# Meta analysis: age-related measurement error can lead to systematic overestimates of childhood stunting rates

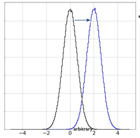
Joseph Grange, PhD for the USAID Bureau for Humanitarian Assistance M&E Team

AAAS Policy Fellow

This study is the sole responsibility of the Joseph Grange and does not necessarily reflect the views of the United States Agency for International Development or the United States Government.

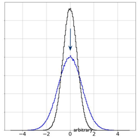
## Measurement error and SDs

- Directional measurement errors:
  - every measurement changes by the same amount - the distribution *shifts*
  - changes the **position (mean)** but not the **width** (SD)




- Examples in children's anthropometry:
  - children wear bulkier clothes in colder climates
  - measuring device systematically flawed
  - all birth records under-estimated

- Non-directional (or, *random*) measurement errors:
  - any given measurement may be incorrect, but there is no *net* effect
  - increases the **width** (SD) of the distribution but not the **position (mean)**




- Examples in children's anthropometry:
  - squirming children
  - transcription errors
  - all birth records equally under- and over-estimated

**Any distribution's standard deviation (SD) value is strongly influenced by measurement data quality.**
(of course, genuinely heterogeneous distributions can also lead to large SDs!)

Box A

## Anthropometry and stunting

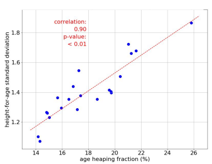
- Accurate nutritional monitoring for the world's youngest citizens is a critical tool in evaluating past, present, and future health trajectories in vulnerable communities
  - nutrition-based indicators must be able to be collected **accurately** and **efficiently** in the context of large-scale surveys
- **Height-for-age** measurements are among the highest-level childhood nutrition indicators supported by **Feed The Future** and plays a key role in the UN's **Sustainable Development Goals**.
- Typically, height-for-age results are distilled into a single quantity: the prevalence of **stunting**. This indicator is defined by the proportion of observations falling 2 or more SDs below a healthy reference population (see **Box C**).

- This work presents simulations and a meta-analysis that demonstrates how a particular **measurement error** type (see **Box F**) has led to **significant historical biases for this stunting indicator**.

Box B

## Datasets included and primary findings

- Included in this meta-analysis are the height-for-age distributions measured in **21 population-based surveys related to USAID-supported resilience activities** collected within the past 10 years. Inclusion criteria was defined prior to any analysis.
- In these datasets, a **strikingly strong correlation** (ρ=0.90, p-value ≪ 0.01) was found between the 'age heaping fraction' variable (defined in **Box F**) and the output height-for-age z score SD. Important implications include:
  - **age heaping is a driving factor in the height-for-age data quality in these survey.**
  - the SD values in this dataset are inflated due to measurement error. As demonstrated in **Box C**, this can lead to **overestimating the true stunting prevalence.**
  - as simulated results show in **Box E**, **this bias can exceed 10 percentage points for some surveys in this set! Improved stunting results could be attained by correcting the historical results by these amounts.**



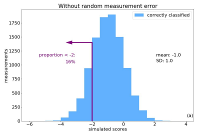

- Study limitations:
  - the corrections simulated in **Box E** do not address of directional error, which can also influence prevalence rates.
  - the corrections simulated in **Box E** assume the SD value should be 1.0 in the absence of non-directional measurement error (a range of values between 0.8 and 1.2 is reasonable).

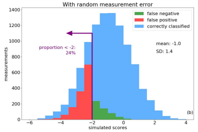
Box D

## SDs and threshold-based prevalence indicators

- As discussed in **Box A**, non-directional measurement error leads to an increase in distributional SD values.
- For the many indicators that track distributional mean values, SD increases do not influence the result.
- **However**, the definition of the **stunting** indicator involves the proportion of observations falling two or more SDs below the healthy reference population. **This box demonstrates how an inflated SD can lead to a bias in the prevalence of stunting**.

- The figures on the right show the results of a simple simulation.
- In (a), a distribution with no measurement error was simulated. All observations are therefore **correctly classified**
  - note the proportion falling below -2 is 16%
- To create the distributions in (b), non-directional measurement error was added to each observation in figure (a). **Note the SD value has increased.** These measurement errors have introduced the mis-classification categories of **false positives** and **false negatives**.
- Note the **false positives** clearly outnumber the **false negatives**, leading to **an overall shift in the overall prevalence rate** (now at 24%)



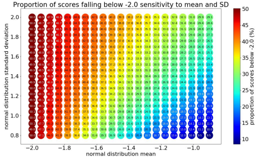

- This simple simulation shows how a realistic increase in SD **due to non-directional measurement error can lead to a significant bias in overall stunting rates** (in this case, 8 percentage points!).
- **Box E** explores the relationship between bias, SD, and mean value in more detail

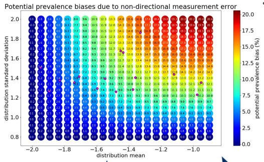
Box C

## Prevalence rate dependence on distribution mean and SD

- Assuming normality, apart from a distribution's SD (see **Box C**), the prevalence for stunting also depends on the distributions mean value (and nothing else!).
  - this is true for **any** indicator defined by the proportion of observations above or below a certain threshold

- To explore this three-dimensional relationship, high-statistics normal distributions were simulated with a variety of mean and SD values. For each of these distributions, the **prevalence** of observations below -2 (*à la stunting*) were calculated and presented to the right as a function of **mean** and **SD** value.




- **So what can we do with this information?** Fortunately, there is guidance in the literature: the vast majority of anthropometric z-score distributions known to have rigorous data quality controls have SD values between 0.8 - 1.2. Therefore, **anthropometric z-score distributions with SD values significantly above 1.2 are likely significantly influenced by non-directional measurement error.**

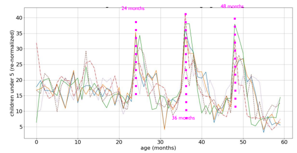
- So, assuming the SD *should have* been close to 1.0 in the absence of random measurement error, we can calculate the **prevalence bias** caused by these data quality issues (graph above).
- Overlaid on this same graph (•) is the height-for-age distribution values for 21 USAID-supported survey height-for-age distributions (see **Box D**). The bias for many are within a few-%, **while the bias for some exceed 10%!**

Box E

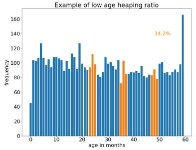
## Clustering near integer ages

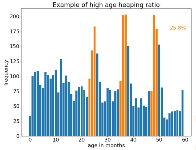
- **Height-for-age data quality review:** as stunting prevalence relies only on measurements of children's height and age, a DQ review of the input distributions showed the following:
  - **no** significant **digit preference** found in the **height distributions**
  - height-for-age z score distributions **sufficiently normal**
  - **however**, strong and consistent peaks in the **age distribution** at **integer-year ages**:

Six age distributions with strong age heaping




- **These features are likely the result of age measurement error.** As described in **Box A**, non-directional measurement error can result in an inflated SD value. **Boxes C** and **E** demonstrate that an inflated SD can result in significant prevalence biases.
- To explore whether these age features are meaningfully related to height-for-age z score SD values, an 'age heaping ratio' quantity was created, defined by the frequency found in the nine months closest to 2, 3, and 4 years (24±1, 36±1, and 48±1 months) divided by the total
  - note that, for a completely age homogeneous distribution, this ratio would be 9/60=15%
  - see **Box D** for results!




Box F

## Conclusions and recommendations

- The simulations and meta analysis presented here have demonstrated that, **due to age-related measurement errors** (see **Boxes F** and **D**), the calculated stunting prevalence rates from the surveys considered are likely **significantly overestimated** (see **Box E**). This challenge, at some level, must be present in other datasets as well.
- The areas of these resilience activities are, by design, challenged. Imperfections in local administrative systems such as birth records should be expected. To recognize these challenges in the future, **a thorough report of the strength of local birth record knowledge and practices should accompany any survey report that relies on their accuracy to calculate high-level outcomes such as stunting rates.**

- Other potential improvement prospects include:
  - include the age distribution and SD value in routine DQ assessments and publish this information in the survey report
  - when stunting results are clearly influenced by measurement error, consider improving the point-estimate via a method such as in **Box E**

Box G

10/21/2020