On assessment of reliability and validity of child anthropometric measurements in India.

Ritankar Chakraborty[1*], Udaya S Mishra[2]

[1] PhD Scholar, International Institute for Population Sciences, Mumbai, India. Email: chakrabortyritankar@gmail.com

[2] Professor, Department of Biostatistics and Epidemiology, International Institute for Population Sciences, Mumbai, India

Introduction:

Despite the rapid economic improvements in India, malnutrition and its associated mortality among children still continues to be widespread (Aurino E, 2016; Coffey et al., 2013a). Even though a large number of intervention programmes are operational in India, the burden of malnutrition still remains a major concern for public health professionals and policy makers across the country. The flagship Integrated Child Development Service (ICDS) programme was launched in India in 1975 to combat the problem of child malnutrition. The ICDS programme focuses on children under 5 years of age, along with pregnant and lactating mothers and provides services such as anthropometric measurements, supplementary nutrition, immunization, health check-ups, referrals, health and nutrition education, and pre-school education to the beneficiaries (Baghel et al., 2018). These services are provided at “Anganwadi” centers (courtyard shelters) through community health workers known as Anganwadi workers (AWWs).

Valid and reliable measurements of anthropometry are indispensable in the context of the ICDS programme, as they play a crucial role in ensuring the effectiveness, credibility, and ethical integrity of these essential services. Among all the services provided by the Anganwadi centers, an important service that often goes unnoticed is the anthropometric measurements of child. These measurements serve towards assessment of growth in early years of life and the nutritional make-up of children. Height and weight measurements are obtained every month to maintain the growth trajectory for each and every child. However, these measurements suffer from multiple errors owing to defective tools of measurement on one hand and inadequate skills of using the measurement device on the other. The ideal expectation on such measurements made at the Anganwadi centers relates to the accuracy and consistency of these anthropometric measurements, as they describe the exact status of the beneficiaries’ early life growth, nutritional make-up, and deficiencies if any for rehabilitation. Errors in these measurements therefore can have serious implications, and can question the overall impact of these services.

This is an attempt at understanding the validity and reliability of anthropometric measurements from Anganwadi centers. The data on anthropometric indicators often are compromised in quality and inherent errors in them leads to their under-utilization in gauging programme performance. The reliability and validity of the data obtained from Anganwadis remains a common concern that remains unattended. This paper therefore attempts to assess the validity and reliability of Anganwadi measurements and tries to find the potential errors and strategies for improved estimation of anthropometric failures.

Data Source

Data on height and weight of children were collected from 420 children from 42 ICDS centers across a particular district in India. A two-stage sampling procedure was adopted to select the 420 children. At the first stage, we selected 42 ICDS centers at random from the sampling frame using Simple Random Sampling (SRS). At the second stage, we selected 10 children from each center using systematic sampling to ensure children from both sexes and of all ages were included in the sample. There were two sets of data. The first set was collected from the Anganwadi workers, and the other set was measured by a trained personnel using standard protocols. Height was measured using a stadiometer with measurements upto 0.1 cm. Weight was measured using a digital weighing machine upto 0.01 kg. Infants were measured using an infantometer.

Methodology

Validity

Validity is closely linked to reliability, however whilst reliability relates to the consistency of a method, validity relates to the accuracy. Validity refers to the extent to which the research instruments really measure what they are supposed to measure. A valid measurement is generally reliable, i.e., if a test produces accurate results, they should be reproducible.

To measure the validity of the anthropometric measurements, we compare the two sets of data, one obtained from the Anganwadi workers, and the other collected separately by a trained personnel. To measure the validity of anthropometric data, it is important to measure differences in absolute measurements. To measure for the normality of the data, we used the **Shapiro-Wilk test** (Shapiro S.S; Wilk M.B 1965). Since the data was not normally distributed, we applied non parametric tests such as **Wilcoxon matched-pairs signed-rank test** to check for the validity of the anthropometric measurements.

Reliability

Reliability is the degree to which a method provides estimates that are stable or constant, as opposed to erratic or variable. It describes the extent to which a method is able to yield reproducible data under the various conditions or contexts for which it has been defined.

To assess the reliability of Anganwadi measurements, we used the **Bland and Altman method**. Bland and Altman (1986) developed an approach to assess agreement in method comparison studies. A Bland-Altman plot has differences, percentage differences, or ratios on the y-axis and a mean of the data pairs on the x-axis, with 95% limits of agreement indicating the central 95% range of differences, percentage differences, or ratios. According to Bland and Altman (1999), 95% limits of agreement (LOA) provide an interval within which 95% of differences between measurements are expected to lie. If these limits are not too large, then the methods can be considered interchangeable. If the points do not lie in a narrow band, it suggests that the variability of the differences are large and indicates the presence of bias between the two methods. In addition to the Limits of agreement, we also used 95% Confidence Intervals for LOA, along with Prediction intervals and Tolerance intervals to give us a better understanding of the reliability of measurements.

Hypothesis

For testing the validity of the two sets of data using Wilcoxon matched-pairs signed-rank test, we used the following hypothesis:

H0: Both sets of data have the same distribution vs H1: The distributions are different.

Results

Table 1 gives us the Wilcoxon matched-pairs sign rank test for weight measurements, where we test that the weight measurements obtained from Anganwadi centers and those obtained from own measurements have the same distribution. Out of the total 420 observations, there are 243 observations where the sign is positive, suggesting that the measurements obtained from AWC are higher than those obtained from own measurements. There are 167 observations where the measurements obtained from AWC are less than those obtained from own measurements, and 10 observations where both the sets of observations are equal. Since the sample size is more than 20, the test uses a normal approximation. The p-value obtained through the normal approximation is similar to the exact p-value, which is less than the significance level of 0.05. So, the test results suggests us to reject the null hypothesis and conclude that the two sets of data have different distributions.

Table 1: Wilcoxon matched-pair Sign Rank test for weight measurements.

Sign	No. of observations	Sum of ranks	Expected
Positive	243	55912.5	44177.5
Negative	167	32442.5	44177.5
Zero	10	55	55
All	420	88410	88410
Unadjusted variance	6196067.5		
Adjustment for ties	-1364.38		

Adjustment for zeros	-96.25		
Adjusted variance	6194606.88		
z=4.715			
Prob > |z| = 0.000			
Exact prob = 0.000			

Table 2 gives us the sign rank test for height measurements. There are 203 comparisons where measurements from AWC is greater than those obtained from own measurements, 209 comparisons where measurements from AWC is less than those obtained from own measurements, and 8 where there were no differences. The p-value computed using the normal approximation is 0.8312, which is similar to the exact p-value 0.8315. Since the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis and conclude that the two sets of height measurements have a similar distribution.

Table 2: Wilcoxon matched-pairs Sign Rank test for height measurements.

sign	No. of observations	Sum of ranks	Expected
Positive	203	44717.5	44187
Negative	209	43656.5	44187
Zero	8	36	36
All	420	88410	88410
Unadjusted variance	6196067.5		
Adjustment for ties	-489.5		
Adjustment for zeros	-51		
Adjusted variance	6195527		
z=0.213			
Prob > |z| = 0.8312			
Exact prob = 0.8315			

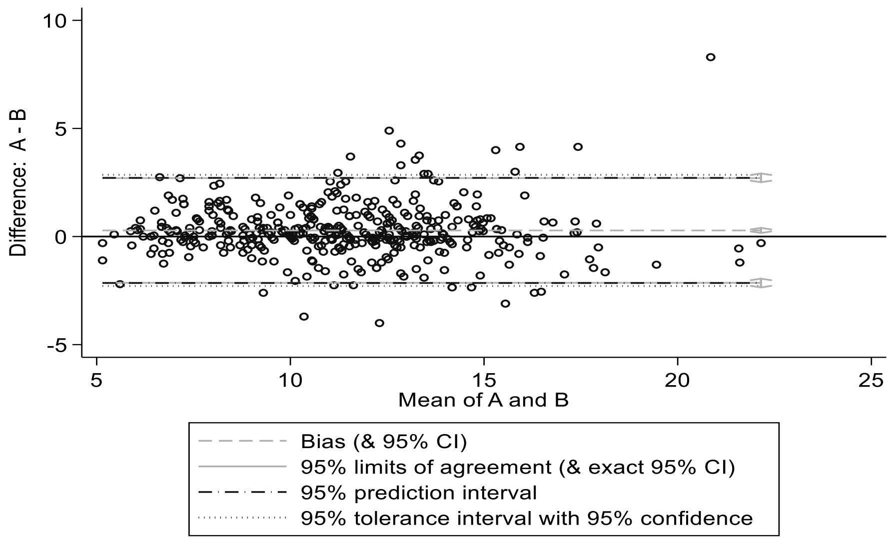

Fig 1: Bland-Altman plot for weight measurement with 95% Prediction Interval and 95% Tolerance interval.

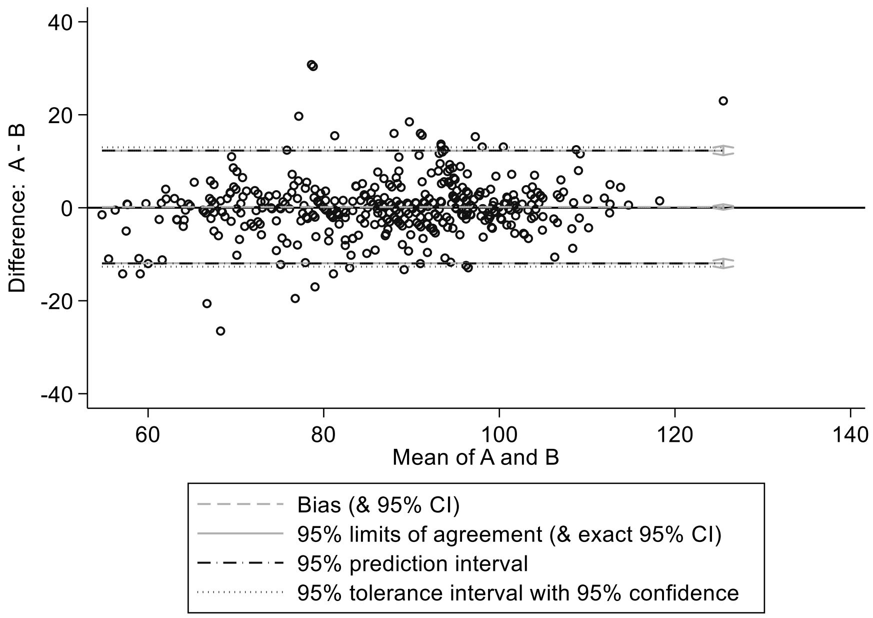

Fig 2: Bland-Altman plot for height measurement with 95% Prediction Interval and 95% Tolerance interval.

Figures 1 and 2 give the Bland-Altman plots for measuring agreement between weight and height of children obtained from two measurements respectively. The difference of the two methods is plotted against the mean of the two methods. As can be seen from the plots, most of the points lie within a narrow band. For plot 1, the mean of the differences of the two methods is 0.28, with a standard deviation of 1.24. The 95% Limits of Agreement ranges from -2.14 (LLOA) to 2.70 (ULOA). For plot 2, the mean of the differences of the two methods is 0.16, with a standard deviation of 6.17. The 95% Limits of Agreement ranges from -11.93 (LLOA) to 12.25 (ULOA). The exact confidence intervals of the LoA, along with 95% Prediction intervals and 95% Tolerance intervals reveal that there is very less variability in the anthropometric measurements obtained from the two methods.

Conclusion

This paper deals with the quality of anthropometric measurements obtained from the Anganwadi centers. Such a quality assessment is based on the validity and reliability of the Anganwadi measurements. Validity was examined using the Wilcoxon-matched pairs signed rank test, whereas reliability was examined using the Bland and Altman plots. The Wilcoxon-matched pairs signed rank test suggested that the weight measurements obtained from the two sets had different distributions. The test results suggest that the weight measurements obtained from the Anganwadi centers lacked validity. However, the height measurements were comparable and the test results show that the height measurements obtained from the Anganwadi centers were valid. Bland-Altman plots show us that both the weight and height measurements are reliable. Most of the data points lie within a narrow band of the 95% Limits of Agreement. The presence of bias is very less for both weight and height measurements, as can be observed from the Bland-Altman plots.

The weight measurements of the children remain an area of concern. Weights are often not measured properly at the Anganwadi centers, and this is mainly due to the faulty devices available to them. Often, these measuring devices have been in use for a very long time and thus give erroneous results. Human errors are also very common in weight measurements. Anganwadi workers often do not measure weights of all children properly due to lack of time and commitment to other activities. They simply record the weights from the previous measurements. But unlike height, weight changes over time, and may increase or decrease due to several biological and environmental factors. Policies and interventions are eventually formed from these erroneous and manipulated data which raises a serious concern on the effectiveness of these policies.

Ethical consideration

Ethical approval for this study was taken from the Student Research Ethics Committee at the International Institute for Population Sciences, Mumbai, India.