ECD Working Paper Series

No. ECD WP010/2024

Measuring stunting and tracking trends in prevalence: Conceptual and methodological considerations with a focus on South Africa

Kate Rich, Liezel Engelbrecht, Gabrielle Wills and Ronelle Burger

August 2024

Stellenbosch University Iyunivesithi Universiteit
forward together sonke siya phambili saam vorentoe

RESEP Research on Socio-Economic Policy

ilifa labantwana

Measuring stunting and tracking trends in prevalence:
Conceptual and methodological considerations with a focus on South Africa

Kate Rich, Liezel Engelbrecht, Gabrielle Wills and Ronelle Burger

Abstract

Estimated stunting prevalence among children under 5 can vary widely across surveys, even over relatively short periods of time, making it challenging to track stunting prevalence over time and monitor progress towards reducing stunting. Measuring linear growth faltering in young children comes with a range of methodological and conceptual challenges, and surveys collecting height-for-age data in children do not follow a uniform methodology. This paper reviews the conceptual limitations of stunting as a measure of linear growth faltering and their implications for measuring stunting, as well as methodological considerations that may limit the comparability of stunting estimates across surveys. Stunting is an imperfect indicator of linear growth faltering at the individual level and underestimates the extent of linear growth faltering. The cut-off used to define stunting is arbitrary, with no biological basis. Measuring stunting in children under 5 underestimates the number of children affected by stunting at the age of peak prevalence (around 2 years of age), in many cases with lasting consequences. The comparability of estimates of stunting prevalence may be limited by differences across surveys in the composition and representativeness of survey samples, the extent of measurement error, and definitions and measurement procedures. We compare some of the more recent South African surveys that include height-for-age data, highlighting some measurement issues that may limit comparability. We conclude with recommendations to improve the measurement of stunting and linear growth faltering to improve the comparability of stunting estimates over time and allow better monitoring of progress towards reducing stunting.

View other papers in this series here.

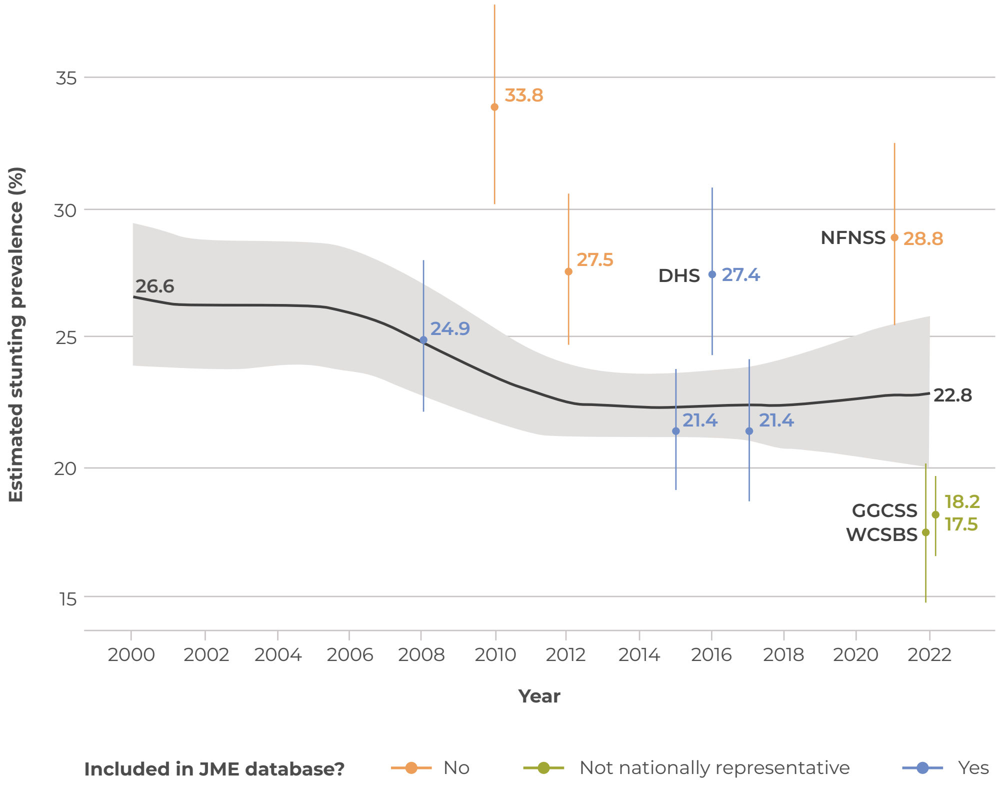
1. Introduction

South Africa is one of the 34 countries with 90 percent of the world's stunted children (Bhutta *et al*., 2013). Stunting – being too short for one's age – in early childhood is a marker of a poor growth environment, and is associated with lower cognitive development, poorer schooling outcomes, and lower health and earnings in adulthood (Black *et al*., 2013; Dewey and Begum, 2011). Childhood stunting can also have intergenerational effects, affecting birth outcomes and linear growth of the following generation (Dewey and Begum, 2011; Martorell and Zongrone, 2012; Victora *et al*., 2008). In recognition of the importance of child stunting for human development, in 2012 the World Health Organization (WHO) set a target of reducing the number of children under 5 who are stunted by 40 percent by 2025 (World Health Organization, 2014). We need accurate estimates of stunting prevalence to monitor South Africa's progress towards these goals.

A systematic review of fifty South African studies covering the period 1970 to 2013 suggested that the prevalence of stunting among South African children under 6 increased after 1993 but declined in the decade from 2003 to 2013 (Said-Mohamed *et al*., 2015). However, differences in age ranges, representativity, and growth standards used in these studies limited their comparability. Since that study was published, several nationally representative surveys and district- or province-specific surveys have collected data on linear growth among children under 5, but estimates from these surveys differ quite widely.

A joint UNICEF, WHO and World Bank team releases estimates of global and regional child malnutrition every two years. These Joint Child Malnutrition Estimates (JME) are based on a model incorporating estimates from nationally representative surveys, with non-linear time trends, regional and country-specific trends, and taking into account a country's socio-demographic index and average health system access over the previous five years, as well as the uncertainty of the survey data source (see UNICEF-WHO-World Bank Joint Child Malnutrition Estimates (JME) Working Group, 2021, for details of this model). Figure 1 plots the JME estimates for South Africa from 2000 to 2022, as well as the nationally representative survey estimates on which their estimates were based (blue) and available estimates from surveys that were excluded from the JME estimates (orange) (possibly due to concerns with data quality). The JME estimates suggest a decline in stunting rates from 2000 to about 2014, corresponding to the decline found by Said-Mohamed *et al*. (2015), but no change in the years since (2014 to 2022). However, the sparsity and variability of the survey data on which these estimates are based means that the confidence intervals are quite wide and overlap to some extent. Estimates from the individual surveys show even greater variability. For instance, the prevalence of stunting among children under 5 was estimated at 25 percent in NIDS 2008, and 34 percent in NIDS 2010. NIDS 2014/5 estimated stunting at 21 percent, the DHS 2016 at 27 percent, and NIDS 2017 at 21 percent. Over the five-year period from 2010 to 2015, we thus have estimates of stunting prevalence differing by about 12 percentage points. How can nationally representative surveys a year or two apart produce such widely varying estimates? Given the variation in estimates and wide confidence intervals of available stunting estimates, it is difficult to say with any certainty what has happened to stunting in South Africa in recent years.

Figure 1: Estimated stunting prevalence among children under 5 in South Africa between 2000 and 2022: UNICEF-WHO-World Bank Joint Child Malnutrition Estimates and South African survey estimates



Source: JME estimates from UNICEF-WHO-World Bank JME database, May 2023 (UNICEF-WHO-World Bank Joint Child Malnutrition Estimates (JME) Working Group, 2023). Estimates for NIDS Waves 2 and 3 (not included in JME database) are based on authors' own calculations.

Note: The black line represents the UNICEF-WHO-World Bank JME estimated stunting prevalence, with confidence intervals in grey. The coloured data points are point estimates with confidence intervals from national surveys as follows: 2008 = NIDS Wave 1, 2010 = NIDS Wave 2, 2012 = NIDS Wave 3, 2015 = NIDS Wave 4, 2016 = DHS, 2017 = NIDS Wave 5, 2021 = National Food and Nutrition Security Survey (NFNSS). Estimates from the Grow Great Community Stunting Survey (GGCSS) and Western Cape Stunting Baseline Survey (WCSBS) are included for comparison, but are not nationally representative. The surveys in blue were included in the database used to produce the Joint Child Malnutrition Estimates (JME); the surveys in orange (NIDS waves 2 and 3 and the NFNSS) were not. The JME estimates suggest a decline in stunting in South Africa since 2000, with the decline concentrated between around 2006 and 2012. However, the sparsity and variability of the surveys on which the estimates are based means that the confidence intervals are wide.

This paper has three main aims. First, it reviews some of the pitfalls in comparing stunting estimates across surveys, as well as the conceptual limitations of stunting as a measure of linear growth faltering and their implications for measuring stunting. Second, we compare some of the more recent estimates of stunting among South African children under 5, highlighting factors that may limit their comparability and lead to large variations in estimated stunting rates. Finally, the review moves towards outlining key considerations for improving the measurement of height-for-age, in turn enabling more comparable estimates of stunting prevalence over time. The paper aims to support improved understanding and practice in the measurement and analysis of stunting, with a particular focus on South Africa. It is intended to promote discussion and support consistency and common understanding in the measurement of height-for-age and stunting across key stakeholders, and in turn promote cohesion among the community concerned with tracking and monitoring stunting.

First, in Section 2 we discuss some of the limitations of stunting as a measure of linear growth faltering. Next, in Section 3 we discuss some of the issues affecting the comparability of stunting estimates – even across surveys designed to be nationally representative. Broadly, these issues can be broken down into sample representativeness and measurement issues. In Section 4 we compare six of the most recent South African surveys measuring stunting in terms of these factors: three nationally representative surveys – the South African Demographic and Health Survey (2016), the National Income Dynamics Study Wave 5 (2017) and the National Food and Nutrition Security Survey (2021) – and three surveys focused on specific districts, provinces or populations – the Grow Great Community Stunting Surveys (2022), the Western Cape Stunting Baseline Survey (2022), and the Thrive by Five Index (2021). Section 5 discusses some implications and recommendations for improving the measurement of stunting and the comparability of stunting estimates across surveys and over time.

2. Limitations of stunting as an indicator of linear growth faltering

The rate of stunting in a population is easy to understand and interpret, making it useful as a goal for policy. However, stunting is an imperfect indicator of linear growth faltering at the individual level and suffers from several conceptual issues.

Linear growth faltering is a failure to reach one's linear growth potential (Leroy and Frongillo, 2019). Stunting is having a height-for-age z-score (HAZ) more than two standard deviations below the median of a reference population. Most (but not all) children who are classified as stunted suffer from linear growth faltering, though 2.3 percent of children in a healthy population have a HAZ below -2 and would be classified as stunted. However, a child can experience linear growth faltering without being stunted; many more children suffer from linear growth faltering than are stunted (Leroy and Frongillo, 2019). Unfortunately, a child's individual linear growth potential, and therefore growth faltering, is very difficult to measure[1].

Height-for-age Z-scores and stunting rates were originally intended as population-level indicators; they are imperfect indicators of growth faltering at the individual level (Perumal *et al*., 2018). The WHO Child Growth Standards (as well as other growth references) reflect the distribution of HAZ for children in an environment conducive to healthy growth. Children are regarded as stunted if they are more than 2 standard deviations below the median of this distribution. On a population level, a high prevalence of children more than 2 standard deviations below the median of this distribution (i.e., stunted) indicates that the HAZ distribution is shifted to the left of the WHO growth standards – how children should grow in a healthy environment – and that many children are not reaching their growth potential. But on the individual level, the median of the growth standards distribution does not reflect an individual child's genetic growth potential. A child may have suffered from greater growth faltering than another child with a lower HAZ if the child's genetic growth potential was higher to begin with (see Figure 2).

The prevalence of stunting likely underestimates the true extent of linear growth faltering (de Onis and Branca, 2016; Leroy and Frongillo, 2019; Perumal *et al*., 2018; Roth *et al*., 2017). In low- and middle-income countries (LMICs) the entire distribution of height-for-age Z-scores is shifted to the left of the WHO Child Growth Standards distribution (de Onis and Branca, 2016; Roth *et al*., 2017), suggesting that many children above the cut-off also experience growth faltering, failing to reach their growth potential.

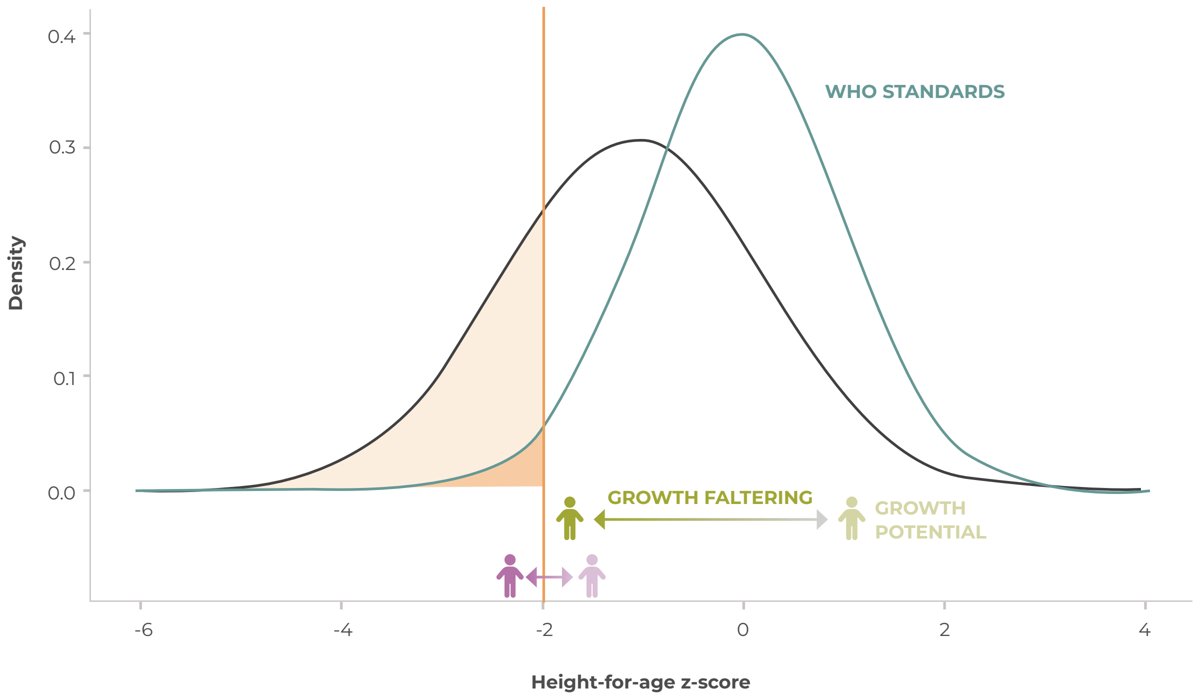
A further issue is that stunting is a binary indicator, marking children with a height-for-age z-score more than 2 standard deviations below the WHO Child Growth Standards median as having experienced linear growth faltering and children above the cut-off as having normal growth. But this cut-off has no biological or clinical

[1]Mid-parental height – the average of the parents' heights – may be a useful indicator of a child's expected adult height. However, if the parents themselves did not reach their growth potential, as is common in contexts of widespread undernutrition, then mid-parental height would be less useful as a predictor of a child's adult height (Garza *et al*., 2013). While much of the variation in height is genetic, a wide range of genes are associated with height, and the genetic basis of height is still poorly understood (McEvoy and Visscher, 2009).

basis (Leroy and Frongillo, 2019; Perumal *et al*., 2018). The risks associated with stunting increase continuously as height-for-age decreases (de Onis and Branca, 2016; Perumal *et al*., 2018), and are unlikely to be much larger for children just below the cut-off than for children just above it.

As a final caveat, Leroy and Frongillo (2019) argue that linear growth faltering or stunting do not directly cause many of the poor outcomes found to be associated with them, such as impaired cognitive development, poorer schooling outcomes, reduced earnings, and higher risk of chronic disease. Rather, linear growth faltering is a marker of a deficient early environment, which causes both linear growth faltering and these other outcomes. Linear growth faltering is a useful marker, they argue, but improving linear growth alone is not sufficient for optimal child development.

Figure 2: An illustration of some conceptual issues with measuring stunting



Source: Own illustration based on concepts discussed in de Onis and Branca, 2016; Leroy and Frongillo, 2019; Perumal *et al*., 2018; Roth *et al*., 2017.
Note: 1) Stunting underestimates the true extent of linear growth faltering: in developing countries, the whole HAZ distribution (black line) is shifted to the left of the WHO Child Growth Standards distribution (blue line, representing how children should grow in a healthy growth environment).

2) HAZ and stunting are imperfect measures of growth faltering at the individual level. The solid green and pink figures represent the actual HAZ of two individual children; the transparent figures represent the growth potential of those children. The green child has a higher HAZ than the pink child but is experiencing greater growth faltering than the pink child – falling further behind how that child would have grown in an optimal growth environment.

3) Problems with stunting as a binary measure: the solid green child has a HAZ not much higher than the solid pink child, but the pink child falls below the -2 cut-off and is considered stunted, while the solid green child is not. The cut-off has no biological basis; the risks associated with linear growth faltering are unlikely to be much larger for children just below the cut-off than for children just above it.

3. Factors limiting comparability of stunting estimates

3.1 Composition and representativeness of survey samples

3.1.1 Age

Stunting prevalence differs markedly with age. The bulk of linear growth faltering happens within the first two years of life, but stunting takes time to develop. Across LMICs mean height-for-age z-scores tend to be lower than the growth standard at birth, but decline sharply between 6 and 18 months of age, reaching their lowest point around 24 months (Leroy *et al*., 2014; Victora *et al*., 2010). The prevalence of stunting also peaks around 24 months. Estimates from a South African urban birth cohort suggested that 90 percent of the deficit in eventual adult height was already established by 24 months of age (Lundeen *et al*., 2014). From 2 to 5 years HAZ scores tend to stabilise or even improve slightly (see Box 1).

Samples with different age distributions may thus produce different stunting estimates. The prevalence of stunting is most commonly reported for children aged 0-59 months (i.e. children under 5 years of age). However, some surveys only include children aged 6 months and above. Others include children up to 6 years (see for example the age ranges in the review by Said-Mohamed *et al*., 2015). Given that stunting prevalence is lowest among children under 6 months, as they are still in the process of becoming stunted, surveys excluding children under 6 months are likely to overestimate stunting prevalence relative to those that include them. Given that height-for-age z-scores tend to improve somewhat after 24 months, surveys targeting older children (e.g. preschool-aged children) are likely to produce lower stunting estimates than those targeting children across the 0-59 months age range.

Even surveys targeting the same age range may have different sample age distributions due to sampling variation or due to the peculiarities of the data collection process. For example, mothers of very young children may be more likely to refuse to allow their children's anthropometric measurements to be taken; older children may be more likely to be at an ECD centre or school during the day when fieldworkers visit and be unavailable for measurement.

Box 1: Can children "catch up" from stunting?

Stunting tends to peak and height-for-age z-scores (HAZ) reach their lowest point at 24 months (see Figure 3). After this point z-scores tend to stabilise or even improve somewhat (Victora *et al*., 2010), and many studies have found that stunting rates are lower at 4 or 5 years than at 2 years. Some reversal of stunting status appears possible even before the age of 15 months, but even when it does occur, many children eventually relapse (Benjamin-Chung *et al*., 2023). These findings have led to debate over whether children can 'catch up' or recover from the effects of stunting.

Part of the debate centres around the definition of catch-up growth. Some studies define catch-up as an improvement in HAZ between early childhood (usually measured at 6-18 months or 24 months) and mid or later childhood (age ranges vary, but often around 5 years). Others define catch-up as being stunted in early childhood and no longer stunted later on, i.e. crossing the -2 threshold. Using these definitions a substantial proportion of children who were stunted at

2 years appear to catch up. In the South African Birth to Twenty urban cohort, 93 percent of children caught up using the first definition and 75 percent using the second (Desmond and Casale, 2017).

But part of the apparent catch up may be driven by the construction of HAZ. HAZ is the absolute height deficit (the cm difference between the child's height and the growth standards median) divided by the standard deviation of height for the child's age and sex. But the standard deviation of height widens with age as height increases. This means that HAZ may improve with age even when the absolute height deficit has not – even when the height deficit has worsened slightly (Leroy *et al*., 2014). While average HAZ tends to stabilise or increase slightly after 24 months, average height deficits continue to widen with age (Leroy *et al*., 2015, 2014; Lundeen *et al*., 2014). For this reason Leroy *et al*. (2015) propose that absolute height deficits should be used to assess catch-up growth. But even using this stricter definition, Desmond and Casale (2017) found that a substantial number of individual children did experience a narrowing of the height deficit with age.

While catch up in linear growth may have direct benefits – for example, children of taller mothers have better birth outcomes (Leroy and Frongillo, 2019) – much of the debate has centred on whether catch up growth helps to remedy the damage to cognitive function associated with early stunting. The first 1000 days of life, from conception to 2 years, are a critical period for brain development. Poor cognitive development in this period may be irreversible, and have long-term consequences for human capital (Victora *et al*., 2008). Findings on whether cognitive function recovers when children 'recover' from stunting have been mixed. Several studies have found that children who catch up do no worse than children who were never stunted (Crookston *et al*., 2013, 2010). But other studies found that children who recover from stunting do better than those who remain stunted, but still worse than children who were never stunted (Casale, 2020; Casale *et al*., 2020).

Casale *et al*. (2020) show that one reason for the differing results may lie in the different age ranges used in the two sets of studies. The studies that have found that children who catch up recover in terms of cognitive function measure stunting between 6 and 18 months, while the studies that have found that children who catch up still do worse measure stunting at 24 months. As stunting only tends to peak around 24 months, children younger than 24 months are still becoming stunted. Casale *et al*. (2020) find that only children who recover from stunting at 2 years to the point that they are within the 'normal' range (HAZ >-1) at 5 years do no worse than children who were never stunted (but very few children caught up to this extent, raising possible concerns around sample size). Adoption and foster care studies support the idea that partial recovery in linear growth and cognitive development is possible when a child's environment is improved dramatically (Leroy *et al*., 2020). But most children who are stunted do not experience such a dramatic improvement in their environment.

More research on the effects of catch-up growth is necessary, but currently the weight of the evidence seems to suggest that being stunted at 2 years is associated with lower cognitive function compared to children who were never stunted, even if children subsequently catch up in height – perhaps unless the catch up is very substantial (and relatively few children do catch up to this extent).

3.1.2 Other sampling issues

3.1.2.1 Sex

Boys are more likely to be stunted than girls. Sampling variation or bias may result in differing sex distributions across surveys. The World Health Organization and United Nations Children's Fund (2019) recommend comparing the sex ratio of children in a survey to the expected sex ratio for that country. If sex ratios differ substantially across surveys, this could limit the comparability of estimated stunting prevalence.

3.1.2.2 Socioeconomic status

Stunting rates are likely to be lower in surveys where the sample is richer on average. If differences in average socioeconomic status (SES) reflect real differences in the population (e.g. improvements in SES over time), then differences in stunting rates may reflect genuine differences in population stunting rates. But if average SES differences across surveys reflect sampling differences (e.g. over-sampling or under-sampling of richer households), then stunting estimates may be biased and differences in stunting prevalence between surveys would not reflect true differences in the population.

Differences in the socioeconomic composition of survey samples could result from differences in sampling strategies, such as sampling children at ECD centres versus in the home. Children from higher-SES backgrounds are more likely to attend ECD facilities (Moses, 2021), so surveys sampling children at ECD centres are likely to produce a sample with a higher average SES than the national population and, therefore, lower estimates of stunting prevalence than nationally representative surveys.

Even if children are successfully sampled (i.e. their caregiver answers a questionnaire about them), this does not mean that their anthropometric (height and weight) measurements are successfully taken. For example, in NIDS Wave 1 (2008), 31 percent of children under 5 who were 'successfully interviewed' were missing height data; in Wave 2 (2010/11) this figure was as high as 50 percent. In the 2016 SA DHS, 26 percent of children had missing height measurements. Caregivers may consent for their children to participate in the survey but not for their height or weight to be measured, or children may not be present when fieldworkers visit (e.g. they may be attending an ECD centre/school). If there is selection into measurement – for example, if children who are less likely to be stunted are also less likely to have their height measurement taken – stunting estimates would be biased. For example, in a household survey higher-SES children could be less likely to be measured because they are more likely to be attending an ECD centre and therefore not at home when fieldworkers visit, or high-SES caregivers may be less likely to consent to measurement.

Survey weights attempt to correct for unequal probabilities of being sampled, but they do not account for the fact that fairly large proportions of sampled children do not have their height measurements taken. Thus, even if the survey is a nationally representative sample of children under 5, the subset of children with non-missing height data may not be. Surveys should ideally report the number and percentage of children whose heights or ages were not recorded and the distribution of these children across key variables (e.g. mother's education, income or wealth quintile, province, urban/rural location) to allow for an assessment of the risk of bias. Surveys should also record if a household or child was sampled but not successfully interviewed (e.g. they were not present or refused participation). There is scope for further research to explore the extent of selection into measurement and the possibility of creating survey weights to correct specifically for differing probabilities of successful anthropometric measurement. Further investigation into the socioeconomic composition of the survey samples on which stunting estimates are based would be valuable, but is beyond the scope of this paper.

3.2 Measurement issues

3.2.1 Definitions of stunting and measurement procedures

Differences in estimated stunting rates across studies could arise from using different definitions of stunting, using different growth references, or from differences in measurement procedures – in particular, whether studies adjust height-for-age for gestational age, and whether they record and adjust for measuring height instead of length in children less than 2 years old. These issues are discussed in more detail below.

3.2.1.1 Definitions of stunting

Children are considered stunted if their height or length is more than 2 standard deviations below the WHO Child Growth Standards' median height for their age and sex (WHO Multicentre Growth Reference Study Group, 2006). Children whose height or length for age is more than 3 standard deviations below the median are considered severely stunted (de Onis and Branca, 2016). Some studies additionally define children as mildly stunted if their height- or length-for-age is more than 1 standard deviation below the median. In the past some studies have used different definitions, such as a height-for-age below 80 percent or 90 percent of the median or below certain percentiles of the reference population (Said-Mohamed *et al.*, 2015), but the definitions given above are now standard. Differences in stunting rates between surveys or studies published since 2006 are thus unlikely to result from differing definitions of stunting.

3.2.1.2 Different growth references

Before 2006, the WHO recommended using the National Center for Health Statistics (NCHS) growth reference to assess stunting in children. This growth reference was based on data for children of European ancestry in the US, most of whom were formula-fed (WHO Multicentre Growth Reference Study Group, 2006). Some countries developed their own growth references, such as the CDC 2000 and British 1990 growth references. In 2006 the WHO published new growth standards based on a study of healthy, breastfed infants born to non-smoking mothers in the USA, Oman, Norway, Brazil, Ghana and India (WHO Multicentre Growth Reference Study Group, 2006). More affluent neighbourhoods and individuals were selected to reflect socio-economic conditions conducive to growth. The WHO 2006 standards thus reflect normal growth patterns – "how children should grow" (WHO Multicentre Growth Reference Study Group, 2006) – in a healthy growth environment, regardless of ethnic or geographic background.

Height-for-age Z-scores and stunting rates calculated using different growth standards are not comparable. Stunting rates based on the WHO 2006 standards are higher than those based on the previous NCHS international growth reference (WHO Multicentre Growth Reference Study Group, 2006). The use of different growth references affects the comparability of estimates before and after 2006, but since 2006 studies have generally used the new WHO 2006 growth standards, so differences in estimates since 2006 are not likely to be driven by the use of different standards.

3.2.1.3 Standing height versus recumbent length

Children younger than 2 years old should be measured lying down (length-for-age), while the children 2 years and older should be measured standing up (height-for-age) (de Onis and Branca, 2016). Recumbent length tends to be greater than standing height by about 0.7 cm (WHO Multicentre Growth Reference Study Group, 2006). If height is measured instead of length or vice versa, this can be adjusted for. But if a survey does not record whether height or length was measured, estimates of stunting rates may be biased.

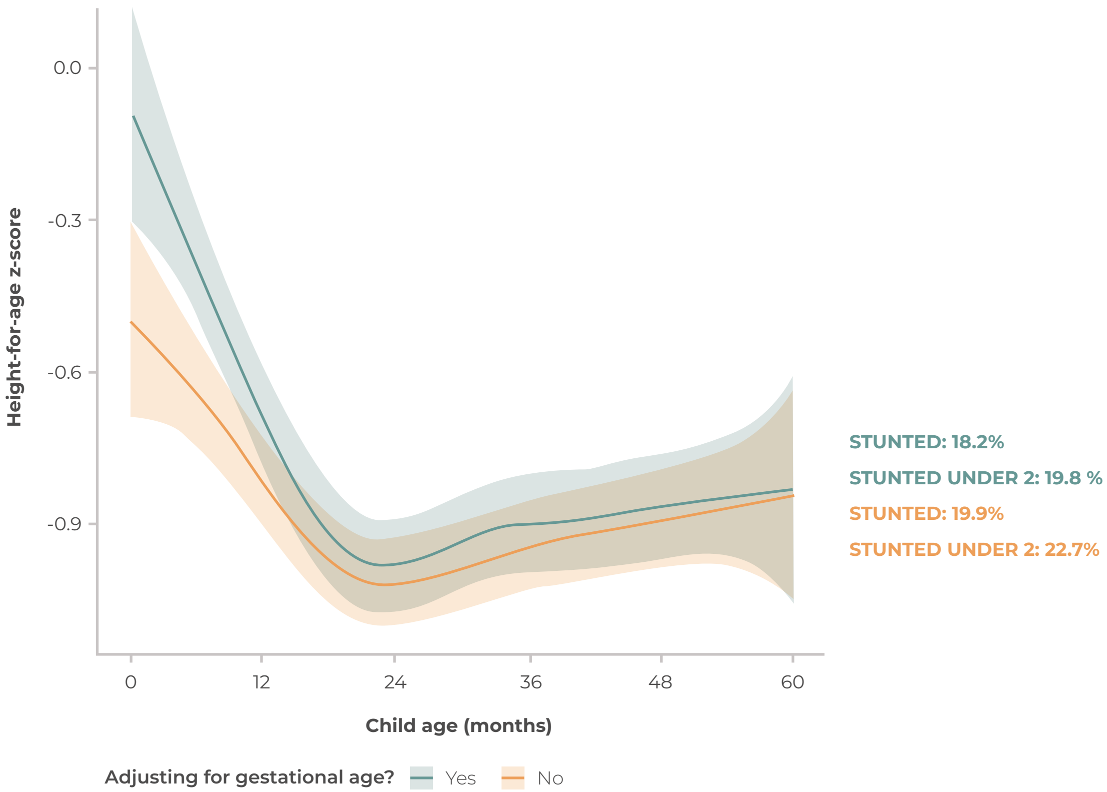
3.2.1.4 Adjusting for gestational age

Adjustment for gestational age is another possible source of differences in stunting estimates. Failing to adjust for gestational age means that one is not comparing apples with apples: two children of the same chronological age (measured as time since birth) may have different corrected or conceptional age (measured as time since conception). For example, at a chronological age of 6 months a child born at 32 weeks would have an adjusted age of 4 months (adjusting for the fact that the child was born 8 weeks premature).

When measuring age from conception, a baby born prematurely is younger (has had less time to grow) when compared with a baby born at full term, even if the two children are of identical age since birth. A child born at 32 weeks cannot be expected to have the same height as a child born at 40 weeks when they are assessed in the first year or two of life. However, adjusting for gestational age may also introduce another source of measurement error if gestational age is inaccurate.

As shown in Figure 3, failing to adjust for gestational age results in lower height-for-age z-scores (and therefore inflated estimates of stunting). The difference is greatest among younger children.

Figure 3: The effect of adjusting for gestational age on height-for-age z-scores

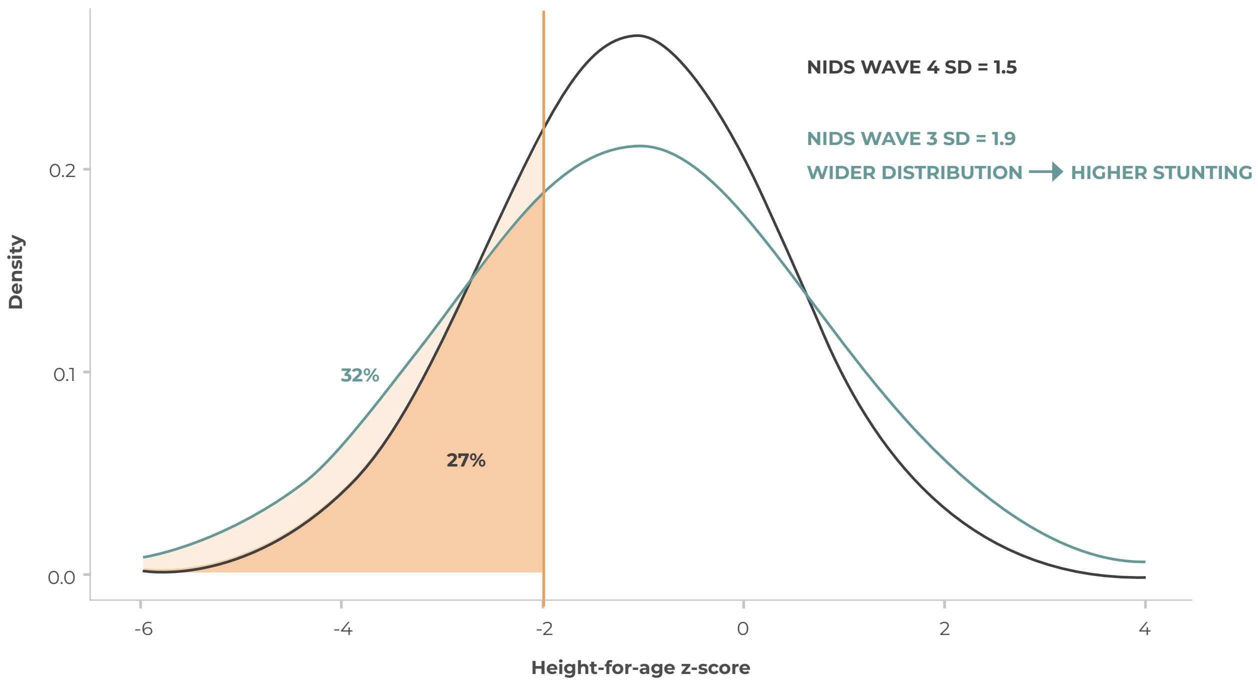


Source: Rich *et al.* (2024, forthcoming), based on Grow Great Community Stunting Survey data.
Note: The blue line represents mean HAZ by age after adjusting for gestational age; the yellow curve mean HAZ without adjusting for gestational age. The transparent bands are 95 percent confidence intervals. Adjusting for gestational age results in higher height-for-age z-scores; the difference is most pronounced below 12 months of age.

3.2.2 Measurement error

Measuring height is a manual process prone to error. Measurement error artificially inflates estimated stunting prevalence (Grellety and Golden, 2016). Differing levels of measurement error across surveys may thus limit the comparability of stunting estimates.

Figure 4: The effect of measurement error on estimated stunting prevalence: an illustration using NIDS



Source: Own calculations using NIDS data, based on concepts in Grellety and Golden (2016)
Note: The black line shows a normal distribution of HAZ with a mean of -1.1 and a standard deviation of 1.5, corresponding to the (unweighted) mean and SD found in NIDS Wave 4. The blue line shows a normal distribution of HAZ with a mean of -1.1 and a standard deviation of 1.9, corresponding to the (unweighted) mean and SD found in NIDS Wave 3. The percentages are the percentage of children with a HAZ < -2 (stunted) in each wave based on these distributions (i.e. the cumulative distribution function). Though mean HAZ is the same in the two waves, the greater spread of HAZ (wider SD) in wave 3 means that more children are classified as stunted. In reality, the distribution of height-for-age z-scores in NIDS is not perfectly normal, so the actual prevalence of stunting in NIDS differs somewhat.

Greater measurement error leads to a height-for-age distribution that is more spread out – reflected in wider standard deviations – as measurement error leads to more children having extreme values of height-for-age (being in the tails of the distribution). This means that a greater percentage of children will fall below a height-for-age z-score of -2 and be classified as stunted (see Figure 4). Estimates of severe stunting are even more sensitive to measurement error (World Health Organization and United Nations Children's Fund, 2019). Larger sample sizes do not reduce the effect of measurement error (Grellety and Golden, 2016).

Measurement error may result from rounding off of height measurements, age heaping (for example due to rounding age to full years), or even data fabrication (World Health Organization and United Nations Children's Fund, 2019). Measurement error may also differ across surveys because they use different procedures for taking repeated measurements on the same child and dealing with differences between them. Surveys allowing a smaller discrepancy before an additional measurement must be taken are likely to be more accurate. Taking more than one measurement may also result in more accurate data. The accuracy of the equipment used to measure height may also differ between surveys. Surveys using fieldworkers with a background in nutrition or experience of taking anthropometric measurements would also be expected to yield more accurate height data than those using fieldworkers with no such background. The extent and quality of fieldworker training may also influence accuracy. At the sub-national level, specific fieldworkers or fieldworker teams operating in specific areas may be more or less precise in taking measurements, leading to different levels of measurement error across regions and thus complicating comparisons across provinces or districts.

The handling of implausible values due to measurement error is another possible source of discrepancies between surveys. The WHO and UNICEF recommend treating height-for-age z-scores with an absolute value greater than 6 (i.e. less than -6 or greater than 6) as biologically implausible and excluding them from stunting estimates (World Health Organization and United Nations Children's Fund, 2019). However, some studies have used different cut-offs. For example, Stata's user-written zanthro macro commonly used to calculate z-scores uses a default cut-off of 5 (scores greater than or equal to 5 standard deviations above or below the mean). Studies using a less conservative cut-off (i.e. a higher absolute value) allow more extreme values and so are likely to have wider standard deviations and classify a greater proportion of children as stunted than studies using a more conservative (lower absolute value) cut-off.

Stunting estimates are likely to be more affected by measurement error among younger children. Standard deviations of height-for-age z-scores tend to be higher among younger children (World Health Organization and United Nations Children's Fund, 2019). Younger children are more difficult to measure accurately, as length is more difficult to measure than height (World Health Organization and United Nations Children's Fund, 2019). Furthermore, because the standard deviation of height (the denominator in the calculation of height-for-age z-scores) widens with age, a given absolute measurement error in centimetres translates into a higher error in height-for-age z-score for younger children. This issue may exacerbate the problems associated with age distributions differing across surveys, as discussed in Section 3.1.1.

3.2.2.1 Possible indicators of poor data quality

The practice of using standard deviations as a blanket indicator of data quality has been criticised (Sandler, 2021), but on the whole, higher standard deviations tend to indicate poorer data quality (World Health Organization and United Nations Children's Fund, 2019). A wider standard deviation may be due to population heterogeneity (for instance, in a population with high economic inequality, where some children face an optimal growth environment and others an environment that constrains their growth, children's heights may be more widely dispersed) rather than measurement error. However, as standard deviations become larger, it becomes more likely this is due to data quality issues[2] (World Health Organization and United Nations Children's Fund, 2019). The standard deviation of height-for-age z-scores in the WHO Child Growth Standards reference population is 1 (by definition), but more heterogeneity is to be expected in malnourished populations, where some children are on track and others malnourished (World Health Organization and United Nations Children's Fund, 2019). The median standard deviation in the JME database of 474 nationally representative household surveys from 112 countries, as of 2019, was 1.54 (World Health Organization and United Nations Children's Fund, 2019). While there is as yet no rule-of-thumb for standard deviations indicating poor data quality (and it is debatable whether there should be), and high standard deviations do not always reflect poor data quality, relatively high standard deviations should at least raise concerns and merit further investigation.

As height-for-age z-scores tend to be approximately normally distributed, a relatively high prevalence of severely stunted (HAZ below -3) relative to moderately stunted (HAZ below -2 but greater than or equal to -3) children may also suggest poor data quality[3]. Based on a normal distribution (with a mean above -2) fewer children would tend to be severely stunted than moderately stunted. More extreme values tend to be less common, so if extreme values such as a HAZ of less than -3 are relatively common, this may indicate that some of the extreme values are due to measurement error. Additional data quality checks are described by the World Health Organization and United Nations Children's Fund (2019, Section 3.1). Comparing estimates from surveys with varying degrees of measurement error may result in an apparent difference in stunting prevalence when there is none.

[2]Particularly in the case of two consecutive waves of a longitudinal survey conducted two years apart, a wider standard deviation in one wave seems more likely to be due to measurement error than due to an actual greater spread of heights in the data.
[3]An exception to this is if mean HAZ is below -2. However, in this case stunting rates would be more than 50 percent (based on an approximately normal distribution), which is far beyond the stunting rates observed in South Africa. The mean HAZ has been below -2 in some DHS surveys in other LMICs, but this is relatively uncommon (Perumal *et al*., 2018).

4. Comparison of recent South African stunting estimates

Table 1 presents results from some recent South African surveys including anthropometric data for children under 5. The table outlines sample sizes by age categories of children and stunting prevalence rates by age category, gender and urban/rural location where possible. It also provides some specificity on the quality training of assessors, measurement setting of each study, how anthropometric measurement was approached in each study and whether parental height and gestational age were considered, where this was mentioned in the report.

Table 1: Comparison of aspects of data collection, sampling and measurement that may influence reported stunting estimates from recent South African surveys

	Grow Great Surveys (GGCSS) (2022)	WC Stunting Baseline (WCSB) (2022)	National Food and Nutrition Security Survey (NFNSS)	Thrive by Five (TBF) (2021)	NIDS Wave 5 (2017)	SA-DHS (2016)
Sample size: 0-59 mths	In 7 districts: 3221 With non-missing HAZ: 3211	In WC: 1214 With non-missing HAZ: 1202	6265		With non-missing HAZ: 3714	Total: 2024 With non-missing HAZ: 1494 In WC: 50
Overall stunting: 0-59 mth	In 7 districts: 18.2%	In SA: - In WC: 17.5%	28.8%		21%	In SA: 27 % In WC: 23%
Moderate stunting (HAZ <-2 & >=-3)	14.9%	Not available	14%	4.6%	13.9%	17.6%
Severe stunting (HAZ <-3)	3.3%	Not available	14.8%	0.5%	7.1%	9.8%
Sample size: approx. 4-5 yrs	495	In WC: 187 (4-5 yrs)	42-53 mths: 1142 54-59 mths: 384	5222 In WC: 559 (50-59 mths)	896	330
Stunting: approx. 4-5 yrs in SA	12.9% (4-5 yrs)	In SA: - In WC: 15,7% (4-5 yrs)	42-53 mths: 20.2% 54-59 mths: 20.2%	In SA: 6% (weighted 5.1%) In WC: 5,2% (50-59 mths)	13.6% (4-5 yrs)	In SA: 15.3% (4-5 yrs)
Sample size: 0 to <6 mths	284	In SA: - In WC: 69	560	NA	NA	132
Stunting: 0 to <6 mths	8.6%	In SA: - In WC: 25,2%	24.5%	-	NA	32%
Age distribution: N (weighted %)	0 to 11 mths = 606 (18.3%) 12 to 23 mths = 765 (24.4%) 24 to 35 mths = 746 (23.8%) 36 to 47 mths = 608 (18.3%) 48 to 59 mths = 495 (15.3%)	0 to 11 mths = 196 (17.5%) 12 to 23 mths = 271 (23.4%) 24 to 35 mths = 315 (20.7%) 36 to 47 mths = 245 (21.6%) 48 to 59 mths = 187 (16.6%)	(unweighted %) 0 to 5 mths = 599 (9.2%) 6 to 17 mths = 1578 (24.1%) 18 to 29 mths = 1447 (22.1%) 30-41 mths = 1288 (19.7%) 42 to 53 mths = 1199 (18.3%) 54 to 59 mths = 434 (6.6%)	50-59 months (only)	6 to 11 mths = 363 (8.9%) 12 to 23 mths = 754 (21%) 24 to 35 mths = 831 (23.8%) 36 to 47 mths = 870 (23.9%) 48 to 59 mths = 896 (22.4%)	0 to 11 mths = 255 (17.1%) 12 to 23 mths = 292 (21.3%) 24 to 35 mths = 299 (18.6%) 36 to 47 mths = 318 (21.1%) 48 to 59 mths = 330 (21.8%)
Stunting prevalence by age:	0 to 11 mths = 14.3% 12 to 23 mths = 24.0% 24 to 35 mths = 18.4% 36 to 47 mths = 18.6% 48 to 59 mths = 12.9%	0 to <6 mths= 25.2% 6 to 11 mths = 7.7% 12 to 23 mths = 23.9% 24 to 35 mths =18.0% 36 to 47 mths = 14.3% 48 to 59 mths = 15.7%	0 to 5 mths = 24.5% 6 to 17 mths = 34.1% 18 to 29 mths = 35.9% 30-41 mths = 25.4% 42 to 53 mths = 20.2% 54 to 59 mths = 20.2%	-	6 to 11 mths = 12.5% 12 to 23 mths = 34.9% 24 to 35 mths = 21.7% 36 to 47 mths = 18.1% 48 to 59 mths = 13.6%	0 to 11 mths = 25.2% 12 to 23 mths = 35.9% 24 to 35 mths = 33.4% 36 to 47 mths = 27.5% 48 to 59 mths = 15.3%

	Grow Great Surveys (GGCSS) (2022)	WC Stunting Baseline (WCSB) (2022)	National Food and Nutrition Security Survey (NFNSS)	Thrive by Five (TBF) (2021)	NIDS Wave 5 (2017)	SA-DHS (2016)
Urban / Rural sample distribution	Urban: 56.7% Rural: 43.3%	Rural towns: 157 Urban informal: 162 Urban formal: 883	Not available	Not available	Urban: 55.6% Rural: 44.4%	Urban: 51% Rural: 49%
Urban/Rural Stunting prevalence	Urban: 17% Rural: 19.8%	Rural towns: 25.8% Urban informal: 20.5% Urban Formal: 16.5%	Not available	Not available	Urban: 19.6% Rural: 22.7%	Urban: 25.6% Non-urban: 29.1%
Gender split	Male: 51.8% Female: 48.2%	Male: 49.6% Female: 50.4%	Male: 49.9% Female: 50.1%	Male: 48% Female: 52%	Male: 49.9% Female: 50.1%	Male: 51.1% Female: 48.9%
Setting of measurement	Home	Home	Home	Early Learning Programmes (ELPs)	Home	Home
Measurement training quality	Training by GG team on anthropometric measures. Organisations provided supervision in ensuring anthropometric measurements were conducted accurately and standard operating procedures adhered to. Preference was given to fieldworkers with experience in taking anthropometric measurements of young children, a nursing background, those who had worked with ikapadata before and/or had attended the first round of fieldworker training in 2019 Fieldworkers were evaluated in the field based on their knowledge of the project, their capacity to conduct measurements correctly, as well as technical factors in conducting fieldwork according to training protocols. A final group of 35 fieldworkers were selected based on their performance during training	An experienced registered dietician oversaw and coordinated all aspects of data collection, including quality control of completed questionnaires and logistics. Comprehensive training and standardisation of fieldworkers by an expert team led by a registered research dietician with comprehensive experience in fieldworker training for anthropometric and dietary surveys; calibration of equipment for anthropometric measures as recommended by Wenhold *et al*. (2022)	-	Data for the Index was collected between September and November 2021 by a team of trained assessors managed by ikapadata Effort was made in the training, monitoring and support of fieldworkers to ensure inter-rater reliability **Fieldworker report:** All fieldworkers underwent ELOM assessment training, of which height-for-age domain forms a part. Practical assessments were done and recruits completed the ELOM interrater reliability test	Does not specifically mention children's anthropometric measurements, but does report the following: Fieldworkers are trained at the same time as the pre-tests to ensure consistency. To improve the quality of data collected, certain key indicators are closely monitored during fieldwork, among others the magnitude of anthropometric measurement differences between current waves and previous waves, as well as flags for extreme BMI measures . These checks are usually taken periodically from about 6 weeks into fieldwork (or when there is enough data to estimate meaningful averages). Where interviewers' performance measures lie outside of ±50% of the mean they are investigated, retrained, moved to different teams for closer supervision or removed. In some cases the households are re-interviewed to include hitherto missed respondents.	Nurses were trained to collect biomarker data, including taking height/ length, weight, and waist measurements The biomarker training consisted of lectures, demonstrations of biomarker measurement or testing procedures, exercises aimed at standardisation of height and weight measurements, and practice with children at a health clinic. The logistics officers trained alongside the nurses to ensure that they would be able to support them
Measurement method: standing and recumbent height	-	0 to 23 mths – length (recumbent length) 24 to 60 mths – height (standing height measured)	-	Standing: Growth status is measured as the child's height-for-age, using a stadiometer.	Assumed that child was measured in the recumbent position if < 24 mths, and heigh is measured standing if >24 mths.	Recumbent length measured for children <2yr Standing height for all other children.

	Grow Great Surveys (GGCSS) (2022)	WC Stunting Baseline (WCSB) (2022)	National Food and Nutrition Security Survey (NFNSS)	Thrive by Five (TBF) (2021)	NIDS Wave 5 (2017)	SA-DHS (2016)
Measurement method: decision rule	Two height and weight measurements were taken, and if these differed by more than a small amount (0.5 cm) a third measurement was taken. The final height and weight variables were constructed from the average of the two most similar measurements.	Recumbent length was taken to the nearest 0.1 cm for children under 2 years of age and height to the nearest 0.1 cm for 2–<5-year-olds using a non-flexible MUAC tape. As standing height is approximately 0.7 cm less than recumbent length, which was considered in developing the WHO Growth Standards, length/height measures were adjusted accordingly. If a child under 2 years of age was not willing to lie down, standing height was measured and 0.7 cm added to convert it to length. If a child 2 years of age or older could not stand, recumbent length was measured, and 0,7 cm subtracted to convert it to height. Repeated all anthropometric measures twice and using the average in statistical analyses.	-	-	Fieldworkers instructed to take two height measures and then a third if the first two measures are more than one centimetre apart. Similarly, a third weight measure is required if the first two weight measures are more than one kilogram apart. In practice, the third measures are very seldom taken. For calculating z-scores, the average of the first two measures were used. In instances were these first two measures differ by more than one centimetre in the case of height and one kilogram in the case of weight, a third measurement was used if available.	Not mentioned but only space to record 1 weight in the biomarker questionnaire
Measurement method: Equipment used	-	Infants/children up to 12 months old were weighed using a Scalerite Micro Digital Table Baby scale. Older children were weighed standing up on a Scalerite Micro Glass Digital bathroom scale. Length: Seca 210 Mobile Measuring Mat for Babies Height: Seca 213 Portable Stadiometer	-	Stadiometer for growth status.	-	Seca 878 digital scales, Seca 417 infantometers (for children under age 2) Seca 213 portable stadiometers (for children age 2 and older and for adults) were used for these measurements.
Measurement considerations: parental height	-	Height and weight of the primary caregiver/mother were taken.	-	-	Asked as part of adult questionnaire	Asked as part of biomarker questionnaire
Measurement considerations: gestational age	Yes. Gestational age is recorded from Road to Health booklets and adjusted for when creating height-for-age z-scores.	Yes. Reports health status indicators other than anthropometric measures included in the WCSBS were gestational age, but does not report adjusting for gestational age.	Not mentioned.	No.	No.	Suggests not. Describes that low-birth weight is based on percentage of births with a reported birth weight below 2,5 kg, regardless of gestational age.
Other	No details about questionnaire	Questionnaire for children not based on WHO/ UNICEF guidelines	-	Food intake not measured. Reports that it may slightly over-represent poorer children and under-represent wealthier children	Questionnaire for children not based on WHO/UNICEF guidelines. Focusses on household consumption in general. Neither adult or child questionnaire asks about breastfeeding	Questionnaire for children's' intake based on WHO/UNICEF guidelines Questionnaire only has space for 1 length/height measurement

	Grow Great Surveys (GGCSS) (2022)	WC Stunting Baseline (WCSB) (2022)	National Food and Nutrition Security Survey (NFNSS)	Thrive by Five (TBF) (2021)	NIDS Wave 5 (2017)	SA-DHS (2016)
Implausible value cut-offs	+/-4 SD	Not available	Not available	None – public release includes raw data.	+/- 6 SD	+/- 6 SD
Standard deviation of z-scores[4]	1.38	Not available	Not available	1.36 (raw) 1.06 (after removing biologically implausible values using +/-6 cut-off)[5]	1.46	1.41

Age

NIDS did not take anthropometric measurements of children under 6 months, whereas the Thrive by Five Index only included children 50-59 months. The DHS, NFNSS, WCSBS and GGCSS, on the other hand, included children 0-59 months. Based on this, one would expect NIDS to overestimate stunting somewhat, as children under 6 months tend to have relatively low stunting prevalence, whereas one would expect considerably lower stunting prevalence in the Thrive by Five data than in surveys including the whole 0-5 age range. However, even in the four surveys covering the whole 0-5 age range, the age distributions do not appear to be perfectly uniform. All four appear to have undersampled children under 1 year slightly, and the WCSBS, GGCSS and NFNSS also have a smaller proportion of 4- to 5-year-olds than expected.

Representativeness

Only NIDS, the DHS and the NFNSS were designed to be nationally representative. As Thrive by Five sampled children in early learning programmes who would on average be on the higher end of the SES spectrum, one would expect lower stunting rates in this dataset (Henry and Giese, 2022). All other surveys were household surveys. The DHS included a higher proportion of children in rural areas relative to NIDS Wave 5, the GGCSS and WCSBS. Poverty tends to be higher in rural areas, and urban children were less likely to be stunted in all these surveys (though not necessarily significantly so), so the higher proportion of rural children in the DHS may have contributed to the higher stunting rate observed in that survey. However, more investigation is needed of the extent to which the socioeconomic composition of these surveys is comparable.

As mentioned above, a fairly high proportion of children who were sampled in NIDS and the DHS have missing height data; this may be correlated with SES and create selection bias. This was not reported in the NFNSS report. In the DHS 2016 higher-SES children were indeed more likely to have missing height data: height data was missing for 12 percent of children whose mothers had no education and 41 percent of children whose mothers had more than secondary education (Department of Health, 2019, p. 424). Urban children were more likely to have missing height data than rural children (35 percent versus 16 percent), as were children in Gauteng and the Western Cape, the country's two most affluent provinces: height data was missing for 42 percent of sampled children in Gauteng and 56 percent of those in the Western Cape, while most other provinces ranged between 18 and 26 percent (Department of Health, 2019, p. 424). This could lead to substantial overestimates of stunting prevalence, and make comparisons across provinces problematic.

[4]Standard deviations were calculated after removal of biologically implausible z-scores, because raw z-scores were not always available in the publicly released datasets. A stricter cut-off for biologically implausible values would result in a smaller standard deviation.
[5]Standard deviations are expected to be lower among older children.

This problem is likely not unique to the DHS.

Unfortunately, two of the surveys with the largest sample sizes (GGCSS and Thrive by Five) are not nationally representative of children under 5. Though nationally representative, the DHS 2016 has an unusually low sample size compared to DHS surveys from other countries.

Measurement issues

All surveys used the standard definition of stunting based on the 2006 WHO Child Growth Standards. The surveys followed the recommended practice of measuring standing height in children 2 years and older and recumbent length in children under 2, but they did not all record whether children were actually measured standing or lying down. In these cases no adjustment could be made for this, which may introduce measurement error. Only the GGCSS and WCSBS reported collecting data on gestational age, but only the GGCSS reported adjusting height-for-age z-scores for gestational age. This is likely to result in stunting estimates that are somewhat lower than in other surveys.

It is difficult to assess the extent of measurement error in surveys. However, given the broad scope of NIDS, it appears that less emphasis was placed on selection of fieldworkers with experience in taking anthropometric measurements or on training of fieldworkers to take anthropometric measurements than in some other surveys. It is thus plausible that height may have been measured with greater error in NIDS than in surveys that emphasised anthropometric training. Standard deviations were not reported in all surveys, and when they were they were not always comparable. The NFNSS reports a higher proportion of severely stunted than moderately stunted children, which could be an indicator of measurement error.

All recent South African stunting surveys have strengths and weaknesses, and provide valuable data for examining child health and its correlates. However, while an exhaustive review was beyond the scope of this paper, some of the factors highlighted here may limit the comparability of stunting estimates from these surveys.

5. Discussion and recommendations

5.1 Caution is needed when comparing stunting estimates across and within surveys

First, we need to be very cautious when comparing estimates of stunting prevalence across surveys. This is particularly the case when comparing surveys with different sampling strategies or targeting different age groups or regions, but even nationally representative surveys ostensibly targeting the same population (even consecutive waves of a repeated survey) can produce very different estimates. As discussed above, differences in measurement procedures and the extent of measurement error and issues with sample representativeness may result in large variations in the estimated prevalence of stunting that do not necessarily reflect actual variations in the population. Simple comparisons across surveys (or time) or across regions may lead to erroneous conclusions. These issues may be exacerbated on a sub-national level (districts, provinces), where small sub-sample sizes may also come into play.

Compare height-for-age z-score distributions and means across surveys in addition to stunting rates

Though it will not eliminate all the problems, limiting the comparability of estimates of linear growth faltering across surveys, it is worth examining mean height-for-age z-scores and their distributions in addition to stunting rates. The mean and particularly the median are less sensitive to extreme values driven by measurement error. The World Health Organization and United Nations Children's Fund (2019) recommend reporting mean z-scores and their standard deviations along with stunting prevalence estimates. There is also an argument for examining the whole distribution of height-for-age z-scores based on one of the criticisms of stunting outlined in Section 2 – that linear growth faltering is a whole-population phenomenon, not limited to children falling below the -2 height-for-age z-score cut-off.

5.2 We need more regular, high-quality, repeated and comparable national surveys

While existing surveys are a valuable source of data on stunting prevalence and its predictors in South Africa, there is a need for more frequent, high-quality data on stunting that are comparable across time to accurately monitor trends. South Africa should consider establishing a national nutrition surveillance system to generate data that would be better suited to tracking trends in stunting, as well as other nutritional indicators such as obesity. To better track progress in reducing stunting, we need nationally representative surveys using comparable methodology and sampling strategies, repeated regularly and ideally with larger sample sizes. As shown in Figure 1, the available estimates have wide confidence intervals (i.e. we do not have a high degree of certainty in the estimates). Larger sample sizes would help to make the confidence intervals smaller, increasing the precision of the estimates. Sentinel surveillance sites could also provide useful data for tracking trends in stunting as part of a national nutrition surveillance system (Tuffrey, 2016). Though they do not provide representative estimates, they may be cheaper than large nationally representative surveys, and data could thus be collected more regularly.

There is a need for more unified data collection efforts to enhance the comparability of surveys to track trends in stunting. Efforts to date have been fairly fragmented, with a range of organisations collecting child growth data using different methods and survey designs. Greater collaboration across organisations and sectors concerned with growth monitoring would allow for the sharing of approaches to measurement and rigorous discussion of challenges, to move towards a more consistent and cohesive set of standards for growth monitoring.

The example of Peru shows what is possible with more frequent comparable nationally representative surveys. The Peru Demographic and Family Health Survey (ENDES) has been conducted nearly every year since 2007/8, with a sample of nearly 8000 children or more with complete height-for-age data every year (UNICEF-WHO-World Bank Joint Child Malnutrition Estimates (JME) Working Group, 2023, 2021). The regular collection of this data allows us to say with a high degree of certainty that Peru has reduced stunting dramatically in the last two decades, from an estimated 28 percent in 2005 to 10 percent in 2022 (UNICEF-WHO-World Bank Joint Child Malnutrition Estimates (JME) Working Group, 2023). This data has also allowed researchers to investigate the drivers of Peru's reduction in stunting (Bhutta *et al*., 2020). Unfortunately, with the available data, it is difficult to say with any certainty what has happened to stunting prevalence in South Africa, and if it has declined, what factors may have contributed to this.

Improve data quality

As outlined above, the comparability of surveys is affected by differences in data quality across surveys. Attention should be given to appropriate training of fieldworkers to improve data quality, and to improved monitoring of data quality during fieldwork. The WHO-UNICEF recommendations for collecting and analysing anthropometric data in children under 5 (World Health Organization and United Nations Children's Fund, 2019) are a useful resource for designers of future surveys and researchers analysing data. Using the same cut-offs for biologically implausible values (preferably using the WHO recommended cut-off of +/- 6 SD) would also improve comparability of estimates. It is also worth exploring alternative ways of measuring children that may be more accurate.

Collect data on gestational age where possible

While it is debatable whether height-for-age z-scores should automatically be adjusted for gestational age, and gestational age may introduce another source of measurement error,

It would be useful for all surveys to record gestational age so that the sensitivity of stunting estimates to adjustment for gestational age can be checked. At the least, we need to be aware that stunting estimates that do adjust for gestational age are likely to be somewhat lower.

5.3 The proportion of children affected by stunting may be better reflected by stunting rates around 2 years of age

Given that linear growth faltering follows a clear age pattern, with stunting peaking around 2 years of age as shown in Section 3.1.1, it is worth considering whether estimating stunting prevalence among children across the 0-5 age range (as is currently common practice) gives us the most accurate picture of the proportion of children affected by stunting. As outlined in Box 1, while children appear to 'catch up' in terms of height to some extent after 2 years of age, the evidence suggests that children who were stunted at 2 years but are no longer stunted later in childhood still experience lingering effects on cognitive function. Children younger than 2 years who are not stunted may still become stunted; children older than 2 years who are not considered stunted may have been stunted previously and may still bear permanent consequences of the poor growth environment that led to them being stunted at 2 years. This implies that stunting rates in children 0-5 underestimate the proportion of children affected by stunting at some point in their lives. The proportion of children stunted at or around 2 years of age would give a fuller picture of the extent of the problem. Using the DHS 2016, Karlsson *et al*. (2023) found an overall under-5 prevalence of stunting of 26 percent, but a peak prevalence at 22 months of 36 percent.

Unfortunately, the sample sizes of children around 2 years of age in existing surveys are often too small to give precise estimates of stunting rates at 2. Furthermore, it may be more difficult to sample children around 2 years of age than children under 5. A possibility worth considering would be to collect data on children at 2 years from clinic visits recorded on their Road to Health cards and use this to monitor stunting at 2. However, this approach comes with its own challenges. Though all children are supposed to be measured during routine clinic visits, Blaauw *et al*. (2017) found that only 16 percent of a sample of children aged 0-36 months attending primary healthcare facilities in the Western Cape had their length/height measured. Even when children are measured, there may be concerns about the reliability of data collected at clinics. A study from Canada found a high degree of agreement between routine measurements taken in primary care facilities and those taken by trained research staff (Carsley *et al*., 2019), but an assessment of the reliability of measurements taken by clinic staff would be needed. In the study by Blaauw *et al*. (2017), even when length or height were measured, correct procedures were followed in only 18 percent of cases.

Box 2: Key considerations to improve the measurement of stunting

- We need nationally representative surveys using comparable methodology and sampling strategies, repeated regularly and ideally with larger sample sizes, to accurately monitor trends in stunting.
- South Africa should consider establishing a national nutrition surveillance system to generate data that would be more suited to tracking trends in stunting, as well as other nutritional indicators such as obesity.
- Attention should be given to appropriate training of fieldworkers to improve data quality, and to improved monitoring of data quality during fieldwork.

- Surveys should ideally report the number and percentage of children whose heights or ages were not recorded and the distribution of these children across key variables to allow for an assessment of the risk of bias.
- Surveys should ideally include a set of survey weights adjusting for non-response in anthropometric modules.
- Comparisons of stunting rates and height-for-age over time require using the same growth reference.
- Surveys should clearly record whether height or length was measured so that necessary adjustments can be made.
- Surveys should capture both gestational and chronological age.
- Taking more than one height measurement may result in more accurate data.
- Consistent approaches to handling biologically implausible values would support improved comparability of stunting rates across studies. The cut-off recommended by the WHO and UNICEF is a height-for-age z-score less than -6 or greater than 6. Data releases should ideally include children with z-scores above or below the cut-off to allow comparisons across surveys using different cut-offs. Improved data validation controls during survey data collection could enable verification of biologically implausible z-scores.
- In the analysis and reporting of data, it is worth examining and reporting mean and median height-for-age z-scores and their distributions in addition to stunting rates.
- Future research could explore the possibility and practicality of stratifying child anthropometric surveys by age or creating age-specific weights to render surveys with differing age distributions more comparable.

5.4 Areas for further research

Future research could examine the sensitivity of estimates to differing age distributions. In theory, differing age distributions may render comparisons across surveys problematic, but it is unclear to what extent this makes a difference. Furthermore, future research could explore the possibility and practicality of stratifying child anthropometric surveys by age, or creating age-specific weights to render surveys with differing age distributions more comparable. Some studies already include some form of age stratification, such as the WCSBS, which stratified by three age groups: under 6 months, 6-23 months, and 24 to 59 months. However, these age categories are probably still too broad to ensure a comparable age distribution across surveys; height-for-age z-scores differ dramatically across the 6-24 months age range.

6. Conclusion

The available South African surveys with anthropometric data on children under 5 have produced stunting estimates that vary considerably. This paper has outlined some of the factors that may limit the comparability of stunting estimates and contribute to their variability. However, despite the variation across surveys, this much is clear: far too many South African children still suffer from linear growth faltering. Measurement error likely inflates stunting estimates, but all available estimates of stunting prevalence are much higher than the 2.3 percent one would expect in a population facing healthy conditions for growth. Available estimates suggest that around 1 in every 4 or 5 South African children is stunted; in a healthy population this would be around 1 in 40. In all available surveys the distributions of height-for-age z-scores are shifted well to the left of the distribution found in the optimal growth environments of the WHO Multicentre Growth Reference Study. Furthermore, stunting underestimates the number of children affected by linear growth faltering more broadly. As argued in this paper, estimates of stunting in children under 5 likely underestimate the proportion of children affected by stunting at some point in their lives and suffering from lasting associated consequences.

It is clear that more needs to be done to reduce stunting among South African children. Much is already known about what works to reduce stunting (Burger *et al*., 2022). But in tracking our progress towards that goal, we are largely flying blind. We need more data, but more attention needs to be given to ensuring this data is of high quality and comparable over time, and more investment is needed in investigating alternative ways to measure and monitor linear growth faltering.

7. References

Benjamin-Chung, J., Mertens, A., Colford, J.M., Hubbard, A.E., van der Laan, M.J., Coyle, J., Sofrygin, O., Cai, W., Nguyen, A., Pokpongkiat, N.N., Djajadi, S., Seth, A., Jilek, W., Jung, E., Chung, E.O., Rosete, S., Hejazi, N., Malenica, I., Li, H., Hafen, R., Subramoney, V., Häggström, J., Norman, T., Brown, K.H., Christian, P., Arnold, B.F., 2023. Early-childhood linear growth faltering in low- and middle-income countries. Nature 621, 550–557. https://doi.org/10.1038/s41586-023-06418-5

Bhutta, Z.A., Akseer, N., Keats, E.C., Vaivada, T., Baker, S., Horton, S.E., Katz, J., Menon, P., Piwoz, E., Shekar, M., Victora, C., Black, R., 2020. How countries can reduce child stunting at scale: lessons from exemplar countries. The American Journal of Clinical Nutrition 112, 894S-904S. https://doi.org/10.1093/ajcn/nqaa153

Bhutta, Z.A., Das, J.K., Rizvi, A., Gaffey, M.F., Walker, N., Horton, S., Webb, P., Lartey, A., Black, R.E., Lancet Nutrition Interventions Review Group, the Maternal and Child Nutrition Study Group, 2013. Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? Lancet 382, 452–477. https://doi.org/10.1016/S0140-6736(13)60996-4

Blaauw, R., Daniels, L., Plessis, L.M. du, Koen, N., Koornhof, H.E., Marais, M.L., Niekerk, E. van, Visser, J., 2017. Assessing the utilisation of a child health monitoring tool. South African Journal of Child Health 11, 174–179.

Black, R.E., Victora, C.G., Walker, S.P., Bhutta, Z.A., Christian, P., De Onis, M., Ezzati, M., Grantham-McGregor, S., Katz, J., Martorell, R., Uauy, R., 2013. Maternal and child undernutrition and overweight in low-income and middle-income countries. The Lancet 382, 427–451. https://doi.org/10.1016/S0140-6736(13)60937-X

Burger, R., Du Plessis, L., Gangaidzo, T., Wills, G., 2022. Priority areas for reducing stunting in South Africa: Examining the implications of recent international evidence (Working Paper No. No. ECD WP 004/2022), Ilifa Labantwana & Resep ECD Working Paper Series.

Carsley, S., Parkin, P.C., Tu, K., Pullenayegum, E., Persaud, N., Maguire, J.L., Birken, C.S., TARGet Kids! Collaboration,

2019. Reliability of routinely collected anthropometric measurements in primary care. BMC Med Res Methodol 19, 84. https://doi.org/10.1186/s12874-019-0726-8

Casale, D., 2020. Recovery from stunting in early childhood and subsequent schooling outcomes: Evidence from NIDS Waves 1–5. Development Southern Africa 37, 483–500. https://doi.org/10.1080/0376835X.2020.1715790

Casale, D., Desmond, C., Richter, L.M., 2020. Catch-up growth in height and cognitive function: Why definitions matter. Economics & Human Biology 37, 100853. https://doi.org/10.1016/j.ehb.2020.100853

Crookston, B.T., Penny, M.E., Alder, S.C., Dickerson, T.T., Merrill, R.M., Stanford, J.B., Porucznik, C.A., Dearden, K.A., 2010. Children Who Recover from Early Stunting and Children Who Are Not Stunted Demonstrate Similar Levels of Cognition. The Journal of Nutrition 140, 1996–2001. https://doi.org/10.3945/jn.109.118927

Crookston, B.T., Schott, W., Cueto, S., Dearden, K.A., Engle, P., Georgiadis, A., Lundeen, E.A., Penny, M.E., Stein, A.D., Behrman, J.R., 2013. Postinfancy growth, schooling, and cognitive achievement: Young Lives. The American Journal of Clinical Nutrition 98, 1555–1563. https://doi.org/10.3945/ajcn.113.067561

de Onis, M., Branca, F., 2016. Childhood stunting: a global perspective. Matern Child Nutr 12, 12–26. https://doi.org/10.1111/mcn.12231

Department of Health, 2019. South Africa Demographic and Health Survey 2016. NDoH, Stats SA, SAMRC, and ICF, Pretoria, South Africa, and Rockville, Maryland, USA.

Desmond, C., Casale, D., 2017. Catch-up growth in stunted children: Definitions and predictors. PLOS ONE 12, e0189135. https://doi.org/10.1371/journal.pone.0189135

Dewey, K.G., Begum, K., 2011. Long-term consequences of stunting in early life. Maternal & Child Nutrition 7, 5–18. https://doi.org/10.1111/j.1740-8709.2011.00349.x

Garza, C., Borghi, E., Onyango, A.W., de Onis, M., WHO Multicentre Growth Reference Study Group, 2013. Parental height and child growth from birth to 2 years in the WHO Multicentre Growth Reference Study. Matern Child Nutr 9 Suppl 2, 58–68. https://doi.org/10.1111/mcn.12085

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

Henry, J., Giese, S., 2022. An exploration of updated Thrive by Five Index stunting rates in children aged 50-59 months enrolled in Early Learning Programmes in South Africa.

Karlsson, O., Kim, R., Moloney, G.M., Hasman, A., Subramanian, S.V., 2023. Patterns in child stunting by age: A cross-sectional study of 94 low- and middle-income countries. Maternal & Child Nutrition n/a, e13537. https://doi.org/10.1111/mcn.13537

Leroy, J.L., Frongillo, E.A., 2019. Perspective: What Does Stunting Really Mean? A Critical Review of the Evidence. Advances in Nutrition 10, 196–204. https://doi.org/10.1093/advances/nmy101

Leroy, J.L., Frongillo, E.A., Dewan, P., Black, M.M., Waterland, R.A., 2020. Can Children Catch up from the Consequences of Undernourishment? Evidence from Child Linear Growth, Developmental Epigenetics, and Brain and Neurocognitive Development. Advances in Nutrition 11, 1032–1041. https://doi.org/10.1093/advances/nmaa020

Leroy, J.L., Ruel, M., Habicht, J.-P., Frongillo, E.A., 2015. Using height-for-age differences (HAD) instead of height-for-age z-scores (HAZ) for the meaningful measurement of population-level catch-up in linear growth in children less than 5 years of age. BMC Pediatr 15, 145. https://doi.org/10.1186/s12887-015-0458-9

Leroy, J.L., Ruel, M., Habicht, J.-P., Frongillo, E.A., 2014. Linear Growth Deficit Continues to Accumulate beyond the First 1000 Days in Low- and Middle-Income Countries: Global Evidence from 51 National Surveys. The Journal of Nutrition 144, 1460–1466. https://doi.org/10.3945/jn.114.191981

Lundeen, E.A., Stein, A.D., Adair, L.S., Behrman, J.R., Bhargava, S.K., Dearden, K.A., Gigante, D., Norris, S.A., Richter, L.M., Fall, C.H., Martorell, R., Sachdev, H.S., Victora, C.G., 2014. Height-for-age z scores increase despite increasing height deficits among children in 5 developing countries. The American Journal of Clinical Nutrition 100, 821–825. https://doi.org/10.3945/ajcn.114.084368

Martorell, R., Zongrone, A., 2012. Intergenerational Influences on Child Growth and Undernutrition. Paediatric and Perinatal Epidemiology 26, 302–314. https://doi.org/10.1111/j.1365-3016.2012.01298.x

McEvoy, B.P., Visscher, P.M., 2009. Genetics of human height. Econ Hum Biol 7, 294–306. https://doi.org/10.1016/j.ehb.2009.09.005

Moses, E., 2021. Enrolment in early childhood care and education programmes in South Africa: challenges and opportunities (Working Paper No. No. ECD WP 002/2021), Ilifa Labantwana & Resep ECD Working Paper Series.

Perumal, N., Bassani, D.G., Roth, D.E., 2018. Use and Misuse of Stunting as a Measure of Child Health. The Journal of Nutrition 148, 311–315. https://doi.org/10.1093/jn/nxx064

Roth, D.E., Krishna, A., Leung, M., Shi, J., Bassani, D.G., Barros, A.J.D., 2017. Early childhood linear growth faltering in low-income and middle-income countries as a whole-population condition: analysis of 179 Demographic and Health Surveys from 64 countries (1993-2015). Lancet Glob Health 5, e1249–e1257. https://doi.org/10.1016/S2214-109X(17)30418-7

Said-Mohamed, R., Micklesfield, L.K., Pettifor, J.M., Norris, S.A., 2015. Has the prevalence of stunting in South African children changed in 40 years? A systematic review. BMC Public Health 15, 534. https://doi.org/10.1186/s12889-015-1844-9

Sandler, A., 2021. Against Standard Deviation as a Quality Control Maxim in Anthropometry. Econ Journal Watch 18.

Tuffrey, V., 2016. Nutrition surveillance systems: their use and value. Save the Children and Transform Nutrition, London.

UNICEF-WHO-World Bank Joint Child Malnutrition Estimates (JME) Working Group, 2023. UNICEF/WHO/World Bank Joint Child Malnutrition Estimates Database.

UNICEF-WHO-World Bank Joint Child Malnutrition Estimates (JME) Working Group, 2021. Technical Notes from the background document for country consultations on the 2021 edition of the UNICEF-WHO-World Bank Joint Malnutrition Estimates. UNICEF, WHO and World Bank.

Victora, C.G., Adair, L., Fall, C., Hallal, P.C., Martorell, R., Richter, L., Sachdev, H.S., 2008. Maternal and child undernutrition: consequences for adult health and human capital. The Lancet 371, 340–357. https://doi.org/10.1016/S0140-6736(07)61692-4

Victora, C.G., De Onis, M., Hallal, P.C., Blössner, M., Shrimpton, R., 2010. Worldwide Timing of Growth Faltering: Revisiting Implications for Interventions. Pediatrics 125, e473–e480. https://doi.org/10.1542/peds.2009-1519

WHO Multicentre Growth Reference Study Group, 2006. WHO Child Growth Standards based on length/height, weight and age. Acta Paediatrica 95, 76–85. https://doi.org/10.1111/j.1651-2227.2006.tb02378.x

World Health Organization, 2014. Global nutrition targets 2025: stunting policy brief (No. WHO/NMH/NHD/14.3). World Health Organization, Geneva.

World Health Organization, United Nations Children's Fund, 2019. Recommendations for data collection, analysis and reporting on anthropometric indicators in children under 5 years old. World Health Organization, Geneva.

ILIFA-RESEP ECD WORKING PAPER SERIES:

Estimating the impact of five early childhood development programmes against a counterfactual:	VIEW HERE
Enrolment in early childhood care and education programmes in South Africa- challenges and opportunities:	VIEW HERE
Supply-side and demand-side approaches to financing early childhood care and education in South Africa:	VIEW HERE
Priority areas for reducing stunting in South Africa: Examining the implications of recent international evidence:	VIEW HERE
Compliance, cost and user fees in the Early Childhood Care and Education Sector in South Africa:	VIEW HERE
The cognitive gains case for ECD: missing pieces of the puzzle	VIEW HERE
Cost, compliance and user fees in the Early Childhood Care and Education Sector in South Africa:	VIEW HERE
Promoting early childhood development through multi-dimensional service delivery in South Africa: A municipal view	VIEW HERE
Post-COVID Early Childhood Care and Education enrolment gaps in South Africa: Exploring Demand and Supply Influences	VIEW HERE

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT | forward together sonke siya phambili saam vorentoe

RESEP Research on Socio-Economic Policy

ilifa labantwana