and age on the means and standard deviations of the Z scores and the prevalence of the problematic outcomes. We also describe how population heterogeneity tends to increase the standard deviations and should not always be interpreted as evidence of measurement error, and discuss the potential role of bias, or systematic error, in the measurement of height, weight, and age.

Keywords: anthropometry, measurement error, simulated error

ACRONYMS AND ABBREVIATIONS

CS	continuous survey
DHS	Demographic and Health Surveys
HAZ	height-for-age score
MICS	Multiple Indicator Cluster Surveys
SDG	sustainable development goal
TEM	technical error of measurement
USAID	United States Agency for International Development
WAZ	weight-for-age score
WHO	World Health Organization
WHZ	weight-for-height score
z score	theoretical normally distributed variable with mean 0 and standard deviation 1
Z score	HAZ, WAZ, or WHZ

1 INTRODUCTION

In many household surveys conducted by The Demographic and Health Surveys (DHS) Program and the Multiple Indicator Cluster Surveys (MICS) program, there is evidence of over-dispersion in the height-for-age (HAZ), weight-for-age (WAZ), and weight-for height (WHZ) scores, or Z scores, which potentially can lead to overestimates of the levels of stunting, underweight, overweight, and wasting for children age 0-4. In recent years, DHS has increased its efforts to reduce errors in the measurement of height, weight, and age that could be the underlying source of over-dispersion in anthropometry scores. As part of that effort, this report will investigate possible sources of over-dispersion. We will use analytical and simulation methods to assess the potential impact that errors in the measurement of height, weight, and age may have on the dispersion of the Z scores and the estimates of anthropometric outcomes.

The HAZ, WAZ, and WHZ scores were originally constructed to have—as nearly as possible—a unit normal distribution in a well-nourished population of children age 0-4. This correspondence with a z or N(0,1) distribution is conveyed by the letter Z in the names of the variables. There is empirical evidence that the Z scores have a standard distribution that is approximately one even when the mean differs from zero. The standard deviations of the Z scores are important indicators of data quality for anthropometric measurement (Mei and Grummer-Strawn 2007).

In this report, Chapter 2 describes the relationships of the Z scores and the anthropometric outcomes to the "true" values of height, weight, and age, in the absence of measurement error. This analysis is based entirely on the formulas for the Z scores and how they are used to calculate the percentage of children who are stunted, underweight, overweight, or wasted. Chapter 2 also describes how the percentage of children who are stunted is related to the mean and standard deviation of the HAZ under the assumption that the HAZ is normally distributed. This correspondence is identical with how the percentage underweight relates to the mean and standard deviation of the WAZ, and how the percentages of overweight or wasted are related to the mean and standard deviation of the WHZ.

Chapter 3 provides a brief overview of the means and standard deviations of the Z scores observed in 74 standard DHS surveys conducted from 2010 to 2018. Two specific surveys are used for illustrative purposes because they include high-quality anthropometric data: the 2012 round of the Peru Continuous Survey and the 2016 survey of Nepal. The two surveys differ in their prevalence of malnutrition, but both surveys have minimal dispersion in the Z scores and generally good data quality.

Chapter 4 investigates the potential effect of heterogeneity on the dispersion of Z scores. When a population is a mixture of homogeneous subpopulations that have Z scores with standard deviations of one, but different means, it is impossible for the overall population to have standard deviations of one. In most DHS surveys, for example, the mean Z scores—and the levels of stunting, etc.—vary across wealth quintiles. Poorer quintiles typically have lower nutritional levels and thus lower means. A criterion of a standard deviation of one within the wealth quintiles is incompatible with a criterion of one for the combined population. Nevertheless, DHS has observed standard deviations closer to one in surveys known to have measurements of better quality.

Chapter 5 uses simulation to investigate the role of bidirectional random errors in the measurements of height, weight, and age, one at a time. This simulates the effect that such errors would have on the dispersion of the Z scores and the estimates of nutritional outcomes.

Chapter 6 includes conclusions and describes potential extensions of these approaches.

2 RELATIONSHIPS AMONG THE HAZ, WAZ, WHZ AND HEIGHT, WEIGHT, AGE

2.1 Anthropometric indicators: HAZ, WAZ, and WHZ

Commonly cited anthropometric indicators are the percentages of children who are stunted, underweight, overweight, or wasted.[1] All four indicators also appear in “severe” versions as the percentage of children who are severely stunted, etc. These binary indicators are calculated from the distributions of the HAZ, WAZ, and WHZ, which in turn are calculated from the child’s height, weight, age in days, and sex. The formulas for the HAZ and WHZ include a fifth input, which is a characteristic of the measurement process: whether the child was lying down or standing when height was measured. The HAZ, WAZ, and WHZ used here and in DHS reports are based on the WHO Child Growth Standards (WHO Multicentre Growth Reference Study Group 2006a).

The HAZ, for example, is a height-for-age “z score.” The empirically based coefficients used to calculate the HAZ were constructed in such a way that, when the HAZ is calculated for all children in a healthy reference population, the mean will be approximately 0, the standard deviation will be approximately 1, and the distribution will be approximately normal. Approximately 2.3% of children in such a population will have HAZ < -2 and approximately 0.1% will have HAZ<-3. Children with HAZ < -2 or HAZ < -3 are described as stunted or severely stunted, respectively. However, even in a healthy population, these extremes can arise through random processes and are not interpreted as evidence of malnutrition. In a malnourished population, substantially higher percentages of stunting or severe stunting will be observed.

Similarly, the WAZ is a weight-for-age z-score. Children with WAZ < -2 or WAZ < -3 are described as underweight or severely underweight. The WHZ is a weight-for-height z-score. Children with WHZ > 2 or WHZ > 3 are overweight or severely overweight, while children with WHZ < -2 or WHZ < -3 are wasted or severely wasted.

We will refer to measured height as “ht”, measured in units of 0.1 cm, or millimeters. In DHS data files, the variable is hc3. Measured weight is “wt”, measured in units of 0.1 kg. In DHS surveys, the variable is hc2. The formulas for the HAZ, WAZ, and WHZ expect height and weight to be on scales of cm and kg. In the application of the formulas, ht and wt as we define them must be divided by 10 as they are stored with one implicit decimal place in hc3 and hc2. Age is measured with “days”—completed days of age, calculated as the difference between the day of measurement and the day of birth. Five full years include 4*365+1*366=1826 days, if the interval includes only one leap year, or 1827 for two leap years. Children’s days of age are numbered 0 through 1827, inclusive.

In a heuristic or metaphorical sense, the HAZ is a ratio of height to age. Thus, among children who have the same day of age, a taller child will have a higher HAZ score, because the “numerator” is larger. Among children who have the same height, an older child will have a lower HAZ score, because the “denominator” is larger. Similarly for the WAZ and WHZ. Height appears once in a numerator and once in a denominator

[1] Among these outcomes, all except the percentage underweight are specified in the Sustainable Development Goals (SDGs). The main reports of all DHS surveys with anthropometry have included all four outcomes.

(in the HAZ and WHZ, respectively), while weight appears only as a numerator (in the WAZ and WHZ) and age appears only as a denominator (in the HAZ and WAZ).

For an analogy with a usual z score, the HAZ is similar to the observed height, minus the expected height for the given days of age, divided by the standard deviation of height (among children in the normative population with the given days of age). This analogy is closer to the actual calculation but is still not complete. The calculation of the HAZ involves three parameters which are specific to each combination of inputs, and are not constants: M (analogous to the expected value); S (a measure of dispersion that is analogous to the standard deviation but only appears as a multiplier of the other two parameters); and L (a shape parameter). These three coefficients are used in a formula for the HAZ in the range of -3 to +3:

$$Z(Y) = \frac{(Y/M)^L - 1}{SL}. \qquad (2.1)$$

In this formula, Y is height in centimeters (for the HAZ) or weight in kilograms (for the WAZ and WHZ). The three coefficients are not constants, but for the HAZ vary according to the sex and age of the child and whether the child was measured lying or standing. In the calculation of the WAZ, the values of M, S, and L depend on the child's sex and weight. In the calculation of the WHZ, the values of M, S, and L depend on the child's sex, height, and whether the child was measured lying or standing.

It is possible to re-state the formula with Y as a function of a specified Z:

$$Y(Z) = M(1 + SLZ)^{1/L}. \qquad (2.2)$$

If Z is greater than 3, a different formula is used that extrapolates from the values at Z = 2 and Z = 3:

$$Z(Y) = 3 + \left[\frac{Y - Y(3)}{Y(3) - Y(2)}\right]. \qquad (2.3)$$

If Z is less than -3, the formula extrapolates from the values at Z = -2 and Z = -3:

$$Z(Y) = -3 + \left[\frac{Y - Y(-3)}{Y(-2) - Y(-3)}\right]. \qquad (2.4)$$

In the reference population for which the framework was devised and M, S, and L were estimated, scores outside the range -3 to +3 are extremely rare. Only about two well-nourished children in 1,000 would be expected to be outside this range—one above and one below.

For the HAZ, L is always 1, so (2.1) simplifies to

$$Z(Y) = \frac{Y - M}{MS} \qquad (2.5)$$

In this case, which applies only to the HAZ, (2.3) and (2.4) also reduce to (2.5). For the WAZ and WHZ, however, (2.3) and (2.4) do not simplify in this way and have discontinuous derivatives at -3 and +3.

The WHO guidelines for full ranges of the HAZ, WAZ, and WHZ that will be used here are -6 to +6 for the HAZ, -6 to +5 for the WAZ, and -5 to +5 for the WHZ. Scores outside those ranges are deemed implausible, even for malnourished populations.

Heights of children can be measured when the child is lying or standing. This difference is captured with a variable LORH, which is coded 1 if the child was lying or 2 if the child was standing. Measurers are trained to measure the child lying if the age is 0-23 months or standing if the age is 24+ months. Because there may be deviations from this rule for individual children, the WHO files for calculating the HAZ or WHZ include this variable over wide range of ages and lengths/heights. The M, S, and L parameters vary slightly depending on this characteristic. The DHS data files include a variable hc15, which is correspondingly coded 1 or 2. In rare instances, hc15 is not coded. Such cases are given default values of 1 if the child's age is 0-23 months and 2 if age 24+ months.

The Stata version of the WHO package "igrowup" includes four data files for the construction of the Z scores: lenanthro.dta, required for the HAZ; weianthro.dta, for the WAZ; and wflanthro.dta and wfhanthro.dta, for the WHZ (LORH is implicitly coded 1 or 2, respectively, in the last two files). For example, lenanthro.dta has one line for each possible combination of sex (1,2) and days. For each line, it gives the values of M, S, and L that would be required to calculate the HAZ for that combination of sex and age. An analyst would program the formula and then substitute an observed child's height (in DHS terms, hc3/10) into it. The result would be the HAZ for that child. Similar steps would be used for the WAZ and WHZ.

The "igrowup" package includes a comprehensive program that manipulates these four data files (and files for other indices, such as upper arm circumference) for interactive use by pediatricians, for example. Both DHS and MICS use a modification to process the child data and produce Z scores with CSPro. We use a different approach in Stata.

We construct three files, based on the four igrowup files described above, which give the appropriate calculated z score for each combination of the relevant inputs. HAZ_reference.dta, for example, gives the calculated HAZ score for each combination of sex, LORH, days, and ht, and is sorted by these variables. It omits any combinations that would produce a HAZ value outside the range -6 to 6, but can be modified easily for a more narrow or wider range. WAZ_reference.dta and WHZ_reference.dta are calculated in a similar way. These files include all specific values of ht, wt, and days (and sex and LORH) that could occur and lead to Z scores in the specified ranges.

With this approach, a DHS data file for children is first sorted by sex, LORH, days, and ht, and then merged with HAZ_reference.dta to obtain the HAZ scores. It is then sorted by sex, day, and wt, and merged with WAZ_reference.dta to obtain the WAZ scores. It is then sorted by sex, LORH, wt, and ht and merged with WHZ_reference.dta to obtain the WHZ scores. The scores obtained in this way agree with the scores generated by the usual DHS approach in CSPro.

The units of measurement of height, weight, and age are arbitrary and not comparable. The ranges of height, weight, and age in the reference files can provide some guidance for comparing the sensitivity of the Z scores to the three measurements. We suggest comparing the scales in terms of 5% of their ranges. Height ranges from 380 to 1380 mm (the units of hc3) with 5% of this range at 50 mm, or 5 cm. Weight ranges from 9 to 380 tenths of a kg (the units of hc2) with 5% of this range approximately 20 tenths of a kg, or 2 kg. Age ranges from 0 to 1827 days with 5% of this range approximately 90 days. Thus, we can crudely say that there is comparability between measurement errors that are proportional to 5 cm in height, 2 kg in weight, and 90 days of age.

2.2 Sensitivity of the HAZ, WAZ, and WHZ formulas to differences in ht, wt, and days

If two children have exactly the same age, and one is taller, what is the difference in their HAZ scores? If two children have exactly the same height, but one is older, what is the difference in their HAZ scores? Similar questions can be asked about the effect of weight and age on the WAZ score, and the effect of weight and height on the WHZ score.

These questions could be answered directly from the formulas, but because the M, S, and L coefficients are different for each value of the denominator measurement, we prefer a graphical approach that uses the reference files for the HAZ, WAZ, and WHZ. The hypothetical differences between two children's height, weight, and age will be 5 cm, 2 kg, and 90 days, respectively.

Figure 2.1 describes the effect of a fixed difference in a height of 5 cm and contains subfigures for the two Z scores affected by height. The subfigure on the left is based on analysis of the HAZ reference file, which simply gives the HAZ for all combinations of height and age that produce a HAZ in the range -6 to +6. This shows that a 5 cm increase in length or height produces an increase in the HAZ that ranges from about 2.5, soon after birth, to about 1.0 just before the child's fifth birthday. The effect is curvilinear and begins to flatten in the last half of the age interval.

There is a small difference between the curves for boys and girls. For our purposes, the difference is not relevant. To avoid repetition, Figures 2.1-2.3 are restricted to boys. One might expect to see different lines for different levels of height, but lines for three levels of height—70, 90, and 110 cm—are indistinguishable because the parameter L is fixed at 1 in the calculation of the HAZ. Other comparisons shown in Figures 2.1-2.3 do have separate lines.

The subfigure on the right side of Figure 2.1 describes the effect on the WHZ of the same fixed change (5 cm) in height. The vertical axis is negative because an increase in height always reduces the WHZ. The lines for the three levels of height are distinct but are almost entirely within a narrow range of effect on the WHZ, between -1 and -1.5.

All the lines for the WAZ and WHZ have three segments. In the subfigure for the WHZ in Figure 2.1, the middle segment is for WHZ in the range -3 to +3, which includes virtually all children in the well-nourished reference population. The segment on the left refers to a WHZ between -5 and -3, while the segment on the right refers to a WHZ between 3 and 5.[2]

Our interpretation of the effects on the HAZ and WHZ of a fixed difference or change in height is that for most children, the effects are comparable in magnitude, although in opposite directions. For children under age 2, the effect on the HAZ is greater than the effect on the WHZ.

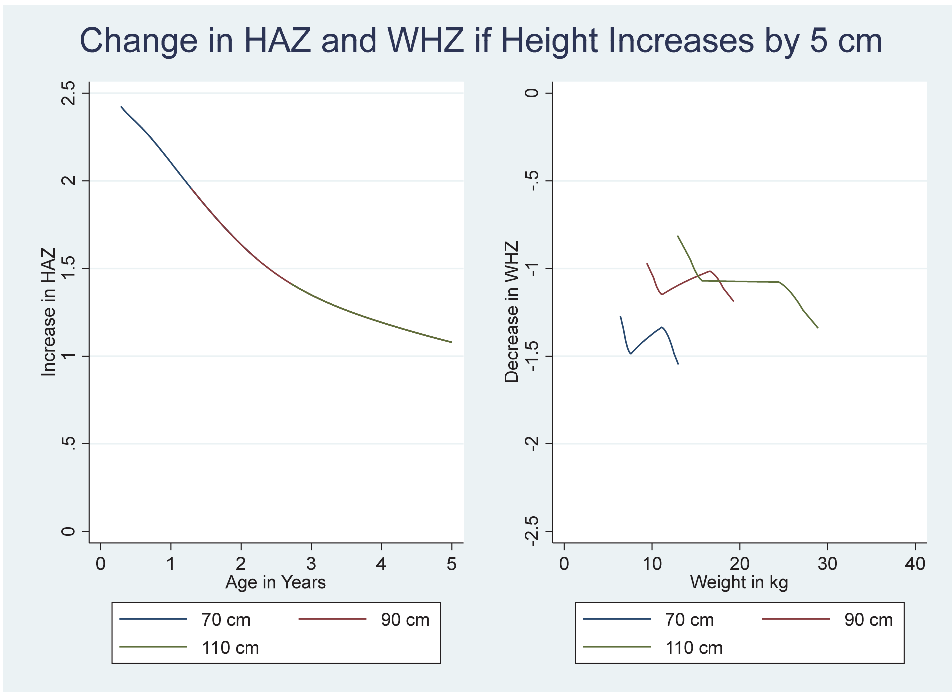
The subfigures in Figure 2.2 describes how the HAZ and the WHZ change if weight is increased by 2 kg. The vertical axis is positive in both figures because an increase in weight (holding the other measurement constant) will always increase these Z scores. Each subfigure has two lines, one for an arbitrary base of 10

[2] As described above, Z scores in the WHO plausible range but less than -3 or greater than +3 are calculated with equations (2.3) and (2.4). For the WAZ and WHZ, the extrapolation produces some discontinuity or even a reversal in slope, but within ranges that are narrow and hypothetical.

kg and the other for a base of 20 kg. The blue line describes the difference between a boy who weighs 12 kg and a boy who weighs 10 kg. The red line describes the difference between 22 kg and 20 kg. Both lines are higher in the subfigure for the WHZ than in the subfigure for the WAZ. That is, the same fixed difference has more impact on the WHZ than on the WAZ. The lower the child's weight, the greater the impact on the WHZ, compared with the WAZ.

The third figure in this group, Figure 2.3, describes the effect on the HAZ and WAZ of a 90-day increase in age. Because age is in the denominator of these two Z scores, both of the vertical axes are negative. Each subfigure has four lines for a 90-day increase past the first, second, third, and fourth birthdays. The lowest line in each subfigure describes the reduction if a child 90 days past the first birthday (age 455 days) is compared with a child on the first birthday (age 365 days). The older the child, the smaller the effect of a 90-day increase in age, but the lines all slope downward. The penalty is slightly greater for taller or heavier children at each specific age. The lines in the subfigure for the HAZ are consistently lower than those for the WAZ, and the effect on the HAZ is about twice the effect on the WAZ.

Figure 2.1 The increase in the HAZ and the decrease in the WHZ due to an increase of 5 centimeters of height, restricted to boys and ignoring whether height is measured lying or standing



In summary, the effect of a fixed difference in height is similar in magnitude for the HAZ and WHZ—although in opposite directions—except for children under age 2, for whom the effect on the HAZ is much greater. The effect of a fixed difference in weight is greater for the WHZ than for the WAZ, except for lighter weight children, for whom the role of height is more pronounced. The effect of a fixed difference in age is much greater on the HAZ than on the WAZ, for children of all ages, heights, and weights.

Figure 2.2 The increase in the WAZ and the WHZ due to an increase of 2 kilograms of weight, restricted to boys and ignoring whether height is measured lying or standing

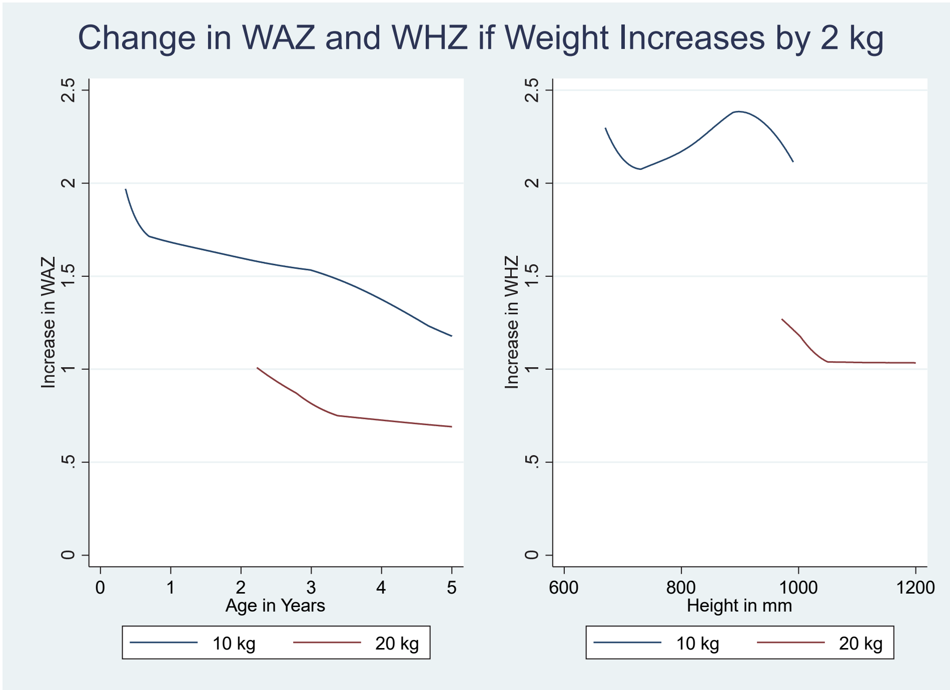
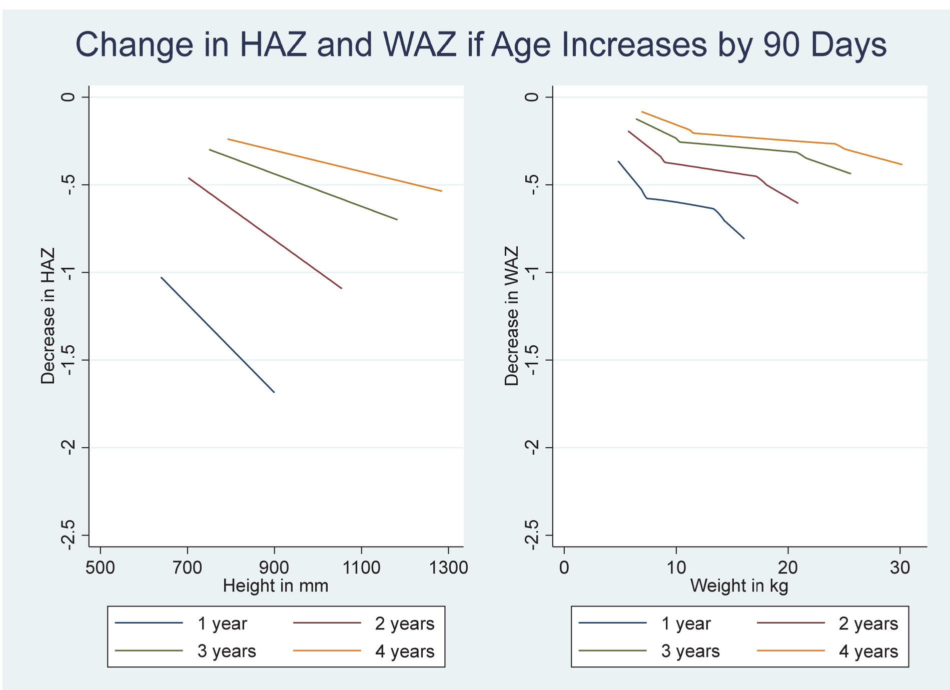


Figure 2.3 The decrease in the HAZ and the WAZ due to an increase of 90 days of age, restricted to boys and ignoring whether height is measured lying or standing



2.3 Correspondences among height, weight, and age implied by the HAZ, WAZ, and WHZ

The WHO files that contain the coefficients for the Z scores (lenanthro, weianthro, wflanthro, and wfhanthro) can be analyzed in their own right, separately from DHS data, to describe the trajectories of child growth. As a child becomes progressively older, taller, and heavier, the Z scores describe a normative pattern of development. The rate of change, or slope, of a growth trajectory can be interpreted as the amount of change in the numerator variable, per unit of increase in the denominator variable, that is required to keep the child on the trajectory. These slopes help calibrate or equilibrate the units of measurement for height, weight, and age.

Figure 2.4 Expected gain in height per day implied by a HAZ score of -3, 0, or +3

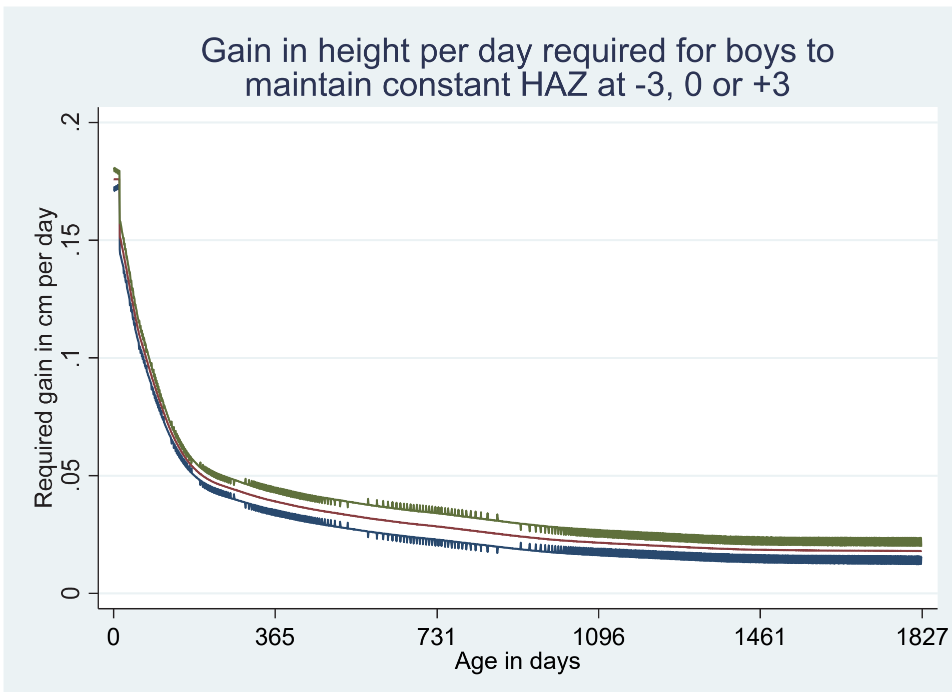


Figure 2.4 includes three lines for the daily gain in height required for a boy to remain consistently with a HAZ of -3, 0, or 3. The lines are not easily distinguishable because the range in daily gain is so small. The highest line is for HAZ=3, the lowest line for HAZ=-3, and the middle (red) line for HAZ=0. Over the course of the first 5 years (1827 days), the height of a well-nourished (HAZ=0) boy will increase from about 50 cm to about 110 cm, an increase of about 60 cm, or about 0.033 cm per day. As Figure 2.4 shows, the daily rate of increase is highest during the first several months and then levels off. On the first, second, third, fourth, and fifth birthdays, the daily increase in height that will keep a child at HAZ=0 will be (expressed in cm) 0.039, 0.028, 0.021, 0.019, and 0.018, respectively. Stated differently, around the third or fourth birthday for a child near HAZ=0, an age that is one day higher than the true value will completely offset a height that is 0.020 cm higher than the true value. Two such displacements will neutralize each other, in terms of their impact on the HAZ. For a child at HAZ=3, the daily increments in height would be about 0.005 higher, across most ages. For child at HAZ=-3, the daily increments in height would be about 0.005 less. The boundary for severe stunting is HAZ=-3.

Figure 2.5 Expected gain in weight per day implied by a WAZ score of -3, 0, or +3

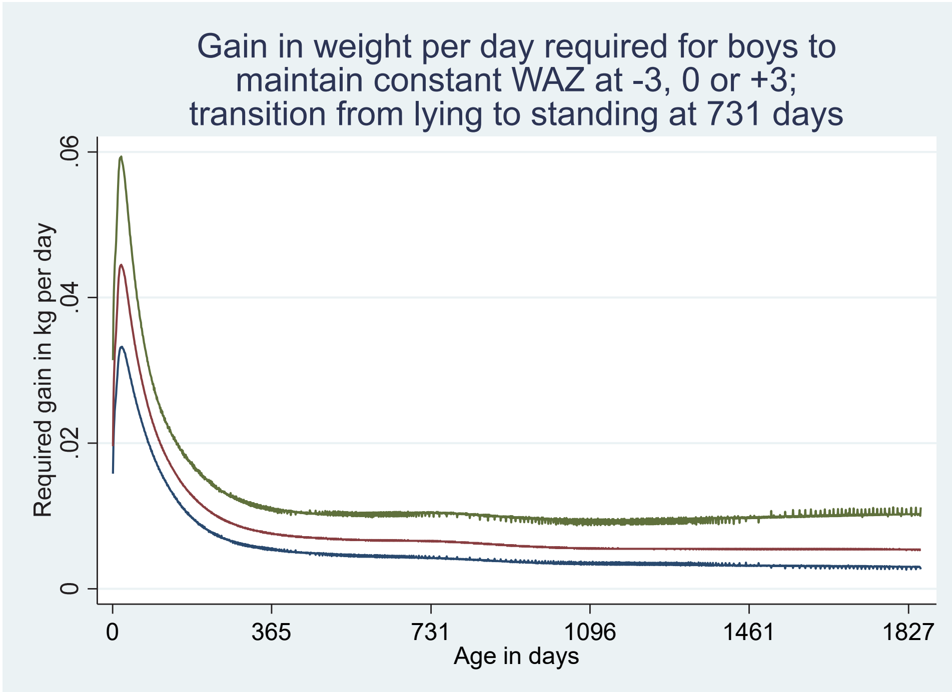


Figure 2.6 Expected gain in weight per unit of height implied by a WHZ score of -3, 0, or +3

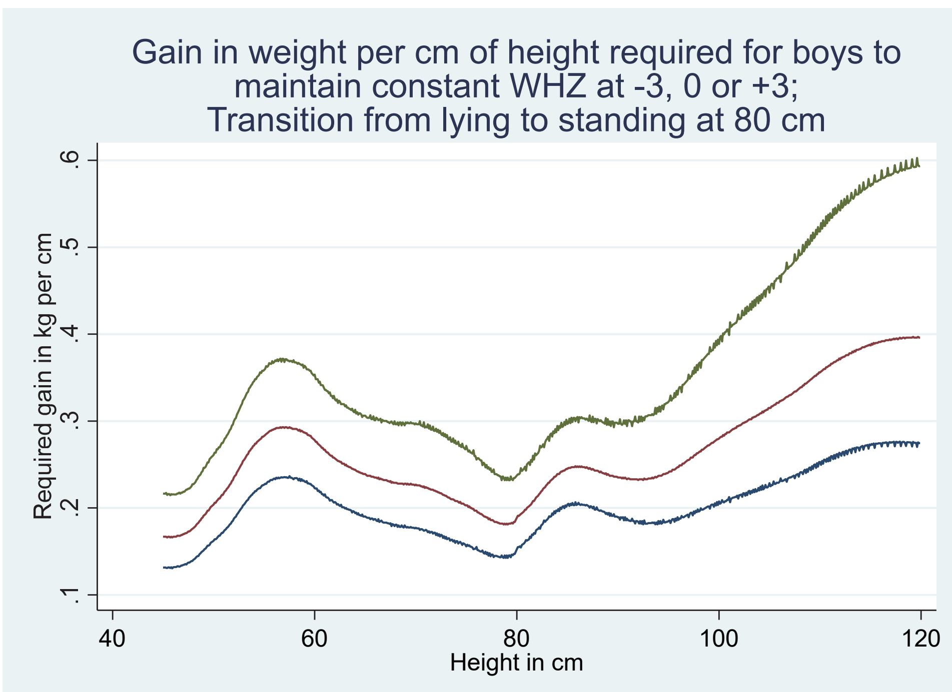


Figure 2.5 shows that after the first birthday, the daily increment in weight required to maintain a child at a fixed level of the WAZ is quite steady. The daily weight gain to remain at WAZ=0 on the successive birthdays would be 0.0075, 0.0066, 0.0055, 0.0055, and 0.0054 kg. These are very small amounts. During

most of the under-5 years, a normal child will gain about 2 kg per year. An increment of one day and an increment of 0.0055 kg will offset each other in terms of weight for age. A child with WAZ=-3 will remain at the same level with about 0.0020 less weight gain per day throughout most of the under-5 age range. Such a child is at the boundary for being severely underweight. A child with WAZ=3 will stay at the same level with about 0.0045 more weight gain per day.

The three lines in Figure 2.6 show the increase in weight, per unit of increase in height, that is required to maintain a child at the same level of the WHZ (for values -3, 0 and +3). The units of weight and height are kg and cm. The pattern is irregular and very different from what was observed for the HAZ and WAZ. The middle (red) line describes well-nourished children with WHZ=0. The upper line is for children with WHZ=3, the boundary for severely overweight children, and the bottom line is for WHZ=-3, the boundary for severely wasted children. Although there is irregularity, the great majority of children are in the range of an increase of 0.2 kg to 0.4 kg per increase of 1 cm in height. The range above and below the middle line becomes wider for taller children, due primarily to the rising boundary for WHZ=3, the boundary for severely overweight children.

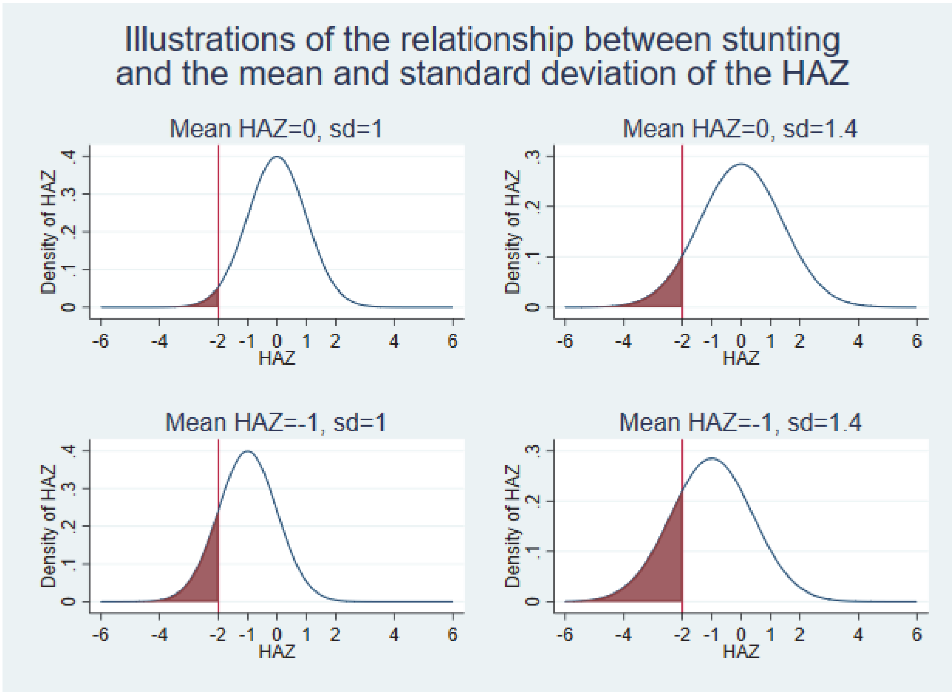
In summary, after the first year, a typical daily gain in height should be 0.020 cm, which is equivalent to about 1 cm in 50 days. A typical daily gain in weight would be 0.0055 kg, which is equivalent to about 1 kg in 180 days. The ratio of 0.0055 kg per day to 0.020 cm per day is about 0.3 kg per cm, which is consistent with the slope of the WHZ during most of the age range. Thus, for most of the age range after the child's first birthday, comparable effects on the Z scores would come from multiples of these pairwise ratios of height in cm, weight in kg, and age in days.

2.4 Relationship between stunting and the mean and standard deviation of the HAZ

As stated earlier, the level of stunting is calculated as the percentage of children whose HAZ score is less than -2. All anthropometry indicators in this report are defined similarly, in terms of Z scores that are more extreme than threshold values of -2 and -3, for underweight and wasting, as well as stunting, or +2 and +3, for overweight. This section discusses the relationship between stunting and the HAZ, although this terminology represents a more general correspondence between all the binary outcomes and the three Z scores.

The transformations of height, weight, and age into the HAZ, WAZ, and WHZ scores was done in such a way that the three Z scores are approximately normal in a homogenous well-nourished population. It is possible to approximate the percentage stunted as the percentage of the area under the curve and to the left of -2 for any normally distributed variable with a specified mean and standard deviation. Figure 3.1 illustrates the relationship for four different combinations. Within the figure, the first row shows two HAZ distributions with mean 0 and the second row shows two distributions with mean -1. The distributions in the first column have a standard deviation of 1 and those in the second column have a standard deviation of 1.4. In each figure, the red area to the left of HAZ=-2 is the proportion stunted.

Figure 2.7 **Relationship between the level of stunting and the mean and standard deviation of the HAZ distribution, if it is exactly normal, for combinations of the mean HAZ (-1 and 0) and standard deviation of the HAZ (1.0 and 1.4)**

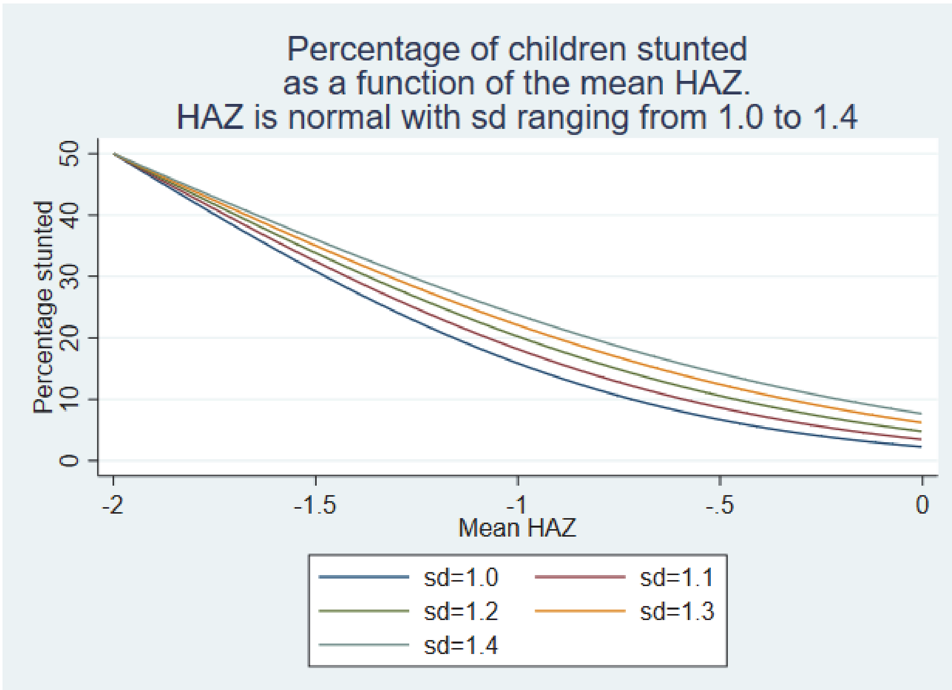


The distribution in the upper left of Figure 2.7, with mean 0 and standard deviation 1, is the normative distribution constructed by WHO. The area to the left is a proportion 0.023, or a percentage of 2.3%, of the area under the curve. As described earlier, this percentage is expected because of random components that affect height, and given age (and sex) in a healthy population. All other figures have larger areas to the left of HAZ=-2. If the mean is 0 and the standard deviation is 1.4, the percentage stunted increases to 7.7%. This increase is completely explained by the increased dispersion or heterogeneity. If the mean is -1 and the standard deviation is 1, the percentage stunted is 15.9%. This increase above 2.3% is completely explained by the lower mean. If the mean is -1 and the standard deviation is 1.4, the percentage stunted is 23.8%, which is nearly a quarter of the cases and about ten times the percentage in the reference distribution. This increase beyond 2.3% is due to the combined effects of a lower mean and increased heterogeneity. The fourth combination is fairly typical for DHS data.

Comparing the fourth estimate of stunting with the third, the ratio 23.8/15.9=1.50 implies that if the mean HAZ is -1, an increase in the standard deviation from 1.0 to 1.4 will produce a 50% increase in the estimate of stunting. If the increase in the standard deviation in the HAZ, from 1.0 to 1.4, were the result of measurement error, rather than underlying heterogeneity, this would be a substantial distortion.

Figures 2.8 and 2.9 explore this joint effect on the estimate of stunting of different values of the mean and standard deviation of the HAZ. In simulations that use the theoretical normal distribution, rather than empirical HAZ data, we allow the mean HAZ to range anywhere between -2 and 0, and the standard deviation to range between 1.0 and 1.4, with increments of 0.1. Figure 2.8 shows the estimated percentage stunted on the vertical axis. The horizontal axis is the mean HAZ. There are five curved lines for standard deviations 1.0 (the lowest line), 1.1, 1.2, 1.3, and 1.4 (the highest line). If the mean is 0 and the standard deviation is 1.0 or 1.4, then the corresponding values on the vertical axis are 2.3% and 7.7%, respectively. These were given as the percentages of the red areas in the upper subgraphs of Figure 2.7. If the mean is -1 and the standard deviation is 1.0 or 1.4, then the estimates are 15.9% and 23.8%, which correspond with the red areas in the lower subgraphs of Figure 2.7.

Figure 2.8 Relationship between the percentage stunted and the mean and standard deviation of the HAZ distribution, if it is exactly normal, for any value of the mean HAZ between -2 and 0 and values of the standard deviation of the HAZ between 1.0 and 1.4



The five lines in Figure 2.8 are farthest apart—that is, the additive effect on the estimate of stunting is greatest—if the mean HAZ is, in fact, in the vicinity of -1. If the mean HAZ were as low as -2, the estimate of stunting would be 50%, regardless of the standard deviation. The level of stunting can only exceed 50% if the mean HAZ drops below -2.

Figure 2.9 shows the additive effect of dispersion above an SD of 1.0 in a different way. As in Figure 2.8, the mean HAZ is shown on the horizontal axis. The vertical axis is the *increase* in the estimated percentage stunted, above the normative level of 2.3%. There are four lines to show the increase if the standard deviation of the HAZ is 1.1, 1.2, 1.3, or 1.4 (the highest line). The horizontal axis serves as the reference if the standard deviation is 1.0. In effect, this figure repeats the separation between the five lines in Figure 2.8. It shows more clearly, with green dots, that the maximum effect of increased dispersion occurs when the mean HAZ is close to (but slightly above) -1. The additive effect of increasing the standard deviation is minimal if the mean HAZ is near 0 or if between -2.0 and about -1.5. The maximum additive increase in the estimated percentage stunted is about 2% if the standard deviation increases from 1.0 to 1.1, and about 2% for every additional increase of 0.1. Thus, if s is the observed standard deviation, and s is greater than 1, because of genuine heterogeneity and/or measurement error, then the *maximum* additive increase in the percentage stunted is approximately (s-1)*20%. Otherwise, the percentage stunted is determined by the mean of the HAZ distribution, with a pattern shown by the bottom line in Figure 2.8.

Figure 2.9 Relationship between the percentage stunted and the mean and standard deviation of the HAZ distribution, if it is exactly normal, for any value of the mean HAZ between -2 and 0, and values of the standard deviation of the HAZ from 1.0 and 1.4 in increments of 0.1

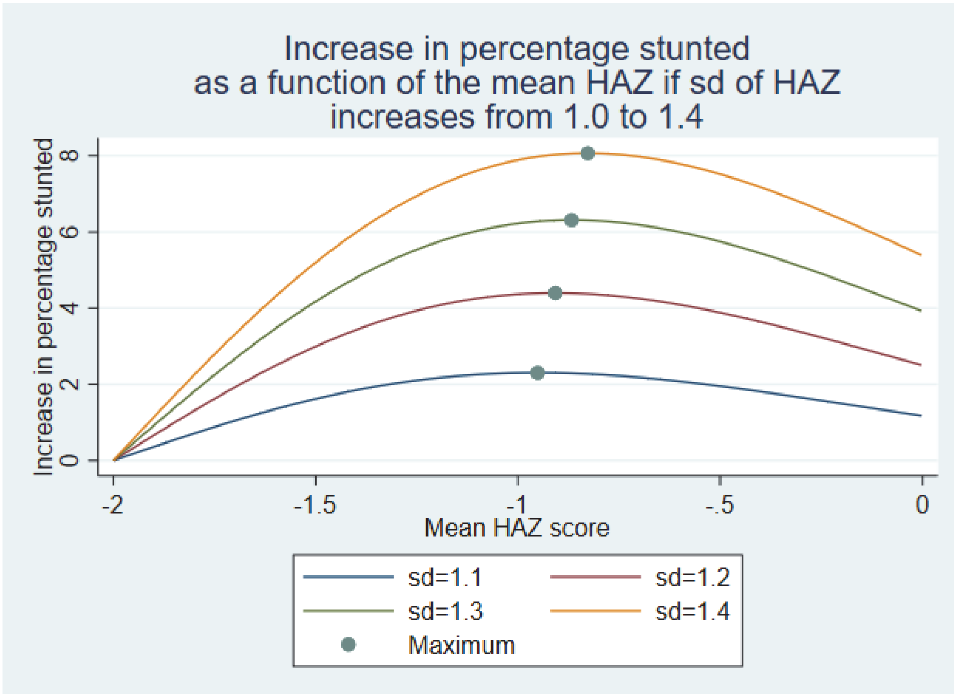


Table 2.1 Percentage of children who are stunted for combinations of the mean (m) and standard deviation (s) of the HAZ under the assumption that the HAZ is normally distributed

	Standard deviation s						
Mean m	**0.9**	**1**	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**
0	1.31	2.28	3.45	4.78	6.20	7.66	9.12
-0.1	1.74	2.87	4.21	5.67	7.19	8.74	10.26
-0.2	2.28	3.59	5.09	6.68	8.31	9.93	11.51
-0.3	2.95	4.46	6.11	7.83	9.55	11.23	12.85
-0.4	3.77	5.48	7.29	9.12	10.92	12.65	14.31
-0.5	4.78	6.68	8.63	10.56	12.43	14.20	15.87
-0.6	5.99	8.08	10.16	12.17	14.08	15.87	17.53
-0.7	7.43	9.68	11.86	13.93	15.87	17.66	19.31
-0.8	9.12	11.51	13.77	15.87	17.80	19.57	21.19
-0.9	11.08	13.57	15.87	17.97	19.87	21.60	23.17
-1	13.33	15.87	18.17	20.23	22.09	23.75	25.25
-1.1	15.87	18.41	20.66	22.66	24.44	26.02	27.43
-1.2	18.70	21.19	23.35	25.25	26.92	28.39	29.69
-1.3	21.84	24.20	26.23	27.98	29.51	30.85	32.04
-1.4	25.25	27.43	29.27	30.85	32.22	33.41	34.46
-1.5	28.93	30.85	32.47	33.85	35.03	36.05	36.94
-1.6	32.84	34.46	35.81	36.94	37.92	38.75	39.49
-1.7	36.94	38.21	39.25	40.13	40.87	41.52	42.07
-1.8	41.21	42.07	42.79	43.38	43.89	44.32	44.70
-1.9	45.58	46.02	46.38	46.68	46.93	47.15	47.34
-2	50.00	50.00	50.00	50.00	50.00	50.00	50.00

Exactly the same relationships apply to the estimates of the percentage of children who are underweight using the WAZ and the percentage who are overweight or wasted using the WHZ.

Table 2.1 shows the percentage of children who would be stunted with specific combinations of the mean and standard deviation of the HAZ, under the assumption that the HAZ is normal. The mean ranges between 0 and -2 in increments of 0.1, and the standard deviation between 0.9 and 1.5, in increments of 0.1. The numbers in the table include values given above for specific combinations. The table extends the range of the standard deviation by 0.1 in each direction.

3 ANTHROPOMETRY DATA IN DHS SURVEYS

3.1 Overview of surveys conducted between 2010 and 2018

DHS has more than 30 years of experience collecting height and weight measurements for children age 0-4 (and women age 15-49). Considerable analysis of the data, some that assesses data quality, has appeared in DHS publications listed in the references. A thorough description of procedures, as well as efforts to monitor data quality during fieldwork, is given by Allen, Croft, Pullum, and Namaste (2019). Current procedures for data collection, analysis, and reporting are consistent with WHO and UNICEF (2019) guidance.

Height and weight are typically measured by a biomarker technician after completion of the household and individual interviews. Age is obtained during the household interview, first in single years, as part of the household roster. Subsequently, birthdates are obtained for children. Birthdates are provided by the mother, if she is in the household, and otherwise are provided by the household respondent.

This chapter provides an overview of the anthropometric results with the means and standard deviations of the three Z scores in 74 standard DHS surveys conducted from 2010 to 2018, inclusive, that included height and weight measurements. We omit rounds of the Continuous Surveys in Peru and Senegal, except for the 2012 round of the Peru CS. A complete list of the surveys is included in an appendix, along with the numerical results that are the basis of three scatterplots, Figures 3.1-3.3. Estimates in this chapter include the adjustment for sample weights.

The scatterplot in Figure 3.1 describes the distribution of the HAZ for these surveys. The horizontal axis shows the mean HAZ for the survey and the vertical axis the standard deviation. A green dot represents the Peru 2012 survey and a blue dot represents the Nepal 2016 survey. Four surveys had a standard deviation for the HAZ greater than 2: Benin 2011-12 (2.33); Egypt 2014 (2.02); Nigeria 2013 (2.01); and Timor-Leste 2016 (2.14). Those four surveys are excluded from Figure 3.1 but are included in Figures 3.2 and 3.3 simply to have the same vertical scale for all three figures.

The three figures have identical ranges of the horizontal and vertical scales. The scale for the mean ranges from -2.5 to 1, and the scale for the standard deviation from 1 to 2. The four excluded surveys are within these ranges for all measures except the standard deviation of the HAZ, and all other surveys are within these ranges for all measures.

The points are clustered in very different ways in the three figures. In Figure 3.1, the mean HAZ is always less than zero. The average of the mean HAZ is -1.30. The average standard deviation of the HAZ is high, 1.54, including the four surveys that were excluded from the figure. In Figure 3.2, the mean WAZ is shifted slightly to the right, with an average value of -0.85. The average standard deviation is lower, 1.18. In Figure 3.3, the mean WHZ is much further to the right and closer to 0, with an average value of -0.15, and the average standard deviation is 1.25.

It is not possible to generalize about how much of the dispersion in Z scores, as measured with the standard deviations, is due to genuine heterogeneity and how much is due to measurement error. However, it is clear from the three scatterplots that the dispersion tends to be greatest for the HAZ and least for the WAZ.

Dispersion is only slightly higher for the WHZ than for the WAZ. As listed above, the average standard deviations for the Z scores are 1.54, 1.18, and 1.25, respectively.

Figure 3.1 Scatterplot of the standard deviation and mean of the HAZ for children under age 5 in all DHS surveys since 2010

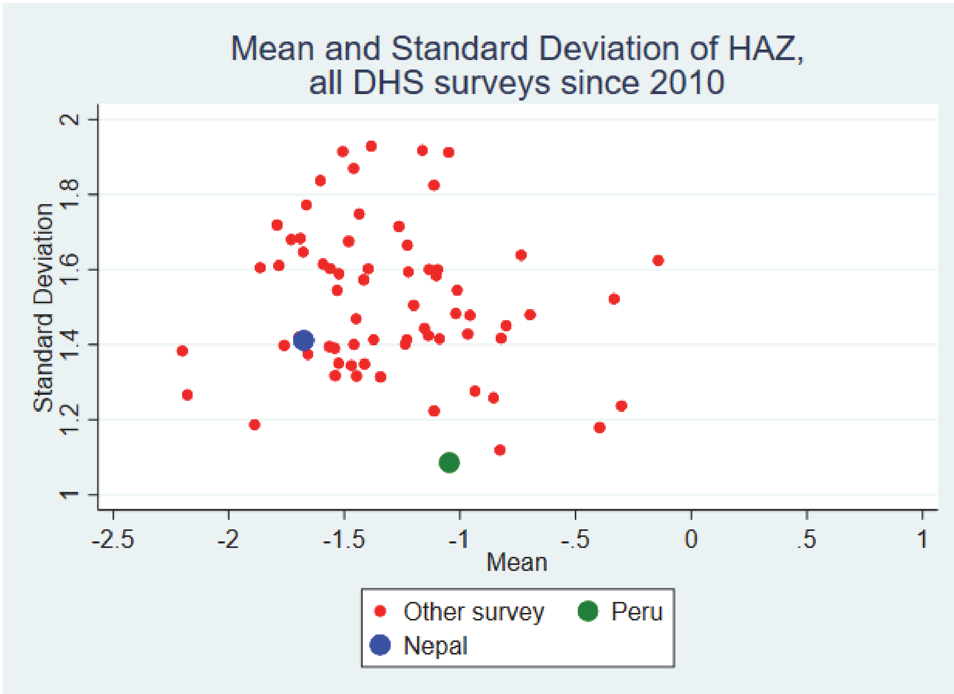


Figure 3.2 Scatterplot of the standard deviation and mean of the WAZ for children under age 5 in all DHS surveys since 2010

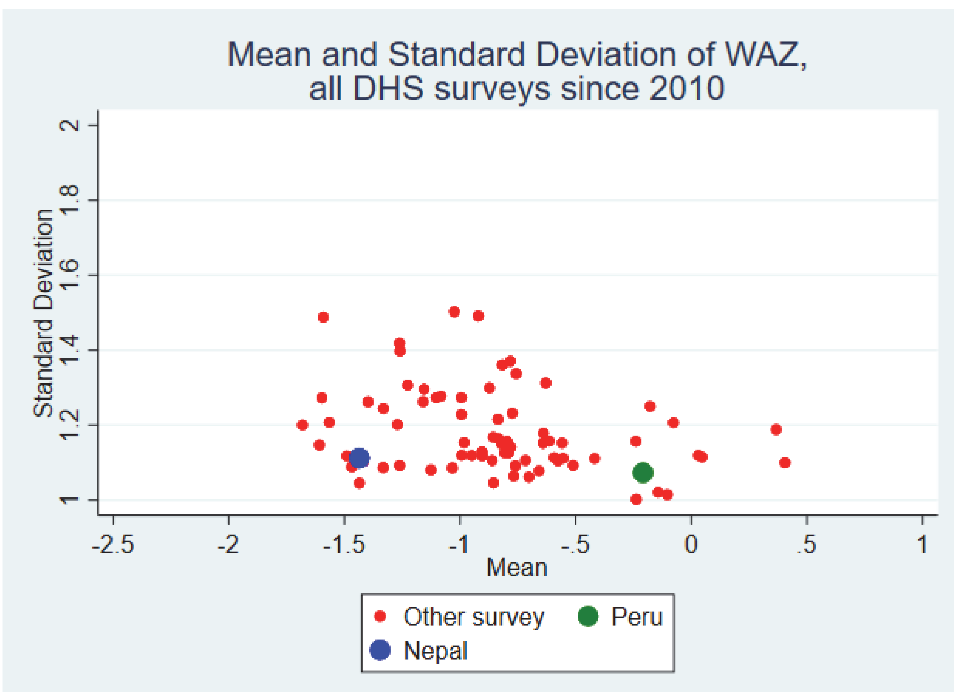
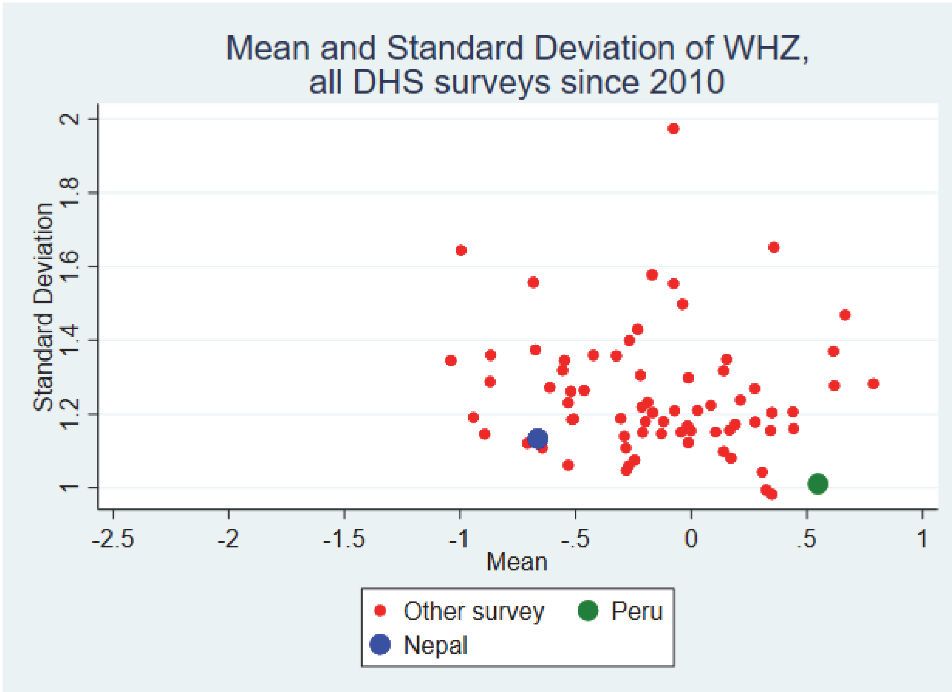


Figure 3.3 Scatterplot of the standard deviation and mean of the WHZ for children under age 5 in all DHS surveys since 2010



After comparing the range in the standard deviation of the HAZ with the ranges in standard deviation of the WAZ and WHZ (the vertical scales in the three figures), it is clear that dispersion in height causes more dispersion in Z scores than dispersion in weight. Height is the denominator of the WHZ, as well as the numerator of the HAZ, and the WHZ scores are not as widely dispersed as the HAZ scores. However, it was seen earlier, in Section 2.2, that a shift in height has a much greater effect on the HAZ than it has on the WHZ. It follows that dispersion in height, whether real or due to measurement error, will have a much greater effect on the HAZ than on the WHZ.

Similar observations could link the structure of the Z transformations discussed in Chapter 2 to the empirical evidence in these 74 surveys. However, the present goal is to provide an overview of the Z distributions in the surveys in order to position the Peru 2012 and Nepal 2016 surveys within that larger set.

3.2 The Peru 2012 and Nepal 2016 surveys

The means and standard deviations of the Z scores for the Peru 2012 and Nepal 2016 surveys are shown in Table 3.1. The mean Z scores for these surveys are well within the overall ranges, but the means for Peru are substantially higher than those for Nepal. In 2012, Peru continued to have a moderate level of stunting, and an increase in overweight, but no other serious nutritional challenges, while all the mean Z scores in Nepal 2016 were among the lowest of all surveys. In terms of dispersion, the standard deviations of Peru's Z scores were very low, and the standard deviation of the HAZ was the lowest observed among all 74 surveys. Nepal's Z scores had higher standard deviations than Peru's, but were lower than the great majority of surveys. Both surveys show little evidence of measurement error.

Table 3.1 Weighted mean and standard deviation of the HAZ, WAZ, and WHZ for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys

	Peru 2012 DHS		Nepal 2016 DHS	
	Mean	St. Dev.	Mean	St. Dev.
HAZ	-1.05	1.09	-1.53	1.35
WAZ	-0.21	1.07	-1.33	1.09
WHZ	0.55	1.01	-0.64	1.11

For children age 0-4 in the Peru survey, the observed mean height is 85.1 cm and the observed mean weight is 12.3 kg. The percentages of children who are stunted, underweight, overweight or wasted are 20.8%, 4.3%, 6.1%, and 0.7%, respectively. For children age 0-4 in the Nepal data, the observed mean height is 83.3 cm and the observed mean weight is 10.7 kg. The original percentages of children who are stunted, underweight, overweight or wasted are 36.2%, 27.0%, 1.3%, and 9.6%, respectively, in Nepal. All values are unweighted.

On average, children age 0-4 are about 1.8 cm taller and 1.6 kg heavier in Peru than in Nepal. Stunting is an issue in both countries, but is much more common in Nepal than in Peru. The profile for the remaining outcomes is much different in the two countries. In Nepal, 28% of children are underweight or overweight. (Underweight is calculated from the WAZ, and overweight from the WHZ, but both reflect extreme levels of weight.) Of the two conditions, far more children are underweight than overweight. In Peru, about 10% of children are underweight or overweight, and more are overweight than underweight. In Nepal, about 10% of children are wasted, but in Peru the prevalence is below 1%.

Figures 3.4 and 3.5 describe the shapes of the Z distributions in these two surveys with both histograms and the best-fitting normal distributions, which have the same means and standard deviations as the Z scores. Within each figure, the subfigure in the upper right shows the distribution of the HAZ and the one in the upper right shows the distribution of the WAZ. Both of the lower subfigures show the distribution of the WHZ. A red vertical line at -2 shows the boundary for stunted (using the HAZ), underweight (using the WAZ), and wasted (using the WHZ). A red vertical line at +2 shows the boundary for overweight (also using the WHZ). The figures include the levels of stunting, etc., calculated directly from the Z distributions, and from the best fitting normal distribution, and the difference between the two (delta=observed – fitted).

In Peru, as mentioned, stunting is the most problematic outcome. Beyond that, compared with a normative level of 2.3% for each outcome, there is some evidence of overweight children. The percentages stunted, underweight, overweight, or wasted in the 2012 survey were 18.2%, 3.4%, 7.1%, and 0.7%, respectively These percentages add to 29.3%. The corresponding fitted percentages were 19.0%, 4.8%, 7.5%, and 0.6%. The deviations or deltas were -0.8%, -1.3%, -0.5%, and 0.1%. The sum of the absolute values of the deviations, which is labeled as the “total delta” was 2.7%.

The low value of the total delta (2.7%), especially when compared with the sum of the problematic outcomes (29.3%), is a way of quantifying the nearness of the observed distributions to normal distributions, at least for calculating the outcomes. The HAZ, WAZ, and WHZ show similar differences from a normal shape—some excess of cases with values just below the mean and some deficits in the upper half, although these differences have little impact on the areas in the tails.

The Nepal data show a correspondence between the observed and fitted distributions that is at least as close as in the Peru data. The total delta is 2.3%, relative to 74.1% of children having at least one problematic outcome, which is primarily stunting (36.3%), underweight (27.3%), or wasting (11.1%). For these surveys, at least, the mean m and the standard deviation s of a Z score are sufficient information to produce close approximations to the prevalences of the problematic outcomes. In Stata, the following two commands will generate the estimate:

*scalar problem=100*normal((-2-m)/s)*

scalar list problem

These commands will produce the percent stunted, under the normal approximation, if m and s refer to the mean and standard deviation of the HAZ. If m and s refer to the WAZ, they will generate the percent underweight. If m and s refer to the WHZ, they will generate the percent wasted. If the first command is replaced with "100*(1-normal((2-m)/s)", and m and s refer to the WHZ, the commands will produce the percent overweight. If "2" is replaced by "3", we obtain estimates of the prevalences of "severe" outcomes.

Figure 3.4 Estimates of stunted, underweight, overweight, and wasted for children under age 5 in the Peru 2012 DHS survey, calculated from both the Z scores and the best-fitting normal distributions

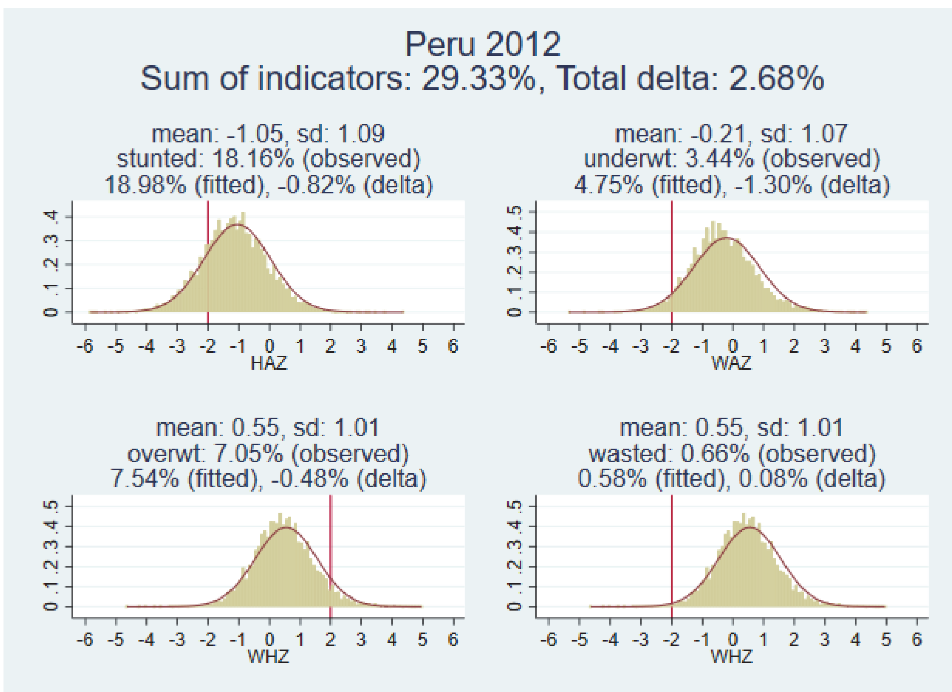
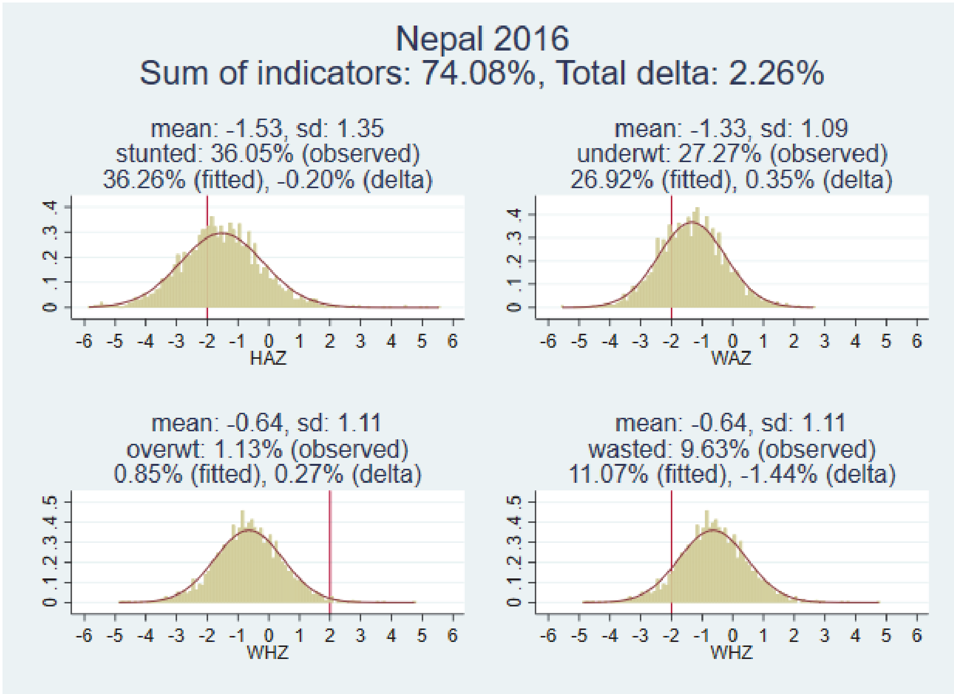


Figure 3.5 Estimates of stunted, underweight, overweight, and wasted for children under age 5 in the Nepal 2016 DHS survey, calculated from both the Z scores and the best-fitting normal distributions



The standard deviation of the HAZ in the Nepal data is 1.35, which is substantially larger than the standard deviation of the HAZ in the Peru data (1.09) or the standard deviation of the WAZ or WHZ within the Nepal data (1.09 and 1.11). Although 1.35 is within an acceptable range that extends to 1.4, a case could be made that this standard deviation is inflated by measurement error. Using a normal distribution, we can easily simulate what the percent stunted would be if the standard deviation of the HAZ were adjusted or "corrected" to 1.00. The Stata commands would change to

*scalar problem=100*normal((-2-m)/1)*

scalar list problem

The observed value of "s" would be replaced by "1". After forcing the standard deviation of the normal distribution to 1, the fitted estimate of stunting declines from 36.3% to 31.7%. If s=1 were taken as the true standard deviation of the HAZ, then the true level of stunting would be 31.7% and the observed 36.3% would be too large by an additive amount of 4.6%. The effect would be somewhat larger if the mean WAZ were about -1 rather than -1.53. As noted earlier, the effect of over-dispersion increases as the mean moves down from 0 to about -1, and then diminishes as the mean moves closer to -2.

The approximation to a normal distribution provides a simple tool to simulate the impact of over-dispersion. For example, the standard deviation of the HAZ in the Peru 2012 data could be artificially increased from the observed value 1.09 to an artificial value 1.4, which would induce an increase in the estimate of stunting, from the observed 18.2% to a simulated level of 24.8%. This is an additive increase of 6.6%. The mean HAZ (-1.05) is almost exactly at the value at which over-dispersion would have its maximum effect. Simulations based on the mean, standard deviation, and the normal approximation are macro simulations.

The anthropometric inputs and outputs for the Peru 2012 and Nepal 2016 surveys will be described in a set of six figures, three for measured height, weight, and age, and three for the HAZ, WAZ, and WHZ. These figures include both boys and girls. The Z values incorporate the adjustments for sex and if the child is lying or standing, where relevant. These figures will help to clarify the relationships among the three measurements, the three Z scores, and the four outcomes, as well as to describe the similarities and differences of the two illustrative surveys.

Figures 3.6-3.11 contain four histograms as subfigures. The upper two subfigures refer to Peru 2012 and the lower two to Nepal 2016. Visual comparisons of upper and lower pairs of figures help to clarify the differences between the two samples. The subfigure on the left in each row includes the sample weights and the subfigure on the right the unweighted. The results of simulations in Chapter 5 will be unweighted, with every case counted in the data equally. The weighted and unweighted distributions are shown side by side to convey their similarity. To estimate population characteristics, DHS samples should be weighted, but for sensitivity analyses, little is lost by ignoring the weights.

Figure 3.6 shows the distribution of height. The distribution is far from uniform, because height gain is most rapid in the first year or two of the age interval. The steep dropoff on the right side is partly due to the truncation of the age range at the fifth birthday. If anthropometry had extended to the tenth birthday, for example, the range of heights would extend farther to the right, but a similar dropoff and tail would be seen at the high end. The distribution of height would not be well approximated by a normal distribution and certainly not by a uniform distribution.

The subfigures for Nepal 2016 in Figure 3.6 are much more erratic than those for Peru 2012. The overall shape of the height distribution for Nepal 2016 is fuzzier, less well defined, with many spikes and valleys. We will not show indices of digit preference, but it is clear from the figures that the measurement of height is better in the Peru 2012 survey than in the Nepal 2016 survey.

Figure 3.7 shows the distribution of weight in the two surveys. Although these distributions are relatively symmetric and could be approximated by normal distributions, the steep dropoff on the right is again related to the truncation of age at the fifth birthday. If the age range were increased, the distribution would move to the right—but there would still be a dropoff and a tail on the right.

The distribution of age, in days, is shown in Figure 3.8. On the x axis, the number of days at the first, second, third, fourth, and fifth birthdays are identified. (Note: There is trivial uncertainty related to leap years.) The age distribution is very close to uniform. In some surveys, there are significant departures from uniformity because of population growth, seasonality of births, under-5 mortality, or displacement across the fifth birthday, but these are difficult to detect in a figure.

Figures 3.9, 3.10, and 3.11 show histograms for the distributions of the HAZ, WAZ, and WHZ scores, respectively. Each subfigure includes a best-fitting normal distribution—that is, the normal curve that matches the mean and standard deviation of the Z scores. The vertical axis shows the percentages for the bars in the histograms. All subfigures have the same vertical scale. The horizontal axis spans the WHO plausible range for the respective Z score. A vertical red line at Z=0 makes it easier to position the distribution relative to the normative well-nourished reference population, for which the distribution would be centered at 0.

Each figure also identifies the Z values for problematic nutritional outcomes. In Figure 3.9, the area to the left of HAZ=-2, for stunting, is highlighted in blue. In Figure 3.10, the area to the left of WAZ=-2, for underweight, is highlighted in red. In Figure 3.11, for the WHZ, there are two problematic areas. The WHZ scores to the right of +2 are shaded in green and represent overweight children. The WHZ scores to the left of WHZ=-2 are highlighted in black and represent wasted children.

These figures convey a great deal of information about the two surveys. There is no visible difference between the weighted and unweighted distributions of the Z scores. All the Z distributions are very close to normal, although the approximation is closer for Peru 2012 than for Nepal 2016. Corresponding with the means in Table 3.1, all the Z distributions are centered to the left of Z=0, except for the distribution of the WHZ in Peru 2012, which has a mean WHZ greater than 0. The distribution of the HAZ in Figure 3.9 is visibly wider in Nepal 2016 than in Peru 2012, which corresponds to the higher standard deviation shown in Table 3.1.

The relative roles of the mean and standard deviation of a normal distribution in determining the area of the tails were reviewed in Chapter 2, where the focus was on the HAZ and stunting. Here we see how the shapes of the three Z scores connect with the all four outcomes. In Figure 3.11, the high mean WHZ in the Peru 2012 survey is the primary reason why overweight (the green area) is an emerging issue, but there is virtually no stunting. Conversely, the low mean WHZ in the Nepal 2016 survey is the primary reason why wasting (the black area) is a serious problem although there is virtually no overweight. The low dispersion of the WHZ in both surveys translates to confidence in inferences based on both tails of the WHZ distribution.

Figure 3.6 The distribution of height, in cm, for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted

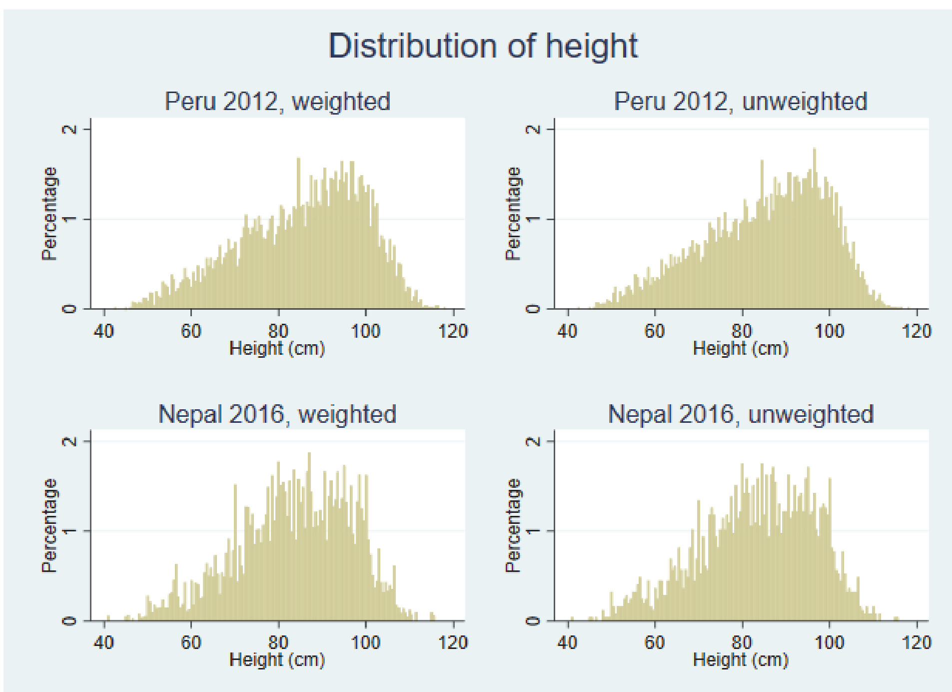


Figure 3.7 The distribution of weight, in kg, for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted

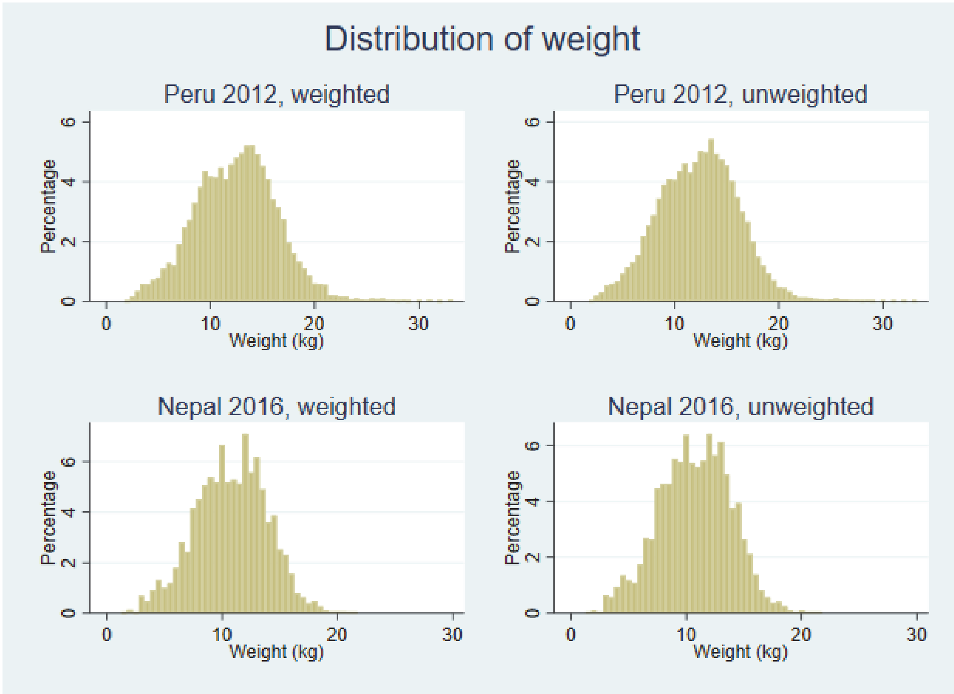


Figure 3.8 The distribution of age, in days, for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted

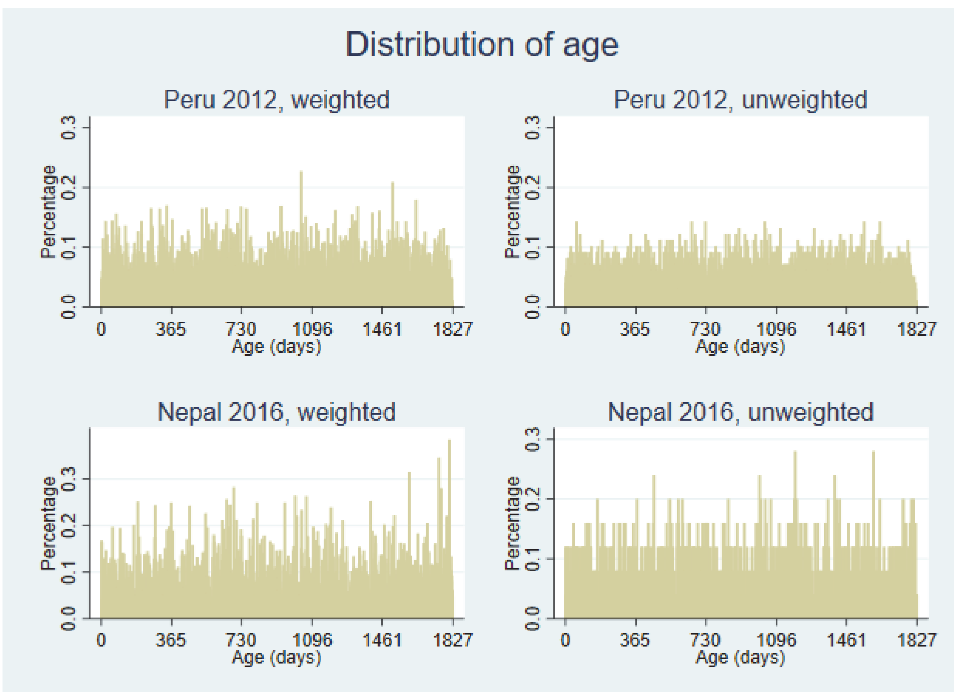


Figure 3.9 The distribution of the HAZ for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted

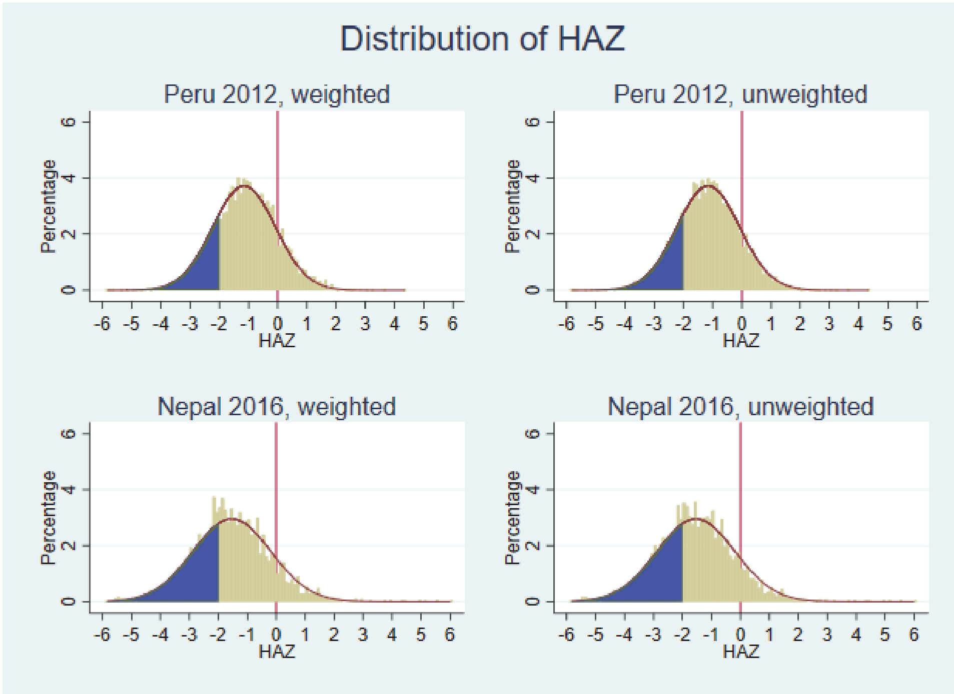


Figure 3.10 The distribution of the WAZ for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted

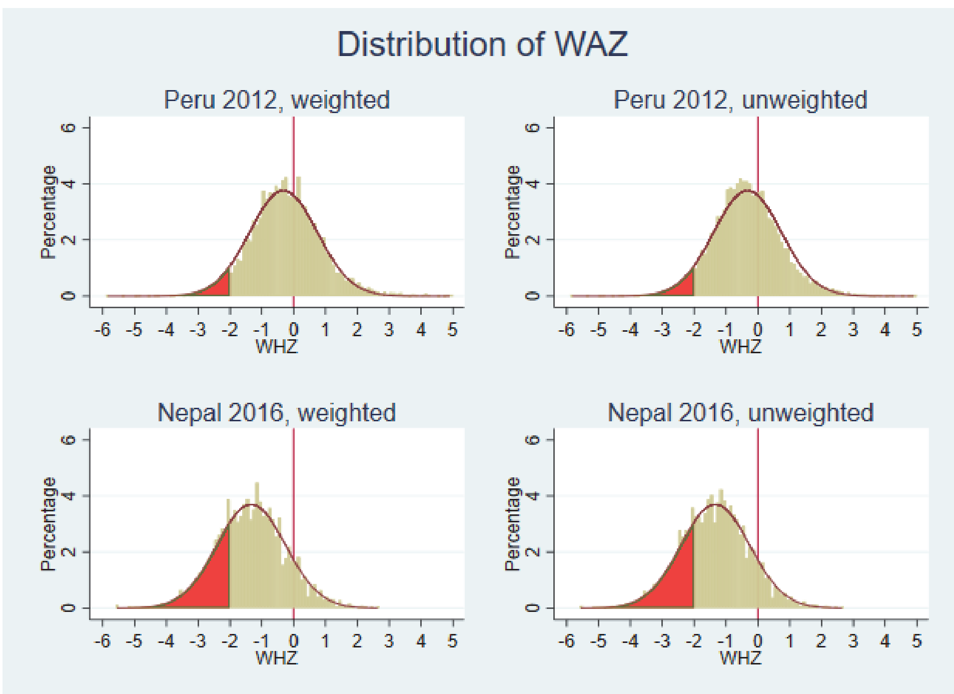
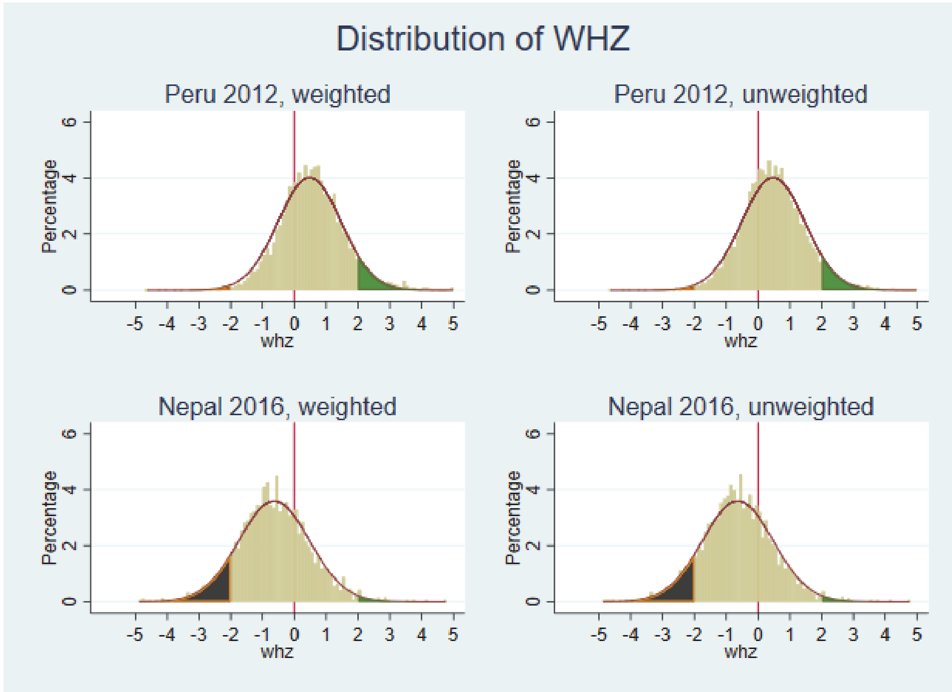


Figure 3.11 The distribution of the WHZ for children age 0-4 in the Peru 2012 and Nepal 2016 surveys, weighted and unweighted



Later in this report, the Peru 2012 and Nepal 2016 surveys will be used to illustrate the potential consequences of disturbances to the measurements of height, weight, and age for the HAZ, WAZ, and WHZ, and the estimates of stunted, underweight, overweight, and wasted. We will treat the observed height, weight, and age in these surveys as baselines, and then add simulated error. We will simulate disturbances to the original individual-level data and will not use the normal approximation to the overall distribution. In those micro (rather than macro) simulations, the sampling weights will be ignored. Individual scores will be displaced probabilistically, and the interpretation will be easier if every case counts equally. The goal is to understand the consequences of measurement error and the degree to which specified types and amounts of error may affect anthropometric outcomes.

4 THE POTENTIAL IMPACT OF POPULATION HETEROGENEITY

The WHO growth standards were based on a population of well-nourished children whose data came from multiple sources. Their joint distribution of height, weight, and age was transformed into the HAZ, WAZ, and WHZ, with normal distributions and mean 0 and standard deviation 1.

The application of this model to real populations raises two questions related to the normative standard deviation of 1. These questions are relevant because the observed standard deviations are the result of true variation, combined with variation induced by measurement error. Over-dispersion due to measurement error is clearly more difficult to assess if we do not know what the dispersion would be in the absence of such error.

The first question focuses on the application of the model to populations that are not well nourished. In nearly all DHS surveys, the mean Z scores are negative, often below -1. If the observed mean of a Z score is substantially different from 0, can it safely be assumed that the correctly measured Z scores will be normal and have a standard deviation of 1? Can the mean shift but the standard deviation and shape of the distribution remain unchanged?

We consider this to be a reasonable question but follow the general practice of accepting that, in a homogeneous population, the true standard deviation is close to 1 and the distribution approximately normal, regardless of the mean.

The second question concerns the application of the model to heterogeneous populations. In virtually all DHS surveys conducted in a generally malnourished population, the *degree* of malnutrition is not consistent within the population. The DHS reports show that in virtually every country, the mean Z scores vary across regions, wealth quintiles, and levels of maternal education. The national population is at best a blending or mixture of subpopulations with different means that are internally homogeneous. Apart from possible measurement error, what is the relationship between the dispersion of Z scores within the component subpopulations and their dispersion within the combined population? This question is easier to answer because it can be approached mathematically.

Almost by definition, Z scores will be more dispersed in a heterogeneous population than in a homogeneous population. If there is heterogeneity in mean Z scores, such as across geographic subdivisions or wealth quintiles, then the amount of dispersion will be greater in the combined population than in the subpopulations. This pattern implies that the criterion for dispersion cannot be the same for the combined population as for the relatively homogeneous subpopulations.

Because the HAZ, WAZ, and WHZ are constructed to mimic normal distributions, the impact of heterogeneity on dispersion can be approached with a generic normally distributed variable z that has a different mean in each subpopulation but has a standard deviation of 1 (and variance of 1) in each subpopulation. Many different combinations are possible, as shown with the following examples.

First, suppose we have two subpopulations that are equal in size but different in their nutritional status. In group 1, the mean of z is μ_1 and in group 2 it is μ_2. In both groups, the standard deviation (and variance) is exactly 1. In the combined population, the variance of z will be $1 + \left(\frac{\mu_1 - \mu_2}{2}\right)^2$. For example, if the difference between the means is 1, the overall variance will be 1.25, and the standard deviation sqrt(1.25)=1.12. If the difference is 2, the overall variance will be 2 and the standard deviation sqrt(2)=1.41. The formula for the variance in this special case is easily derived mathematically, and can be confirmed by simulating z scores for subpopulations.

More generally, we can assume that there are K subpopulations—for example, regions—of equal size. Let μ_k and σ_k be the true mean and standard deviation, respectively, in subpopulation k (k=1,…,K). Let μ and σ be the true mean and standard deviation, respectively, in the combined population. Because (by assumption) the subpopulations are equal in size, the mean for the combined population will be the arithmetic mean of the subpopulation means: $\mu = \overline{\mu_k}$. The variance in the combined population can be shown to be $\sigma^2 = \overline{\sigma_k^2} + \delta$, or the arithmetic mean of the variances in the subpopulations, plus a quantity δ, where $\delta = (1/K^2)\sum_{i<j}\left(\mu_i - \mu_j\right)^2$, implying that $\delta \geq 0$. If the subpopulations are not equal in size, the only modification will be the inclusion of weights for the sizes of the subpopulations. If the standard deviation in each subpopulation was exactly 1, then the standard deviation of the pooled population would be the square root of 1+ δ, which must be greater than 1. This formula does not require an assumption of normality.

Table 4.1 Observed means and standard deviations of the HAZ in the Nepal 2016 DHS, by region

Region	Mean	Std. Dev.	Freq.
Eastern Mountain	-1.7309	1.2957	152
Central Mountain	-1.8688	1.1740	113
Western Mountain	-2.2650	1.3458	208
Eastern Hill	-1.7513	1.3040	209
Central Hill	-1.3448	1.3519	157
Western Hill	-1.4092	1.3474	195
Mid-Western Hill	-2.0284	1.5273	227
Far-Western Hill	-2.0782	1.3416	223
Eastern Terai	-1.3766	1.2902	211
Central Terai	-1.5909	1.5443	244
Western Terai	-1.6511	1.4087	172
Mid-Western Terai	-1.6702	1.3640	219
Far-Western Terai	-1.2202	1.2216	143
All Nepal	-1.7081	1.3946	2,473

As an example, Table 4.1 shows the means and standard deviations of the HAZ for children under age 5 in the Nepal 2016 DHS, for all Nepal and 13 regions. All means are negative, but they vary from a low of -2.2650, an extremely low value in the Western Mountain region, to a high of -1.2202 in the Far-Western Terai. The standard deviations vary from a low of 1.1740 in Central Mountain to a high of 1.5443 in Central Terai. All observed standard deviations are considerably greater than 1.

Table 4.1 includes the unweighted number of children with valid HAZ scores in each region. There is variation in sample sizes, but in this simplified example, we assume that the number of cases is the same in each region. The average variance, obtained by squaring the observed standard deviations, adding, and dividing by 13, is 1.8255. The square root of this, 1.3511, would be the pooled standard deviation if the means were all the same. Calculating δ with the above formula gives δ=0.0894. Therefore, taking into account the heterogeneity in means, the pooled standard deviation is the square root of 1.8255 + 0.0894 = 1.3838. This is very close to the pooled standard deviation of the HAZ, 1.3946, given in Table 4.1, even though we ignored the inequality in the size of the groups. In this example, heterogeneity in means across regions has increased the standard deviation by approximately 0.03.

If the standard deviations in the regions had all been 1, then the pooled standard deviation would be the square root of 1+0.0894, which is 1.0437. The increase in the standard deviation would be approximately 0.04. In any case, the pooled standard deviation would necessarily be greater than 1.

Of course, within each region of Nepal, there are administrative subdivisions, and within those are urban and rural areas. If there is heterogeneity in the underlying nutritional status of the various levels of aggregation, down to some level at which the standard deviation could be assumed to be 1, then there will be increasing dispersion as each level is aggregated up to the next level. The fact that each region has a standard deviation greater than 1 is at least partly the result of heterogeneity within each region. It cannot safely be inferred that measurement error is the only, or even the main, reason for a standard deviation of a Z score greater than 1.

We have also examined heterogeneity across interview teams in the Nepal survey. There were only 16 teams in this survey. Their geographic assignments were generally within four or five regions. There was much less variability in the means for the teams than in the means for the regions. If the HAZ is regressed on region ID as a categorical variable, the adjusted (and unweighted) R-squared implies that 4.1% of the total variation is due to variation in the region means. If the HAZ is regressed on interview team ID as a categorical variable, the adjusted R-squared implies that only 1.4% of the total variation is due to variation in the team means. That is, team means are much more similar than region means. Thus, heterogeneity across teams has a much smaller effect on overall dispersion than heterogeneity across regions.

Another example is based on variation in the HAZ in the same survey. The overall mean HAZ rounds to -1.71. The mean unweighted HAZ scores in the five wealth quintiles are -2.13, -1.91, -1.60, -1.33, and -1.09, in sequence from the lowest to the highest wealth quintile. There is slightly more than one full point of difference between the lowest and highest quintiles. There is a strong, monotonic gradient of improvement in height-for-age, although even in the highest quintile, the mean is negative. The standard deviations in the five quintiles are 1.42, 1.34, 1.34, 1.28, and 1.25, respectively. The overall standard deviation is 1.39, which is larger than the standard deviation of all but one of the separate quintiles.

The number of children declines sharply in the higher quintiles because there is a negative relationship between wealth and fertility. For simplification, we give equal weight to each quintile. If there were no variation in the means, we would expect the overall standard deviation of the HAZ to be the square root of the average variance in the five subgroups, which is sqrt(1.7617) = 1.3273. Because of the dispersion in means, we have δ = 0.1433, which inflates the standard deviation to sqrt(1.7617+.1433) = 1.3802, which is close to the observed standard deviation, 1.3946. That is, variation in the mean HAZ across wealth quintiles has induced an increase of about 0.05 in the overall standard deviation. If the within-quintile standard deviations were all exactly 1, the pooled standard deviation would be sqrt(1.1433)=1.0700, and heterogeneity by wealth quintile, alone, would have increased the pooled standard deviation by 0.07.

Other examples could be developed that involve working with wealth quintiles nested in regions. However, it is clear that heterogeneity across subpopulations will increase the standard deviations of Z scores in the pooled population.

5 THE POTENTIAL IMPACT OF ERRORS IN THE MEASUREMENT OF HEIGHT, WEIGHT, OR AGE

5.1 Systematic errors and random errors

Potential measurement errors are very different for height, weight, and age, and can be classified as either systematic and primarily introducing bias, or random and primarily introducing over-dispersion. Systematic error is directional and random error is bidirectional. Most types of errors probably incorporate both bias and randomness, but the distinction is helpful.

For example, if a child is wearing more than minimal clothing, her weight may be overestimated. This is more likely in cold weather. As described earlier, an increase in the child's weight, whether real or spurious, will increase the WAZ and WHZ scores, and may reduce the estimates of underweight and wasting and increase the estimate of overweight. Roche et al. (2015), in a study of Ecuadorean children age 0-23 months, estimated the weight of clothing to range from 120 g to 789 g, with a mean of 331 g, which is about one-third of a kilogram. The ranges and means increased gradually with age within the interval. Tuan et al. (2003), who obtained similar estimates in a study of Vietnamese children age 6-42 months, found that the weight of clothing varied according to temperature. The mean weight declined from 357 g when the outdoor temperature was 6-10 ^{o}C to 82 g when it was 30-34 ^{o}C.

In another example, it is possible that the arrangement of hair on top of the head leads to an overestimate of height, which will increase the HAZ score, decrease the WHZ score, and potentially reduce the estimates of stunting and overweight and increase the estimate of wasting.

Systematic or directional error in the measurement of height or weight leads to a measurement that is too high. Usually this error has a random component, because it occurs in varying degrees. For example, if extra clothing is an issue, the *amount* of additional clothing is probably not the same for all children, even after adjusting for sex and age.

Age in days is simply the difference between the date of measurement and the date of birth. Ideally, the child's date of birth is known and provided, and the calculation of age is simple, with much less uncertainty than the measurement of either weight or height. If a birthdate cannot be provided immediately, DHS surveys use probing and a calendar with reference dates, such as recent election dates, to improve the estimate of the birthdate.

There are systematic patterns of age displacement that can lead to upward bias and other patterns that can lead to downward bias. Whether positive or negative, the bias is restricted near the boundaries by the limitation to ages in the range 0-1827 days. Age bias has no effect on the WHZ or estimates of overweight or wasting. If the age in the survey is too high, then the HAZ and WAZ will be biased downwards and estimates of stunting and underweight will be biased upwards. If age is biased downwards, the estimates of stunting and underweight will be biased downwards.

In many surveys, there is evidence of upward displacement at the upper end of the age 0-4 interval, because some interviewers have a tendency to move some birthdates across the date of eligibility for the child health questions (usually January 1 of the fifth calendar year before the beginning of fieldwork) (Pullum 2006).

Some children who are actually age 4 are reported by the interviewer to be age 5 in order to avoid the child health questions. Other adjacent ages are probably affected, and not just age 4 and 5.

There are special issues in simulating the consequences for anthropometry of this type of displacement. If a child's recorded age were artificially increased from 4 to 5, for example, then the child would be out of range for anthropometry and the Z scores. The case would be lost and the only effect would be to thin out the sample for age 4. The distribution of Z scores for the remaining children would be unaffected. Whatever the motivation, any simulation of displacement across the fifth birthday translates into a lost case, rather than a modified HAZ or WAZ score.

There are other issues when simulating a *downward* displacement across the fifth birthday, for example, shifting a child from age 5 to age 4. The birthdate of a 5-year-old child is only known if the child also appears in the birth history of a woman in the household. Even for those children, we do not have a height or weight measurement, so we cannot calculate the HAZ or WAZ.

Sometimes the birthdate is derived from an estimate of current age, which can also lead to errors. As part of the household listing, which precedes the selection of children for anthropometry, all household members are listed and everyone's age is recorded in years. This is a reversal of the ideal sequence, in which birthdate would be provided, and age would then be calculated during data processing.

In several DHS surveys, children tend to be reported with a later birth year and an earlier birth month than would be correct (Pullum 2019, Chapter 2). This displacement is probabilistic, and can only be described in the aggregate, rather than for individual children. The effect can be traced to a tendency for year of birth to be calculated, by either the respondent or the interviewer, as "year of interview minus years of age," which is correct if the child has already had her birthday in the calendar year of the interview, but incorrect if she has not. The shift tends to displace the date of birth toward the date of interview, slightly biasing age downwards in months or days, but staying within the same year of age. The probability of the shift depends primarily on the timing of the interview within the calendar year. When it is observed, this pattern appears to affect the calculation of birthdate for children and adults at any age. Implications of this pattern of displacement have been described by Larsen, Headey, and Masters (2019) and Finaret and Masters (2019).

In some settings, even in the absence of child health questions, there is evidence that children are systematically reported as older than they actually are, even as infants. For examples, see Shah, Pullum and Irfan (1986), Pullum (1990), or Pullum and Stokes (1997) for analysis of this pattern in Pakistan.

This report will not attempt to quantify the effect of systematic error on the Z distributions and the percentages of children who are stunted, etc. There are many potential scenarios, which could be indexed by the amount of bias and the probability distribution of the bias if it is in a range. The analysis of displacement is deferred to another report.

5.2 Assumptions and model for random errors in the measurement of height, weight, and age

Another potentially important type of measurement error is random error that is bidirectional and may introduce over-dispersion, but not bias. Likely sources of such errors are carelessness, due to poor training or poor supervision, and/or working too quickly. Rounding or other kinds of digit preference will introduce errors that are probably equally likely to be too high or too low. In recent surveys, these kinds of errors are probably more likely with the measurement of height than weight, because weight is measured with a digital scale and height (or length) depends on position and posture. However, even the weight measurement requires copying a number from the scale. Most of the recent surveys still use a paper questionnaire for the biomarkers, and only a few have used tablets. There is a possibility of transcription error.

The literature on anthropometry includes analysis of repeated measurements of the same child (Ulijaszek and Kerr 1999; WHO Multicentre Growth Reference Study Group 2006b). The technical error of measurement (TEM) for height can be estimated when two measurers record independent assessments of the heights of a group of children. If d is the difference between the two measurements of n children, calculated for each child, then the TEM is defined as the square root of the sum of the squared deviations divided by 2n; that is,

$$TEM = \sqrt{\frac{\sum d^2}{2n}} \qquad (5.1)$$

The value of the TEM for height varies considerably and is somewhat different for children who are lying or standing. The optimal or minimal value is as low as 0.25 cm. The WHO-UNICEF (2019) guidance has established acceptable TEM values for use during training which are 0.6 cm for precision and 0.8 cm for accuracy. Estimates from field studies range from 0.3 to 0.9 cm for children who are standing and 0.4 to 1.2 for children who are lying. The lower bound is the average inter-observer TEM from the WHO Growth reference study sites, and the upper bound is three times that of the WHO Growth Reference Study Sites.

The TEM is an estimate of the standard deviation of random measurement error that has a mean of zero. That is, if the observed height is equal to the true height, plus an error term e with mean 0 and standard deviation σ, then the TEM is an estimate of σ. It is plausible to assume that the error term e is normally distributed, in which case the TEM is a maximum likelihood estimate of σ, if there is no bias in height from extra hair, for example. We can refer to this unmeasured error as *baseline error*, because in an actual survey we do not know the TEM, which can be calculated during training and standardization at the beginning of the survey. During fieldwork, there is normally only one measurer on the interview team and only one measurement of height is taken. However, for subsamples, DHS may estimate the TEM during fieldwork by having an anthropometrist measure the same child twice. The measurer is blinded as to whether or not they are doing a remeasurement. In other contexts (outside of DHS), the comparison is between the supervisor and the measurer for a subsample of children.

Greatest use of the TEM has been related to the measurement of height, although the concept also extends to weight and age. For weight, since recent surveys use scales with a digital readout, measurement error should not arise. However, there are small variations in a child's weight even within a day that depend on factors such as liquid intake, and repeated measurements at different times of the day might have different results.

Certain other types of errors occur occasionally, and effectively at random, but they are difficult to include in a model. For example, transcription errors can occur, although more frequently in older surveys than current ones. Examples include reversing the height and weight measurements, reversing digits, dropping a leading digit, or dropping a final digit. We analyzed a non-DHS survey in which nearly all children in a cluster were recorded with the same weight because the digital scale was stuck at that weight. Such problems may produce a Z score that is out of range and will be excluded from all calculations. Some will be in range but extreme, which potentially inflates the estimates of problematic outcomes. Some may also move a child from a true extreme value to a more central value, which potentially reduces the estimates of problematic outcomes. We will not attempt to simulate these kinds of errors. They are virtually impossible to identify unless they are out of range, and we have no basis for attempting to replicate them.

Since height and weight are not affected by specific limits, it is a plausible hypothesis that the magnitude of the error does not depend on the level. We will assume that the magnitude of error for height does not depend on the child's height and that the magnitude of error in weight does not depend on the child's weight. There is evidence that children who are lying down are more difficult to measure than children who are standing (Assaf, Kothari, and Pullum 2015), but we will not make that distinction in the simulations. Measurement error, adding to what is already present in the baseline data, will be introduced by adding a disturbance that is normally distributed with mean zero and a standard deviation that is independent of the observed height or weight. The standard deviation of the errors will be specified, at levels that range from a minimum of zero to a maximum of 5 cm for height and 2 kg for weight.

As stated earlier, measurement error for a child's age is more difficult to model. Some of the complexities are described briefly here. There are boundary effects for age that do not apply to height or weight. Obviously, the birthdate of the child cannot be in the future, and the child cannot have a negative age. A plausible model for bidirectional misreporting must be constrained to non-negative ages. Our simulation procedure will prevent negative values in the following way. Suppose that the reported age of a child is d days. In a specific simulation, the randomly generated error or disturbance can be called e. The child's simulated age will be d+e. If e is negative, and relatively large in magnitude, then d+e could be less than 0, a possibility that must be avoided. The procedure replaces e with 0 if this happens. However, in order to avoid displacing the mean of the simulated values upwards, compared with the observed values, the procedure will also drop a simulated value if e is positive but large enough in magnitude that *if it were negative*, then d+e would be negative. That is, if |e|>d, then e will be replaced with 0 and the original d will be maintained. This algorithm results in fewer imputed errors in the youngest ages, and no systematic introduction of bias. The procedure has a secondary effect of reducing the probability of any displacement at the youngest ages. Otherwise, we do not build in an assumption, although it could be plausible that the amount of error in age tends to increase as true age increases. Implementation of such a pattern would require arbitrary assumptions for which we have no empirical basis.

The truncation of age at the fifth birthday produces a ceiling effect that was discussed in relation to upward bias in the previous section. Our simulations of reporting error can move children who were reported at age 58 or 59 months, for example, to ages 60 or 61 months, which would move them completely out of age 0-4. We cannot balance that by moving children who were reported at 60 or 61 months down to 58 or 59 months, because children originally reported at 60 or 61 months were not measured for height and weight. As a result, induced errors near the upper limit will have the unavoidable effect of reducing the sample size and slightly reducing the mean age of the sample.

Define M to be the observed, or measured, value of an anthropometric characteristic and T to be the true value. Use subscripts 1, 2, or 3 for height, weight, or age, respectively.

Define Z to be the calculated, value of the HAZ, WAZ, or WHZ, and ZT to be the true value, which would be calculated if its components were measured accurately. Use subscripts 1, 2, or 3 for the HAZ, WAZ, or WHZ, respectively.

We will assume that measurement errors for height, weight, and age are additive, normally distributed, and have a mean of zero. That is,

$$M_1 = T_1 + a_1$$

$$M_2 = T_2 + a_2 \qquad (5.2)$$

$$M_3 = T_3 + a_3$$

Here each a is normal with mean 0 and has its own standard deviation s or variance s^2. We also assume that the error terms are independent across cases and across types of measurement. The true Z scores are functionally related to the true values of height, weight, and age as follows:

$$ZT_1 = Z_1(T_1, T_3)$$

$$ZT_2 = Z_2(T_2, T_3) \qquad (5.3)$$

$$ZT_3 = Z_3(T_2, T_1)$$

When errors are introduced to the measurements, the calculated Z scores become the following:

$$Z_1 = Z_1(M_1, M_3)$$

$$Z_2 = Z_2(M_2, M_3) \qquad (5.4)$$

$$Z_3 = Z_3(M_2, M_1)$$

The relationship between the calculated Z scores and the true Z scores is mathematically complex, but need not be made explicit because our focus will be narrowly on how measurement error affects the dispersion (the variance or the standard deviation) of each Z score. We hypothesize that over-dispersion of the Z scores is a consequence of over-dispersion of the measurements, and arises from the addition of random and normally distributed error to the measurements. Specifically,

$$Var(Z_1) = Var(ZT_1) + f_1(s_1, s_3)$$

$$Var(Z_2) = Var(ZT_2) + f_2(s_2, s_3) \qquad (5.5)$$

$$Var(Z_3) = Var(ZT_3) + f_3(s_2, s_1)$$

Here s_1, s_2, and s_3 are the standard deviations of the error distributions of height, weight, and age, respectively. At this point it is not necessary to specify the functional forms f_1, f_2, or f_3.

As described above, we know that some degree of measurement error is inherent in the data. The standard deviation of baseline error would be estimated with the TEM, but is not known. We simulate additional random error, which by design is normally distributed with mean 0 and a specified standard deviation s. This additional simulated error is independent of the baseline error. Therefore, the errors are additive. About half of the time, the errors partially cancel each other out and about half of the time they combine to produce a greater deviation from the mean. The variance of the combined error is the variance of the baseline error plus the variance of the simulated error, $\text{TEM}^2 + s^2$, and the standard deviation of the combined error is $\sqrt{TEM^2 + s^2}$. The simulation of different values of s is essentially a simulation of hypothetical increases in baseline error, a larger TEM. The new or simulated TEM is $\sqrt{TEM^2 + s^2}$ and the simulated increase in the TEM is $\sqrt{TEM^2 + s^2} - TEM$.

Figure 5.1 Values of the new or simulated TEM that are produced by combinations of the true TEM and the standard deviation (s) of additional random error

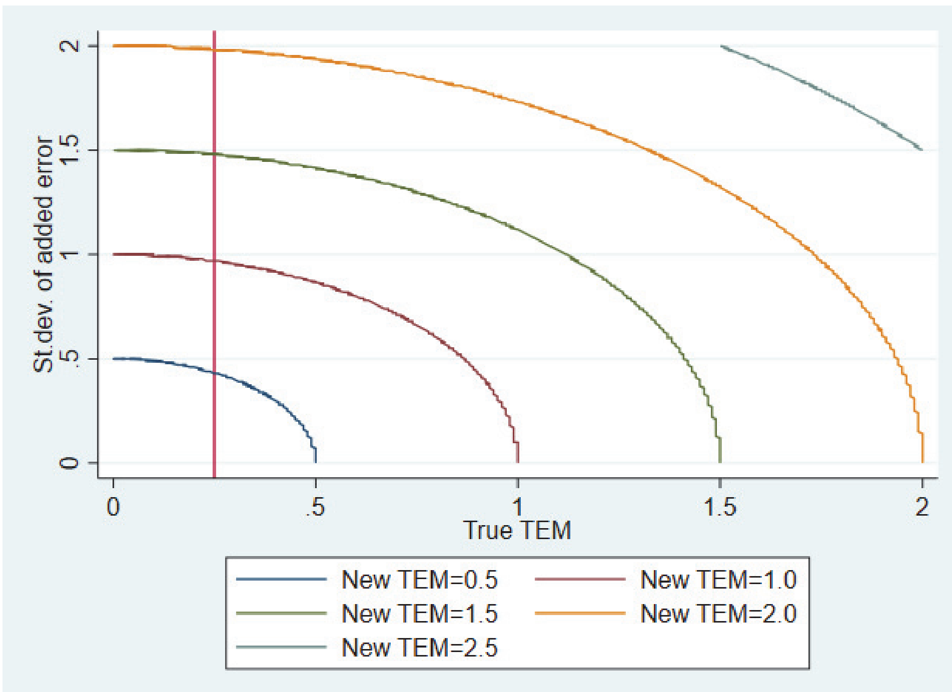


Figure 5.1 shows how the TEM, s, and $\sqrt{TEM^2 + s^2}$ are related. The horizontal axis describes a potential range for the true value of the TEM, for a wide range between 0 and 2 cm, but with a vertical line at TEM=0.25 to mark the minimal value. Values to the left of the vertical line are implausibly low. High values of the TEM, on the right side of the figure, are implausibly high for a survey in which the measurers are well-trained and supervised. The vertical axis is s, the standard deviation of the simulated additional error, which is in a wide range from 0 to 2 cm. Convex curved lines show the combinations of TEM and s for which the simulated TEM is 0.5, 1.0, 1.5, 2.0, or 2.5. In the figure, the vertical scale is more compressed than the horizontal scale. If the two scales were the same, the lines would be arcs of circles centered on (0,0) and with radius 0.5, etc. The true TEM and the simulated s contribute equally to the simulated TEM.

If the value of the TEM is known, or a good estimate is available, then it is possible to calculate the value of the simulated TEM for a given value of s. The results in this chapter are indexed by the value of s, but if we knew the actual value of the TEM for a survey, the results could be indexed with a new or simulated TEM.

5.3 Simulating the impact of random errors of measurement on the dispersion of the Z scores

The simulation procedure is described in more detail in Appendix 1, but can be summarized briefly as follows. Steps 1 and 2 are repeated with each simulation and Step 3 compiles the results. The cases are limited to children with *de facto* residence in the household, as in the standard tabulations.

Step 1. The original data file (children age 0-4 in the household data) is opened. The measured height, weight, or age of the child is disturbed by the addition of an error term. The error term is generated to have a normal distribution with mean 0 and standard deviation σ, where σ is specified and gradually increases from one simulation to the next. The error terms are rounded (to 0.1 cm or 0.1 kg or 1 day) before they are added to the observed height, weight, or age, because they are the units of measurement in the data. The standard deviation of the error term, s, is calculated for each simulation and saved. (s is used in place of the theoretical standard deviation that was used to generate the distribution, σ.)

Step 2. Each affected Z score is recalculated for each child. For example, if height has been disturbed, the HAZ is recalculated with the original measurement of age and the disturbed measurement of height. The WHZ is recalculated with the original measurement of weight and the disturbed measurement of height. The mean and standard deviation of the HAZ and WHZ are recalculated and saved. The percentages of stunted, overweight, and wasted are recalculated and saved. Similar calculations are made if weight or age have been disturbed.

Step 3. After repeated loops through Steps 1 and 2, a new file is constructed with a line that retains the results for each loop or simulation. The line contains the standard deviation of the errors, the recalculated means and standard deviations of the relevant Z scores, and the recalculated levels of the outcomes. The file summarizes the simulations and is the basis for establishing correspondences between the dispersion of the simulated errors and the additional dispersion of the Z scores.

Nearly all DHS surveys include some degree of error in the measurement of height, weight, and age. In effect, the simulations have added *more* error, of a specific type—additive random disturbances. As stated earlier, we assume that this type of measurement error is normally distributed with a mean of zero, and that the observed Z scores include some amount of over-dispersion that can be attributed to such measurement error. By identifying the relationship between induced measurement error (in height, weight, and age) and the degree of additional over-dispersion (in the HAZ, WAZ, and WHZ) that it produces, we hope to gain leverage on the amount of measurement error in the original measurements of height, weight, and age.

For illustrative purposes, we apply the procedure to the Peru 2012 and Nepal 2016 DHS surveys, which were introduced in Chapter 3. Any DHS survey could be used, but these surveys appear to be minimally distorted by measurement error and their nutritional outcomes are quite different. The 2012 round of the Peru Continuous Survey, in particular, had relatively little evidence of measurement error. Rankings of surveys, including these two, in terms of data quality, can be found in Perumal et al. (2020).

The specific data files are PEPR6IFL.dta and NPPR7HFL.dta. We use all children age 0-59 months with *de facto* residence in the household. All the analysis is unweighted, and all cases are counted equally. As a result, the means, standard deviations, and percentages will differ from those in the reports on these surveys, which are weighted.

There are similarities between the approach used here and that of Grellety and Golden (2016). The main methodological difference is that they simulated the addition of error to populations of children (boys and girls separately) that were artificially constructed and conformed exactly to the WHO Growth Standards. Their baseline population for simulations of the effect of errors in height or weight on the WHZ was artificially constructed to have a uniform distribution of height between 60 and 110 cm (in intervals of 0.5 cm). Each case was assigned a randomly generated z score. Weight was then calculated for each case using equation (2.2), with M, S, and L for the specified height coming from the wflanthro.dta file and treating the z score as equivalent to the WHZ. For simulations of the effect of errors in height or age on the WAZ, they constructed a separate baseline population with a uniform distribution of age, between 6 and 59 months (in integers). Again, each case was assigned a randomly generated z score. Height was then calculated from the reference HAZ function, using the appropriate M, S, and L values and the randomly generated z, treating it as equivalent to the HAZ.

During the simulations of error, random disturbances were added to the height, weight, or age in these artificial populations and the effects were described. In contrast, we use real data from two DHS surveys as baselines and add random error to whatever error may have been induced during data collection.

The approach of Grellety and Golden has an advantage in that their baseline data are guaranteed to be error-free. A disadvantage is that the starting distribution of height is uniform in the first file just described, which makes it very different from an observed distribution. In the second file, the uniform distribution of age is less of a potential issue, because observed age distributions for ages 6-59 months, apart from typical seasonality of births, are close to uniform.

Unlike Grellety and Golden (2016), who construct separate baseline populations for each assessment that do not have a joint distribution of height, weight, and age, and do not have nutritional deficiencies, our baseline data come from real populations with nutritional deficiencies.

Simulating the net effect of random disturbances to baseline files constructed in this way has theoretical value, analogous to Chapter 2 in this report. However, we would argue that the approach used in this chapter comes closer to simulating the consequences of different levels of the TEM in a DHS or MICS survey.

We now estimate the increase in dispersion that would result from the hypothetical addition of random measurement error in the three underlying variables.

We use graphics to develop functional forms for the relationships between dispersion in height, weight, age and dispersion in the HAZ, WAZ, WHZ. The figures are constrained as follows:

- The maximum standard deviation of random error in height is 5 cm.
- The maximum standard deviation of random error in weight is 2 kg.
- The maximum standard deviation of random error in age is 90 days.

The maximum simulated standard deviations of height, weight, and age are about 5% of the observed ranges of the respective inputs. The higher values of these errors, especially in height or weight, could only occur in a survey with poor training and/or supervision of the measurers.

5.4 The impact of random measurement error on the standard deviations of the Z scores

The additional dispersion in the Z scores is measured with the increase in the standard deviation, above its observed level. Additional dispersion in height, weight or age can only increase the standard deviations of the Z scores and increase the percentages in both tails, which reflect the levels of problematic outcomes.

Figures 5.2-5.4 describe the increased levels of the standard deviations of the Z scores that are induced by adding random error to the measurements of height, weight, and age in the Peru 2012 and Nepal 2016 surveys. Within each figure, two subfigures for Peru 2012 are in the top row and two subfigures for Nepal 2016 in the bottom row.

Appendix 3 provides numerical values that correspond with these figures. Equation (5.4) included f_1, f_2, and f_3, possible functions that would describe the increases in the variances of the HAZ, WAZ, and WHZ that correspond with the standard deviations of the simulated error. Those hypothesized functions, or others using the variances (rather than the standard deviations) of the Z scores, which would parameterize the lines in Figures 5.2-5.4, can be approximated but will not be fitted here.

The horizontal axes of the subfigures are labeled consistently for increments in the dispersion of height, weight, and age, although the vertical axes vary considerably, depending on the initial observed level and the amount of induced increase.

The fuzziness or jitter in Figures 5.2-5.7 is due to the randomness of the induced changes in the heights, weights, and ages. Each child's height, weight, or age is artificially displaced, upwards or downwards in 1,000 simulations with increasing values of the standard deviation, s, of the displacements. The expected deviation is zero, although the mean induced error from a specific simulation typically varies slightly. The variation in simulated Z scores increases as s increases. Induced error in age appears to produce the fuzziest patterns, largely because the vertical axes in the figures for age displacement are the narrowest.

Figure 5.2 Change in the standard deviation of the HAZ and WHZ when height is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys

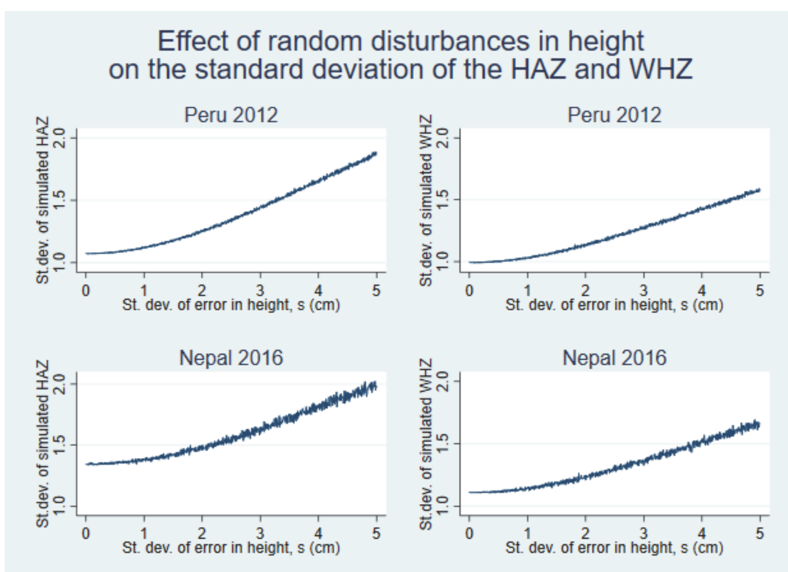


Figure 5.3 Change in the standard deviation of the WAZ and WHZ when weight is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys

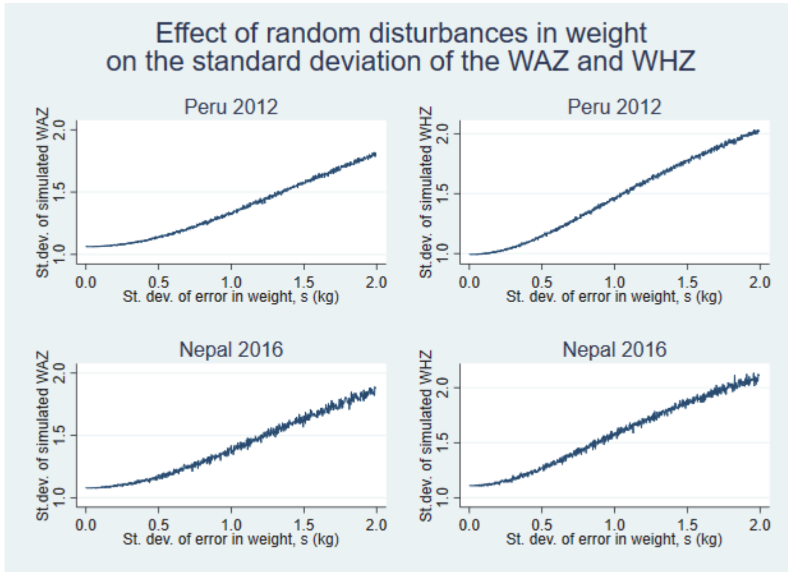
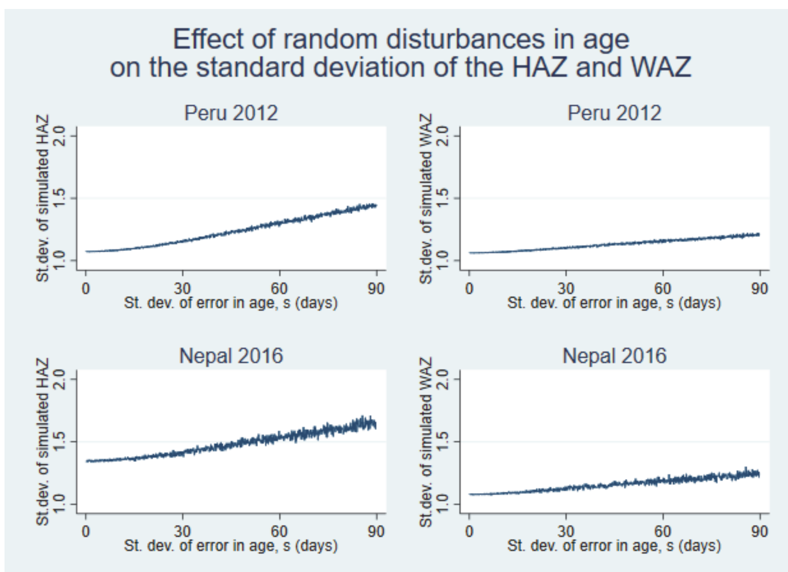


Figure 5.4 Change in the standard deviation of the WAZ and HAZ when age is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 and Nepal 2016 DHS surveys



Using Figures 5.2-5.4 and the corresponding tables in Appendix 3 (Tables A3.1-A3.3), we can make the following generalizations:

- Random displacements in height, weight, or age have little effect on dispersion in the Z scores if s is less than about 1 cm or 0.25 kg or 15 days, respectively.

- Beyond those levels, the standard deviation of the Z scores increases nearly linearly with increases in the standard deviation (s) of the random displacements in height, weight, or age.

- Random displacement in height has more effect on the standard deviation of the HAZ than the WHZ in Peru 2012, but in Nepal 2016 the effects on the dispersion of the HAZ and WHZ are similar.

- Random displacement in weight has similar effects on the standard deviations of the WAZ and WHZ in both Peru 2012 and Nepal 2016.

- Random displacement in age has more effect on the standard deviation of the HAZ than the standard deviation of the WAZ in both Peru 2012 and Nepal 2016.

The most conspicuous difference between the simulations in Peru 2012 and Nepal 2016 is described in the third bullet above, However, in both datasets, a standard deviation of random displacement in height of s=5 cm will increase the standard deviation of the HAZ to about 2.0 and the standard deviation of the WHZ to about 1.6. The *change* is smallest for the HAZ in Nepal 2016 because the baseline standard deviation of the WHZ is higher, 1.35. All other baseline standard deviations of the Z scores are in a narrow range of about 1.0 to 1.1. In both datasets, a standard deviation of random displacement in weight of s=2 kg will also increase the standard deviation of both the WAZ and the WHZ to about 2.0 (a little less in Peru 2012).

In both datasets, standard deviation of random displacement in age of s=90 days will increase the standard deviation of the HAZ by about 0.4 and the standard deviation of the WHZ by about 0.2. A standard deviation of s=90 days of age has much less effect on the standard deviation of the HAZ than a standard deviation of s=5 cm of height, and much less effect on the standard deviation of the WAZ than a standard deviation of s=2 kg of weight.

The simulations show how it is possible for random displacement in height, up to s=5 cm, or weight, up to s=2 kg, interpreted as random measurement error, to induce standard deviations of Z scores up to 2.0. Standard deviations of that magnitude would almost always be suspicious. Measurement errors averaging s=5 cm or s=2 kg would be enormous. It is helpful that the simulations have established a degree of correspondence between specific levels of random measurement error in height and weight and specific large standard deviations in the Z scores. Figure 5.4 (and Table A3.3) implies that the amount of error in age that would increase the standard deviation of the HAZ or WAZ to 2.0 would have to be much greater than the maximum of 90 days that has been simulated here.

5.5 The impact of random measurement error on the anthropometric outcomes

Figures 5.5-5.7 describe the increased levels of the four nutritional outcomes that are induced by adding random error to the measurements of height, weight, and age in the Peru 2012 and Nepal 2016 surveys. Each figure has an “a” panel for Peru 2012 and a “b” panel for Nepal 2016. As with the previous figures, the horizontal axes of the subfigures are labeled consistently, for increments in the dispersion of height, weight, and age. The vertical axes vary considerably, depending on the initial observed level and the amount of induced increase. Again, Appendix 3 provides numerical values that correspond with these figures.

All the vertical scales for the figures in this section of the report begin with 0%. They extend to 40% for stunted and underweight and to 30% for overweight and wasted. For specific subfigures, the low end is lower than necessary and the high end is higher than necessary to capture the range of the line, but these scales facilitate comparisons between countries, outcomes, and the input measurements.

Figure 5.5a Change in the estimates of stunting, overweight, and wasting when height is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 DHS survey

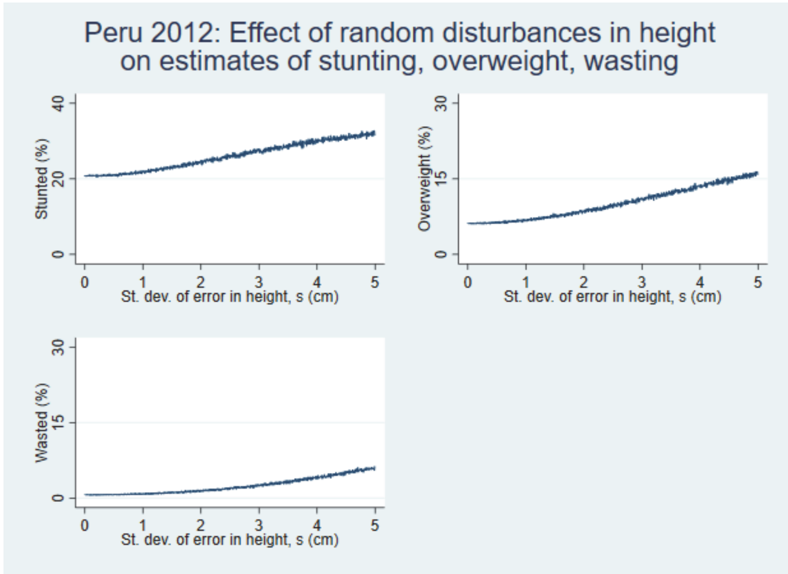


Figure 5.5b Change in the estimates of stunting, overweight, and wasting when height is disturbed by random error with standard deviation s for children under age 5 in the Nepal 2016 DHS survey

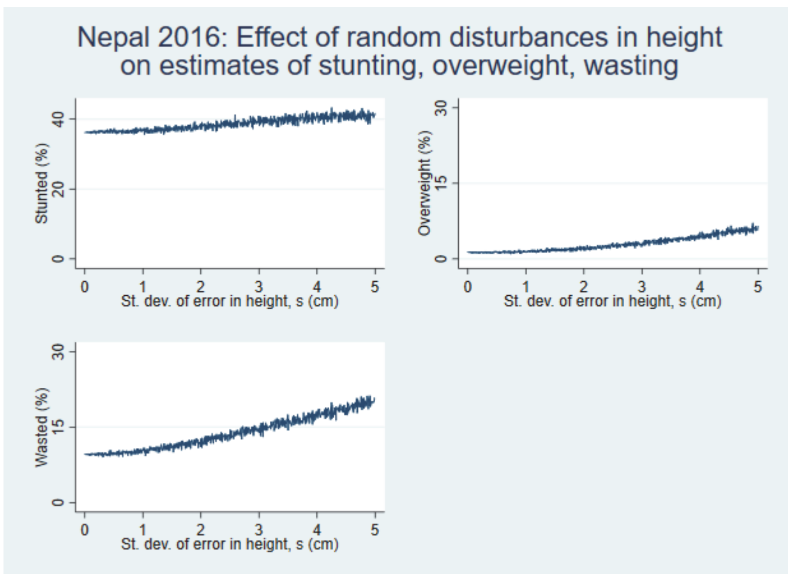


Figure 5.6a Change in the estimates of underweight, overweight, and wasting when weight is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 DHS survey

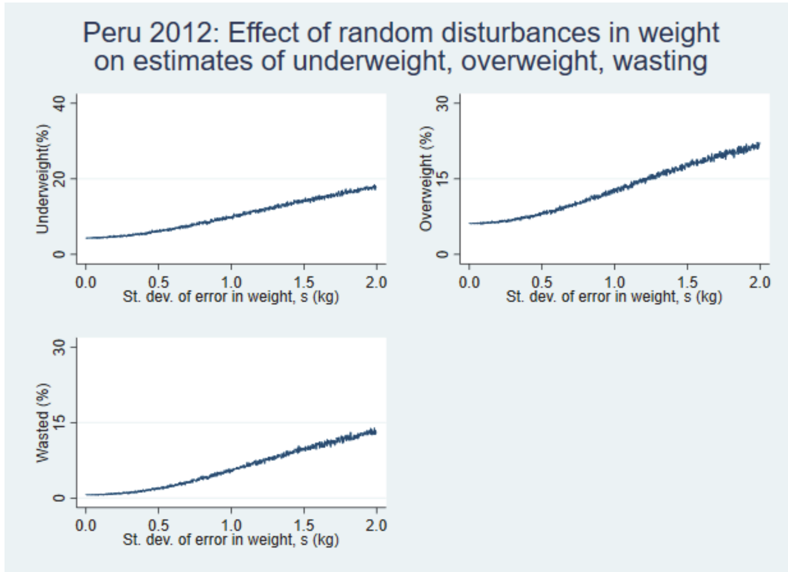


Figure 5.6b Change in the estimates of underweight, overweight, and wasting when weight is disturbed by random error with standard deviation s for children under age 5 in the Nepal 2016 DHS survey

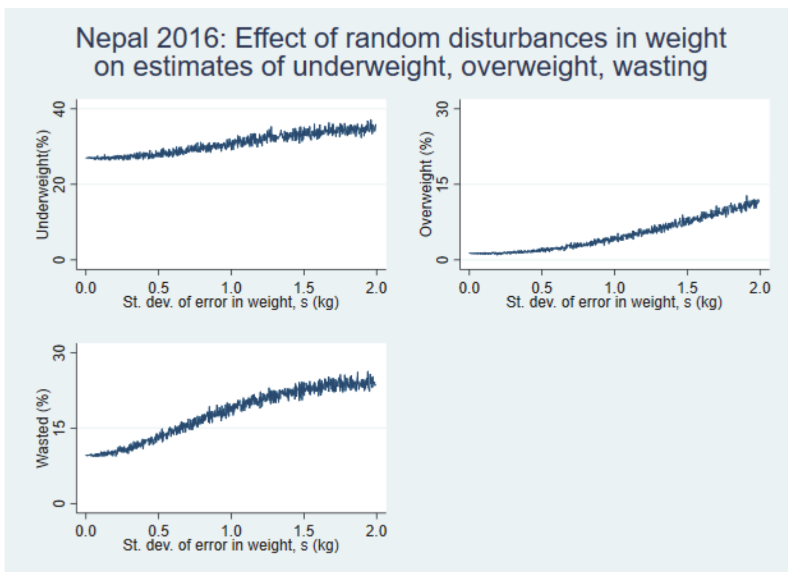


Figure 5.7a Change in the estimates of stunting and underweight when age is disturbed by random error with standard deviation s for children under age 5 in the Peru 2012 DHS survey

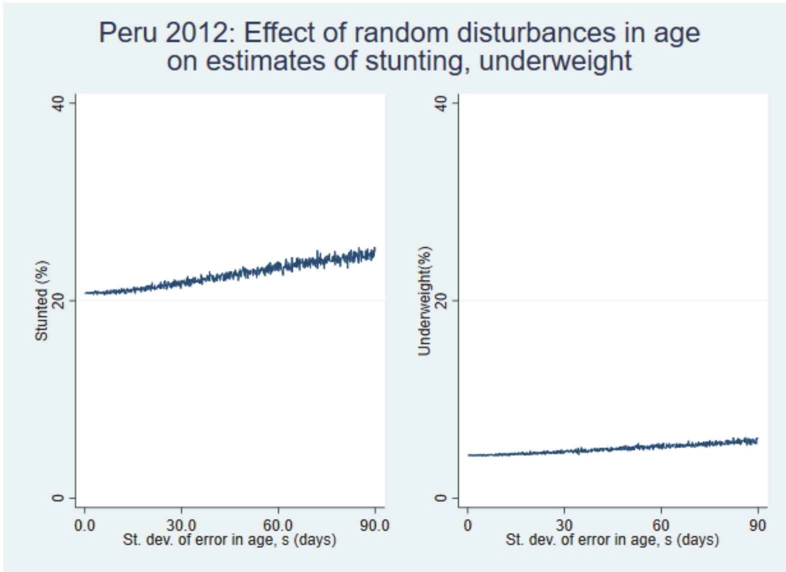
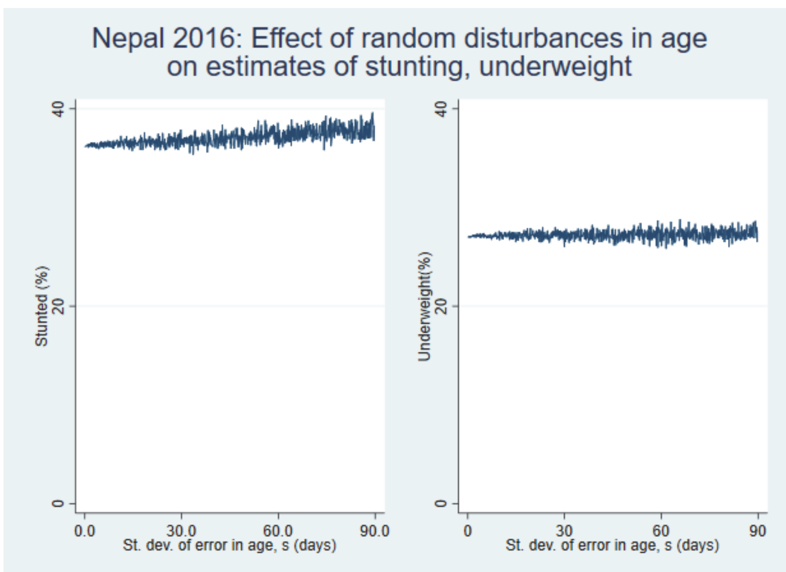


Figure 5.7b Change in the estimates of stunting and underweight when age is disturbed by random error with standard deviation s for children under age 5 in the Nepal 2016 DHS survey



Using Figures 5.5-5.7 (a and b) and the corresponding tables in Appendix 3 (Tables A3.4-A3.6), we can make the following generalizations:

- The effect of random dispersion in height, weight, or age on the percentages stunted, underweight, overweight, or wasted varies greatly and depends on the original dispersion and the original levels of the outcomes.

- As with the standard deviations of the Z scores, small values of s for height, weight, and age, less that approximately 0.5 cm or 0.25 kg of 15 days, respectively, have virtually no effect on the prevalence of the outcomes.

- The effects on the outcomes are not as linear as the effects on the standard deviations. Several of the lines in the figures have curvature that depends on both the outcome and the dataset, i.e. on the baseline levels of the outcomes.

- Several of the effects on the outcomes, within the range of simulated error in the measurements, are small. In particular, the effect on stunting and underweight of additional random dispersion in age, shown in Figure 5.7 (a and b), is in a very narrow range.

To better understand the simulations in this chapter, it may be helpful to review the interpretation of the model for random errors and the preceding figures, particularly the interpretation of the standard deviation of the error, s, and the fuzziness in the figures.

The usual interpretation of the standard deviation is that it is the average deviation from the mean that ignores the direction of the deviation. The standard deviation is not actually the same as the mean absolute deviation, but this interpretation is convenient. For example, when s=0.5 kg is the standard deviation of the normal distribution (with mean 0) from which the draws are being made, we can think of 0.5 kg as the average amount of error that is being added or subtracted. However, with the bell shape of the normal distribution, for most individual children in the data file, the error is less than 0.5. For some children, the error is more than 0.5, and occasionally, it is much more.

The maximum dispersion in these simulations, indexed by the standard deviation, s, is very large relative to plausible mechanisms and sources of measurement error; 5 cm is approximately 5% of the full range of height, 2kg is approximately 5% of the full range of weight, and 90 days is 5% of the full 5-year age range. A normally distributed error in height with mean 0 and standard deviation s=5 cm would imply that about 34% of measurements are too high by 0 to 5 cm, 34% are too low by the same amount, about 14% are too high by 5-10 cm, 14% are too low by the same amount, about 2% are too high by more than 10 cm, and about 2% are too low by the same amount. Although the error at the high ends of the ranges in s would be massive, the high ends provide a sense of the overall pattern. It is useful to know, for example, that random over-dispersion in height with s=5 cm would increase the standard deviation of the HAZ to about 2.0 in both Peru 2012 and Nepal 2016. A value of 2.0 for the standard deviation of a Z score in any survey would be clear evidence of measurement error. These simulations provide evidence that such high dispersion in the HAZ is consistent with a specific hypothetical pattern of dispersion in height. Over-dispersion in the HAZ could also be due to error in age (although with a different gradient) or to a combination of error in both height and age.

The lines in the figures are fuzzy in a vertical direction. Each line is essentially a band, because of the randomness in the induced error. Each line represents 1,000 simulations. In each simulation with weight, for example, a random amount is added to, or subtracted from, the recorded weight for each child. The WHZ is recalculated for each child, the standard deviation of the WHZ in the sample is recalculated, and the increase over the original standard deviation is saved. In each simulation, the amount added or subtracted is generated as random draws from a normal distribution with mean 0 and standard deviations suggest.

6 CONCLUSIONS

The goal of this report has been to improve our understanding of how changes in measurement of height, weight, and age can induce changes in the means and standard deviations of the HAZ, WAZ, and WHZ and the estimates of stunting, underweight, overweight, and wasting. To the extent that these changes can be interpreted as errors of measurement, this is an analysis of the sensitivity of the Z scores and outputs to errors of measurement.

The three methodological strategies we used were analysis, macro simulation, and micro simulation. In Chapter 2, the formulas and WHO reference files for the calculation of Z scores and stunting, wasting, etc. were described in detail. Separately from any empirical data, we examined the change in Z scores that would be induced by a change or difference in height, weight, or age, conditional on the level of those inputs, as appropriate. The changes were 5 cm in height, 2 kg in weight, or 90 days in age, amounts that are arbitrary but are approximately 5% of the ranges in the respective inputs for children age 0-4. This is an analytical strategy based on the formulas for the Z scores. We described the impact of the changes for the two Z scores that each input is used for, and the difference between an input for the numerator or the denominator of the Z score.

Chapter 2 also examined the sensitivity of the percentages stunted, underweight, overweight, or wasted to changes in the mean and standard deviation of the Z score. This strategy is a macro simulation, because it involves manipulating the shape of a population distribution rather than individual cases. A great deal can be inferred from the original construction of the Z scores to have a normal distribution with mean 0 and standard deviation 1 in a reference population of well-nourished children. If an observed Z distribution can be approximated more generally by a normal distribution with mean m and standard deviation s, then the percentages stunted, wasted etc. can be calculated easily with a link to the standard normal distribution. The sensitivity of the percentages stunted, wasted, etc. to displacements of m and s can be easily simulated. In particular, it was shown that the sensitivity of the percentages to any increase in s depends on the value of m. If m is near -1, the percentage stunted has maximum sensitivity to s, increasing by approximately 2 percentage points for each increase of 0.1 in the standard deviation of the HAZ. However, if the mean HAZ is m=-1, for example, the population has a relatively high percentage of stunting regardless of the value of s. One interpretation is that if a substantial percentage of children are stunted, or have HAZ scores that are low but are not below -2, there will be maximum sensitivity to any mechanism that could displace children in either direction across HAZ=-2, the threshold for stunting.

Chapter 3 reviewed the standard DHS surveys conducted between 2010 and 2018 that included anthropometry, plus one round of the Peru Continuous Survey for 2012. That survey and the Nepal 2016 survey are among those with the least amount of dispersion in their Z scores and different nutritional profiles. In the Peru survey, the level of stunting is relatively high. Overweight has become an issue, although it is less common than stunting. In the Nepal survey, children age 0-4 had high levels of stunting and underweight, and wasting is an issue, although less prevalent than the other two problematic outcomes. We examined these two surveys in detail, graphically describing their distributions of height, weight, age, Z scores, and prevalence of stunting, wasting, underweight and overweight.

Chapter 4 described the possibility for variation in mean Z scores, across subpopulations, to increase the dispersion of Z scores at the national level, using the Nepal 2016 survey for empirically based illustration. This kind of analysis could be extended to hierarchies of nested subpopulations, using DHS samples or simulation, to develop a more complete picture of the importance of genuine heterogeneity as a determinant of over-dispersion of Z scores. However, there is no question that very high dispersion of Z scores, particularly the HAZ in some surveys, is the result of measurement error in height or age. Measurement error in weight is probably the least serious.

Chapter 5 uses microsimulation, in which individual measurements of height, weight, and age are disturbed by some amount and then the Z scores and outcomes are recalculated for the entire sample. We use the Peru 2012 and Nepal 2016 surveys as examples. By the very nature of anthropometric data collection, the surveys are assumed to include some measurement error. Our approach is to add *more* measurement error, but in controlled amounts. The question is how the additional induced error will lead to changes in the Z scores and then to changes in the level of stunting, wasting, etc. in the data.

Chapter 5 drew a distinction between systematic or directional error, and random error. Directional error primarily induces bias in height, weight, or age measurements—and thus in the means of the Z scores. Random or bidirectional error primarily affects the dispersion of the Z scores. Directional error in height can be caused by additional hair on children's heads, while directional error in weight can be caused by additional clothing, beyond the minimum. Some known types of errors in age reporting are consistently upward, while others are consistently downward. It is difficult to model the magnitude of these biases in the basic measurements, their relation to age, or their relative prevalence even within a single population. This report recognizes the relevance of such errors but does not attempt to replicate them because of the variability in potential patterns of displacement. In future research, we hope to simulate a number of specific scenarios.

Bidirectional random error is simpler to specify. In Chapter 5, we applied this to the two selected surveys. The additional error terms are modeled to be normally distributed with mean 0 and with a standard deviation that increases from 0 to 5 cm, 2 kg, or 90 days. For height, weight, and age, one at a time, we conducted 1,000 simulations of error in this range. Each simulation involved replicating the entire data set, recalculating the Z values, and then recalculating the percentages of children who are wasted, stunted, etc. We simulated 3,000 versions of the Peru 2012 data and 3,000 versions of the Nepal 2016 data. For each, we calculated the mean and standard deviation of the Z scores and the percentages stunted, wasted, etc. Figures show the variability in the simulations, and tables show the average outputs indexed to the standard deviation of the additional error.

Because of the shape of the normal distribution, errors generated with up to 5 cm, 2 kg, and 90 days for the standard deviations of the error term will occasionally be much larger than those levels. We consider the index levels, and the displacements that may be much larger than those levels, to be highly unlikely because we intentionally extended beyond what is acceptable or likely in real data collection. Chapter 5 includes a discussion of how random errors relate to the Technical Error of Measurement, or TEM. We extend far beyond the plausible levels of the TEM because the simulations provide a tool to envision the implications of hypothetical bidirectional error of this magnitude.

This type of analysis could be extended in other ways. First, the modeling of systematic or unidirectional error would be helpful, especially with a random component. For example, if additional clothing were hypothesized to add 0.3 to 1.5 cm of height, it would be possible to add a term that is randomly distributed in that range—although it would be necessary to specify if the probability distribution for the additional error was uniform, or predominantly in either the lower or upper end of the range.

Second, we could combine directional sources of error with bidirectional random error. Third, it is possible to combine errors in height, weight, and age. For example, how would the WHZ distribution, and the percentages of children in the tails of that distribution be affected if errors were simulated in both height and weight? Many errors would offset each other because as we noted, an increase in height and an increase in weight, whether real or spurious, can lead to the same WHZ score as the original measurements. Simulating combinations of multiple types of error, while possible, would amplify the difficulty and arbitrariness of constructing probability distributions that are plausibly consistent with what may happen during fieldwork.

Finally, and most ambitiously, these methods could be used in reverse to estimate the amount of measurement error in height, weight, and age that is consistent with observed levels of over-dispersion in the Z scores. There are three Z scores and three inputs—height, weight, and age—and although the transformation is complex, the relationship can be analyzed further. A challenge is that the true level of dispersion, in the absence of any measurement error at all, is unknown and therefore, the excess dispersion is unknown. Nevertheless, it may be possible to estimate ranges of measurement error in height, weight, and age that are compatible with the observed standard deviations for the three Z scores.

REFERENCES

Allen, C. K., T. N. Croft, T. W. Pullum, and S. M. L. Namaste. *Evaluation of Indicators to Monitor Quality of Anthropometry Data during Fieldwork.* DHS Working Paper No. 162. Rockville, Maryland, USA: ICF International. https://www.dhsprogram.com/pubs/pdf/WP162/WP162.pdf.

Assaf, S., M. T. Kothari, and T. W. Pullum. 2015. *An Assessment of the Quality of DHS Anthropometric Data, 2005-2014*. DHS Methodological Reports No. 16. Rockville, MD, USA: ICF International. https://www.dhsprogram.com/pubs/pdf/MR16/MR16.pdf.

Finaret, A. B. and W. A Masters. 2019. "Correcting for Artifactual Correlation between Misreported Month of Birth and Attained Height-For-Age Reduces but Does Not Eliminate Measured Vulnerability of Season of Birth in Poorer Countries." *American Journal of Clinical Nutrition* 110 (2): 485–497. https://doi.org/10.1093/ajcn/nqz111.

Grellety, E. and M. H. Golden. 2016. "The Effect of Random Error on Diagnostic Accuracy Illustrated with the Anthropometric Diagnosis of Malnutrition." PLoS ONE 11 (12): e0168585. https://doi.org/10.1371/journal.pone.0168585.

Larsen, A. F., D. Headey, and W. A. Masters. 2019. "Misreporting Month of Birth: Diagnosis and Implications for Research on Nutrition and Early Childhood in Developing Countries." *Demography* 56 (2): 707–728. https://doi.org/10.1007/s13524-018-0753-9.

Mei, Z., and L. M. Grummer-Strawn. 2007. "Standard Deviation of Anthropometric Z-Scores as a Data Quality Assessment Tool Using the 2006 WHO Growth Standards: A Cross Country Analysis." *Bulletin of the World Health Organization* 85(6):441-8. https://doi.org/10.2471/blt.06.034421.

Perumal, N., S. Namaste, H. Qamar, A. Aimone, D. G Bassani, D. E Roth. 2020. "Anthropometric Data Quality Assessment in Multisurvey Studies of Child Growth." *The American Journal of Clinical Nutrition.* nqaa162. https://doi.org/10.1093/ajcn/nqaa162.

Pullum, T. W. 1990. "Statistical Methods to Adjust for Date and Age Misreporting to Improve Estimates of Vital Rates in Pakistan." *Statistics in Medicine* 10 (2): 191-200. https://doi.org/10.1002/sim.4780100205.

Pullum, T. W. 2006. *An Assessment of Age and Date Reporting in The Demographic and Health Surveys, 1985-2003*. DHS Methodological Report No. 5. Calverton, MD, USA: Macro International. https://dhsprogram.com/publications/publication-mr5-methodological-reports.cfm.

Pullum, T. W. 2008. *An Assessment of The Quality of Data on Health and Nutrition in The Demographic and Health Surveys, 1993-2003*. DHS Methodological Report No. 6. Calverton, MD, USA: Macro International. https://dhsprogram.com/publications/publication-mr6-methodological-reports.cfm.

Pullum, T. W. 2019. *Strategies to Assess the Quality of DHS Data*. DHS Methodological Reports No. 26. Rockville, MD, USA: ICF. https://dhsprogram.com/pubs/pdf/MR26/MR26.pdf.

Pullum, T. W. and S. L. Stokes. 1997. "Identifying and Adjusting for Recall Error, with Application to Fertility Surveys." In *Survey Measurement and Process Quality*, edited by Lars Lyberg and Paul Biemer, et al., 711-732. New York, USA: John Wiley and Sons.

Roche, M. L., T. W. Gyorkos, J. Sarsoza, and H. V. Kuhnlein. 2015. "Adjustments for Weighing Clothed Babies at High Altitude or in Cold Climates." *Global Public Health* 10 (10): 1227-1237. https://doi.org/10.1080/17441692.2015.1037326.

Shah, I. H., T. W. Pullum, and M. Irfan. 1986. "Fertility in Pakistan During the 1970s." *Journal of Biosocial Science* 18 (2): 215-29. https://doi.org/10.1017/S002193200001614X.

Tuan, T., D. R. Marsh, T. T. Ha, D. G. Schroeder, T. D. Thach, V. M. Dung, and N. T. Huong. 2003. "Weighing Vietnamese Children: How Accurate are Child Weights Adjusted for Estimates of Clothing Weight?" *Food and Nutrition Bulletin* 12 (4 Suppl): 48-52. https://pubmed.ncbi.nlm.nih.gov/12503231/.

Ulijaszek, S. J. and D. A. Kerr. 1999. "Anthropometric Measurement Error and The Assessment of Nutritional Status." *British Journal of Nutrition* 82(3):165-77. https://doi.org/10.1017/S0007114599001348.

WHO Multicentre Growth Reference Study Group. 2006a. *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age.* Geneva, Switzerland: World Health Organization. https://www.who.int/childgrowth/standards/technical_report/en/.

WHO Multicentre Growth Reference Study Group. 2006b. "Reliability of Anthropometric Measurements in the WHO Multicentre Growth Reference Study." *Acta Pædiatrica* Suppl 95 (450): 38-46. https://doi.org/10.1111/j.1651-2227.2006.tb02374.x.

WHO and UNICEF. 2019. *Recommendations for Data Collection, Analysis and Reporting on Anthropometric Indicators in Children Under 5 Years Old.* Geneva: World Health Organization (WHO) and the United Nations Children's Fund (UNICEF). https://apps.who.int/iris/bitstream/handle/10665/324791/9789241515559-eng.pdf?ua=1.

APPENDIX 1 MEANS AND STANDARD DEVIATIONS OF Z SCORES IN STANDARD DHS SURVEYS CONDUCTED DURING 2010-18

Table A1.1 Weighted mean and standard deviation of the HAZ, WAZ, and WHZ for children under age 5 in DHS surveys conducted during 2010-18 that included measurements of height and weight

	HAZ		WAZ		WHZ	
Survey	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Albania 2017	-0.33	1.52	0.41	1.10	0.79	1.28
Angola 2015-16	-1.53	1.54	-0.99	1.27	-0.13	1.15
Armenia 2010	-0.74	1.64	0.05	1.11	0.67	1.47
Armenia 2015-16	-0.14	1.62	0.37	1.19	0.61	1.37
Bangladesh 2011	-1.67	1.41	-1.61	1.15	-0.94	1.19
Bangladesh 2014	-1.54	1.32	-1.49	1.12	-0.89	1.15
Benin 2011-12	-1.60	2.33	-0.92	1.49	-0.08	1.97
Benin 2017-18	-1.45	1.32	-1.03	1.09	-0.27	1.06
Burkina Faso 2010	-1.40	1.60	-1.27	1.20	-0.67	1.37
Burundi 2010	-2.20	1.38	-1.42	1.10	-0.21	1.15
Burundi 2016-17	-2.18	1.27	-1.47	1.09	-0.28	1.05
Cambodia 2010	-1.66	1.37	-1.43	1.05	-0.71	1.12
Cambodia 2014	-1.41	1.35	-1.26	1.09	-0.67	1.13
Cameroon 2011	-1.26	1.72	-0.63	1.31	0.14	1.32
Chad 2014-15	-1.51	1.91	-1.26	1.40	-0.56	1.32
Colombia 2010	-0.83	1.12	-0.24	1.00	0.32	0.99
Comoros 2012	-1.16	1.92	-0.76	1.34	-0.17	1.58
Congo 2011-12	-1.02	1.48	-0.72	1.11	-0.20	1.18
Congo Democratic Republic 2013-14	-1.60	1.84	-1.08	1.28	-0.22	1.31
Cote d'Ivoire 2011-12	-1.22	1.59	-0.83	1.16	-0.19	1.23
Dominican Republic 2013	-0.30	1.24	0.03	1.12	0.28	1.18
Egypt 2014	-0.57	2.02	-0.08	1.21	0.36	1.65
Ethiopia 2011	-1.69	1.68	-1.33	1.24	-0.51	1.19
Ethiopia 2016	-1.44	1.75	-1.16	1.26	-0.46	1.26
Gabon 2012	-0.70	1.48	-0.24	1.16	0.21	1.24
Gambia 2013	-1.01	1.55	-0.99	1.12	-0.61	1.27
Ghana 2014	-0.94	1.28	-0.70	1.06	-0.25	1.07
Guatemala 2014-15	-1.89	1.19	-0.85	1.05	0.35	0.98
Guinea 2012	-1.11	1.82	-0.87	1.30	-0.32	1.36
Guinea 2018	-1.05	1.91	-0.78	1.37	-0.27	1.40
Haiti 2012	-0.97	1.43	-0.64	1.18	-0.12	1.18
Haiti 2016-17	-0.96	1.48	-0.56	1.15	-0.01	1.12
Honduras 2011-12	-1.11	1.22	-0.42	1.11	0.31	1.04
India 2015-16	-1.48	1.68	-1.57	1.21	-1.04	1.35
Jordan 2012	-0.40	1.18	-0.10	1.01	0.17	1.08
Kenya 2014	-1.14	1.42	-0.64	1.15	0.00	1.15
Kyrgyz Republic 2012	-0.80	1.45	-0.14	1.02	0.44	1.21
Lesotho 2014	-1.47	1.34	-0.59	1.11	0.35	1.20
Liberia 2013	-1.23	1.66	-0.84	1.22	-0.17	1.20
Malawi 2010	-1.78	1.61	-0.81	1.13	0.27	1.27
Malawi 2015-16	-1.54	1.39	-0.79	1.13	0.14	1.10
Maldives 2016-17	-0.85	1.26	-0.77	1.23	-0.42	1.36
Mali 2012-13	-1.46	1.87	-1.23	1.31	-0.55	1.35
Mali 2018	-1.10	1.58	-0.99	1.23	-0.52	1.19
Mozambique 2011	-1.68	1.65	-0.86	1.17	0.15	1.35
Myanmar 2015-16	-1.34	1.31	-1.13	1.08	-0.53	1.06
Namibia 2013	-1.09	1.42	-0.78	1.14	-0.21	1.22
Nepal 2011	-1.68	1.41	-1.44	1.11	-0.66	1.13
Nepal 2016	-1.53	1.35	-1.33	1.09	-0.64	1.11
Niger 2012	-1.73	1.68	-1.60	1.27	-0.87	1.36

Table A1.1 Weighted mean and standard deviation of the HAZ, WAZ, and WHZ for children under age 5 in DHS surveys conducted during 2010-18 that included measurements of height and weight *(continued)*

	HAZ		WAZ		WHZ	
Survey	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**
Nigeria 2013	-1.38	2.01	-1.26	1.42	-0.68	1.56
Nigeria 2018	-1.52	1.59	-1.10	1.27	-0.29	1.14
Pakistan 2012-13	-1.79	1.72	-1.40	1.26	-0.52	1.26
Pakistan 2017-18	-1.56	1.60	-1.16	1.30	-0.30	1.19
Papua New Guinea 2016-18	-1.66	1.77	-1.03	1.50	-0.08	1.55
Peru 2012	-1.05	1.09	-0.21	1.07	0.55	1.01
Rwanda 2010	-1.76	1.40	-0.77	1.06	0.34	1.16
Rwanda 2014-15	-1.57	1.39	-0.58	1.10	0.44	1.16
Senegal 2010-11	-1.10	1.60	-0.98	1.15	-0.53	1.23
Sierra Leone 2013	-1.38	1.93	-0.82	1.36	-0.04	1.50
South Africa 2016	-1.15	1.44	-0.18	1.25	0.62	1.28
Tajikistan 2012	-1.13	1.60	-0.80	1.16	-0.23	1.43
Tajikistan 2017	-0.82	1.42	-0.51	1.09	-0.07	1.21
Tanzania 2010	-1.70	1.42	-0.95	1.12	0.03	1.21
Tanzania 2015-16	-1.46	1.40	-0.86	1.11	-0.04	1.15
Timor-Leste 2016	-1.52	2.14	-1.59	1.49	-1.00	1.64
Togo 2013-14	-1.23	1.41	-0.90	1.13	-0.28	1.11
Uganda 2011	-1.42	1.57	-0.82	1.15	-0.02	1.17
Uganda 2016	-1.20	1.51	-0.61	1.16	0.11	1.15
Yemen 2013	-1.86	1.61	-1.68	1.20	-0.87	1.29
Zambia 2013-14	-1.59	1.61	-0.90	1.12	-0.01	1.30
Zambia 2018	-1.45	1.47	-0.76	1.09	0.08	1.22
Zimbabwe 2010-11	-1.37	1.41	-0.66	1.08	0.17	1.16
Zimbabwe 2015	-1.24	1.40	-0.55	1.11	0.19	1.17

APPENDIX 2 SIMULATING RANDOM ERRORS OF MEASUREMENT

The reference files make it very easy to calculate the HAZ, WAZ, and WHZ with a DHS data file that gives the values of height, weight, and age. Using the PR file, the labels are ht for height in units of 0.1 cm, hc3, wt for weight in units of 0.1 kg, hc2, and age in days, hc1a. In three steps, the PR datafile is sorted appropriately for each reference file, merged with that file, and the HAZ, WAZ, or WHZ is copied onto the PR file. The processing time to retrieve the three Z scores for the children age 0-4 in a typical PR file is a fraction of a second.

To simulate random disturbances or errors in the measurement of height, weight, and/or age, we repeatedly generate random errors, which may be positive or negative; add them to the observed values; recalculate height, weight, and/or age; recalculate the Z scores; and calculate the standard deviation and variance of the simulated Z scores. These steps can be carried out for as many simulations as desired. In order to make this as clear as possible, we repeat the steps with more detail, with specification of some commands in Stata.

Setup. A file is constructed, based on children age 0-4 in a specific PR file, that includes ID, sex, LORH, ht, wt, and days for each child. Other variables may be included because they are relevant for subsequent analysis, but they are not directly involved in the simulations. The "original" coded values of ht, wt, and days are saved as original_ht, original_wt, and original_days. A number of simulations is specified, as Nsim. The maximum amount of error to be simulated for the three measurements is specified as the maximum standard deviations: max_sd_error_ht, max_sd_error_wt, and max_sd_error_days. For example we could specify Nsim=100, max_sd_error_ht=30 (3 cm), max_sd_error_wt=30 (3 kg), max_sd_error_days=30 (30 days). {Note: "sd_error" is an abbreviation for "standard deviation of the error term", not "standard error".)

Simulation. On each repetition of an iterative process, an increasing amount of error is added to the observed values of height, weight, and age. There are alternative ways to accomplish this. The procedure used here followed three scenarios, as follows:

- Scenario 1: Error is introduced to height; weight and age are kept at their original values.
- Scenario 2: Error is introduced to weight; height and age are kept at their original values.
- Scenario 3: Error is introduced to age; height and weight are kept at their original values.

The procedure for introducing measurement error will be described in detail for height. We generate a random variable with a uniform distribution between 0 and 1 with "gen rn=uniform()". (This function does not require an argument between the parentheses for a uniform distribution between 0 and 1.) These values are converted to a normally distributed random variable with mean 0 and standard deviation (and variance) 1 with "gen z=invnorm(rn)".

The standard deviation of the error is designed to increase gradually from one iteration to the next by incorporating a multiplier. On the first iteration the multiplier is M=1/Nsim; on the second it is M=2/Nsim; on the last iteration it is M=Nsim/Nsim=1. That is, on the last iteration the target standard deviation of the error for height is M*max_sd_error_ht=max_sd_error_ht. M is multiplied by z, is rounded to the nearest integer, and then is added to the observed measurement.

More compactly, on iteration n, the simulated distribution of height is given by this command: "gen ht_sim=original_ht+round((n/Nsim)*max_sd_error_ht*invnorm(uniform()))". The standard distribution of these disturbances will not necessarily be (n/Nsim)*max_sd_error_ht. We must calculate the standard deviation and variance of the disturbances on each iteration. It is this calculated value, rather than the target value, that is used in the analysis.

For the replications done within Scenario 1, no error is introduced to weight and age. The variables ht, wt, and days are set, respectively, to the simulated value of ht and the original values of wt and days. Within each iteration, exactly the same process is followed for weight and days. Because we generate a different "rn=uniform()" for height, weight, and days, the disturbances are independent of one another.

There are Nsim repetitions of the process, or simulations, for each scenario.

For each of the Nsim simulations done under Scenario 1, for example, we calculate the following:

- The standard deviations of ht, wt, and days
- The standard deviations of the HAZ, WAZ, and WHZ
- The proportions of HAZ, WAZ, and WHZ scores that are outside the WHO ranges
- The proportions of children, within the WHO ranges, who are stunted, underweight, overweight, or wasted.

These values are saved in another file, called "simulation_results.dta".

The saved values from the simulations are analyzed with regressions. The objective is to describe the effect of random disturbances, of specified magnitude, to height, weight, and age, on the Z scores. For example, to assess the effect of disturbances to height, when weight and age are measured accurately, on the HAZ and WHZ, we can plot the simulated standard deviations of the HAZ and WHZ against the standard deviation of the error in height. We will estimate six "one-predictor" relationships of this kind:

- Dispersion in the HAZ and WHZ that is induced by error in height alone
- Dispersion in the WAZ and WHZ that is induced by error in weight alone
- Dispersion in the HAZ and WAZ that is induced by error in age alone.

APPENDIX 3 RESULTS OF THE MICRO SIMULATIONS IN CHAPTER 5

Table A3.1 The average effect of simulated random error in height, with mean zero and standard deviation s, on the standard deviation of the HAZ and WHZ, and on the estimates of stunting, overweight, and wasting. Children age 0-4 in the Peru 2012 DHS survey.

	Simulated st.dev. of Z scores				Simulated prevalence of outcomes (%)					
	HAZ		WHZ		Stunted		Overweight		Wasted	
s	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change
0.0	1.08	0.00	0.99	0.00	20.6	0.0	6.0	0.0	0.7	0.0
0.1	1.07	0.00	0.99	0.00	20.8	0.2	6.2	0.1	0.7	0.0
0.2	1.07	0.00	0.99	0.00	20.8	0.2	6.2	0.2	0.7	0.0
0.3	1.08	0.00	1.00	0.00	20.9	0.2	6.2	0.2	0.7	0.0
0.4	1.08	0.00	1.00	0.01	20.9	0.3	6.2	0.2	0.7	0.0
0.5	1.08	0.00	1.00	0.01	21.0	0.4	6.3	0.3	0.7	0.0
0.6	1.09	0.01	1.01	0.02	21.1	0.5	6.4	0.4	0.7	0.0
0.7	1.10	0.02	1.01	0.02	21.3	0.7	6.5	0.4	0.7	0.1
0.8	1.10	0.02	1.02	0.03	21.4	0.8	6.5	0.5	0.8	0.1
0.9	1.11	0.03	1.02	0.03	21.6	1.0	6.7	0.7	0.8	0.1
1.0	1.12	0.04	1.03	0.04	21.8	1.2	6.8	0.8	0.8	0.1
1.1	1.13	0.05	1.04	0.05	22.1	1.4	6.9	0.9	0.9	0.2
1.2	1.14	0.06	1.05	0.06	22.3	1.7	7.1	1.1	0.9	0.2
1.3	1.15	0.07	1.05	0.06	22.6	1.9	7.2	1.2	1.0	0.3
1.4	1.16	0.08	1.07	0.07	22.8	2.2	7.4	1.4	1.0	0.4
1.5	1.18	0.09	1.08	0.08	23.1	2.5	7.6	1.6	1.1	0.4
1.6	1.19	0.11	1.09	0.09	23.2	2.6	7.7	1.7	1.1	0.5
1.7	1.21	0.12	1.10	0.11	23.6	3.0	7.9	1.8	1.2	0.5
1.8	1.22	0.14	1.11	0.12	23.9	3.3	8.1	2.1	1.3	0.6
1.9	1.23	0.15	1.12	0.13	24.1	3.5	8.3	2.3	1.4	0.7
2.0	1.25	0.17	1.13	0.14	24.4	3.8	8.5	2.5	1.5	0.8
2.1	1.27	0.19	1.15	0.16	24.8	4.2	8.8	2.7	1.5	0.9
2.2	1.29	0.20	1.16	0.17	25.1	4.4	9.0	2.9	1.6	0.9
2.3	1.30	0.22	1.17	0.18	25.2	4.5	9.1	3.1	1.7	1.0
2.4	1.32	0.24	1.19	0.19	25.7	5.0	9.4	3.3	1.8	1.1
2.5	1.34	0.26	1.20	0.21	25.9	5.3	9.7	3.7	1.9	1.2
2.6	1.36	0.28	1.22	0.23	26.2	5.6	9.9	3.9	2.1	1.4
2.7	1.38	0.30	1.23	0.24	26.4	5.8	10.2	4.1	2.1	1.5
2.8	1.40	0.32	1.24	0.25	26.8	6.2	10.4	4.4	2.2	1.5
2.9	1.42	0.34	1.26	0.27	27.2	6.6	10.7	4.6	2.4	1.7
3.0	1.44	0.36	1.27	0.28	27.4	6.8	10.9	4.9	2.5	1.9
3.1	1.46	0.38	1.29	0.30	27.4	6.8	11.2	5.2	2.7	2.0
3.2	1.48	0.40	1.30	0.31	27.9	7.3	11.4	5.4	2.8	2.1
3.3	1.50	0.42	1.32	0.33	28.2	7.5	11.7	5.7	2.9	2.3
3.4	1.52	0.44	1.33	0.34	28.4	7.8	11.9	5.9	3.1	2.4
3.5	1.55	0.47	1.35	0.35	28.7	8.0	12.2	6.1	3.2	2.5
3.6	1.57	0.49	1.37	0.37	29.1	8.5	12.5	6.5	3.4	2.7
3.7	1.59	0.51	1.38	0.38	29.2	8.6	12.6	6.6	3.6	2.9
3.8	1.61	0.53	1.39	0.40	29.5	8.9	12.9	6.9	3.8	3.1
3.9	1.64	0.55	1.41	0.42	29.8	9.2	13.3	7.2	3.9	3.2
4.0	1.65	0.57	1.43	0.43	29.9	9.3	13.5	7.5	4.1	3.4
4.1	1.68	0.60	1.44	0.45	30.2	9.6	13.8	7.8	4.3	3.6
4.2	1.70	0.62	1.46	0.46	30.5	9.9	14.1	8.0	4.4	3.7
4.3	1.72	0.64	1.47	0.48	30.6	10.0	14.3	8.3	4.7	4.0
4.4	1.74	0.66	1.49	0.49	30.8	10.2	14.6	8.6	4.8	4.1
4.5	1.76	0.68	1.50	0.51	30.9	10.3	14.7	8.7	5.0	4.4
4.6	1.78	0.70	1.51	0.52	31.2	10.6	15.0	9.0	5.2	4.6
4.7	1.80	0.72	1.53	0.54	31.3	10.7	15.3	9.3	5.5	4.8
4.8	1.83	0.75	1.55	0.56	31.7	11.1	15.6	9.6	5.6	4.9
4.9	1.85	0.77	1.56	0.57	31.8	11.2	15.8	9.7	5.8	5.1
5.0	1.87	0.79	1.58	0.58	32.1	11.5	16.0	10.0	5.9	5.3

Table A3.2 The average effect of simulated random error in height, with mean zero and standard deviation s, on the standard deviation of the HAZ and WHZ, and on the estimates of stunting, overweight, and wasting. Children age 0-4 in the Nepal 20162 DHS survey.

	Simulated st.dev. of Z scores				Simulated prevalence of outcomes (%)					
	HAZ		WHZ		Stunted		Overweight		Wasted	
s	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change
0.0	1.34	0.00	1.11	0.00	36.1	0.0	1.3	0.0	9.5	0.0
0.1	1.35	0.00	1.11	0.00	36.3	0.2	1.3	0.0	9.6	0.1
0.2	1.34	0.00	1.11	0.00	36.3	0.2	1.3	0.0	9.6	0.1
0.3	1.35	0.01	1.11	0.00	36.4	0.3	1.3	0.0	9.6	0.1
0.4	1.35	0.01	1.11	0.00	36.5	0.4	1.3	0.0	9.6	0.2
0.5	1.35	0.01	1.12	0.01	36.6	0.5	1.3	0.1	9.7	0.3
0.6	1.36	0.02	1.12	0.01	36.4	0.3	1.3	0.0	9.8	0.4
0.7	1.36	0.02	1.12	0.01	36.5	0.4	1.3	0.1	9.9	0.4
0.8	1.37	0.03	1.13	0.02	36.5	0.4	1.4	0.1	10.0	0.5
0.9	1.37	0.03	1.13	0.02	36.7	0.6	1.4	0.1	10.2	0.7
1.0	1.38	0.04	1.14	0.03	36.8	0.6	1.5	0.2	10.2	0.8
1.1	1.38	0.04	1.15	0.04	36.8	0.6	1.5	0.3	10.5	1.0
1.2	1.39	0.05	1.16	0.05	36.8	0.7	1.5	0.2	10.6	1.1
1.3	1.40	0.06	1.16	0.05	37.1	1.0	1.6	0.3	10.6	1.2
1.4	1.41	0.07	1.17	0.06	37.0	0.9	1.7	0.4	11.0	1.6
1.5	1.42	0.08	1.18	0.07	37.3	1.2	1.8	0.5	11.1	1.6
1.6	1.43	0.09	1.19	0.08	37.4	1.3	1.8	0.5	11.4	2.0
1.7	1.45	0.11	1.20	0.09	37.3	1.2	1.8	0.5	11.6	2.1
1.8	1.46	0.12	1.21	0.10	37.6	1.5	1.9	0.6	11.6	2.2
1.9	1.46	0.12	1.22	0.11	37.8	1.7	2.0	0.7	12.1	2.6
2.0	1.47	0.13	1.23	0.12	38.0	1.9	2.1	0.8	11.9	2.4
2.1	1.49	0.15	1.25	0.14	37.9	1.8	2.1	0.8	12.5	3.1
2.2	1.50	0.16	1.26	0.15	38.0	1.9	2.2	0.9	12.8	3.3
2.3	1.52	0.18	1.27	0.16	38.0	1.9	2.3	1.0	13.0	3.5
2.4	1.54	0.20	1.28	0.17	38.7	2.6	2.4	1.2	13.1	3.7
2.5	1.55	0.21	1.30	0.19	38.7	2.6	2.6	1.4	13.3	3.9
2.6	1.56	0.21	1.31	0.20	38.8	2.7	2.6	1.3	13.7	4.2
2.7	1.58	0.24	1.32	0.21	38.8	2.6	2.8	1.5	14.0	4.6
2.8	1.60	0.25	1.34	0.23	38.8	2.7	2.9	1.7	14.3	4.8
2.9	1.62	0.27	1.35	0.24	39.5	3.3	2.9	1.7	14.5	5.0
3.0	1.63	0.29	1.36	0.25	39.3	3.2	3.0	1.7	14.5	5.1
3.1	1.65	0.31	1.38	0.27	39.6	3.5	3.2	1.9	14.8	5.4
3.2	1.67	0.33	1.39	0.28	39.4	3.3	3.3	2.0	15.3	5.8
3.3	1.68	0.34	1.41	0.30	39.5	3.4	3.5	2.3	15.7	6.3
3.4	1.71	0.37	1.42	0.31	39.8	3.7	3.5	2.2	15.8	6.3
3.5	1.72	0.38	1.43	0.32	40.1	4.0	3.7	2.4	16.0	6.5
3.6	1.73	0.39	1.45	0.34	39.9	3.8	3.9	2.6	16.3	6.8
3.7	1.76	0.42	1.46	0.35	40.0	3.9	4.0	2.7	16.4	6.9
3.8	1.77	0.43	1.48	0.37	40.3	4.2	4.1	2.8	16.6	7.2
3.9	1.80	0.46	1.50	0.39	40.5	4.4	4.2	3.0	17.1	7.7
4.0	1.82	0.47	1.52	0.41	40.5	4.4	4.5	3.2	17.5	8.1
4.1	1.83	0.48	1.52	0.41	40.5	4.4	4.5	3.2	17.6	8.2
4.2	1.85	0.51	1.54	0.43	40.5	4.4	4.7	3.4	17.8	8.4
4.3	1.87	0.53	1.56	0.45	40.9	4.8	5.1	3.8	18.1	8.7
4.4	1.89	0.55	1.57	0.46	40.6	4.5	5.1	3.8	18.5	9.1
4.5	1.90	0.56	1.59	0.48	40.9	4.7	5.5	4.2	18.4	8.9
4.6	1.92	0.58	1.61	0.50	41.0	4.9	5.5	4.2	19.0	9.5
4.7	1.94	0.60	1.62	0.51	40.9	4.7	5.6	4.4	19.5	10.1
4.8	1.96	0.61	1.64	0.53	41.1	5.0	5.6	4.3	19.8	10.3
4.9	1.98	0.64	1.65	0.54	40.6	4.5	5.9	4.6	20.0	10.5
5.0	1.98	0.64	1.66	0.55	41.2	5.1	6.0	4.7	20.1	10.6

Table A3.3 The average effect of simulated random error in weight, with mean zero and standard deviation s, on the standard deviation of the WAZ and WHZ, and on the estimates of underweight, overweight, and wasting. Children age 0-4 in the Peru 2012 DHS survey.

	Simulated st.dev. of Z scores				Simulated prevalence of outcomes (%)					
	WAZ		WHZ		Underweight		Overweight		Wasted	
s	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change
0.0	1.05	0.00	0.99	0.00	4.3	0.0	6.0	0.0	0.7	0.0
0.1	1.07	0.01	1.00	0.01	4.5	0.2	6.2	0.2	0.7	0.0
0.2	1.08	0.02	1.02	0.03	4.7	0.4	6.5	0.5	0.9	0.2
0.3	1.09	0.04	1.05	0.06	5.0	0.8	6.9	0.8	1.1	0.4
0.4	1.11	0.06	1.09	0.10	5.5	1.2	7.4	1.4	1.5	0.8
0.5	1.14	0.08	1.15	0.15	6.2	1.9	8.1	2.0	1.9	1.2
0.6	1.17	0.12	1.20	0.21	6.8	2.5	8.8	2.8	2.5	1.8
0.7	1.21	0.15	1.27	0.28	7.5	3.2	9.8	3.8	3.2	2.5
0.8	1.25	0.19	1.33	0.34	8.3	4.0	10.7	4.7	3.9	3.2
0.9	1.29	0.23	1.40	0.41	9.1	4.8	11.6	5.6	4.7	4.1
1.0	1.33	0.28	1.46	0.47	9.9	5.6	12.6	6.6	5.6	4.9
1.1	1.38	0.33	1.53	0.54	10.8	6.5	13.6	7.6	6.4	5.7
1.2	1.43	0.37	1.59	0.60	11.6	7.3	14.7	8.7	7.2	6.6
1.3	1.47	0.42	1.65	0.66	12.4	8.2	15.7	9.7	8.1	7.4
1.4	1.52	0.47	1.71	0.72	13.3	9.0	16.6	10.6	8.9	8.2
1.5	1.57	0.52	1.77	0.78	14.2	9.9	17.6	11.6	9.7	9.0
1.6	1.62	0.57	1.83	0.84	15.0	10.7	18.5	12.5	10.5	9.8
1.7	1.67	0.62	1.88	0.89	15.7	11.4	19.4	13.4	11.2	10.5
1.8	1.72	0.66	1.93	0.94	16.5	12.2	20.2	14.2	11.9	11.2
1.9	1.76	0.71	1.98	0.99	17.2	12.9	20.9	14.9	12.6	11.9
2.0	1.80	0.74	2.02	1.03	17.7	13.4	21.6	15.6	13.1	12.4

Table A3.4 The average effect of simulated random error in weight, with mean zero and standard deviation s, on the standard deviation of the WAZ and WHZ, and on the estimates of underweight, overweight, and wasting. Children age 0-4 in the Nepal 2016 DHS survey.

	Simulated st.dev. of Z scores				Simulated prevalence of outcomes (%)					
	WAZ		WHZ		Underweight		Overweight		Wasted	
s	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change
0.0	1.08	0.00	1.11	0.00	26.8	0.0	1.3	0.0	9.5	0.0
0.1	1.08	0.00	1.12	0.01	27.0	0.2	1.3	0.0	9.8	0.3
0.2	1.09	0.02	1.14	0.03	27.1	0.4	1.3	0.1	10.3	0.8
0.3	1.11	0.03	1.17	0.06	27.3	0.6	1.5	0.2	11.0	1.6
0.4	1.14	0.06	1.22	0.11	27.7	0.9	1.7	0.4	12.1	2.6
0.5	1.17	0.09	1.27	0.16	28.0	1.2	1.9	0.6	13.2	3.7
0.6	1.20	0.12	1.32	0.21	28.4	1.6	2.2	1.0	14.3	4.9
0.7	1.25	0.17	1.39	0.28	29.1	2.3	2.7	1.4	15.6	6.1
0.8	1.29	0.21	1.45	0.34	29.7	3.0	3.1	1.8	16.9	7.4
0.9	1.33	0.25	1.51	0.40	30.1	3.3	3.7	2.4	17.7	8.3
1.0	1.38	0.30	1.58	0.47	30.8	4.0	4.2	2.9	18.9	9.4
1.1	1.43	0.36	1.64	0.53	31.5	4.7	4.9	3.6	19.9	10.4
1.2	1.48	0.41	1.70	0.59	32.1	5.3	5.5	4.2	20.7	11.2
1.3	1.53	0.45	1.75	0.64	32.7	5.9	6.2	4.9	21.4	11.9
1.4	1.59	0.51	1.81	0.70	32.8	6.0	7.0	5.7	22.0	12.6
1.5	1.64	0.56	1.87	0.76	33.4	6.6	7.8	6.5	22.7	13.2
1.6	1.68	0.60	1.91	0.80	33.9	7.1	8.5	7.3	23.1	13.6
1.7	1.73	0.65	1.96	0.85	34.1	7.4	9.2	7.9	23.7	14.2
1.8	1.77	0.70	2.01	0.90	34.6	7.8	10.1	8.8	23.8	14.3
1.9	1.82	0.74	2.05	0.94	34.7	7.9	10.8	9.5	24.0	14.5
2.0	1.86	0.78	2.08	0.97	35.1	8.3	11.4	10.2	24.1	14.6

Table A3.5 The average effect of simulated random error in age, with mean zero and standard deviation s, on the standard deviation of the HAZ and WAZ, and on the estimates of stunting, and underweight. Children age 0-4 in the Peru 2012 DHS survey.

	Simulated st.dev. of Z scores				Simulated prevalence of outcomes (%)			
	HAZ		WAZ		Stunted		Underweight	
s	Level	Change	Level	Change	Level	Change	Level	Change
0.0	1.08	0.00	1.05	0.00	20.6	0.0	4.3	0.0
5.0	1.08	-0.01	1.06	0.01	20.8	0.2	4.3	0.1
10.0	1.08	0.00	1.07	0.02	20.9	0.3	4.4	0.1
15.0	1.10	0.02	1.08	0.02	21.1	0.5	4.5	0.2
20.0	1.11	0.03	1.08	0.03	21.3	0.7	4.5	0.2
25.0	1.13	0.05	1.09	0.04	21.5	0.9	4.6	0.3
30.0	1.15	0.07	1.10	0.05	21.7	1.1	4.7	0.4
35.0	1.18	0.10	1.11	0.06	22.0	1.4	4.8	0.5
40.0	1.20	0.12	1.12	0.06	22.3	1.6	4.9	0.6
45.0	1.23	0.15	1.13	0.08	22.7	2.1	4.9	0.7
50.0	1.25	0.17	1.14	0.08	22.9	2.2	5.1	0.8
55.0	1.28	0.20	1.15	0.10	23.2	2.6	5.1	0.8
60.0	1.30	0.21	1.16	0.10	23.5	2.9	5.2	0.9
65.0	1.33	0.25	1.17	0.11	23.6	3.0	5.3	1.0
70.0	1.34	0.26	1.17	0.12	23.9	3.3	5.4	1.1
75.0	1.38	0.29	1.18	0.13	24.1	3.5	5.6	1.3
80.0	1.40	0.31	1.19	0.14	24.5	3.8	5.7	1.4
85.0	1.43	0.35	1.20	0.15	24.5	3.9	5.7	1.4
90.0	1.45	0.37	1.21	0.16	24.8	4.1	5.9	1.6

Table A3.6 The average effect of simulated random error in age, with mean zero and standard deviation s, on the standard deviation of the HAZ and WAZ, and on the estimates of stunting, and underweight. Children age 0-4 in the Nepal 20162 DHS survey.

	Simulated st.dev. of Z scores				Simulated prevalence of outcomes (%)			
	HAZ		WAZ		Stunted		Underweight	
s	Level	Change	Level	Change	Level	Change	Level	Change
0.0	1.34	0.00	1.08	0.00	36.1	0.0	26.8	0.0
5.0	1.35	0.01	1.08	0.00	36.4	0.2	27.2	0.4
10.0	1.36	0.02	1.09	0.01	36.5	0.4	27.2	0.4
15.0	1.36	0.02	1.09	0.01	36.4	0.3	27.1	0.3
20.0	1.38	0.04	1.10	0.03	36.7	0.6	27.3	0.5
25.0	1.40	0.06	1.12	0.04	36.3	0.2	27.1	0.4
30.0	1.41	0.07	1.12	0.04	36.8	0.7	27.2	0.4
35.0	1.44	0.10	1.13	0.06	36.6	0.5	27.1	0.4
40.0	1.45	0.11	1.14	0.07	36.7	0.6	27.2	0.5
45.0	1.48	0.14	1.17	0.09	36.9	0.8	27.1	0.4
50.0	1.50	0.16	1.16	0.09	37.0	0.9	27.2	0.4
55.0	1.52	0.18	1.18	0.10	37.2	1.1	27.2	0.5
60.0	1.54	0.20	1.19	0.11	37.2	1.1	27.2	0.5
65.0	1.55	0.21	1.19	0.11	37.6	1.4	27.3	0.5
70.0	1.57	0.23	1.20	0.12	37.8	1.6	27.5	0.7
75.0	1.59	0.25	1.21	0.14	37.9	1.8	27.3	0.5
80.0	1.60	0.26	1.22	0.15	37.9	1.8	27.4	0.6
85.0	1.64	0.29	1.24	0.16	37.4	1.3	27.2	0.5
90.0	1.64	0.30	1.24	0.16	37.7	1.6	27.5	0.7