INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

IFPRI

IFPRI Discussion Paper 00000

February 2017

Misreporting Month of Birth

Implications for Nutrition Research

Anna Folke Larsen

Derek Headey

William A. Masters

Poverty, Health, and Nutrition Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

AUTHORS

Anna Folke Larsen (afl@econ.ku.dk) is a postdoctoral fellow in the Department of Economics at the University of Copenhagen.

Derek Headey (d.headey@cgiar.org) is a senior research fellow in the Poverty, Health, and Nutrition Division of the International Food Policy Research Institute, Washington, DC.

William A. Masters (william.masters@tufts.edu) is a professor in the Friedman School of Nutrition and Department of Economics at Tufts University, Medford, MA, US.

Contents

Tables

Figures

ACKNOWLEDGMENTS

This paper was implemented under the grant “Advancing Research on Nutrition and Agriculture” (ARENA) from The Bill & Melinda Gates Foundation. Anna Folke Larsen received funding from the Danish Council for Independent Research.

ABSTRACT

Height-for-age z-scores (HAZs) and stunting status (HAZ<−2) are widely used to measure child nutrition and population health. However, accurate measurement of age is nontrivial in populations with low levels of literacy and numeracy, limited use of formal birth records, and weak cultural norms surrounding birthdays and calendar use. In this paper we use Demographic and Health Surveys data from 62 countries over the period 1990–2014 to describe two statistical artifacts indicative of misreporting of age. The first artifact consists of lower HAZs for children reported to be born earlier in each calendar year (resulting in implausibly large HAZ gaps between January- and December-born children), which is consistent with some degree of randomness in month of birth reporting. The second artifact consists of lower HAZs for children with a reported age just below a round age (and hence implausibly large HAZ gaps between children with reported ages just below and just above round ages), which is consistent with survey respondents rounding ages down more than they round ages up. Using simulations, we show how these forms of misreporting child age can replicate observed patterns in the data, and that they have small impacts on estimated rates of stunting but important implications for research that relies on birth timing to identify exposure to various risks, particularly seasonal shocks. Moreover, the misreporting we identify differs from conventional age-heaping concerns, implying that the metrics described above could constitute useful markers of measurement error in nutrition surveys. Future research should also investigate ways to reduce these errors.

Keywords: nutrition; height-for-age; stunting; measurement error; child age

1. INTRODUCTION

Child height and stunting rates are a significant public health concern in developing countries. Around one-quarter of the world's preschool-age population is stunted (UNICEF 2015). Worryingly, stunting has been shown to have numerous short-term and long-term consequences, including increased childhood morbidity and mortality (Black et al. 2008, 2013), delayed gross and motor development (Grantham-McGregor et al. 2007), and long-term educational and economic consequences (Dewey and Begum 2011; Hoddinott et al. 2013). A wide range of research spanning multiple disciplines addresses the causes and consequences of stunting: a Google Scholar search of the terms "stunted" and "stunting" returns 160,000 and 110,000 results, respectively.[1] Stunting is also the preferred policy and program indicator for monitoring changes in undernutrition, and it is a widely used development target, including for the Sustainable Development Goals.

Given widespread attention to stunting in different contexts, it is clearly important to measure height-for-age accurately, and to understand any biases introduced by errors in existing data. Children are classified as stunted if their height (or length) is low relative to the World Health Organization's (WHO's) worldwide reference population of healthy children of the same age and sex (WHO 2006). Differences are measured as z-scores, in units of standard deviation relative to the mean height of healthy children at that age and sex. A height-for-age z-score (HAZ) of less than −2 standard deviations (SD) is considered stunted, while an HAZ<−3 SD is considered severely stunted.

Measuring length or height especially for young children is challenging and has received considerable attention from Demographic and Health Surveys (DHS) analysts—in particular, Assaf, Kothari, and Pullum (2015). However, the measurement of a child's exact age is also difficult and could lead to more substantial errors in HAZs and stunting rates, especially among children 0 to 24 months of age, since this is when young children tend to fall rapidly behind international growth standards (Leroy et al. 2014; Shrimpton et al. 2001; Victora et al. 2010). Moreover, it is highly likely that the actual age of many children in developing countries is not known to their parents because of low numeracy and literacy, lack of birth registration, and limited celebration of birthdays or regular use of conventional calendars. At the same time, statistical agencies rightly try to avoid the selection biases that would emerge if child ages were reported as missing (Croft 1991). Instead, survey enumerators are strongly encouraged to work with respondents to identify plausible ages. In the DHS Program, enumerators are trained to elicit an age in years or a birth year for each child, and to use salient events, festivals, or seasons to narrow down toward the best available estimate of birth month (ICF-Macro 2009). But although that approach avoids the aforementioned selection bias, it nevertheless creates scope for both random and systematic misreporting of age, both of which could bias HAZs and stunting rates.

To our knowledge, the only studies to examine errors in age reporting for these kinds of nutrition studies are those that have been implemented by DHS analysts, notably Pullum (2006) and Assaf, Kothari, and Pullum (2015). Those studies find that a few DHS surveys have nonresponse rates for child age as high as 30 percent but that other surveys in similar settings report birth years and months for 98 to 99 percent of respondents, implying that enumerators sometimes make considerable efforts to enter plausible dates in settings where true age is unlikely to be known. The only systematic bias identified by Pullum (2006) refers to when children are falsely reported as older than 59 months, presumably to speed up the interview process. For adult ages, surveys are often subject to heaping at round numbers or other cognitive anchors, but Assaf, Kothari, and Pullum (2015) use Myers's index to show that no more than 10 percent of children's ages would need to be reallocated to eliminate age heaping. As a result, age errors in children under five years of age are generally treated as random, with limited heaping of the type detectable using standard diagnostics such as Myers's or Whipple's indexes.

[1] "Chronic undernutrition" returns 120,000 items and "growth faltering" 52,600 results. "Height-for-age" returns 30,000 results and "length-for-age" 5,000 results. These searches were conducted on September 23, 2016.

In this study we use DHS data to show how measurement error in age can in fact introduce systematic artifacts in HAZ results. Those artifacts offer testable predictions and new ways to estimate the frequency with which errors occur and the magnitude of bias they introduce. Specifically we describe two potentially related artifacts involving month of birth (hereafter, MOB). The first might be characterized as random or quasi-random estimation of MOB with a given birth year, while the second is characterized by respondents seemingly rounding their children's ages down toward a round number (for example, two years) more than they round them up. Both artifacts are characterized by implausibly large discontinuities; in the first case in the relationship between HAZ and MOB (Artifact 1), and in the second case HAZ and age relative to a round age and a distinctly asymmetric age heaping around round ages (Artifact 2). We show that these systematic errors are prevalent in almost all DHS data (though more so among poorer and less educated populations), that they typically result in small biases in stunting estimates (except in extreme cases), that they can potentially lead to attenuation biases in studies that use exact birth timing to identify a child's exposure to shocks, and that they can lead to erroneous influences on seasonality's effects on child nutrition. These important implications for measurement and research clearly warrant more systematic efforts both to regularly document these errors and to reduce the extent of such errors through improved survey practices.

Section 2 of the paper provides an intuitive conceptual discussion of errors and artifacts in child ages and briefly discusses some resulting analytical implications. Section 3 describes our data and methods. Section 4 focuses on identifying evidence of the two biases described above using both graphical and regression-based tests. Using simulations, Section 5 illustrates how the suggested measurement errors can reproduce the empirical patterns found in Section 4, and how they potentially affect stunting rates. Section 6 discusses in greater depth the implications of these findings for nutrition research and measurement.

2. ERRORS AND ARTIFACTS IN CHILD AGES

In this section we aim to describe in more detail the two statistical artifacts described above, including how those errors might come about and how they would be manifest in data from conventional nutrition surveys. Whereas both of the artifacts likely stem from a common underlying cause—that respondents (usually mothers) do not know the exact age of their children, and that enumerators have no ideal method for extracting precise birth dates in such circumstances—each artifact is nevertheless statistically distinguishable from the other, as we describe below.

Artifact 1: Random or Quasi-random Estimation of Month of Birth

While survey approaches to measuring child age likely vary, DHS enumerators are generally encouraged first to ascertain a child's age in years and then to use salient events to get a (hopefully) more precise estimation of the specific MOB. Although many respondents may be confident in estimating their child's age in years (though errors in years of age cannot be ruled out), we believe that a large proportion of respondents have no solid basis for estimating a child's MOB, and hence they (or the enumerator) may calculate the year of birth correctly but submit a random or almost-random estimate of MOB. Although a random estimation of MOB is apparently symmetric and uninformative, any error of this type would generate an observable artifact: a linear gradient in HAZ by calendar MOB and a gap in HAZ between children whose reported MOB falls at the end of one calendar year and the start of the next. In the conventional Gregorian calendar this would correspond to an HAZ difference between January-born and December-born children, though in non-Gregorian calendars, such as in Nepal or Ethiopia, the gap could exist in April (Nepal) or September (Ethiopia).

To see how random MOB traces out a gap in HAZ between December- and January-born children, consider a child actually born in midyear (for example, June). If she is erroneously recorded as born earlier in the year (for example, January) she is actually younger than reported to be and therefore likely to be short for the reported age. If recorded as born later (for example, December) she is actually older and likely to be tall for the reported age. In addition, children who are actually born early in the year are likely to be assigned later months, and vice versa. The net result is an artifactual linear gradient in HAZ by reported birth month, with each successive month having increasingly large downward bias in age and hence upward bias in HAZ.

Whereas actual MOB could have genuinely causal relationships with height due to seasonality and exposure to idiosyncratic shocks, the purely random MOB error described above would produce an anomaly associated only with calendar dates. This anomaly has been reported in published work such as Lokshin and Radyakin (2012) and Dorelien (2015), though both studies attribute their observed gradients solely to exposure to climatic shocks. However, the former finds that HAZs across India rise quasi-linearly from the start to the end of the calendar year, for a cumulative December–January gap of 0.37 standard deviations. A gap of that magnitude is equivalent to the HAZ difference between children whose mothers have no education at all and those whose mothers completed secondary education. Similarly, across 30 countries in Africa south of the Sahara (SSA) with very different agroclimatic seasons, Dorelien (2015) also finds a gradient in HAZ from start to end of the calendar year that sums to 3.1 percentage points of difference in stunting prevalence between children reported to have been born in December of one year rather than January of the next.[2] Clearly, these large December–January gaps are unlikely to be related to genuine seasonality.

In addition to confounding attempts to uncover genuine seasonality effects, randomness in MOB estimation can create two additional problems. First, imprecision in birth dates can create a "weak instruments" problem for econometric analyses that rely on MOB to identify exposure to various shocks, policies, or programs. Indeed, in one study of exposure to conflict and drought in Rwanda the authors

[2] This is the median of 97 regression coefficients reported in her appendix, for three different models. The mean coefficient is 2.87.

acknowledged that they might be underestimating the true impacts of such shocks on HAZ because of mismeasurement of child age (Akresh, Verwimp, and Bundervoet 2011). Second, MOB errors will cause some upward bias in estimates of the prevalence of stunting, simply by increasing the spread of the HAZ distribution.

Artifact 2: Asymmetric Heaping around Round Numbers

As in Artifact 1, we motivate Artifact 2 by a process in which enumerators first ask respondents about the child's age in years and then prompt for information on likely MOB. However, if the respondent cannot provide an MOB, the enumerator might use the survey month to work backward to an estimated MOB. For example, a child reported as being two years old and surveyed in June 2016 could be assigned an MOB of June 2014. Classical heaping of this form would result in larger numbers of children being assigned "round" ages such as 12, 24, or 36 months, and it would produce a correlation between month of survey and MOB. However, the pattern we actually observe in age distribution in the DHS data is a sawtooth with the peak number of births reported being at ages just above round numbers—such as 13, 25, and 37 months—and then declining linearly with each successive month. This indicates that age is misreported for a share of the children with a reported age just above a round number. Furthermore, the HAZ is also peaking at ages just above round numbers and then declining to the age just below the next round age. Hence, the children with misreported ages must on average be older than they are reported to be. We believe that this asymmetric rounding stems from enumerators tending to assume that a child's age in years is correctly estimated, and then proceeding to prompt the respondent to estimate whether their child is exactly two years old, or somewhat older than two. In reality, some children who are 29 months old may be incorrectly classified as two years old. Thus, many children in a survey characterized by these problems will be classified as younger than they actually are, and hence less likely to be stunted. To our knowledge there is no published work that documents this phenomenon in DHS or other nutrition surveys.

3. DATA AND METHODS

Although the phenomena we describe are not confined solely to the DHS surveys, in this study we confine our analysis to DHS data, details of which can be found in ICF International (2015), for several important reasons. First, the DHS Program is the single largest source of nationally representative nutritional data in the developing world, and the surveys are widely used for nutrition monitoring and analysis by the World Health Organization, the Global Nutrition Report, and many other institutions and individual researchers. Second, the availability of DHS surveys for multiple countries allows us to draw comparisons across countries with different agroclimatic seasons, different levels of education and birth documentation, and different cultural norms. Third, the DHS surveys are relatively standardized, with enumerators receiving similar training on topics such as age measurement.

In this paper we set out to test for MOB errors in all available DHS surveys that collected anthropometric indicators for children 0 to 59 months old (surveys that only sampled children 0 to 36 months were excluded), resulting in the sample sizes described in Table 3.1. In total this sample includes just under 1 million children from 183 surveys covering 62 countries, with almost half the children coming from SSA. Most regions have sizable samples, though we do point out that several regions are dominated by just a few countries. The East Asia and Pacific (EAP) region, for example, includes only Cambodia and East Timor, and the large Middle East and North Africa (MNA) sample is heavily dominated by Egypt and Jordan. However, DHS coverage of SSA and South Asia—the two regions with the highest undernutrition burdens—is excellent.

Table 3.1 Sample sizes for anthropometric data by region for 62 countries

Region	Countries	Child observations	Frequency
East Asia and Pacific	2	19,447	2.0%
Europe and Central Asia	7	18,653	1.9%
Latin America and Caribbean	10	222,255	22.4%
Middle East and North Africa	5	145,081	14.7%
South Asia	5	98,260	9.9%
Africa south of the Sahara	33	486,535	49.1%
Total	62	990,231	100.0%

Source: ICF International (2015).

Anthropometric data collection has been a key component of DHS surveys since 1986, focusing on the heights and weights of children under the age of five who stayed in the household the night before the survey. DHS surveys incorporate a number of steps to improve data quality. Interviewers are typically national staff from private or government statistical agencies (or some mix thereof) who receive extensive additional training on how to obtain and record height and weight measurements as well as the birth date; additional measures include field check tables, multiple layers of supervision, and field visits as part of the standard DHS protocol, as well as occasional research analyses of data quality (Assaf, Kothari, and Pullum 2015; Pullum 2006).

In terms of anthropometric indicators, a child's age is invariably asked of the woman taking care of a child, usually the mother. Child age is computed from the interview date and the birth date (year and MOB, and exact date if known), and length (0–24 months) or height (25–59 months) measures are recorded for any children classified by an enumerator as falling within the 0-to-59-month age range. While height and length measurement is relatively transparent, it is less clear how enumerators measure age if respondents are uncertain about birth dates and MOB. In principle, enumerators first ask respondents for the month and year of birth of all children born in a household, living or dead, as well as the exact date (day) of birth and the age at last birthday of all living children (though this last variable is not directly used to estimate ages). If no day in the month is recorded for birth, the DHS assigns the number 15. If no month is recorded, enumerators have the option of recording this as missing, but since relatively few children in the DHS have missing data on MOB it would appear that enumerators are strongly encouraged to extract an estimate from respondents. As a result relatively few children have imputed MOB data. For example, in India only 0.61 percent of children have no birth month directly recorded by an enumerator, and in Nigeria just 1.36 percent have no month recorded. Only Guinea (14.7 percent), Benin (8.7 percent), and Burkina Faso (5.6 percent) have notable proportions of imputation on birth month. As shown below, these imputations are not the source of the MOB artifacts discussed in this paper.

In the DHS and other studies, each child's HAZ is calculated from their age and height based on the following formula:

$$\textit{z-score} = (\textit{individual height} - \textit{median height in reference population})/\textit{SD of reference population} \quad (1)$$

where the reference population is specific to the child's sex and age in months and SD refers to standard deviation. Children with HAZs of less than −2 SD and less than −3 SD are classified as stunted and severely stunted, respectively. In the DHS, two main flags are often used to exclude outliers: one is to drop HAZ values that fall outside the WHO-recommend limits of −6 and +6, and the second is to drop absolute heights outside of plausible ranges, which are specified to be 45–110 centimeters for children measured lying down and 65–120 centimeters for children measured standing up (Assaf, Kothari, and Pullum 2015). Neither of these screens or other outlier-detection methods can explain or address the MOB artifacts discussed below.

Our analysis focuses on comparing child HAZ scores with their reported birth month within each calendar year and completed age, as well as the number of reported births by age in months. While there may be actual seasonal or idiosyncratic shocks explaining MOB–HAZ relationships, we limit their influence by pooling datasets with very different weather patterns, calendar systems, and cultural norms. Pooling the data in this way is expected to turn any actual seasonality into random noise, especially because our data span the northern and southern hemispheres with offsetting solar exposure. This leads to three specific null hypotheses of no real or artifactual month effects: (1) no association between HAZ and MOB, with no sawtooth between the start and the end of the calendar year; (2) no association between HAZ and round ages, with no sawtooth between the month before and the month after a round age; and (3) no association between number of births and child ages, with no asymmetric rounding around round ages.

To test these predictions we first use graphical methods to describe bivariate relationships between HAZ and MOB, and HAZ and child age, and then use statistical tests to control for any genuine socioeconomic or biological determinants of birth timing and attained heights by month or age in months. These regressions test for significant associations between the residual variation in HAZ and reported calendar MOB, as well as asymmetric rounding bias relative to round age, using the following generic form:

$$H_i = \sum_{m=1}^{11} \beta_m \boldsymbol{MOB}_{m,i} + \sum_{d=1}^{11} \gamma_d \boldsymbol{months}_{d,i} + \delta \boldsymbol{D}_i + \theta \boldsymbol{X}_i + \boldsymbol{\mu}_s + \varepsilon_i, \quad (2)$$

where H_i is the nutrition indicator (HAZ or stunting) for child i, $\boldsymbol{MOB}_{m,i}$ are the 11 month-of-birth dummies where December is the reference category, and $\boldsymbol{months}_{d,i}$ are the 11 dummies for the age in months in addition to a round age where zero (that is, a round age) is the reference category (for example, a 27-month-old girl is three months in addition to her round age of two years). $\boldsymbol{D}_i$ is child demographics (child gender, age, and age squared), $\boldsymbol{X}_i$ is a series of parental/household control variables (household assets, parental education, total number of children, total number of adults, toilet availability, water source, a rural dummy), $\boldsymbol{\mu}_s$ refers to survey dummies for each country and survey wave, and ε_i is the error term that we cluster at enumeration areas. While (2) is the generic form of the equation we estimate, we also estimate regressions excluding $\boldsymbol{X}_i$ and excluding $\boldsymbol{X}_i$, $\boldsymbol{D}_i$, and $\boldsymbol{\mu}_s$ as well as regressions where either $\boldsymbol{MOB}_{m,i}$ or $\boldsymbol{months}_{d,i}$ is excluded, to see whether these biases are orthogonal to each other or interdependent in some way.

4. RESULTS

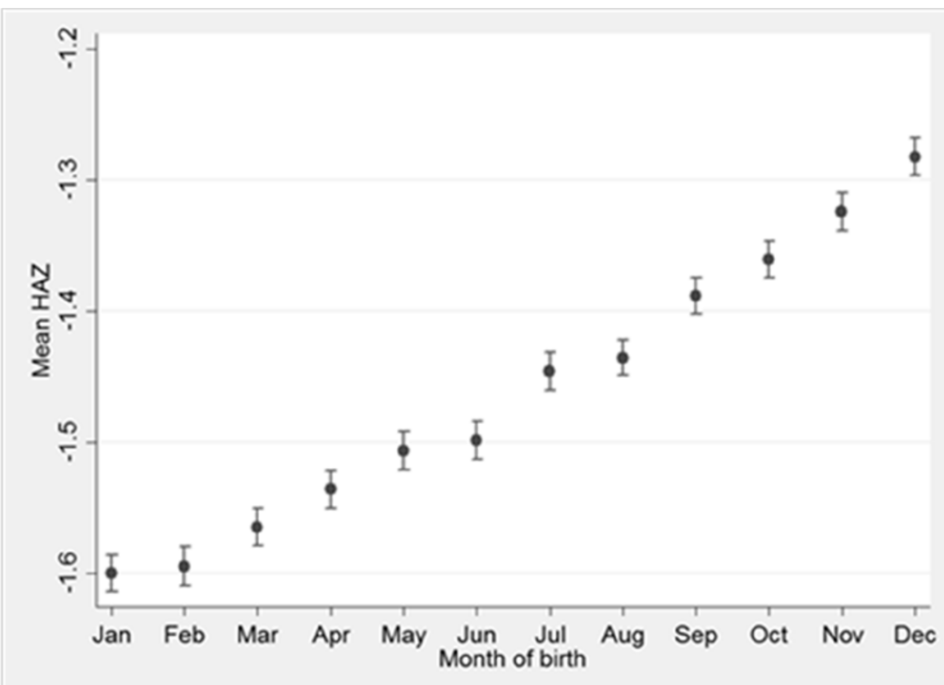
Our model of errors in recording child age aims to predict and explain three specific artifacts in DHS data or other anthropometric surveys: (1) a positive association between HAZ and calendar MOB, visible as a discrete gap in HAZ between the start and end of each calendar year; (2) a negative association between HAZ and age in months after completed years, leading to a discrete gap in HAZ between ages just below and just above a round age; and (3) a negative association between number of births and age in months after completed years, with a similar gap around the round ages.

Random or Quasi-random Estimation of Month of Birth

As discussed above, if calendar year of birth is recalled more accurately than calendar month, then even symmetric errors that are equally likely to overstate as understate age will produce a systematic artifact in the association between HAZ and MOB. Children who are misclassified as being born later in the calendar year will actually be older and hence taller, so their HAZ scores will be overstated, and vice versa for those misclassified as being born earlier in the calendar year. In the benchmark case when reported month is randomly drawn from a uniform distribution, there would be an upward slope in HAZ from the start to the end of each calendar year.

Figure 4.1 reports this association for all DHS data included in this study, which covers 62 countries. Consistent with the random MOB bias, HAZ rises with MOB in an approximately linear fashion, and produces a January–December differential of around 0.32 SD.

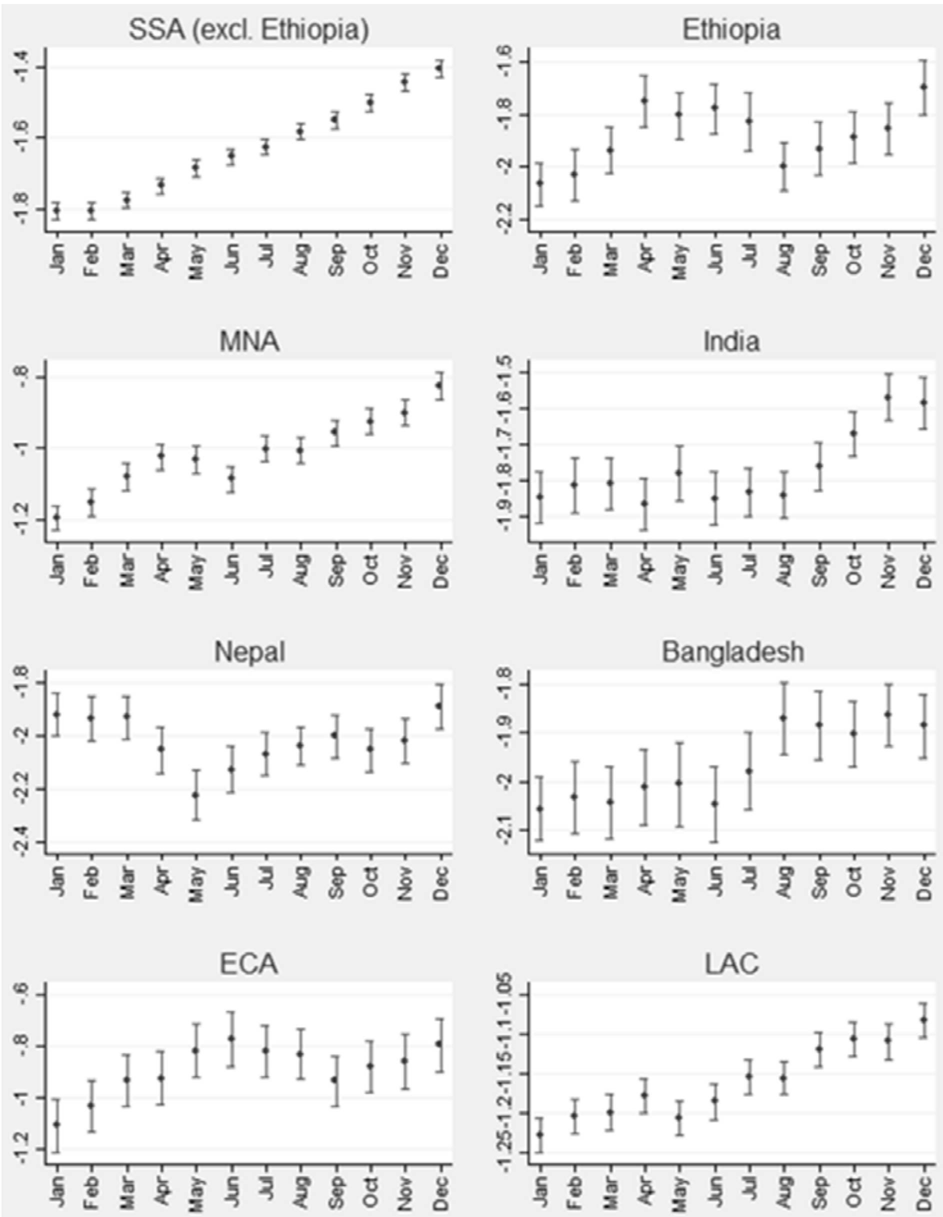
Figure 4.1 Average HAZ by MOB as a test for random MOB bias



Source: DHS data for 990,231 children from 62 countries, various years.
Note: HAZ = height-for-age z-scores; MOB = month of birth. The vertical bars indicate standard errors of the mean HAZ.

Figure 4.2 shows that the gradient exists in all major regions but differs among individual countries presumably due to their particular circumstances. In general, the January-to-December gradients are steeper in poorer countries, where respondents are less likely to be literate and numerate, less likely to have birth certificates for their children, and less likely to celebrate birthdays for cultural and socioeconomic reasons. In SSA and the Middle East and North Africa (MNA) the January–December gap rises to a large 0.4 SD, while the gradients in the two Southeast Asian (EAP) countries (not shown) are particularly large at about 0.5 SD. But as one might expect, the gradients in Eastern Europe and Central Asia (ECA) and Latin America and the Caribbean (LAC) are much more modest (less than 0.2 SD in the case of LAC), though the January–December gaps are still statistically significant. These two regions have much higher maternal education and wealth levels than the remaining regions, much stronger birth registration systems, and cultures that have stronger norms around celebration of birthdays.

Figure 4.2 HAZ–MOB gradients for major regions and selected countries



Source: DHS data for 960,012 children from 58 countries, various years.
Note: SSA = Africa south of the Sahara; MNA = Middle East and North Africa; ECA = Eastern Europe and Central Asia; LAC = Latin America and Caribbean.

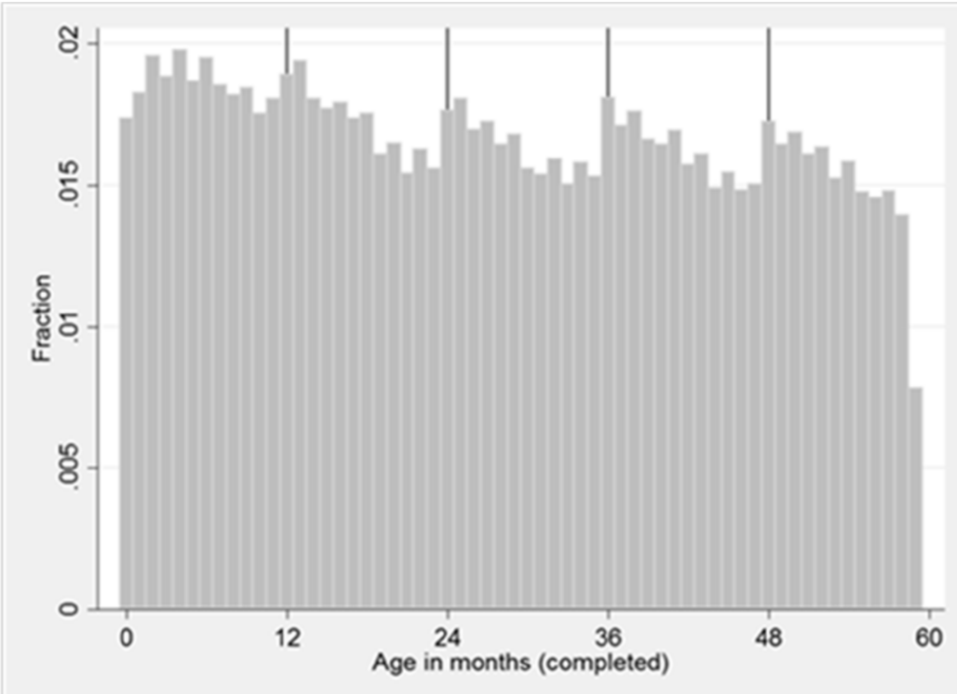
In addition to these major regions we also report HAZ–MOB gradients for some specific countries whose patterns are of special interest because of their unique calendars as well as underlying seasonality. First, the data for Nepal and Ethiopia provide additional corroboration of the MOB errors because of their different calendars. The Nepal DHS survey follows a Hindu calendar where each year generally begins in April, while Ethiopia follows an Orthodox calendar where each year begins in August. Errors in MOB based on their calendars would create an upward slope starting in those months, which is what we observe, although in Ethiopia we also observe a January–December gap, suggesting some respondents/enumerators use the conventional Gregorian calendar. For India, the figure closely replicates the result reported by Lokshin and Radyakin (2012), with an observed gap of 0.3 SD between December and the following January, but substantial nonlinearity in the relationship between MOB and HAZ for the rest of the year, perhaps pointing to genuine seasonality factors.

In the subsection "Econometric Tests of the Two Artifacts," we present regression results where we include the control variables described in Section 2. In appendix Figure A.1, we also report regression results for the major regions and selected countries in addition to the raw means (in blue): one that controls for child demographics and survey fixed effects (red) and one that also controls for parental and household characteristics (green). The described patterns of HAZ by MOB are not affected by the inclusion of these control variables. We also report a number of additional results in Appendix A. First, we show that these biases are often even larger in country-specific DHS surveys. In an extreme but important case—given its large population—Nigeria has a December–January HAZ difference of around 0.7 SD (Figure A.2). Second, we also show that these gradients are not explained by DHS imputation, since steep December–January gaps exist when imputed data are excluded (Figure A.3). However, consistent with our hypotheses, HAZ–MOB gradients are substantially flatter for children who (1) have more educated or literate mothers (Figures A.4 and A.5) and (2) have birth certificates (perhaps surprisingly, however, the gradients do not entirely disappear with birth certification) (Figure A.6). A final result of some note is that the steep HAZ–MOB gradient reported in Figure 4.1 does not vary much by age group (we report separate gradients for children 0–12, 12–24, 24–36, 36–48, and 48–59 months in Figure A.7); nor does it vary strongly by gender (Figure A.8), number of children of the mother (Figure A.9), or location or assets of the household (Figures A.10 and A.11).

Asymmetric Heaping around Round Numbers

If parents do not know exact birth dates, DHS survey manuals encourage enumerators to elicit an age in years and then use salient events to uncover a more precise age in months. In practice, once the enumerator has found the age in completed years he or she may be prone to prompt for a low number of additional months. Asking for whether the child is an additional nine to 11 months (leading to an age in months just below the next age in years) may be quite uncommon. Figure 4.3 reports the number of births that are reported to have occurred at each age in months, relative to the survey date, over all available DHS data. The vertical lines represent the round ages (in years) around which one might expect heaping. Instead of symmetrical errors around that cognitive anchor, what we observe is another asymmetric sawtooth pattern. There are peaks at or just after round ages, and then a declining number of births recorded at each successive age in months thereafter, to the next discontinuity at each completed age in years. In the appendix Figure A.12, we also show that this sawtooth pattern exists when we replace the *x*-axis variable in Figure 4.3 with the age in months in addition to a round age.

Figure 4.3 Asymmetric age heaping for aggregate DHS data

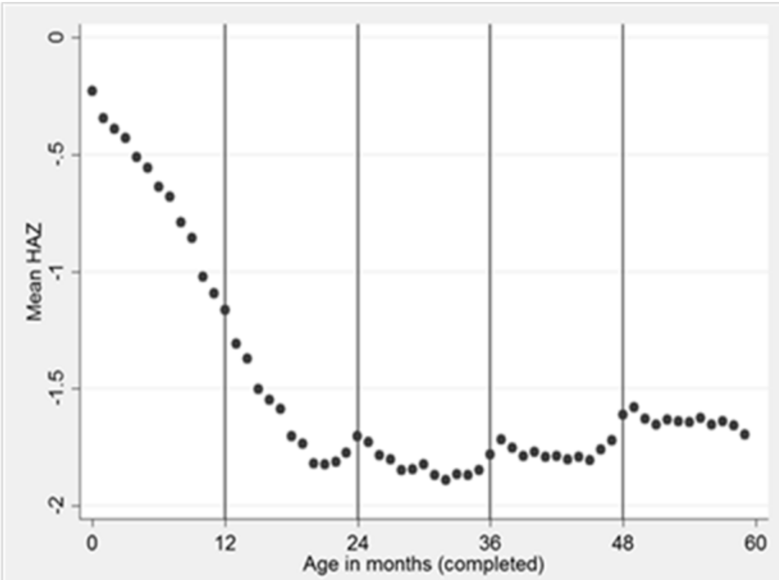


Source: DHS data for 990,231 children from 62 countries, various years.
Note: DHS = Demographic and Health Surveys. Thin red lines show round numbers.

The pattern shown in Figure 4.3 can be explained by what we term asymmetric rounding, in which respondents and enumerators first establish an age in years and then work backward to identify the additional months of age. Asymmetry arises when this process stops too early, thereby adding too few months, with peaks occurring where the largest share of children have a misreported age. The ages with the fewest estimation errors are the troughs of the age distribution, which are round ages plus 11 months. Consistent with this, our appendix Figure A.13 shows that the child age is significantly correlated with mother's education: mothers who report a round child age are less educated than those who report a round age plus 11 months.

The consequences of asymmetric rounding for research on child heights can be illustrated using graphs of HAZ by age. These graphs have become commonplace since the seminal works of Shrimpton et al. (2001) and Victora et al. (2010), who documented that most growth faltering takes place in the first 20 months after birth. Figure 4.4 reproduces this type of visualization, revealing how children at or just above two, three, and four years of age are artificially taller for their age, with systematically lower HAZ scores in each successive month thereafter and a discrete gap between the months just below and just above a round age. Since DHS surveys are often conducted in waves with similar survey dates in a given region, these peaks in HAZ scores might be interpreted as seasonality when they are more likely due to measurement error.

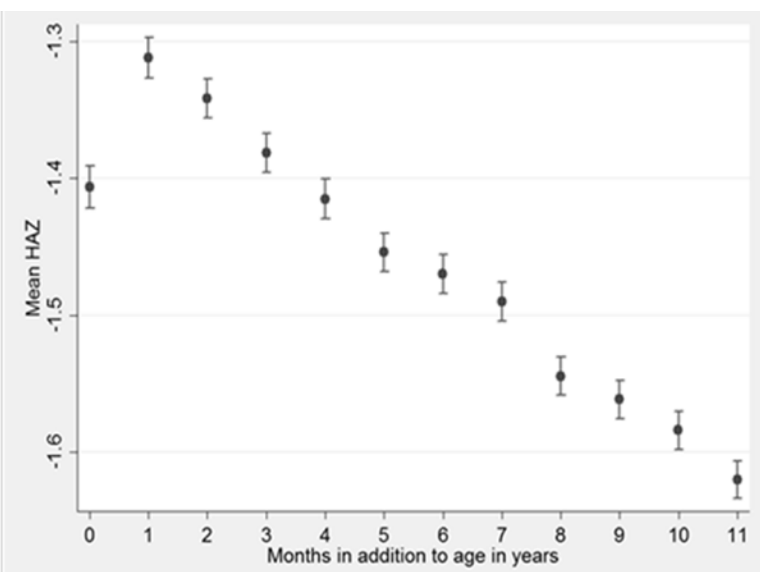
Figure 4.4 Mean HAZ by age in months



Source: DHS data for 990,231 children from 62 countries, various years.
Note: HAZ = height-for-age z-scores.

A different way to illustrate how asymmetric rounding affects analysis of HAZ is to calculate each child's age in months in addition to his or her round age in years. For example, being 11 additional months corresponds to an age just below a round age, for example, 23 months. This calculation allows us to draw a graph that is analogous to Figure 4.1, averaging over all children under five years of age. Results are in Figure 4.5, showing a linear pattern of discrepancy between one and 11 additional months of about 0.3 SD, which is the same magnitude as the discrepancy from Figure 4.1 between December- and January-born children.

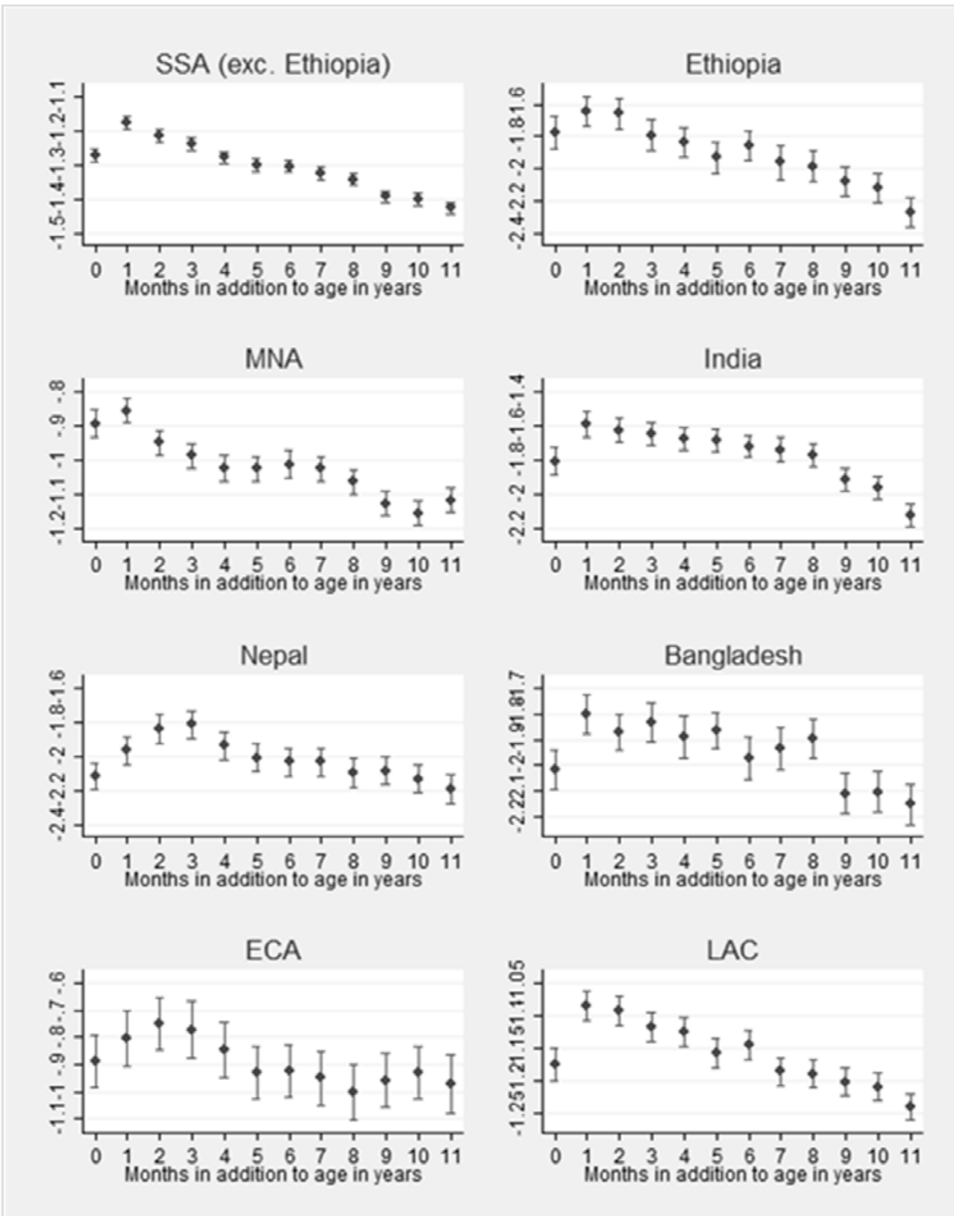
Figure 4.5 Mean HAZ by months in addition to age in years



Source: DHS data for 990,231 children from 62 countries, various years.
Note: HAZ = height-for-age z-scores. The vertical bars indicate standard errors of the mean HAZ.

Figure 4.6 shows how the pattern of HAZ in months beyond completed years varies across the major regions and countries depicted in Figure 4.2. A notable difference is that Ethiopia and Nepal, the two countries for which national calendar systems affected the gradient in Figure 4.2, are not particularly distinctive in Figure 4.6. That observation is consistent with errors due to asymmetric rounding being substantially independent from random MOB errors because the former emerge relative to round ages rather than calendar months. Appendix Figure A.14 also shows these results after controlling for additional covariates.

Figure 4.6 Mean HAZ by additional months for major regions and selected countries



Source: DHS data for 960,012 children from 58 countries, various years.
Note: HAZ = height-for-age z-scores. SSA = Africa south of the Sahara; MNA = Middle East and North Africa; ECA = Eastern Europe and Central Asia; LAC = Latin America and Caribbean. Corresponding graphs where controls are included can be found in appendix Figure A.14.

The patterns shown in Figures 4.5 and 4.6 and Figures 4.1 and 4.2 are analogous and similar in magnitude, but an important difference is that recorded months beyond a child's age in years are collinear with the actual age in months of the child. As a result, the negative HAZ gradient observed in Figures 4.5 and 4.6 could reflect a true decline in HAZ with each successive month of age, especially during a child's first two or three years after birth as illustrated in Figure 4.4. To address this issue we turn to statistical tests, controlling for age along with other covariates.

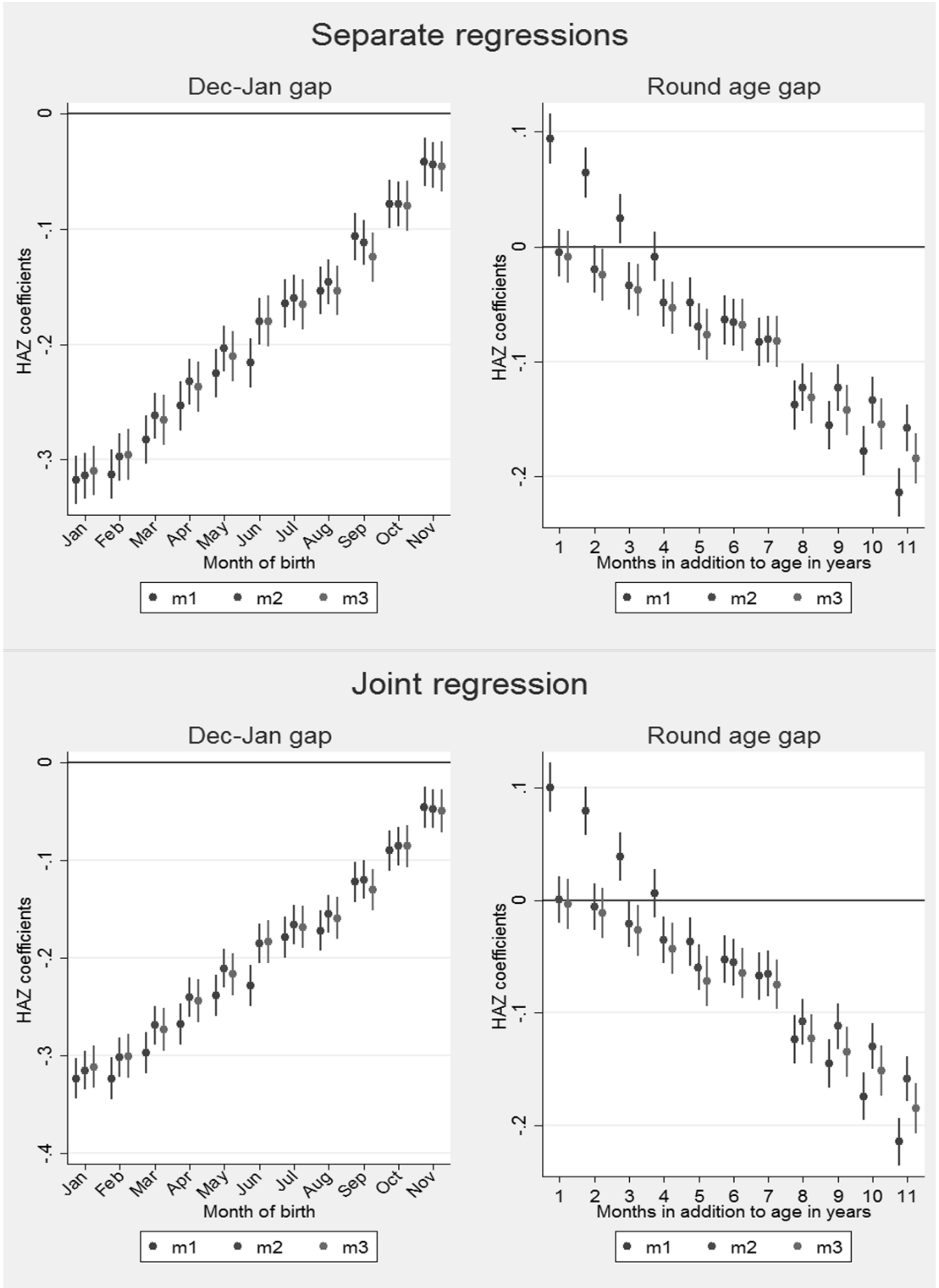
Econometric Tests of the Two Artifacts

Results of our multivariate econometric tests are presented using the same type of visualization as the bivariate relationships, first in terms of calendar months (and hence the December–January gap) and then in terms of months in addition to a round age (and hence the round age gap). Regressions use the specification from equation (2), with robustness tests that vary the statistical controls included in the model. We estimate separate models for each kind of artifact, and then a combined model that includes both calendar months and months in addition to age in years. For each of these two specifications we initially add no controls (model 1), then control for child age, age squared, sex, and survey fixed effects (model 2), and then also control for all available socioeconomic factors (model 3). We summarize the results of these regression estimates in Figure 4.7.

The main conclusions are threefold. First, the calendar MOB bias is entirely robust to confounding factors: a December–January gap of around 0.3 SD persists even after controlling for age and socioeconomic factors. Second, the asymmetric rounding bias appears to be partly explained by socioeconomic factors: the magnitude of the bias falls from 0.3 SD to 0.2 SD once we control for child age. Third, these two biases appear to exist independently of each other: specifying both sets of dummy variables in the model has no effect on the coefficients of either set.

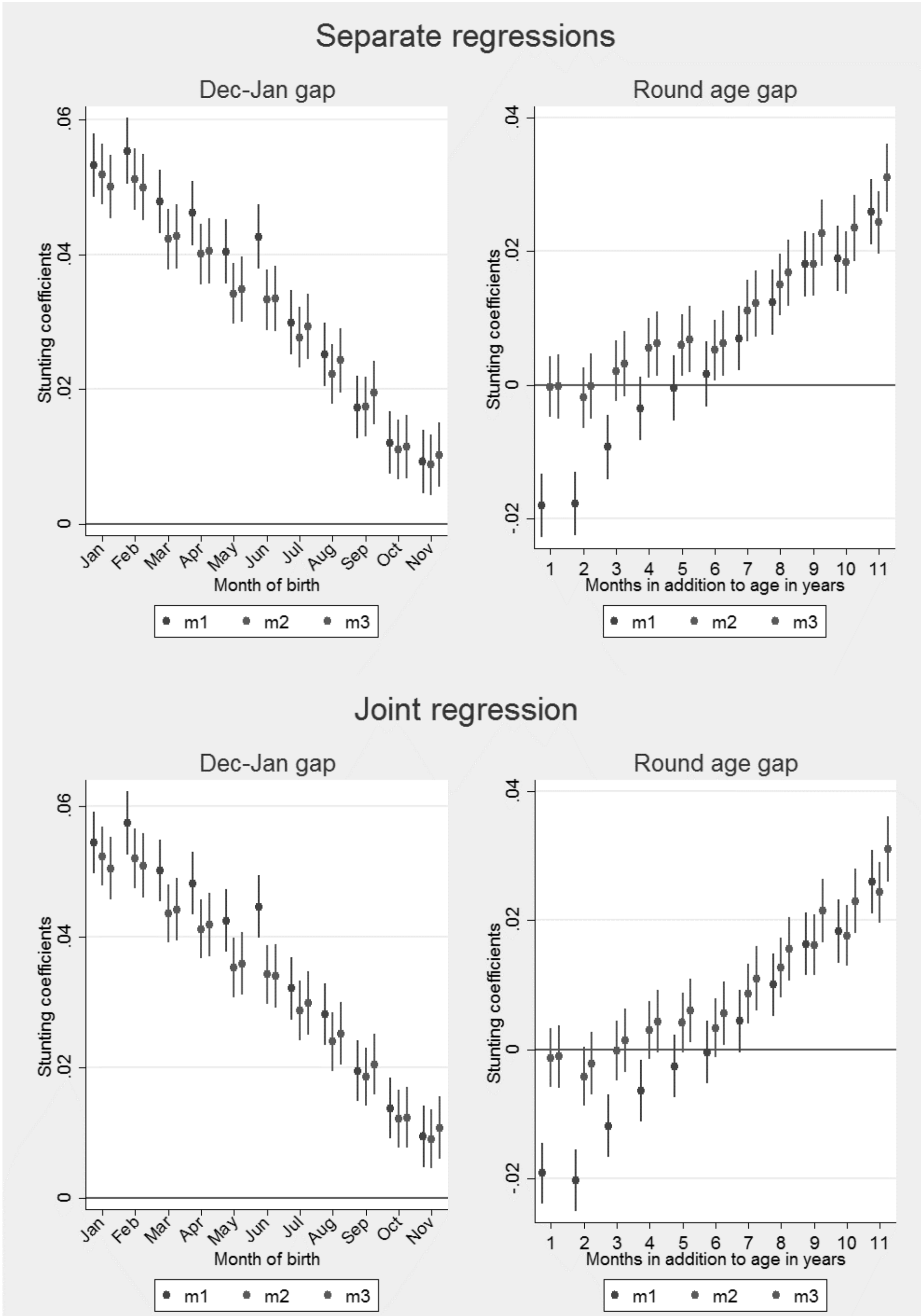
In Figure 4.8 we repeat this exercise for whether a child can be classified as stunted ($HAZ<-2$). The artifacts are again large, with December–January differences of just over 5 percentage points, irrespective of the model, while children reported being 11 months older than a round age are 2 to 3 points more likely to be classified as stunting compared with children with a round age.

Figure 4.7 Tests for month-of-birth and round age biases in HAZ scores



Source: DHS data for 990,231 children from 62 countries, various years.
Note: HAZ = height-for-age z-scores. See text for description of regression coefficients.

Figure 4.8 Tests for month-of-birth and round age biases in stunting status



Source: DHS data for 990,231 children from 62 countries, various years.
Note: See text for description of regression coefficients.

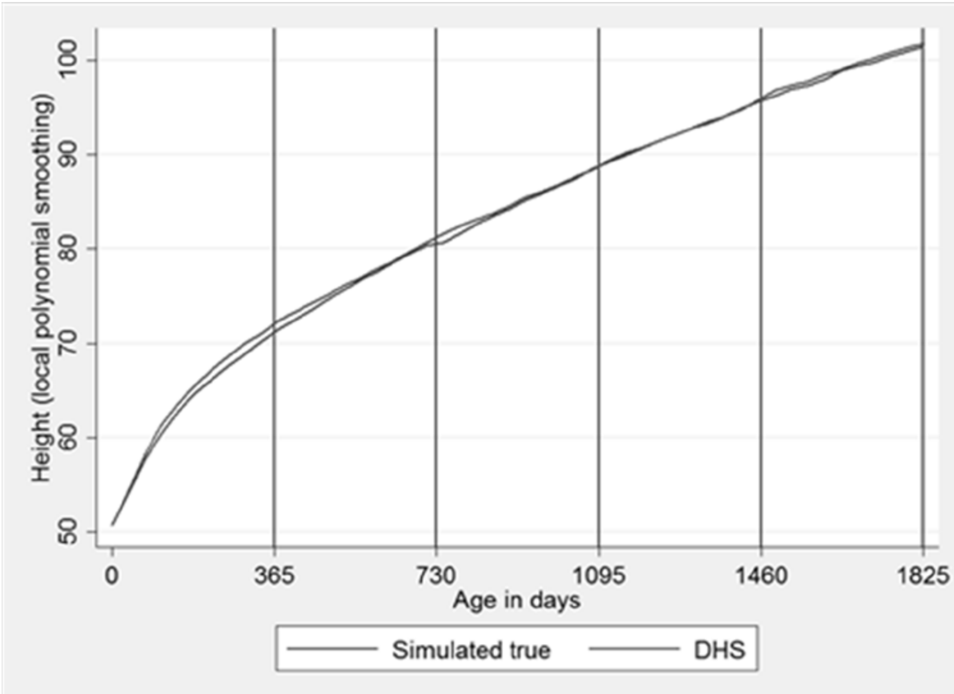
5. SIMULATIONS

We undertake two different simulation exercises to illustrate how measurement errors can lead to the two artifacts we found in the DHS data. First, we replicate the artifacts from a known error process, then we quantify how much of each type of error would be needed to explain the observed data, and finally we estimate how that error would affect mean HAZ and stunting rates. Since the regression analysis has shown that the two types of error are independent, we assess each artifact in two separate simulations. Both simulations start with a reference distribution of child heights by age, to establish a benchmark for the true levels of HAZ and stunting at each age. Next, we use each hypothesized error process to reassign some children to a mismeasured birth month, and calculate the resulting levels of HAZ and stunting at each reported age. This allows us to calculate what fraction of children would be required to have an erroneous age response to generate the patterns observed in the DHS data. Comparison of the reference distribution with the error-laden distributions yields the resulting magnitude of bias in mean HAZ and stunting rates.

Appendix B contains a detailed protocol for these simulations, so here we provide only the basic details and intuition of the simulation approach. The benchmark heights used in the simulations are drawn from normal distributions for a female population at each age in days. The mean and the standard deviation of the distribution at each age are derived from the WHO child growth standards for girls (WHO 2006), transformed to fit the observed level and variation of actual heights in the DHS. To replicate smoothed growth velocities observed in the DHS, our benchmark population has its growth velocity slowed relative to WHO standards by 7 percent less growth each day from 0 to six months, then 21 percent less growth each day from six to 24 months of age, and finally 10 percent less growth from 24 months onward. This benchmark growth trajectory corresponds well to the means in the DHS data as shown in Figure 5.1.[3] To adjust the SDs of height, we simply add 2 to the WHO SDs to approximate a realistic spread in the DHS HAZ distribution. We motivate this adjustment as a combination of overall measurement error (that is, not related to the artifacts described above) and increased dispersion due to variation in nutritional status of the children in the sample. We furthermore multiply the WHO SDs by 0.85 to account for the fact that the SDs in the DHS data increase less with age than the WHO SDs (a potential explanation of this phenomenon could be that errors in height measurements decrease with age). Figure 5.2 illustrates that the simulated SDs fit the slope of the SDs in the DHS data, although they are in general 1.5 times lower. We have chosen this specification to fit the simulated stunting rates to those in the DHS data and because the measurement errors that we simulate also increase the SD of height. The resulting simulated true HAZ by age is shown in Figure 5.3. It appears smooth compared with the DHS HAZ by age.

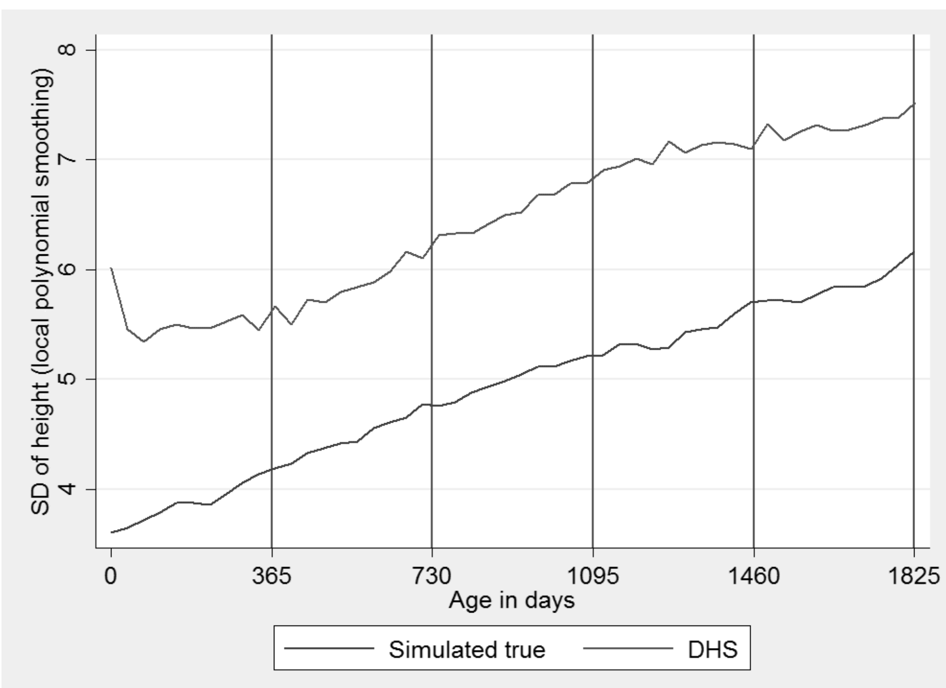
[3] The negative bump in the simulated heights just after 730 days of age (two years) is due to the fact that children are no longer measured recumbent, but standing. According to the WHO growth standards, standing height is 0.7 centimeters shorter than recumbent length.

Figure 5.1 Mean height by age (local polynomial smoothing), DHS data and simulated data



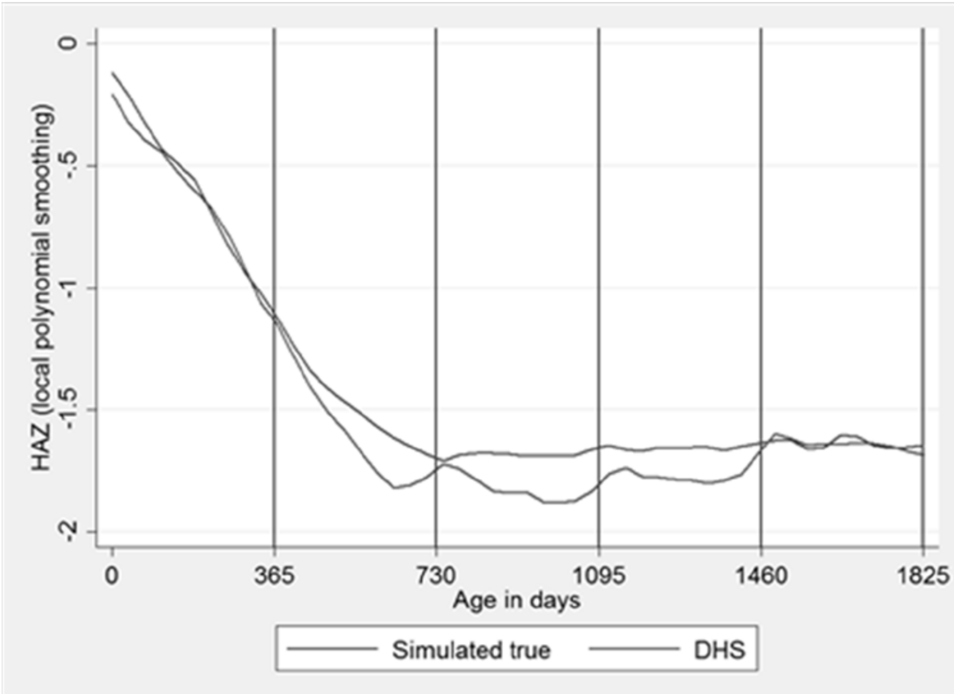
Source: DHS data for 488,307 girls from 62 countries and simulated data, various years.
Note: DHS = Demographic and Health Surveys.

Figure 5.2 SD of height by age (local polynomial smoothing), DHS data and simulated data



Source: DHS data for 488,307 girls from 62 countries and simulated data, various years.
Note: DHS = Demographic and Health Surveys; SD = standard deviation.

Figure 5.3 Mean HAZ by age (local polynomial smoothing), DHS data and simulated data



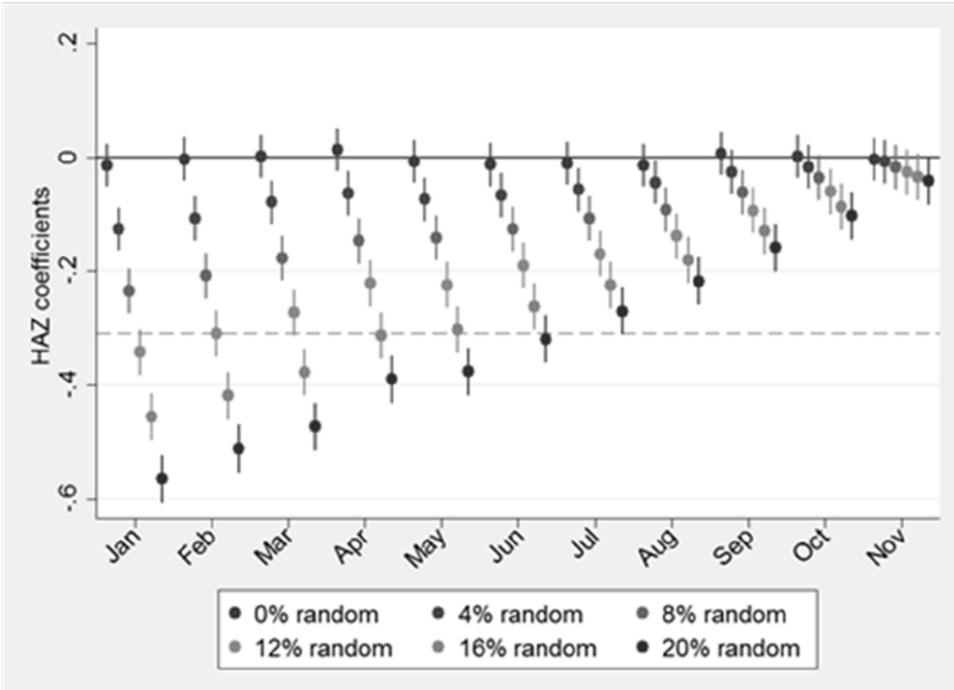
Source: DHS data for 488,307 girls from 62 countries and simulated data, various years.
Note: HAZ = height-for-age z-scores; DHS = Demographic and Health Surveys.

Replication of Random MOB Errors

The first step in replicating random MOB errors involves drawing a random birth month and birth day for each child in the simulated data. We then calculate the age with measurement error based on the random month and day along with a true birth year, and we use this random MOB age to calculate an HAZ with error. (Note that we include only children with an erroneous age below 60 months since the sample selection in the DHS surveys is based on the reported age; this implies that some children that are truly older than 60 months are included in the sample while some children truly below 60 months of age are excluded from the sample). We can then map the true simulated HAZ and the HAZ with error by the reported birth month (either true or random).

This comparison is shown in Figure 5.4, and it is clear that assigning random birth months yields almost exactly the same pattern observed in the DHS data, with a large December–January gap in mean HAZ. However, the gap is almost 10 times as large as in the DHS data shown in Figure 4.7, indicating that most parents in the DHS do not provide random MOBs. Hence, we vary the share of children in the simulated data who report the true birth month and who report a random birth month to find the share that fits the December–January gap to −0.31 SD as in the DHS data. This is also illustrated in Figure 5.4 and detailed in Table B.1 in Appendix B. A gap of −0.32 SD is produced by assigning 11 percent of the children with a random MOB. The 11 percent random MOB sample has a moderate stunting rate of 35.7 percent compared with 35.2 percent in the true data, and a severe stunting rate of 15.2 percent compared with 14.5 percent. Although these errors are small, we note that for particular surveys the December–January gap is much larger (0.7 SD in Nigeria), implying substantially higher degrees of randomness (25 percent in Nigeria) and somewhat larger biases in stunting estimates (0.9 percentage points and 1.6 percentage points for moderate and severe stunting in Nigeria). Moreover, this random MOB artifact will generally bias stunting estimates upward in a poorer and less educated population, slightly inflating the stunting differences between different socioeconomic groups.

Figure 5.4 Simulated HAZ with varying shares of children with random month of birth



Source: Simulated data.
Note: HAZ = height-for-age z-scores. Dashed line represents December–January gap in DHS data.

Replication of Asymmetric Rounding for Age in Months beyond Completed Years

In the DHS data we found that ages appear to be rounded down asymmetrically, toward age in completed years, resulting in many children being reported as younger than they actually are. To simulate this kind of measurement error, we reassign a fraction of the benchmark population a randomly generated age in the interval between their true age in completed years and their true age in days. This implies that the age calculated in complete years is still correct, but the reported number of months in addition to age in years is lower than the true number of months. Hence true age is thereby understated. As in the previous subsection, we use this (mis)reported age to calculate the HAZ with error.

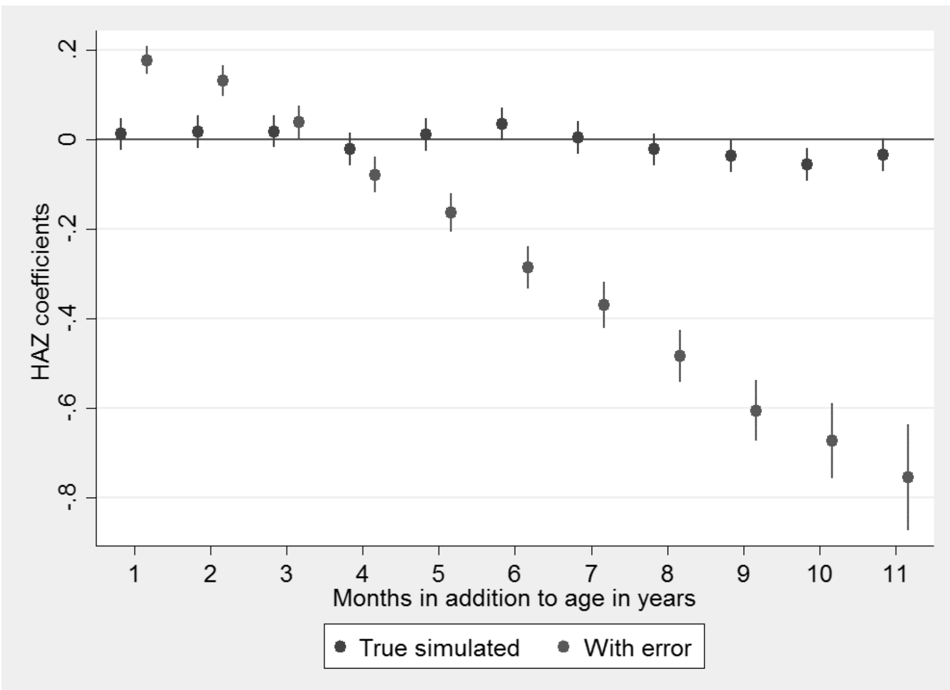
In Figure 5.5 we illustrate how mean HAZ depends on the number of months of age in addition to the age in years. For instance, the category of three months captures all children age 3 months, 27 months, 39 months, and 51 months. Since average age is increasing over the month categories, we control for quadratic age to account for the overall age pattern in HAZ. Figure 5.5 illustrates the HAZ coefficients for the month dummies relative to a round age in years for both the true simulated data and the data with this asymmetric rounding bias. It is clear this type of measurement error can generate the descending pattern in HAZ from one to 11 additional months that we found in the DHS data presented in Figure 4.5. The true simulated HAZ does not show the same descending pattern when we control for age.

As before, we can use our simulation to compute the fraction of children affected by this type of measurement error that would replicate the HAZ gap of 0.18 observed in DHS data between the month just below (that is, 11 months above) and just above a round age, when age is controlled for. We illustrate this calibration in Figure 5.6 and detail it in Table B.2 in Appendix B. We find that when 7 percent of the children have their age rounded down toward the age in completed years, we can reproduce the 11-month gap of 0.18 in the DHS data. Using the calibrated share of 7 percent affected by asymmetric rounding, we can draw the resulting age distribution to compare with the actual distribution in DHS data. Figure 5.7 illustrates this. The distribution has heaps at the round ages and then slowly declines down to round ages plus 11 months before it jumps up again at the next round age, just as we saw in the DHS age distribution depicted in Figure 4.3.

A final question is how asymmetric rounding affects stunting rates. Since it causes age to be understated on average, mean HAZ at each age will be too high, and we therefore expect the mismeasured population's stunting rate to be lower than its true rate. This is also what we find: The moderate stunting rate decreases almost 1 percentage point from 35.2 to 34.3 percent when 7 percent of the sample has an asymmetric rounding error in the reported age. The severe stunting rate is reduced from 14.5 to 14.1 percent. That result for Artifact 2 contrasts with the bias introduced by errors in calendar MOB (Artifact 1), which inflate stunting rates by symmetrically increasing the dispersion of HAZ. In contrast, Artifact 2 increases the mass in the upper tail of the HAZ distribution.

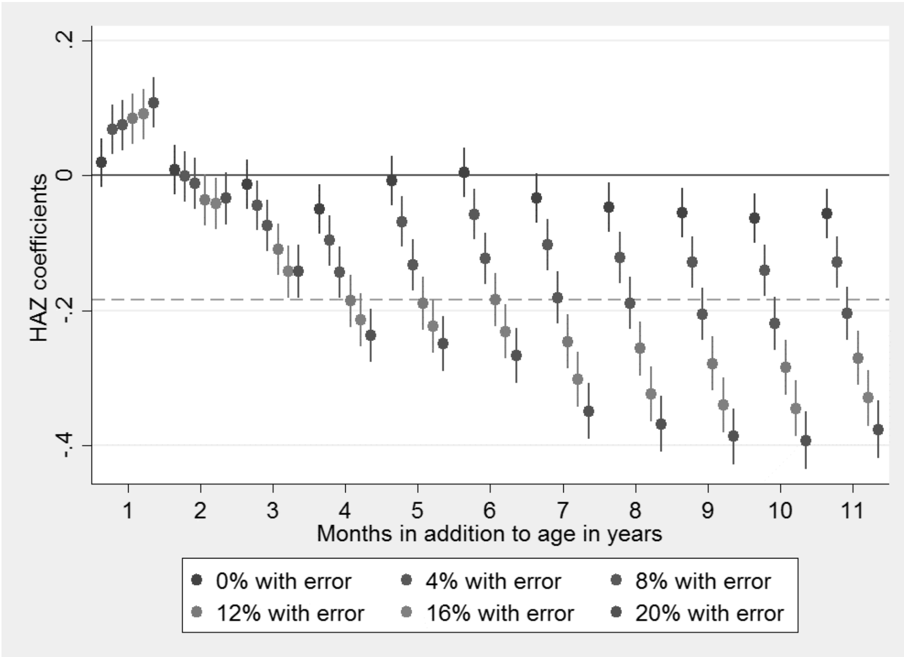
In summary, the two types of measurement error distort aggregate stunting rates in opposite directions, and when taken together, the simulations suggest that the moderate stunting rate is slightly too low in the DHS survey data, while the severe stunting rate is slightly too high.

Figure 5.5 HAZ by age in months in addition to age in years, true simulated data, with asymmetric rounding error



Source: Simulated data.
Note: HAZ = height-for-age z-scores.

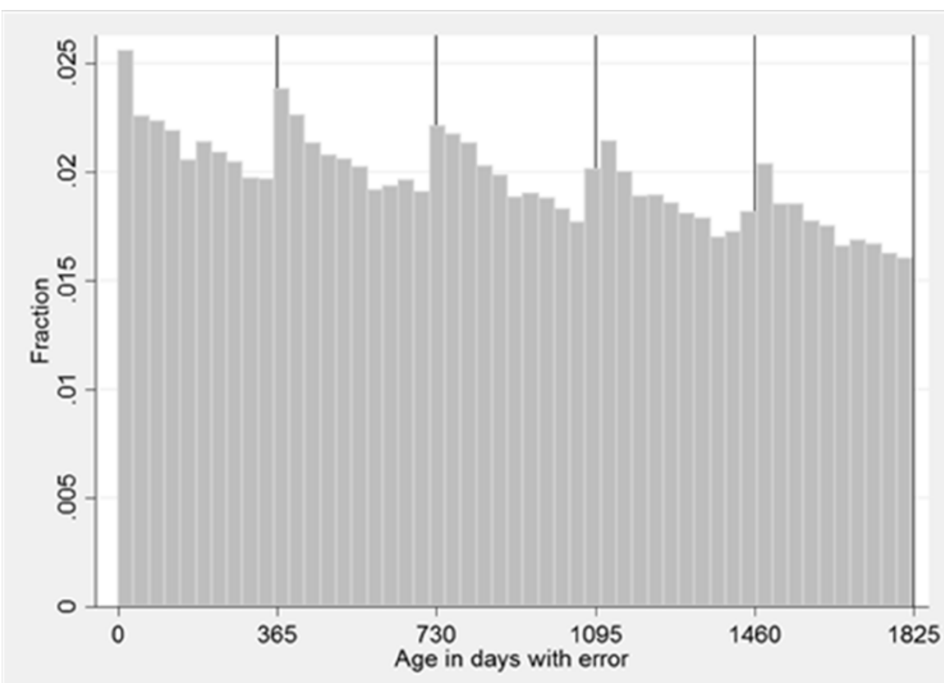
Figure 5.6 Simulated HAZ with varying shares of children with asymmetric rounding error



Source: Simulated data.

Note: HAZ = height-for-age z-scores. Dashed line represents 11-month gap in DHS data.

Figure 5.7 Simulated age distribution with 7 percent asymmetric rounding error



Source: Simulated data.

6. DISCUSSION

This paper examines the consequences of mismeasured month of birth resulting in misreported age. Using all suitable DHS data from 1990 to 2014 we find two puzzling anomalies that can readily be explained by this kind of error, especially in countries and regions with low levels of maternal education, low socioeconomic status, and underused birth registration systems. The first bias appears to be explained by random or quasi-random estimation of MOB, while the second appears to stem from a tendency for enumerators/respondents to round down ages more than they round up. Moreover, the two biases appear to operate independently.

What are the implications of these findings for nutritional research? We identify three areas of concern. First, studies that estimate stunting rates will be affected by the two kinds of error, but in opposite directions. Errors in calendar MOB do not affect median HAZ but do increase the spread of the HAZ distribution, leading to fatter tails and more children recorded to the left of the −2 SD cutoff. Errors in age relative to completed years lead to underreporting of age, on average, and hence overestimation of HAZ scores. For the most part these two errors cancel out, which in some sense is comforting news. However, we caution against ignoring these errors, since particular surveys may be more predisposed to one type of error than the other, and since the prevalence of these errors can be very high in particular surveys (for example, Nigeria).

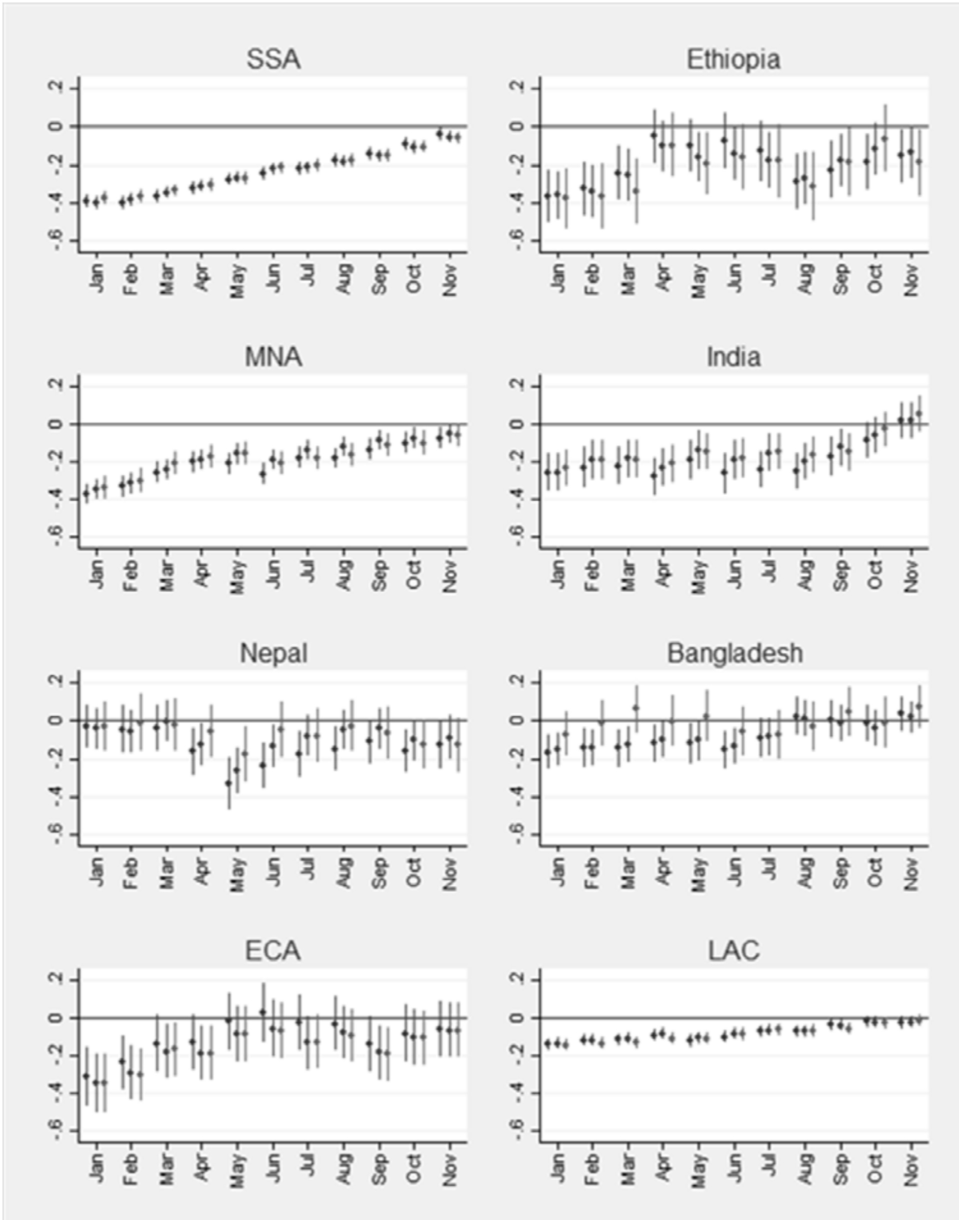
Second, the many studies that use birth timing relative to climatic or other shocks for estimation of causal effects on child height and stunting should take these errors into account. Here there are potentially two types of concerns. The first is attenuation bias, which emerges from the use of age in months to identify exposure to shocks, such as climatic shocks (Mulmi et al. 2016; Tiwari, Jacoby, and Skoufias 2017) or economic or political shocks (Akresh, Verwimp, and Bundervoet 2011). As a result, type II errors are a concern here. A second bias is perhaps more severe, stemming as it does from a simultaneity problem. Work such as Lokshin and Radyakin (2012) and Dorelien (2015) implicitly assume that MOB errors merely attenuate the effects that they find. In contrast the results in this paper show that MOB misreporting introduces a simultaneity problem since misreported age implicitly appears on both the left-hand side of the equation in HAZ and on the right-hand side as a reported MOB (as in Lokshin and Radyakin 2012 and Dorelien 2015). These papers erroneously tend to find that births in earlier months in the calendar lead to worse HAZ outcomes, and therefore warrant season-specific interventions such as safety nets or additional public health interventions. While there may indeed be true seasonality in HAZ outcomes, the errors in DHS and similar surveys arguably create an insurmountable confounding problem.

This is unfortunate, because studies from countries with highly accurate birth registration exploit timing of exposure to shocks very effectively (Currie and Rossin-Slater 2013; Deschenes, Greenstone, and Guryan 2009; Messias et al. 2006) and often demonstrate significant associations between climatic shocks and long-run health and nutrition outcomes. Clearly, one would expect developing-country populations to be far more vulnerable to seasonal insults to nutrition because of the inability of poor populations to effectively protect themselves against adverse shocks (Chambers 1982). Moreover, studies of rural Gambian communities, which circumvented age-misreporting problems by focusing on indicators of birth outcomes, suggest the existence of strong seasonal determinants of children being born prematurely or being short for gestational age, and of poor maternal weight and anemia during pregnancy (Prentice et al. 1981; Rayco-Solon, Fulford, and Prentice 2005). Although there are a handful of similar but much earlier biological studies in other developing countries—see Rayco-Solon, Fulford, and Prentice (2005) for a review—surprisingly little is known about the long-term nutrition, health, and cognitive impacts of birth seasonality across different ecologies and socioeconomic contexts. Hence, this would still seem an area where much more research is needed, albeit with improved survey instruments.

On this last point, our study draws attention to a significant source of measurement error in DHS surveys, but one that exists in other similar multipurpose surveys conducted in underdeveloped populations. (Though not reported here, we find similar errors in the Living Standards Measurement Study surveys implemented by the World Bank and the Multiple Indicator Cluster Surveys implemented by UNICEF.) This study provides simple, implementable markers of measurement error in age that could be used to identify the extent of these biases and to gauge the effectiveness of attempts to reduce measurement error in child age. Measuring children's age more accurately in these settings ultimately requires further experimentation in the field, but some avenues for exploration include the adoption of more sophisticated event calendars, the questioning of both mothers and fathers, and the training of enumerators to be more aware of the serious consequences of age misreporting documented in this paper.

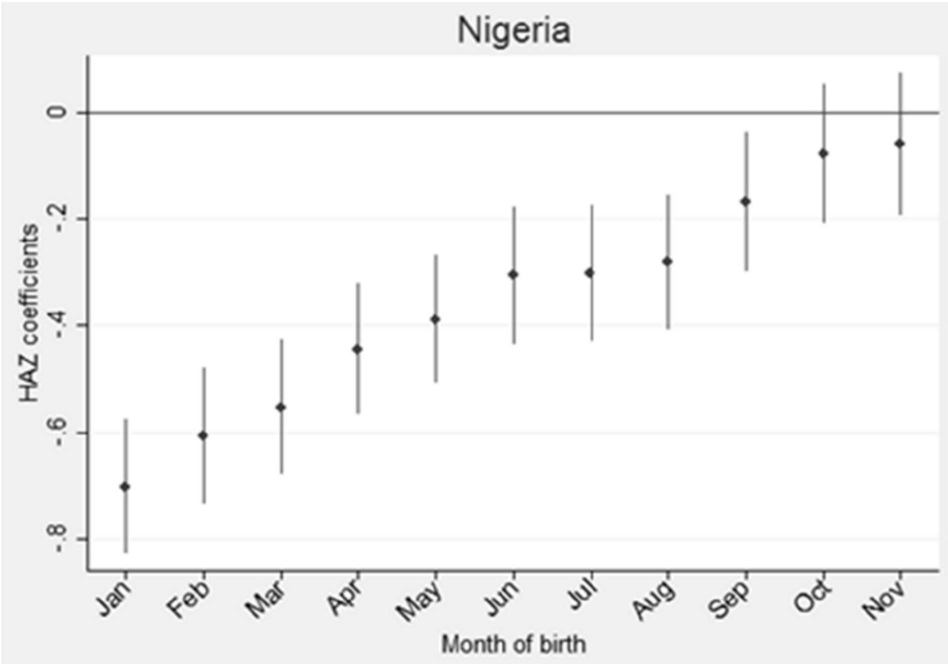
APPENDIX A: SUPPLEMENTARY FIGURES

Figure A.1 HAZ–MOB gradients for major regions and selected countries with controls



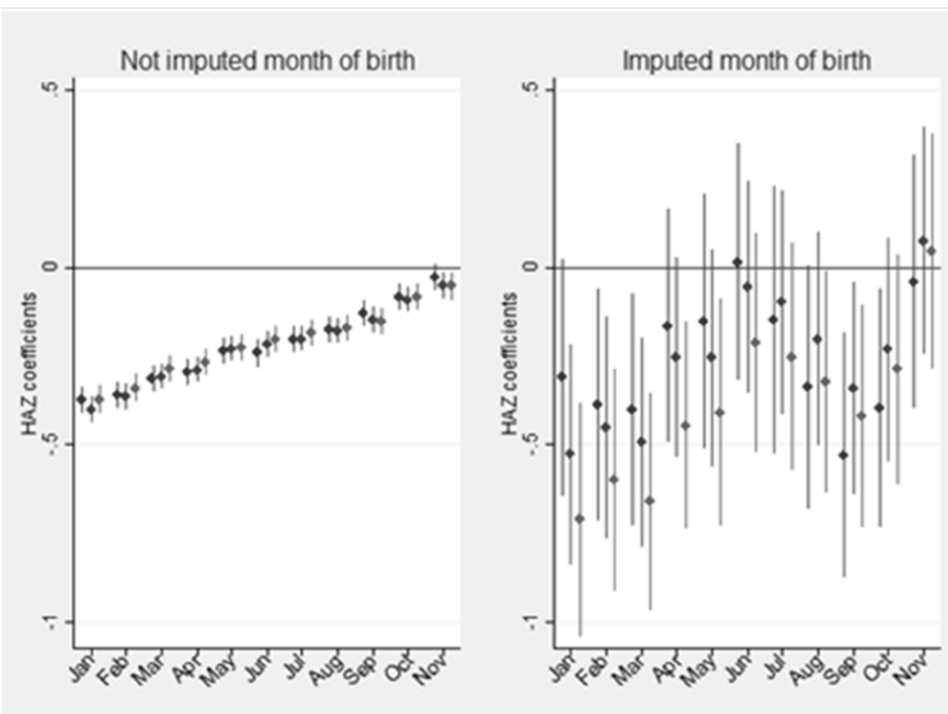
Source: DHS data for 960,012 children from 58 countries, various years.
Note: HAZ = height-for-age z-scores; SSA = Africa south of the Sahara; MNA = Middle East and North Africa; MOB = month of birth; ECA = Eastern Europe and Central Asia; LAC = Latin America and Caribbean.

Figure A.2 HAZ by MOB for Nigeria



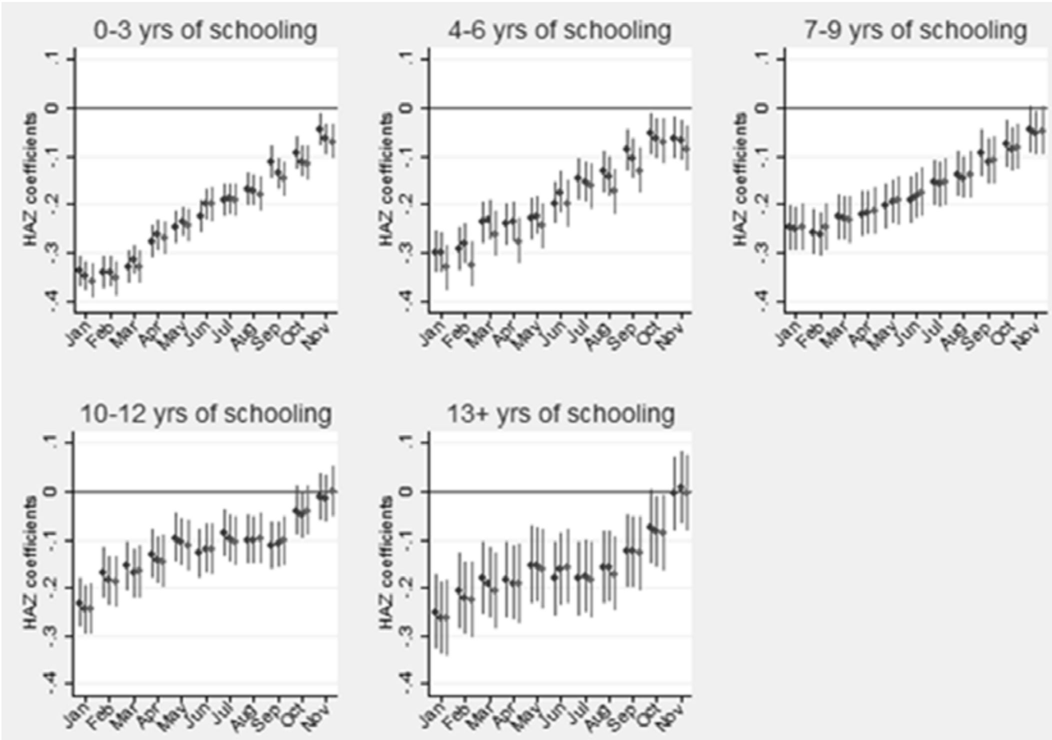
Source: DHS data from 60,893 children from surveys in 1990, 2003, 2008, and 2013.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Coefficients are from regressions where all control variables are included.

Figure A.3 HAZ by MOB for children with and without imputed birth month



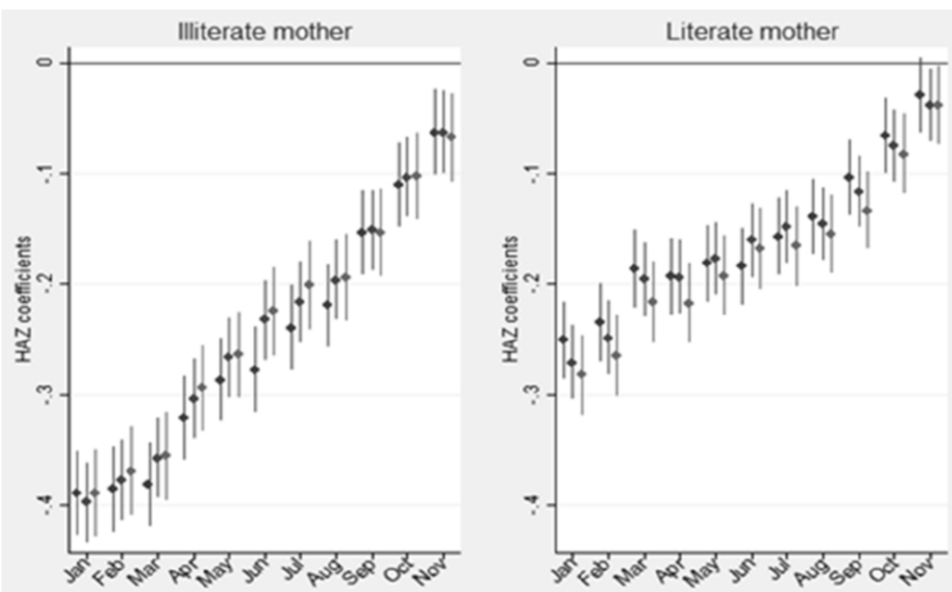
Source: DHS data from 396,299 children in 17 countries in Africa south of the Sahara and Egypt and India.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Two percent of children have imputed month of birth.

Figure A.4 HAZ by MOB depending on the mother's education



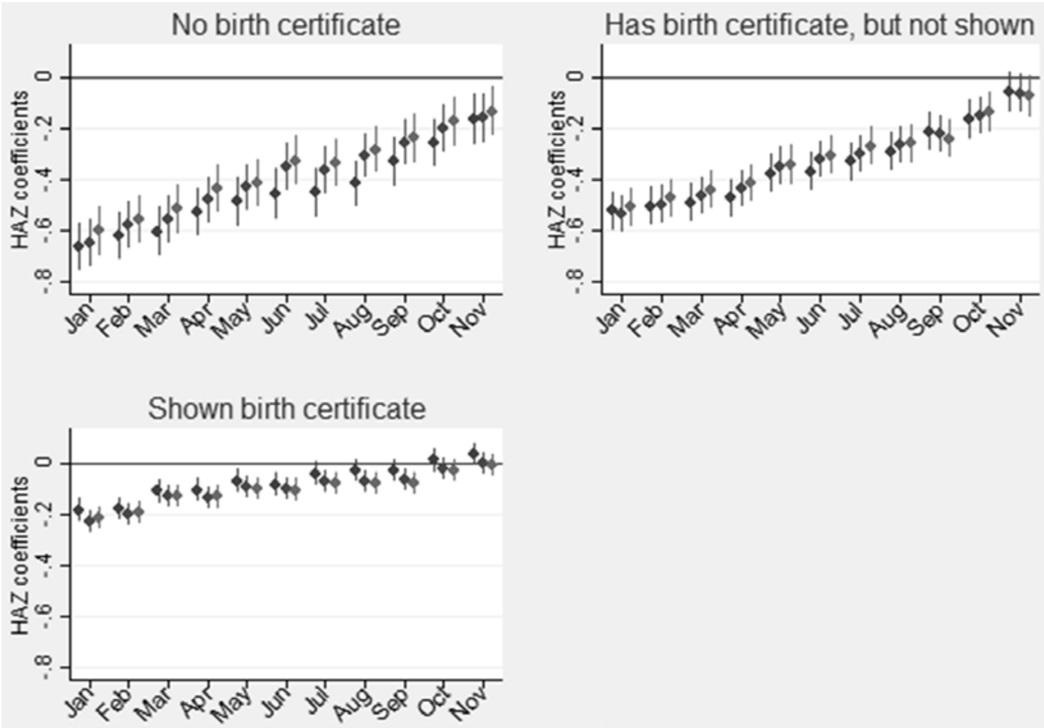
Source: DHS data from 975,534 children in 62 countries.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Forty-four percent of children have mothers with 0–3 years of schooling; 19 percent have mothers with 4–6 years of schooling; 16 percent have mothers with 7–9 years of schooling; 14 percent have mothers with 10–12 years of schooling, and 6 percent have mothers with 13 or more years of schooling.

Figure A.5 HAZ by MOB depending on the mother's literacy



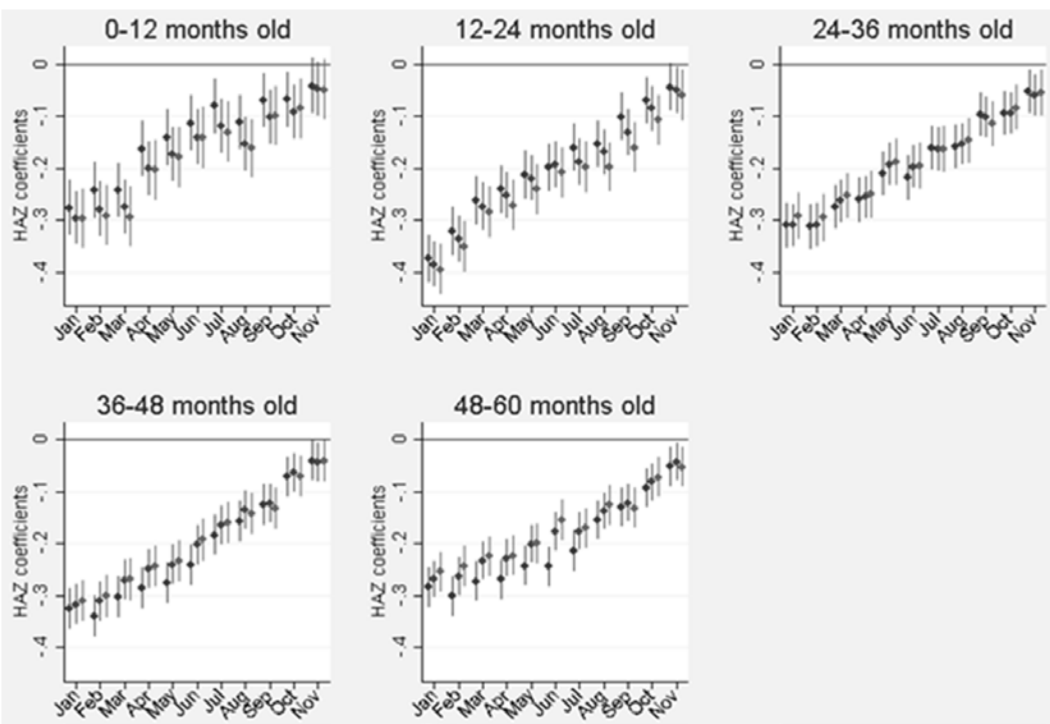
Source: DHS data from 395,347 children in 50 countries, various years.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Fifty-three percent of the children have illiterate mothers; 47 percent have literate mothers.

Figure A.6 HAZ by MOB depending on whether the mother has shown the child's birth certificate



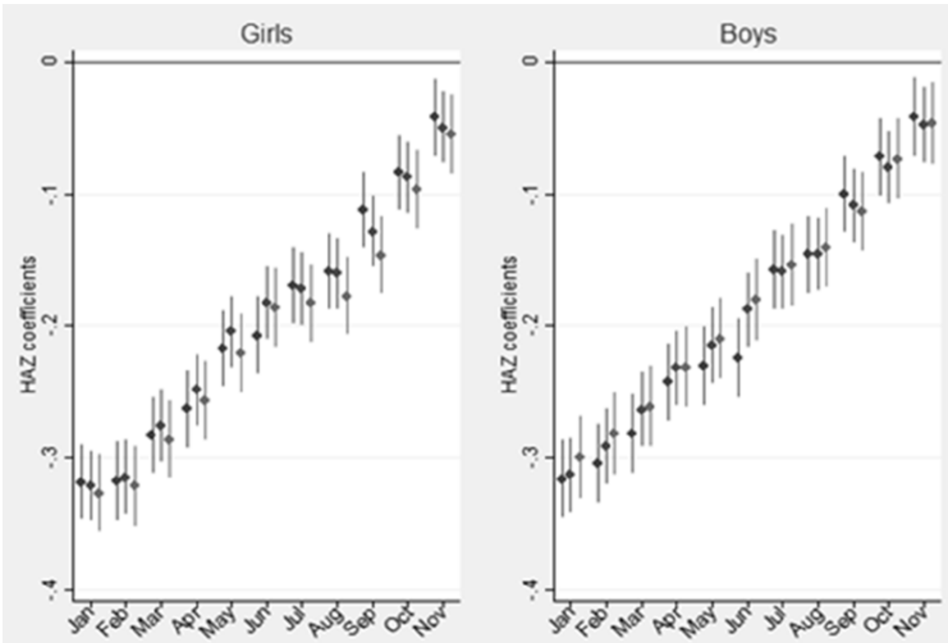
Source: DHS data from 396,299 children from 17 countries in Africa south of the Sahara and Egypt and India.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Twenty-one percent of children have no birth certificate; 26 percent have a birth certificate but it is not shown to the enumerator; and 53 percent show the birth certificate.

Figure A.7 HAZ by MOB by age group



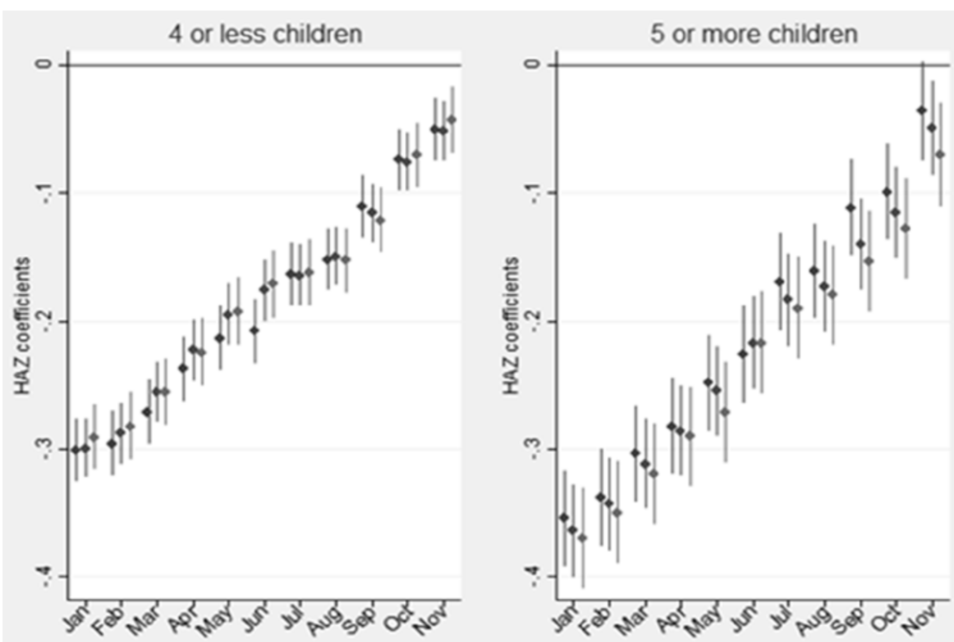
Source: DHS data from 990,231 children in 62 countries, various years.
Note: HAZ = height-for-age z-scores; MOB = month of birth.

Figure A.8 HAZ by MOB by gender of the child



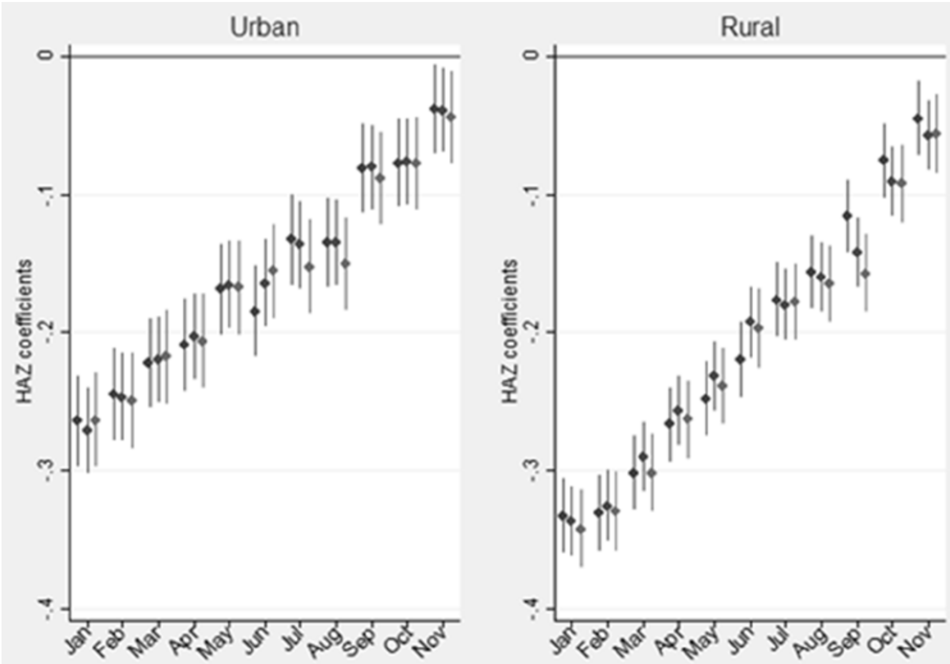
Source: DHS data from 990,231 children in 62 countries.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Fifty-one percent of the children are boys.

Figure A.9 HAZ by MOB by number of children in the household



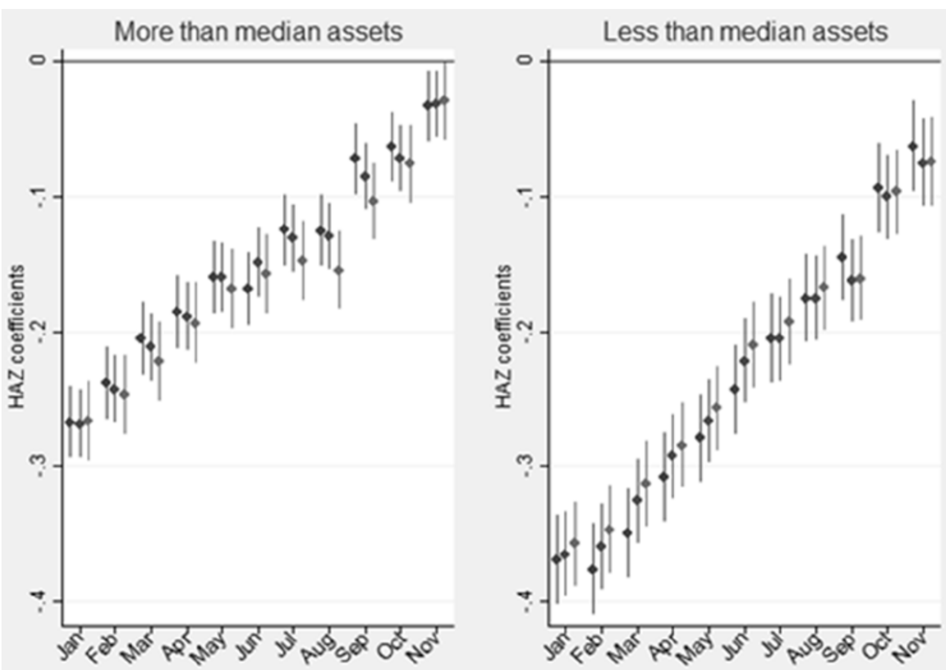
Source: DHS data from 990,231 children in 62 countries.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Sixty-nine percent of children have three or fewer siblings on their mother's side.

Figure A.10 HAZ by MOB by location of the household



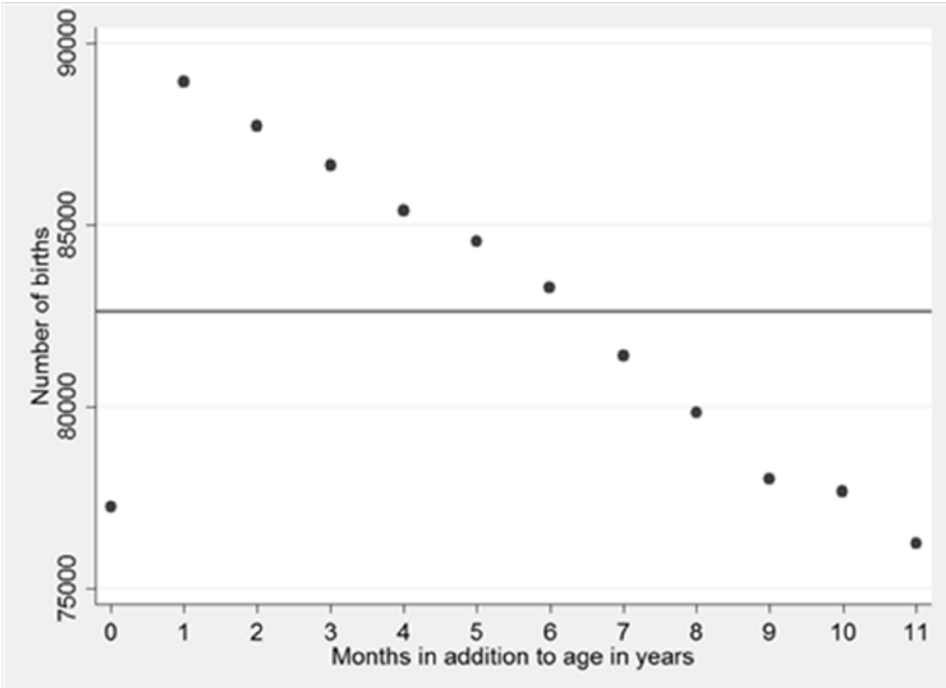
Source: DHS data from 990,231 children in 62 countries.
Note: HAZ = height-for-age z-scores; MOB = month of birth. Thirty-six percent of children live in urban households.

Figure A.11 HAZ by MOB depending on whether the household has above or below median assets



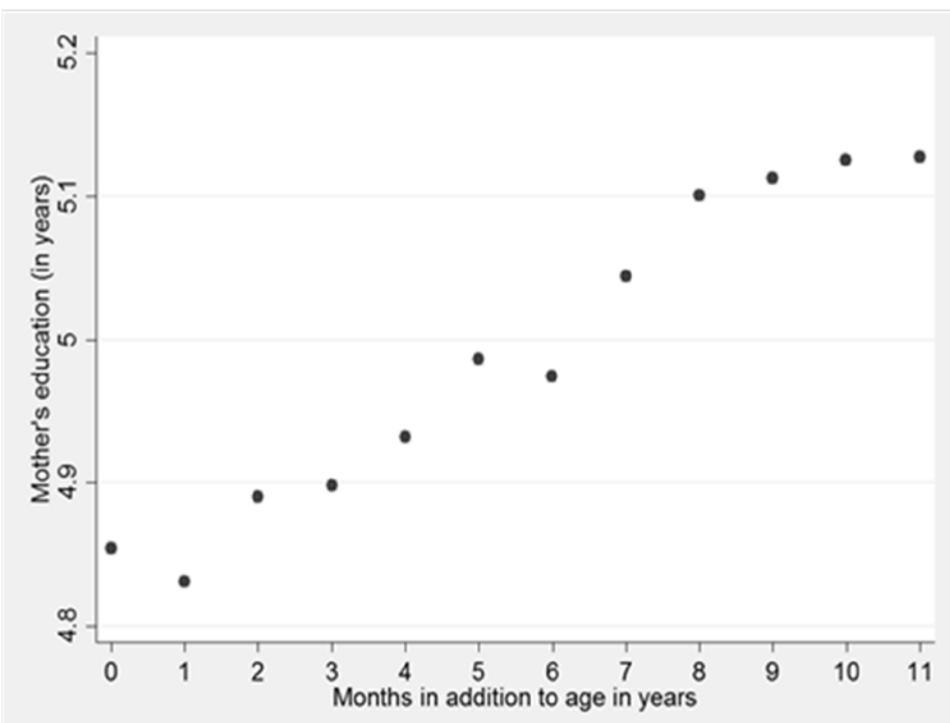
Source: DHS data from 866,450 children in 59 countries.
Note: HAZ = height-for-age z-scores; MOB = month of birth.

Figure A.12 Number of births by months in addition to round age



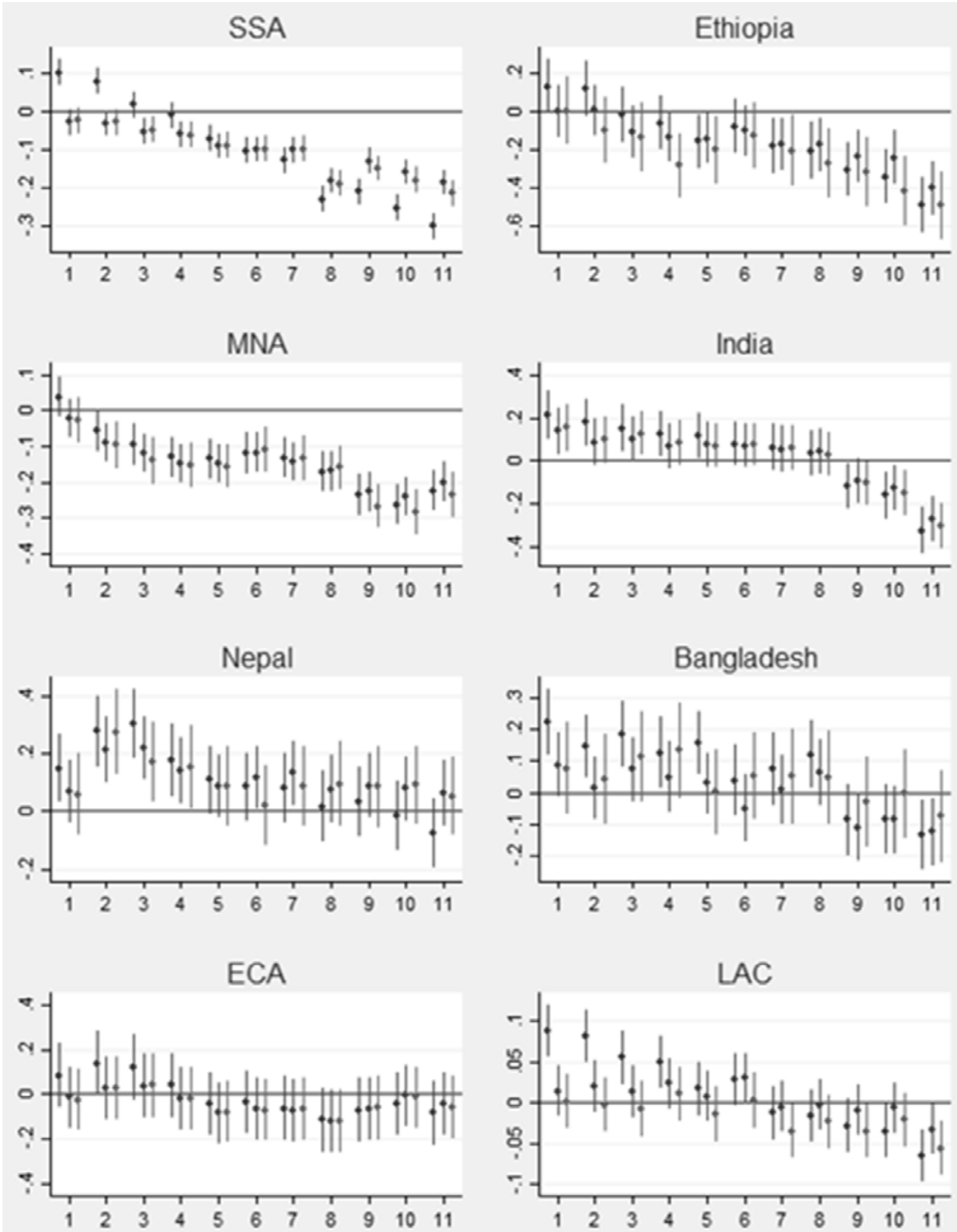
Source: DHS data from 987,028 children in 62 countries.

Figure A.13 Mother's education in years by the difference between interview and birth month



Source: DHS data from 972,337 children in 62 countries, various years.

Figure A.14 HAZ by additional months for major regions and selected countries including controls



Source: DHS data for 960,012 children from 58 countries, various years.
Note: HAZ = height-for-age z-scores. SSA = Africa south of the Sahara; MNA = Middle East and North Africa; ECA = Eastern Europe and Central Asia; LAC = Latin America and Caribbean.

APPENDIX B: SIMULATION PROTOCOL

Data-Generating Process

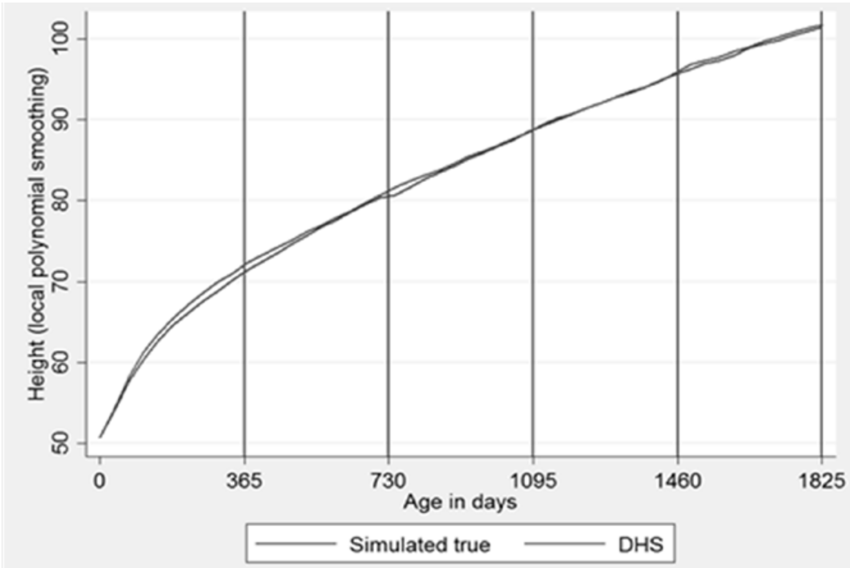
To simulate the true underlying height data, we implement the following data-generating process. We use Stata 14 for the simulations with the seed 1159 for the random number generator.

1. The observations consist of 100 girls born on each day between January 1, 2010, to December 31, 2015 (219,100 observations in total).
2. Assign a random day of measurement for each observation within the time span January 1, 2015, to December 31, 2015.
3. Calculate the true age (in days) as the difference between the birth date and the day of measurement. This leads to an age range from almost −1 year to 6 years of age. The reason to include children with ages greater than five years is that the measurement error in age may cause children to be included in the sample who are truly too old to be included. We disregard children with negative age (that is, born later than the day of measurement, 18,333 observations). Furthermore, we mirror the increasing attrition with age that we see in the DHS data by dropping a number of observations that increase linearly with age up to 24 percent for those 58 months old as found in the DHS data. Now, the total number of observations is 175,030.
4. Merge the data with age-specific synthetic length/height medians and standard deviations (SDs). These are constructed the following way:
   a) We use World Health Organization (WHO) length/height medians and SDs by age in days for girls as a starting point (WHO MGRSG 2006).4 These are available up to 1,856 days of age. For older children, WHO provides means and SDs by age in months (de Onis et al. 2007). We make a linear interpolation to obtain means and SDs by age in days for children older than 1,856 days.
   b) The WHO reference data are based on well-nourished children. To illustrate the measurement error in an environment with a plausible amount of stunting, we adjust the medians and SDs to correspond in a smooth way to the empirical pattern from the DHS data.
   c) The height medians are adjusted by changing the growth velocities such that children up to six months grow 7 percent less each day than well-nourished children; children from six months to two years of age grow 21 percent less each day than the growth standards; and children older than two years grow 10 percent less each day than the growth standards. Figure B.1 illustrates how these adjustments calibrate the synthetic mean heights well to the DHS mean heights.
   d) We add 2 to the height SDs to account for overall measurement error and increased dispersion due to variation in nutritional status of the children in the sample. In the DHS data, the SDs of height increase less with age than the WHO SDs, so we multiply the WHO SDs with 0.85 to have the same age gradient in the synthetic data as in the DHS data. Figure B.2 illustrates the SDs of heights by age in days in the DHS data and in the simulated data. We chose SDs that are below the SDs in the DHS data to better fit the moderate and severe stunting rates of the simulated data with the stunting rates in the DHS data.

4

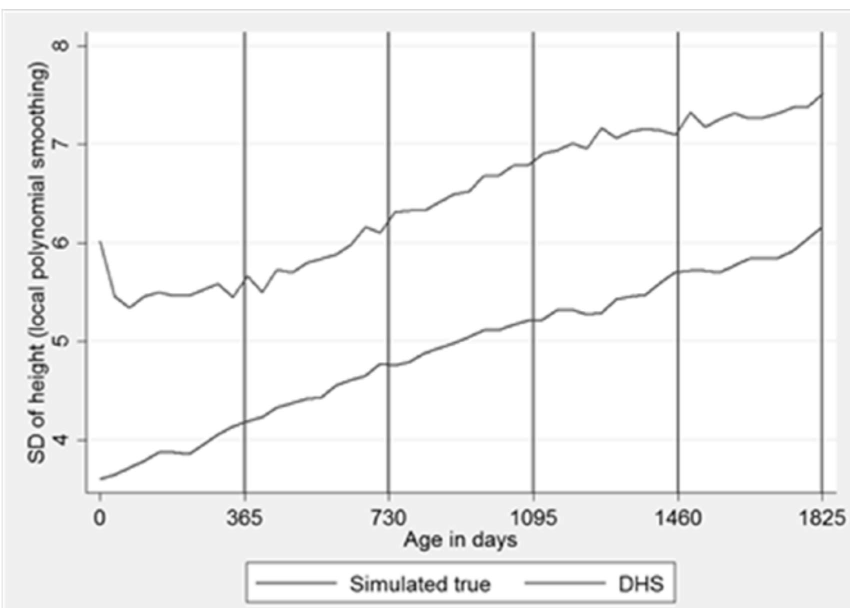
2. Draw heights for each observation from a normal distribution using the synthetic medians and SDs.
3. Calculate the true HAZ based on the simulated data for the children who are younger than 1,826 days (five years). Figure B.3 illustrates how the simulated true HAZ compares to the HAZ in the DHS data.

Figure B.1 Mean height by age (local polynomial smoothing), DHS data and simulated data



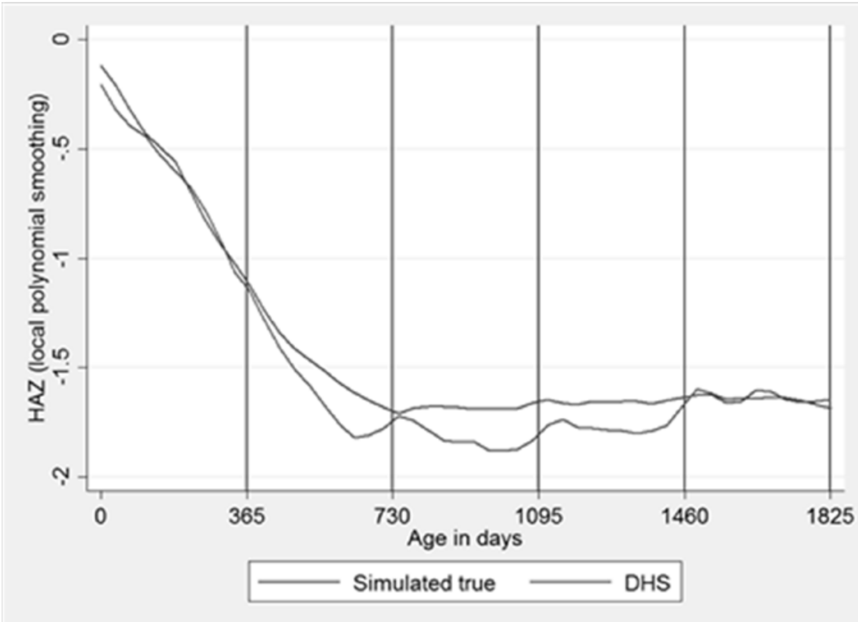
Source: Simulated data and DHS data for 960,012 children from 58 countries, various years.
Note: DHS = Demographic and Health Surveys.

Figure B.2 Standard deviation of height by age (local polynomial smoothing), DHS data and simulated data



Source: Simulated data and DHS data for 960,012 children from 58 countries, various years.
Note: DHS = Demographic and Health Surveys.

Figure B.3 Mean HAZ by age (local polynomial smoothing), DHS data and simulated data



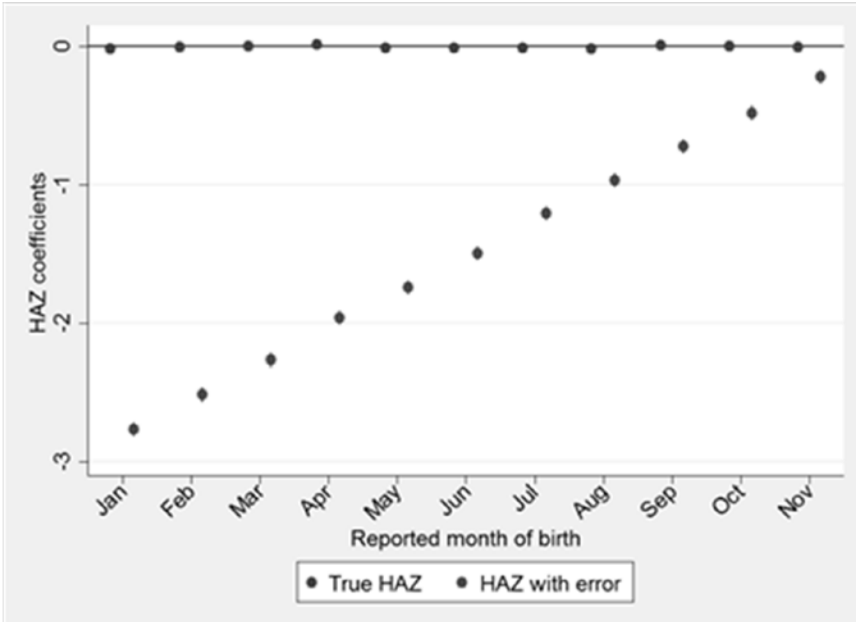
Source: Simulated data and DHS data for 960,012 children from 58 countries, various years.
Note: DHS = Demographic and Health Surveys.

Measurement Error: Random Month of Birth

To illustrate how measurement error in the month of birth can lead to a discontinuity in mean HAZ between December and January and to quantify the impact on stunting rates, we simulate the random month measurement error in the following way:

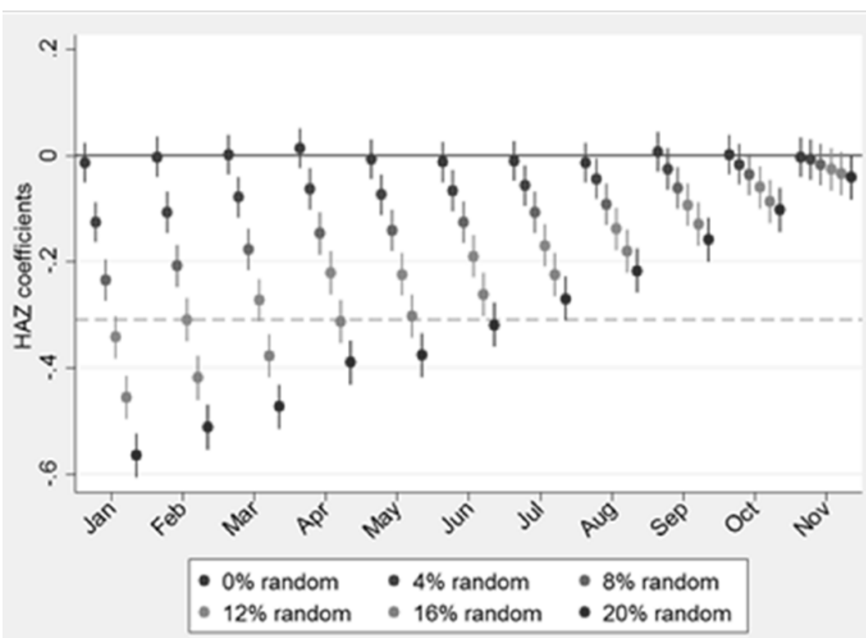
1. Draw random day and month of birth for each observation from a uniform distribution and calculate reported age based on the random day and month and the true birth year. For children born in 2015, the random month is restricted such that they cannot draw a random month of birth after the month of measurement.
2. Calculate the HAZ with random month of birth error for the children with a reported age below 1,826 days (five years).
3. Show how HAZ with error exhibits qualitatively the same pattern over month of birth as in the DHS data. This is illustrated in Figure B.4.
4. Randomly assign whether a child has measurement error in month of birth or not. We vary the share of children with measurement error to find the share that matches the December–January gap in the simulated mean HAZ with the corresponding gap in the DHS data. This is shown in Figure B.5 and Table B.1.
5. Calculate moderate and severe stunting rates for simulated data with varying shares of measurement error in month of birth. These are also included in Table B.1.

Figure B.4 Simulated true HAZ and HAZ with random month of birth



Source: Simulated data and DHS data for 960,012 children from 58 countries, various years.
Note: DHS = Demographic and Health Surveys.

Figure B.5 Simulated HAZ with varying shares of children with random month of birth



Source: Simulated data.
Note: HAZ = height-for-age z-scores. Dashed line represents December–January gap in DHS data.

Table B.1 Share of children with random month of birth, the simulated December–January gap, and stunting rates

Share random	Dec–Jan gap	Moderate stunting		Severe stunting	
DHS	*-0.313*	*0.354*		*0.160*	
		Rate	Difference	Rate	Difference
0.000	**-0.014**	**0.352**	**-**	**0.145**	**-**
0.010	-0.038	0.353	+0.001	0.146	+0.001
0.020	-0.064	0.354	+0.002	0.147	+0.002
0.030	-0.095	0.354	+0.002	0.147	+0.002
0.040	-0.126	0.355	+0.003	0.148	+0.003
0.050	-0.155	0.355	+0.003	0.148	+0.003
0.060	-0.181	0.355	+0.003	0.149	+0.004
0.070	-0.209	0.356	+0.004	0.149	+0.004
0.080	-0.235	0.356	+0.004	0.150	+0.005
0.090	-0.259	0.356	+0.004	0.150	+0.005
0.100	-0.287	0.357	+0.005	0.151	+0.006
0.110	**-0.320**	**0.357**	**+0.005**	**0.152**	**+0.007**
0.120	-0.343	0.358	+0.006	0.153	+0.008
0.130	-0.368	0.358	+0.006	0.153	+0.008
0.140	-0.402	0.359	+0.007	0.154	+0.009
0.150	-0.428	0.359	+0.007	0.155	+0.010
0.160	-0.456	0.359	+0.007	0.155	+0.010
0.170	-0.483	0.360	+0.008	0.156	+0.011
0.180	-0.511	0.360	+0.008	0.156	+0.011
0.190	-0.539	0.360	+0.008	0.157	+0.012
0.200	-0.565	0.361	+0.009	0.158	+0.013
0.210	-0.592	0.361	+0.009	0.158	+0.013
0.220	-0.620	0.362	+0.010	0.159	+0.014
0.230	-0.649	0.362	+0.010	0.160	+0.015
0.240	-0.680	0.362	+0.010	0.160	+0.015
0.250	-0.707	0.363	+0.011	0.161	+0.016

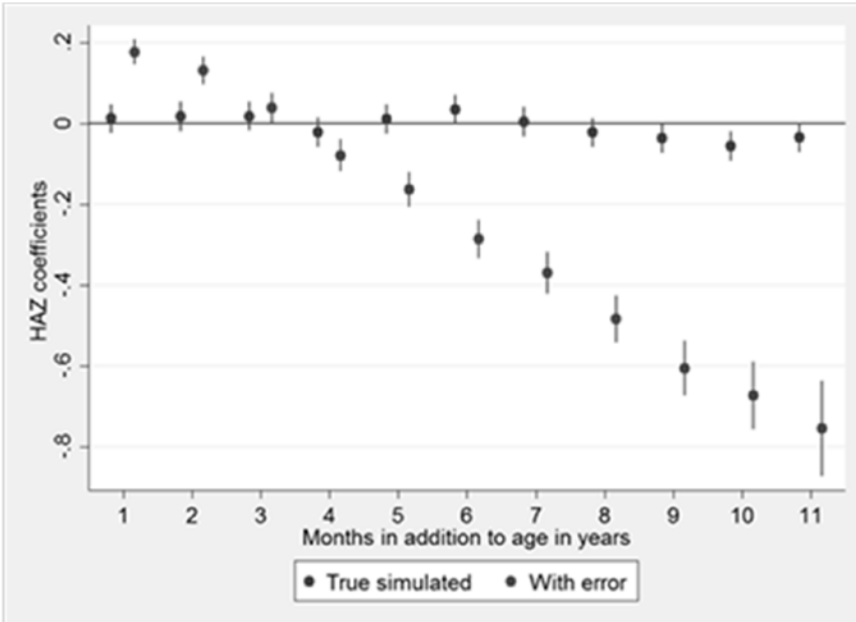
Source: Simulated data.
Note: HAZ = height-for-age z-scores.

Measurement Error: Asymmetric Rounding Error

To illustrate how an asymmetric rounding error in age can lead to a discontinuity in mean HAZ between children that are one month below a round age and the round age and to quantify the resulting impact on stunting rates, we simulate the asymmetric rounding error in the following way:

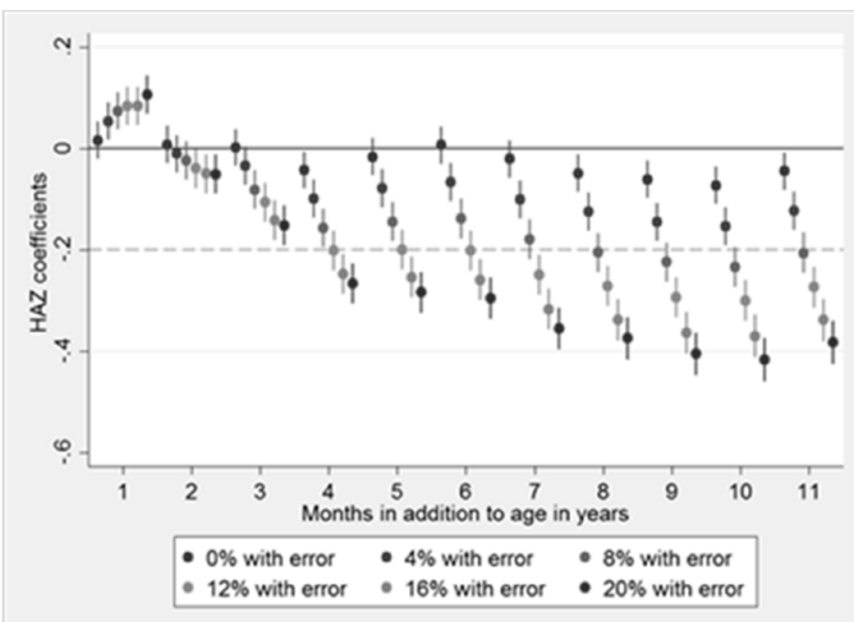
1. Assign an age (in days) randomly drawn from the uniform distribution over the interval from the age in completed years and the true age in days. This implies that the age in completed years is correct but that the number of additional months of age is lower than the true age. Based on this reported age, calculate the age in months in addition to the age in completed years. For example, if a child is reported to be 27 months old, this corresponds to two years and three months.
2. Calculate the HAZ with asymmetric rounding error based on the reported age.
3. Show that HAZ with error by the number of months in addition to age in years exhibits qualitatively the same pattern as in the DHS data. This is illustrated in Figure B.6.
4. Find the share of children with a rounding error in age that results in a gap of 0.18 between a round age and a round age and 11 months. This is illustrated in Figure B.7 and in Table B.2. The latter also includes the resulting stunting rates for different shares of children with rounding error in age.
5. Show the age distribution and the HAZ by age for the simulated data where 7 percent of children have an asymmetric rounding error. This is illustrated in Figures B.8 and B.9.

Figure B.6 HAZ by age in months in addition to age in years for the true simulated data and with asymmetric rounding error



Source: Simulated data.
Note: HAZ = height-for-age z-scores.

Figure B.7 Simulated HAZ with varying shares of children with asymmetric rounding error



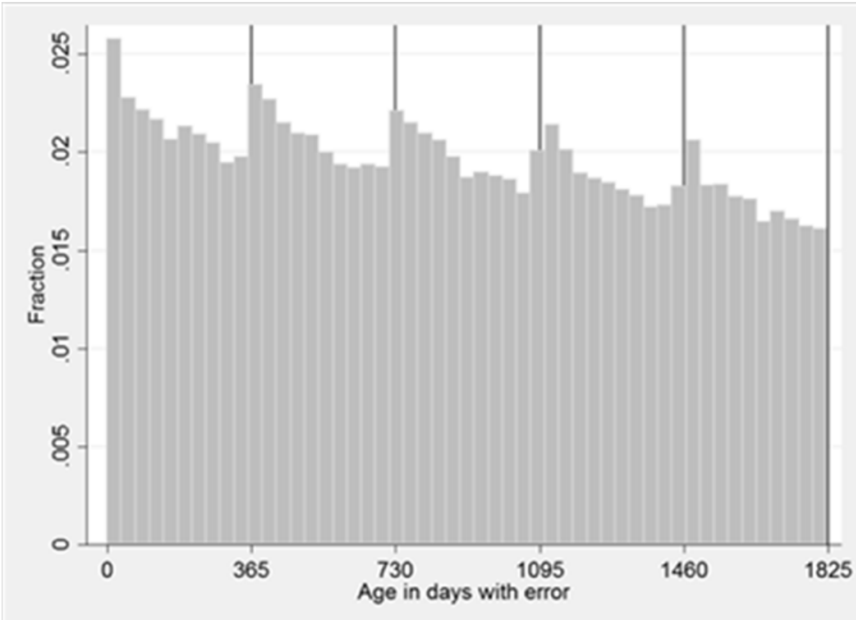
Source: Simulated data.
Note: HAZ = height-for-age z-scores. Dashed line represents 11-month gap in DHS data.

Table B.2 Share of children with asymmetric rounding error, the simulated 11-month gap, and stunting rates

Share with error	11-month gap	Moderate stunting		Severe stunting	
DHS	-0.184	*0.354*		*0.160*	
		Rate	**Difference**	**Rate**	**Difference**
0.000	**-0.044**	**0.352**	**-**	**0.145**	**-**
0.010	-0.072	0.351	-0.001	0.145	0.000
0.020	-0.088	0.350	-0.002	0.144	-0.001
0.030	-0.106	0.348	-0.004	0.143	-0.002
0.040	-0.123	0.347	-0.005	0.143	-0.002
0.050	-0.145	0.346	-0.006	0.142	-0.003
0.060	-0.164	0.344	-0.008	0.142	-0.003
0.070	**-0.185**	**0.343**	**-0.009**	**0.141**	**-0.004**
0.080	-0.206	0.342	-0.010	0.140	-0.005
0.090	-0.222	0.340	-0.012	0.140	-0.005
0.100	-0.237	0.339	-0.013	0.139	-0.006
0.110	-0.258	0.338	-0.014	0.138	-0.007
0.120	-0.274	0.337	-0.015	0.138	-0.007
0.130	-0.287	0.335	-0.017	0.137	-0.008
0.140	-0.308	0.334	-0.018	0.137	-0.008
0.150	-0.324	0.332	-0.020	0.136	-0.009
0.160	-0.339	0.331	-0.021	0.135	-0.010
0.170	-0.344	0.329	-0.023	0.135	-0.010
0.180	-0.352	0.328	-0.024	0.134	-0.011
0.190	-0.365	0.327	-0.025	0.133	-0.012
0.200	-0.383	0.325	-0.027	0.132	-0.013
0.210	-0.395	0.324	-0.028	0.132	-0.013
0.220	-0.406	0.323	-0.029	0.131	-0.014
0.230	-0.423	0.321	-0.031	0.130	-0.015
0.240	-0.437	0.320	-0.032	0.130	-0.015
0.250	-0.445	0.319	-0.033	0.129	-0.016

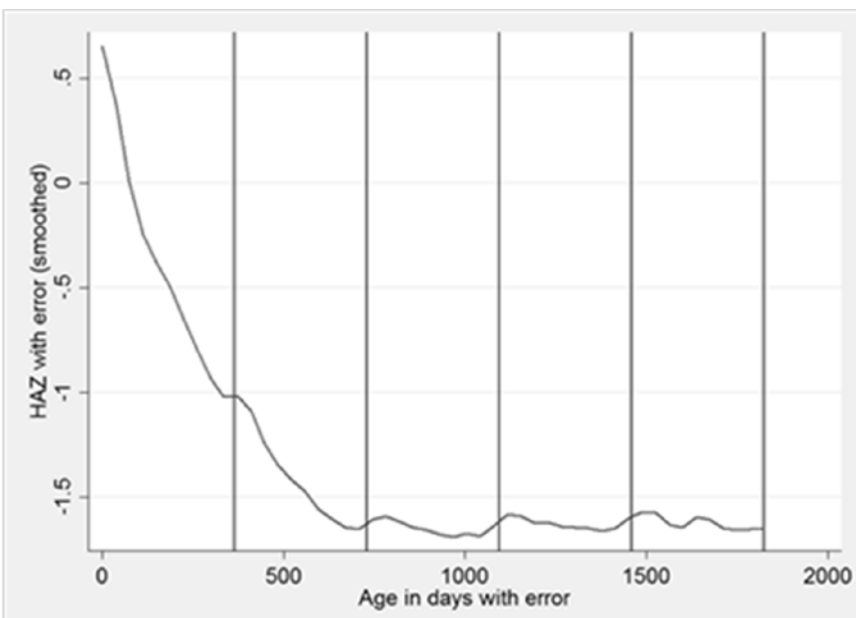
Source: Simulated data.
Note: HAZ = height-for-age z-scores.

Figure B.8 Distribution of simulated ages with 7 percent asymmetric rounding error



Source: Simulated data.
Note: HAZ = height-for-age z-scores. Red lines correspond to ages in round numbers (i.e. 1, 2, 3, 4, 5 years).

Figure B.9 Simulated HAZ by age with 7 percent asymmetric rounding error



Source: Simulated data.
Note: HAZ = height-for-age z-scores. Red lines correspond to ages in round numbers (i.e. 1, 2, 3, 4, 5 years).

REFERENCES

Akresh, R., P. Verwimp, and T. Bundervoet. 2011. "Civil War, Crop Failure, and Child Stunting in Rwanda." *Economic Development and Cultural Change* 59 (4): 777–810.

Assaf, S., M. T. Kothari, and T. Pullum. 2015. *An Assessment of the Quality of DHS Anthropometric Data, 2005–2014.* DHS Methodological Reports, No. 16. Rockville, MD, US: ICF International.

Black, R. E., L. H. Allen, Z. A. Bhutta, L. E. Caulfield, M. de Onis, M. Ezzati, C. Mathers, and J. Rivera. 2008. "Maternal and Child Undernutrition: Global and Regional Exposures and Health Consequences." *Lancet* 371 (9608): 243–260.

Black, R. E., C. G. Victora, S. P. Walker, Z. A. Bhutta, P. Christian, M. de Onis, M. Ezzati, S. Grantham-McGregor, J. Katz, R. Martorell, and R. Uauy. 2013. "Maternal and Child Undernutrition and Overweight in Low-Income and Middle-Income Countries." *Lancet* 382 (9890): 427–451.

Chambers, R. 1982. "Health, Agriculture, and Rural Poverty: Why Seasons Matter." *Journal of Development Studies* 18 (2): 217–238.

Croft, T. 1991. *DHS Data Editing and Imputation.* Calverton, MD, US: Macro International.

Currie, J., and M. Rossin-Slater. 2013. "Weathering the Storm: Hurricanes and Birth Outcomes." *Journal of Health Economics* 32 (3): 487–503.

de Onis, M. A. Onyango, E. Borghi, A. Siyam, C. Nishida, and J. Siekmann. 2007. "Development of a WHO Growth Reference for School-Aged Children and Adolescents," *Bulletin of the World Health Organization* 85: 661–668. http://www.who.int/growthref/en/.

Deschenes, O., M. Greenstone, and J. Guryan. 2009. "Climate Change and Birth Weight." *American Economic Review* 99 (2): 211–217.

Dewey, K. G., and K. Begum. 2011. "Long-Term Consequences of Stunting in Early Life." *Maternal and Child Nutrition* 7 (suppl. 3): 5–18.

Dorelien, A. M. 2015. "Effects of Birth Month on Child Health and Survival in Sub-Saharan Africa." *Biodemography and Social Biology* 61 (2): 209–230.

Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp, and International Child Development Steering Group. 2007. "Developmental Potential in the First 5 Years for Children in Developing Countries." *Lancet* 369 (9555): 60–70.

Hoddinott, J., J. R. Behrman, J. A. Maluccio, P. Melgar, A. R. Quisumbing, M. Ramirez-Zea, A. D. Stein, K. M. Yount, and R. Martorell. 2013. "Adult Consequences of Growth Failure in Early Childhood." *American Journal of Clinical Nutrition* 98 (5): 1170–1178.

ICF International. 2015. "The Demographic and Health Surveys Program." Accessed May 22, 2016. http://dhsprogram.com/Where-We-Work/Country-List.cfm.

ICF-Macro. 2009. *Demographic and Health Survey Interviewer's Manual.* Documentation No. 2. Calverton, MD, US: ICF-Macro.

Leroy, J. L., M. Ruel, J.-P. Habicht, and E. A. Frongillo. 2014. "Linear Growth Deficit Continues to Accumulate beyond the First 1,000 Days in Low- and Middle-Income Countries: Global Evidence from 51 National Surveys." *Journal of Nutrition*. doi: 10.3945/jn.114.191981.

Lokshin, M., and S. Radyakin. 2012. "Month of Birth and Children's Health in India." *Journal of Human Resources* 47 (1): 174–203.

Messias, E., C. Mourao, J. Maia, J. P. Campos, K. Ribeiro, L. Ribeiro, and B. Kirkpatrick. 2006. "Season of Birth and Schizophrenia in Northeast Brazil: Relationship to Rainfall." *Journal of Nervous and Mental Disease* 194 (11): 870–873.

Mulmi, P., S. A. Block, G. E. Shively, and W. A. Masters. 2016. "Climatic Conditions and Child Height: Sex-Specific Vulnerability and the Protective Effects of Sanitation and Food Markets in Nepal." *Economics and Human Biology* 23: 63–75.

Prentice, A. M., R. G. Whitehead, S. B. Roberts, and A. A. Paul. 1981. "Long-Term Energy Balance in Child-Bearing Gambian Women." *American Journal of Clinical Nutrition* 34 (12): 2790–2799.

Pullum, T. 2006. *An Assessment of Age and Date Reporting in the DHS Surveys, 1985–2003.* DHS Methodological Reports, No. 5. Calverton, MD, US: Macro International.

Rayco-Solon, P., A. J. Fulford, and A. M. Prentice. 2005. "Differential Effects of Seasonality on Preterm Birth and Intrauterine Growth Restriction in Rural Africans." *American Journal of Clinical Nutrition* 81 (1): 134–139.

Shrimpton, R., C. G. Victora, M. De Onis, R. Costa Lima, M. Blössner, and G. Clugston. 2001. "The Worldwide Timing of Growth Faltering: Implications for Nutritional Interventions." *Pediatrics* 107 (5): e75.

Tiwari, S., H. G. Jacoby, and E. Skoufias. 2017. "Monsoon Babies: Rainfall Shocks and Child Nutrition in Nepal." *Economic Development and Cultural Change* 65 (2): 167–188.

UNICEF. 2015. *Progress for Children—Beyond Averages: Learning from the MDGs.* New York: UNICEF. http://data.unicef.org/resources/progress-for-children-report.html.

Victora, C. G., M. de Onis, P. Curi Hallal, M. Blössner, and R. Shrimpton. 2010. "Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions." *Pediatrics* 125 (3): e473–e480.

WHO (World Health Organization). 2006. "WHO Child Growth Standards Based on Length/Height, Weight, and Age." *Acta Paediatrica* 95 (suppl. 450): 76–85.

WHO MGRSG (World Health Organization Multicentre Growth Reference Study Group). 2006. *WHO Child Growth Standards: Length/height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body Mass Index-for-age: Methods and Development.* Geneva: WHO. http://www.who.int/childgrowth/standards/en/.

RECENT IFPRI DISCUSSION PAPERS

For earlier discussion papers, please go to www.ifpri.org/publications/discussion_papers.
All discussion papers can be downloaded free of charge.

1613. *Cost-effectiveness of community-based gendered advisory services to farmers: Analysis in Mozambique and Tanzania.* Tewodaj Mogues, Valerie Mueller, and Florence Kondylis, 2017.

1612. *The European Union–West Africa Economic Partnership Agreement: Small Impact and new questions.* Antoine Bouët, David Laborde, and Fousseini Traoré, 2017.

1611. *The returns to empowerment in diversified rural households: Evidence from Niger.* Fleur Wouterse, 2017.

1610. *Prospects for the Myanmar rubber sector: An Analysis of the viability of smallholder production in Mon State.* Joanna van Asselt, Kyan Htoo, and Paul A. Dorosh, 2017.

1609. *How do agricultural development projects aim to empower women?: Insights from an analysis of project strategies.* Nancy Johnson, Mysbah Balagamwala, Crossley Pinkstaff, Sophie Theis, Ruth Meinzen-Dick, and Agnes Quisumbing, 2017.

1608. *Stimulating agricultural technology adoption: Lessons from fertilizer use among Ugandan potato farmers.* Lydia Nazziwa-Nviiri, Bjorn Van Campenhout, and David Amwonya, 2017.

1607. *Strengthening and harmonizing food policy systems to achieve food security: A case study and lessons from Ghana.* Suresh Chandra Babu and Sylvia Blom, 2017.

1606. *Trade and economic impacts of destination-based corporate taxes.* Will Martin, 2017.

1605. *Can better targeting improve the effectiveness of Ghana's Fertilizer Subsidy Program? Lessons from Ghana and other countries in Africa south of the Sahara.* Nazaire Houssou, Kwaw Andam, and Collins Asante-Addo, 2017.

1604. *The impact of Ethiopia's Productive Safety Net Programme on the nutritional status of children: 2008–2012.* Guush Berhane, John Hoddinott, and Neha Kumar, 2017.

1603. *Economic transformation in Africa from the bottom up: Evidence from Tanzania.* Xinshen Diao, Josaphat Kweka, and Margaret McMillan, 2017.

1602. *Liquid milk: Cash constraints and day-to-day intertemporal choice in financial diaries.* Xin Geng, Wendy Janssens, and Berber Kramer, 2017.

1601. *A chicken and maize situation: The poultry feed sector in Ghana.* Kwaw S. Andam, Michael E. Johnson, Catherine Ragasa, Doreen S. Kufoalor, and Sunipa Das Gupta, 2017.

1600. *Gender justice and food security in India: A review.* Nitya Rao, Mamata Pradhan, and Devesh Roy, 2017.

1599. *Cities and rural transformation: A spatial analysis of rural youth livelihoods in Ghana.* Xinshen Diao, Peixun Fang, Eduardo Magalhaes, Stefan Pahl, and Jed Silver, 2017

1598. *The changing structure of Africa's economies.* Xinshen Diao, Kenneth Harttgen, and Margaret McMillan, 2017.

1597. *Perspectives on the role of the state in economic development: Taking stock of the "Developmental State" after 35 years.* Jordan Kyle, 2017.

1596. *Imputing nutrient intake from foods prepared and consumed away from home and other composite foods: Exploring extensions of the Subramanian–Deaton cost per calorie approach.* Dena Metili Mwangi, John L. Fiedler, and Celeste Sununtnasuk, 2017.

1595. *Estimating spatial basis risk in rainfall index insurance: Methodology and application to excess rainfall insurance in Uruguay*. Francisco Ceballos, 2016.

1594. *The effect of land inheritance on youth employment and migration decisions: Evidence from rural Ethiopia.* Katrina Kosec, Hosaena Ghebru, Brian Holtemeyer, Valerie Mueller, and Emily Schmidt, 2016.

1593. *How headquarters relocation is affected by rising wages and ownership: Evidence from China's Annual Survey of Industrial Enterprises, 1999–2008.* Qingtao Wang, Kevin Chen, Longwen Chiang, and Xuanli Xie, 2016.

1592. *External shocks, food security, and development: Exploring scenarios for Central America.* Eugenio Diaz-Bonilla, Valeria Piñeiro, and Pablo Elverdin, 2016.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

www.ifpri.org

IFPRI HEADQUARTERS

2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org