Misreporting Month of Birth: 
Implications for nutrition research

11 January 2017

Anna Folke Larsen, Department of Economics, University of Copenhagen
(afl@econ.ku.dk)

Derek Headey, Poverty, Health & Nutrition Division, International Food Policy Research Institute
(d.headey@cgiar.org)

William A. Masters, Friedman School of Nutrition and Department of Economics, Tufts University
(william.masters@tufts.edu)

Abstract
Height-for-age z scores (HAZ) and stunting status (HAZ< -2) are widely used to measure child nutrition and population health. However, accurate measurement of age is non-trivial 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 DHS data from 62 countries over the period 1990-2014 to describe two statistical artifacts indicative of misreporting of age. The first artifact is lower HAZ 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 is lower HAZ for children with a reported age just below a round age (and hence implausibly large HAZ gaps between children with reported age 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 pre-school 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; Black et al. 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 is a widely used development target, including for the Sustainable Development Goals (SDGs).

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 WHO’s worldwide reference population of healthy children at 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 a HAZ<-3 SD is considered severely stunted.

Measuring length or height especially for young children is challenging and has received considerable attention from Demographic Health Survey (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 HAZ scores 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. 2009). 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 DHS, 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 towards the best

¹“Chronic undernutrition” returns 120,000 items, “growth faltering” 52,600 results. “Height-for-age” returns 30,000 results and length-for-age 5000 results. These searches were conducted on 23rd of September 2016.
available estimate of birth month (ICF-Macro 2009). But while this approach avoids the aforementioned selection bias, it nevertheless creates scope for both random and systematic misreporting of age, both of which could bias HAZ scores 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 et al. (2015). These studies find that a few DHS surveys have nonresponse rates for child age as high as 30%, but that other surveys in similar settings report birth years and months for 98-99% 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 et al. (2015) use Myers’ index to show that no more than 10% of children’s ages would need to be reallocated to eliminate age heaping. As a result, age errors in children under 5 years of age are generally treated as random, with limited heaping of the type detectable using standard diagnostics such as Myers’ or Whipple’s indexes.

In this study we use DHS data to show how measurement error in age can in fact introduce systematic artifacts in HAZ results. These 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. The first might be characterized as random or quasi-random estimation of month of birth with a given birth year, while the second is characterized by respondents seemingly rounding their child ages down towards a round number (e.g. 2 years) more than they round them up. Both artifacts are characterized by implausibly large discontinuities in the relationship between HAZ and month of birth (Artifact 1) and HAZ and age relative to a round age (Artifact 2), while Artifact 2 is also characterized by a distinctly asymmetric age heaping around round ages. 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 to both regularly document these errors, and to reduce the extent of these errors through improved survey practices.
The remainder of this paper is structured as follows. Section 2 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 these errors might come about, and how these errors would be manifest in data from conventional nutrition surveys. While both these 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 these circumstances – each artifact is nevertheless statistically distinguishable from the other, as we describe below.

2.1 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 to first ascertain a child’s age in years, and then to use salient events to get a (hopefully) more precise estimation of the specific month of birth (hereafter MOB). While 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 month of birth, 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 a 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 (e.g., June). If she is erroneously recorded as born earlier in the year (e.g. January) she is actually younger than reported to be and therefore likely to be short for reported
age. If recorded as born later (e.g. December) she is actually older and likely to be tall for 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 artefactual 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.

While 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 HAZ scores 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 sub-Saharan African countries 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, one study of exposure to conflict and drought in Rwanda acknowledged that they might be under-estimating the true impacts of these 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.

2.1 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 is unable to provide an MOB, the enumerator might use the survey month to work backwards to an estimated MOB. For example, a child reported as being 2 years old and surveyed in June 2016 could be

\(^2\) This is the median of 97 regression coefficients reported in her appendix, for 3 different models. The mean coefficient is 2.87.
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 month of birth. 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 there 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 2 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 since then that documented this phenomenon in DHS or other nutrition surveys.

3. Data and methods
While the phenomenon we describe are not confined solely to the DHS, 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 are the single largest source of nationally representative nutritional data in the developing world, and 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 for multiple countries allows us to draw comparisons across countries with different agroclimatic seasons, levels of education and birth documentation, and different cultural norms. Third, the DHS 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 aged 0-59 months (surveys that only sampled children 0-36 months were excluded), resulting in the sample sizes described in Table 1. In total this sample includes just under one million children from 183 surveys covering 62 countries, with almost half the children coming from the sub-Saharan Africa (SSA). Most regions have sizeable samples, though we do point out that several regions are dominated by just a few countries. East Asia and Pacific (EAP), for example, only includes Cambodia and East Timor, and the large Middle East & North Africa (MNA) sample is heavily
dominated by Egypt and Jordan. However, DHS coverage of SSA and South Asia (SAS) – the two regions with the highest undernutrition burdens – is excellent.

Table 1. Sample sizes for anthropometric data by regions for 62 countries

<table>
<thead>
<tr>
<th>Region</th>
<th>Countries</th>
<th>Child observations</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia &amp; Pacific</td>
<td>2</td>
<td>19,447</td>
<td>2.0%</td>
</tr>
<tr>
<td>Europe &amp; Central Asia</td>
<td>7</td>
<td>18,653</td>
<td>1.9%</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>10</td>
<td>222,255</td>
<td>22.4%</td>
</tr>
<tr>
<td>Middle East &amp; North Africa</td>
<td>5</td>
<td>145,081</td>
<td>14.7%</td>
</tr>
<tr>
<td>South Asia</td>
<td>5</td>
<td>98,260</td>
<td>9.9%</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>33</td>
<td>486,535</td>
<td>49.1%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>62</td>
<td>990,231</td>
<td>100.0%</td>
</tr>
</tbody>
</table>


Anthropometric data collection has been a key component of DHS 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 incorporate a number of steps to improve data quality. Interviewers are typically national staff from private and/or government statistical agencies, who receive extensive additional training on how to obtain and record height and weight measurements as well as the birthdate, and 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 et al. 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 month of birth, 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 0-59 month age range. While height and length measurement is relatively transparent, it is less clear how enumerators measure age if respondents are uncertain about birthdates 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% of children have no birth month directly recorded by an enumerator, and in Nigeria just 1.36% have no month recorded. Only Guinea (14.7%), Benin (8.7%) and Burkina Faso (5.6%) 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 score is calculated from their age and height based on the following formula:

\[
(1) \quad Z\text{-score} = \frac{\text{Individual height} - \text{median height in reference population}}{\text{SD of reference population}}
\]

where the reference population is specific to the child’s sex and age in months, and SD refers to standard deviation. Children with HAZ scores 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 cm for children measured lying down and 65-120 cm for children measured standing up (Assaf et al. 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 to 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 spans the northern and southern hemispheres with offsetting solar exposure. This leads to three specific null hypotheses of no real or artefactual month effects: (1) no association between HAZ and MOB, with no sawtooth between the start and 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:

\[
(2) \quad H_i = \sum_{m=1}^{11} \beta_m MOB_{m,i} + \sum_{d=1}^{11} \gamma_d months_{d,i} + \delta D_i + \theta X_i + \mu_s + \epsilon_i
\]
Where $H_i$ is the nutrition indicator (HAZ or stunting) for child $i$, $MOB_{m,i}$ are the 11 month of birth dummies where December is the reference category, and $months_{d,i}$ are the 11 dummies for the age in months in addition to a round age where zero (i.e. a round age) is the reference category (e.g. a 27-month-old girl is three months in addition to her round age of two years). $D_i$ is child demographics (child gender, age and age squared), $X_i$ 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), $\mu_s$ refers to survey dummies for each country and survey wave and $\epsilon_i$ to the error term which we cluster at enumeration areas. While (1) is the generic form of the equation we estimate, we also estimate regressions excluding $X_i$ and excluding both $X_i$, $D_i$, and $\mu_s$ as well as regressions where either $MOB_{m,i}$ or $months_{d,i}$ 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 the round ages.

4.1 Random or quasi-random estimation of month of birth

As discussed above, if calendar year of birth is recalled more accurately than calendar months, 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 1 reports this associated 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 1. Average HAZ by MOB as a test for random MOB bias

Source: DHS data for 990,231 children from 62 countries. Notes: The vertical bars indicate standard errors of the mean HAZ.

Figure 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 sub-Saharan Africa (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 South-East 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 2. HAZ-MOB gradients for major regions and selected countries

Source: DHS data for 960,012 children from 58 countries. Notes: SSA=sub-Saharan Africa; MNA=Middle East & North Africa; ECA=Eastern Europe & Central Asia; LAC=Latin America & 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 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 non-linearity in the relationship between MOB and HAZ for the rest of year, perhaps pointing to genuine seasonality factors.

In section 3.3 we present regression results where we include the control variables described in Section 2. In appendix Figure A1, we also report regression results for the major regions and selected countries in addition to the raw means (in blue): one which controls for child demographics and survey fixed effects (red) and one which 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 the appendix. 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 A2). 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 A3). However, consistent with our hypotheses, HAZ-MOB gradients are substantially flatter for children who: (1) have more educated or literate mothers (Figure A4 and A5); and (2) have birth certificates (perhaps surprisingly, however, the gradients do not entirely disappear with birth certification) (Figure A6). A final result of some note is that the steep HAZ-MOB gradient reported in Figure 1 does not vary much by age group (we report separate gradients for children 0-12, 12-24, 24-36, 36-48, 48-59 months in Figure A7), nor does it vary strongly by gender (Figure A8), number of children of the mother (Figure A9) or location or assets of the household (Figure A10 and A11).

4.2 Asymmetric heaping around round numbers

If parents do not know exact birthdates, 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 they may be prone to prompt for a low number of
additional months. Asking for whether the child is an additional nine to eleven months (leading to an age in months just below the next age in years) may be quite uncommon. Figure 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 A12 we also show that this saw-tooth pattern exists when we replace the x-axis variable in Figure 3 with the age in months in addition to a round age.

**Figure 3. Asymmetric age heaping for aggregate DHS data**

![Figure 3](image)

Source: DHS data for 990,231 children from 62 countries. Thin red lines show round numbers.

The pattern shown in Figure 3 can be explained by what we term asymmetric rounding, in which respondents and enumerators first establish an age in years, and then work backwards 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 eleven.
months. Consistent with this, our appendix Figure A13 shows that the child age is significantly correlated with mothers’ education: mothers who report a round child age are less educated than those who report a round age plus eleven 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. (2009), who documented that most growth faltering takes place in the first 20 months after birth. Figure 4 reproduces this type of visualization, revealing how children at or just above 2, 3 and 4 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. Mean HAZ by age in months**

![Graph showing mean HAZ by age in months](image.png)

Source: DHS data for 990,231 children from 62 countries.

A different way to illustrate how asymmetric rounding affects analysis of HAZ is to calculate each child’s age in months in addition to their round age in years. For example, being eleven additional months corresponds to an age just below a round age, e.g. 23 months. This calculation allows us to draw a graph
that is analogous to Figure 1, averaging over all children under five years of age. Results are in Figure 5, showing a linear pattern of discrepancy between one and eleven additional months of about 0.3 SD which is the same magnitude as the discrepancy from Figure 1 between December and January born children.

**Figure 5. Mean HAZ by months in addition to age in years**

Source: DHS data for 990,231 children from 62 countries. Notes: The vertical bars indicate standard errors of the mean HAZ.

Figure 6 shows how the pattern of HAZ in months beyond completed years varies across major regions and the countries depicted in Figure 2. A notable difference is that Ethiopia and Nepal, the two countries for which national calendar systems affected the gradient in Figure 2, are not particularly distinctive in Figure 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 A14 also shows these results after controlling for additional covariates.
Figure 6. Mean HAZ by additional months for major regions and selected countries

Source: DHS data for 960,012 children from 58 countries. Notes: SSA=sub-Saharan Africa; MNA=Middle East & North Africa; ECA=Eastern Europe & Central Asia; LAC=Latin America & Caribbean. Corresponding graphs where controls are included can be found in appendix in Figure A14.
The patterns shown in Figures 5-6 and Figures 1-2 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 5-6 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. To address this issue we turn to statistical tests, controlling for age along with other covariates.

4.3 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), 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). The results of these regression estimates are summarized in Figure 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.
Figure 7. Tests for month-of-birth and round age biases in HAZ scores

Source: DHS data for 990,231 children from 62 countries. See text for description of regression coefficients.
In Figure 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 2-3 points more likely to be classified as stunting compared to children with a round age.

**Figure 8. Tests for month-of-birth and round age biases in stunting status**

Source: DHS data for 990,231 children from 62 countries. See text for description of regression coefficients.
5. Simulations

We undertake two different simulation exercises in order 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.

A detailed protocol for these simulations is provided in Appendix B, so here we only provide 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 standard deviation (SD) of the distribution at each age is 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 6 months, then 21 percent less growth each day from 6 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 9. 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 (i.e. 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 error in height measurements decrease with age). Figure 10 illustrates that the simulated SDs fits the slope of the SDs in the DHS data while they are in general 1.5 times lower. We have chosen this specification to fit the simulated stunting rates to those in

---

3 The negative bump in the simulated heights just after 730 days of age (2 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 cm shorter than recumbent length.
the DHS data and because the measurement errors that we simulate also increases the SD of height. The resulting simulated true HAZ by age is shown in Figure 11. It appears smooth compared to the DHS HAZ by age.

**Figure 9. Mean height by age (local polynomial smoothing), DHS data and simulated data**

![Figure 9](image)

Source: DHS data for 488,307 girls from 62 countries and simulated data.

**Figure 10. SD of height by age (local polynomial smoothing), DHS data and simulated data**
Figure 11. 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.

5.1 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 a HAZ with error. (Note that we only include 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 13, 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 ten times as large as in the DHS data shown in Figure 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 reports the true birth month and who reports 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 12 and detailed in Table B1 in Appendix B. A gap of -0.32 SD is produced by assigning 11 percent of the children with a random MOB. The 11% random MOB sample has a moderate stunting rate of 35.7 percent compared to 35.2 percent in the true data, and a severe stunting rate of 15.2 percent compared to 14.5 percent. Whilst 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 upwards in poorer and less education population, slightly inflating the stunting differences between different socioeconomic groups.
5.2 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, towards 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 13 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 13 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 eleven additional months that we found in the DHS data presented in Figure 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 (i.e. 11 months above) and just above a round age, when age is controlled for. This calibration is illustrated in Figure 14 and detailed in Table B2 in Appendix B. We find that when 7 percent of the children have their age rounded down towards the age in completed years, we can reproduce the 11 month gap of -0.18 in the DHS data. Using the calibrated share of seven percent affected by asymmetric rounding, we can draw the resulting age distribution to compare with the actual distribution in DHS data. This is illustrated in Figure 15. The distribution has heaps at the round ages and then slowly declines down to round ages plus eleven months before it jumps up again at the next round age, just as we saw in the DHS age distribution depicted in Figure 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 one percentage point from 35.2 to 34.3 percent when seven 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 13. HAZ by age in months in addition to age in years, true simulated data and with asymmetric rounding error

Source: Simulated data
Figure 14: Simulated HAZ with varying shares of children with asymmetric rounding error

Source: Simulated data. Notes: Dashed line represents 11 months gap in DHS 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 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 month of birth, 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 month of birth 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 cut-off. Errors
in age relative to completed years lead to under-reporting of age, on average, and hence over-estimation 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 (e.g. 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 Forthcoming) or economic or political shocks (Akresh et al. 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 show that there are significant associations between climatic shocks and long run health and nutrition outcomes. Clearly one would expect that developing country populations are 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 that there are 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 et al. (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, but one that exists in other similar multi-purpose surveys conducted in underdeveloped populations (Though not reported here, we find similar errors in the Living Standards Measurement Study (LSMS) surveys implemented by the World Bank and the Multiple Indicator Cluster Surveys (MICS) 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, questioning of both mothers and fathers, and training enumerators to be more aware of the serious consequences of age misreporting that are documented in this paper.
References


ICF-International. 2015. "The Demographic and Health Surveys Program."


Appendix A

Figure A1. HAZ-MOB gradients for major regions and selected countries with controls

Source: DHS data for 960,012 children from 58 countries. Notes: SSA=sub-Saharan Africa; MNA=Middle East & North Africa; ECA=Eastern Europe & Central Asia; LAC=Latin America & Caribbean.
Figure A2. HAZ by MOB for Nigeria

Source: DHS data from 60,893 children from surveys in 1990, 2003, 2008 and 2013. Coefficients are from regressions where all control variables are included.
Figure A3: HAZ by MOB for children with and without imputed birth month

Source: DHS data from 396,299 children in 17 countries in Sub-Saharan Africa and Egypt and India. 2% of children have imputed month of birth.
Figure A4: HAZ by MOB depending on the mother’s education

Source: DHS data from 975,534 children in 62 countries. 44% of children have mothers with 0-3 years of schooling; 19% have mothers with 4-6 years of schooling; 16% have mothers with 7-9 years of schooling; 14% have mothers with 10-12 years of schooling, and 6% have mothers with 13 or more years of schooling.
Figure A5: HAZ by MOB depending on the mother’s literacy

Source: DHS data from 395,347 children in 50 countries. 53 % of the children have illiterate mothers, 47 % have literate mothers.
Figure A6: 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 Sub-Saharan Africa and Egypt and India. 21% of children has no birth certificate; 26% has a birth certificate, but it is not shown to the enumerator; and 53% show the birth certificate.
Figure A7: HAZ by MOB by age group

Source: DHS data from 990,231 children in 62 countries
Figure A8: HAZ by MOB by gender of the child

Source: DHS data from 990,231 children in 62 countries, 51% of the children are boys.
Figure A9: HAZ by MOB by number of children in the household

Source: DHS data from 990,231 children in 62 countries. 69% of children have 3 or less siblings on their mother’s side.
Figure A10: HAZ by MOB by location of the household

Source: DHS data from 990,231 children in 62 countries. 36 % of children live in urban households
Figure A11: HAZ by MOB depending on whether the household has above or below median assets

Source: DHS data from 866,450 children in 59 countries.
Figure A12: Number of births by months in addition to round age

Source: DHS data from 987,028 children in 62 countries.
Figure A13: Mother’s education in years by the difference between interview and birth month

Source: DHS data from 972,337 children in 62 countries.
Figure A14: HAZ by additional months for major regions and selected countries including controls

Source: DHS data for 960,012 children from 58 countries. Notes: SSA=sub-Saharan Africa; MNA=Middle East & North Africa; ECA=Eastern Europe & Central Asia; LAC=Latin America & 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 Jan 1 2010 to Dec 31 2015 (219,100 observations in total).

2) Assign a random day of measurement for each observation within the time span January 1st, 2015 to December 31st, 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 larger than 5 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 (i.e. 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 which increase linearly with age up to 24 percent for the 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 WHO length/height medians and SDs by age-in-days for girls as a starting point. These are available up to 1856 days of age. For older children, WHO provides means and SDs by age-in-months. We make a linear interpolation to obtain means and SDs by age-in-days for children older than 1856 days.

b) The WHO reference data is 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 6 months grow 7 percent less each day than well-nourished children; children from 6 months to 2 years of age grow 21 percent less each day than the growth standards; and children older than 2 years grow 10 percent less each day than the growth standards. Figure B1 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 standard 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 B2 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.

5) Draw heights for each observation from a normal distribution using the synthetic medians and SDs.
6) Calculate the true HAZ based on the simulated data for the children who are younger than 1826 days (5 years). Figure B3 illustrates how the simulated true HAZ compares to the HAZ in the DHS data.

Figure B1: Mean height by age (local polynomial smoothing), DHS data and simulated data
Figure B2: SD of height by age (local polynomial smoothing), DHS data and simulated data

Figure B3: Mean HAZ by age (local polynomial smoothing), DHS data and simulated data
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 1826 days (5 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 B4.

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 B5 and Table B1.

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 B1.
Figure B4: Simulated true HAZ and HAZ with random month of birth

Figure B5: Simulated HAZ with varying shares of children with random month of birth. Dashed line represents December-January gap in DHS data
Table B1: Share of children with random month of birth, the simulated December-January gap and stunting rates

<table>
<thead>
<tr>
<th>Share random</th>
<th>Dec-Jan gap</th>
<th>Moderate stunting</th>
<th>Severe stunting</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHS</td>
<td>-0.313</td>
<td>0.354</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>Rate</td>
<td>Difference</td>
<td>Rate</td>
</tr>
<tr>
<td>0.000</td>
<td>0.352</td>
<td>-</td>
<td>0.145</td>
</tr>
<tr>
<td>0.010</td>
<td>-0.014</td>
<td>0.353</td>
<td>+0.001</td>
</tr>
<tr>
<td>0.020</td>
<td>-0.038</td>
<td>0.354</td>
<td>+0.002</td>
</tr>
<tr>
<td>0.030</td>
<td>-0.095</td>
<td>0.354</td>
<td>+0.002</td>
</tr>
<tr>
<td>0.040</td>
<td>-0.126</td>
<td>0.355</td>
<td>+0.003</td>
</tr>
<tr>
<td>0.050</td>
<td>-0.155</td>
<td>0.355</td>
<td>+0.003</td>
</tr>
<tr>
<td>0.060</td>
<td>-0.181</td>
<td>0.355</td>
<td>+0.003</td>
</tr>
<tr>
<td>0.070</td>
<td>-0.209</td>
<td>0.356</td>
<td>+0.004</td>
</tr>
<tr>
<td>0.080</td>
<td>-0.235</td>
<td>0.356</td>
<td>+0.004</td>
</tr>
<tr>
<td>0.090</td>
<td>-0.259</td>
<td>0.356</td>
<td>+0.004</td>
</tr>
<tr>
<td>0.100</td>
<td>-0.287</td>
<td>0.357</td>
<td>+0.005</td>
</tr>
<tr>
<td>0.110</td>
<td>-0.320</td>
<td>0.357</td>
<td>+0.005</td>
</tr>
<tr>
<td>0.120</td>
<td>-0.343</td>
<td>0.358</td>
<td>+0.006</td>
</tr>
<tr>
<td>0.130</td>
<td>-0.368</td>
<td>0.358</td>
<td>+0.006</td>
</tr>
<tr>
<td>0.140</td>
<td>-0.402</td>
<td>0.359</td>
<td>+0.007</td>
</tr>
<tr>
<td>0.150</td>
<td>-0.428</td>
<td>0.359</td>
<td>+0.007</td>
</tr>
<tr>
<td>0.160</td>
<td>-0.456</td>
<td>0.359</td>
<td>+0.007</td>
</tr>
<tr>
<td>0.170</td>
<td>-0.483</td>
<td>0.360</td>
<td>+0.008</td>
</tr>
<tr>
<td>0.180</td>
<td>-0.511</td>
<td>0.360</td>
<td>+0.008</td>
</tr>
<tr>
<td>0.190</td>
<td>-0.539</td>
<td>0.360</td>
<td>+0.008</td>
</tr>
<tr>
<td>0.200</td>
<td>-0.565</td>
<td>0.361</td>
<td>+0.009</td>
</tr>
<tr>
<td>0.210</td>
<td>-0.592</td>
<td>0.361</td>
<td>+0.009</td>
</tr>
<tr>
<td>0.220</td>
<td>-0.620</td>
<td>0.362</td>
<td>+0.010</td>
</tr>
<tr>
<td>0.230</td>
<td>-0.649</td>
<td>0.362</td>
<td>+0.010</td>
</tr>
<tr>
<td>0.240</td>
<td>-0.680</td>
<td>0.362</td>
<td>+0.010</td>
</tr>
<tr>
<td>0.250</td>
<td>-0.707</td>
<td>0.363</td>
<td>+0.011</td>
</tr>
</tbody>
</table>
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 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. E.g. if a child is reported to be 27 months old, this corresponds to 2 years and 3 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 the qualitatively the same pattern as in the DHS data. This is illustrated in Figure B6.

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 B7 and in Table B2. 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 Figure B8 and Figure B9.
Figure B6: HAZ by age in months in addition to age in year for the true simulated data and with asymmetric rounding error

Figure B7: Simulated HAZ with varying shares of children with asymmetric rounding error. Dashed line represents 11 months gap in DHS data.
Table B2: Share of children with asymmetric rounding error, the simulated 11 months gap and stunting rates

<table>
<thead>
<tr>
<th>Share with error</th>
<th>11 months gap</th>
<th>Moderate stunting</th>
<th>Severe stunting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate</td>
<td>Difference</td>
<td>Rate</td>
</tr>
<tr>
<td>DHS</td>
<td>0.000</td>
<td>-0.044</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>-0.072</td>
<td>0.351</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
<td>-0.088</td>
<td>0.350</td>
</tr>
<tr>
<td></td>
<td>0.030</td>
<td>-0.106</td>
<td>0.348</td>
</tr>
<tr>
<td></td>
<td>0.040</td>
<td>-0.123</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>0.050</td>
<td>-0.145</td>
<td>0.346</td>
</tr>
<tr>
<td></td>
<td>0.060</td>
<td>-0.164</td>
<td>0.344</td>
</tr>
<tr>
<td></td>
<td>0.070</td>
<td>-0.185</td>
<td>0.343</td>
</tr>
<tr>
<td></td>
<td>0.080</td>
<td>-0.206</td>
<td>0.342</td>
</tr>
<tr>
<td></td>
<td>0.090</td>
<td>-0.222</td>
<td>0.340</td>
</tr>
<tr>
<td></td>
<td>0.100</td>
<td>-0.237</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>0.110</td>
<td>-0.258</td>
<td>0.338</td>
</tr>
<tr>
<td></td>
<td>0.120</td>
<td>-0.274</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>0.130</td>
<td>-0.287</td>
<td>0.335</td>
</tr>
<tr>
<td></td>
<td>0.140</td>
<td>-0.308</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td>0.150</td>
<td>-0.324</td>
<td>0.332</td>
</tr>
<tr>
<td></td>
<td>0.160</td>
<td>-0.339</td>
<td>0.331</td>
</tr>
<tr>
<td></td>
<td>0.170</td>
<td>-0.344</td>
<td>0.329</td>
</tr>
<tr>
<td></td>
<td>0.180</td>
<td>-0.352</td>
<td>0.328</td>
</tr>
<tr>
<td></td>
<td>0.190</td>
<td>-0.365</td>
<td>0.327</td>
</tr>
<tr>
<td></td>
<td>0.200</td>
<td>-0.383</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td>0.210</td>
<td>-0.395</td>
<td>0.324</td>
</tr>
<tr>
<td></td>
<td>0.220</td>
<td>-0.406</td>
<td>0.323</td>
</tr>
<tr>
<td></td>
<td>0.230</td>
<td>-0.423</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td>0.240</td>
<td>-0.437</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>0.250</td>
<td>-0.445</td>
<td>0.319</td>
</tr>
</tbody>
</table>
Figure B8: Distribution of simulated ages with 7 percent asymmetric rounding error

Figure B9: Simulated HAZ by age with 7 percent asymmetric rounding error