

Reliability of Mid-Upper Arm Circumference Measurements taken by Community Health Nurses

HIBBAH ARABA SAEED^{1,3}, JOSEPH BIRUNDU MOGENDI^{2,5},
ROBERT AKPARIBO³ and PATRICK KOLSTEREN^{2,4}

¹Noguchi Memorial Institute for Medical Research, Ghana.

²Faculty of Biosciences Engineering, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium.

³University of Sheffield, School of Health and Related Research.

⁴Institute of tropical Medicine (ITM), Antwerp, Belgium.

⁵Department of Nutrition and Dietetics, School of Health Sciences,
Mount Kenya University, P.o. Box 342-01000, Thika, Kenya.

<http://dx.doi.org/10.12944/CRNFSJ.3.1.04>

(Received: March 24, 2015; Accepted: April 25, 2015)

ABSTRACT

Mid upper arm circumference-based diagnosis of severe acute malnutrition is seen as very attractive because it is assumed to be simple to use in measuring the arm and also requires a single cut-off without having to compute an index or compare to a reference population. The circumference of the left upper arm measured at the mid-point between the tip of the shoulder and the tip of the elbow is referred to as mid upper arm circumference (MUAC). MUAC was originally used in emergency settings where huge populations had to be screened within a short time. However, in the past few years the development of community therapeutic care of acute malnutrition implied a need for a diagnostic tool. The general objective of the study was to assess the inter- and intra-observer variability in MUAC measurements of under-five children by community health nurses in selected communities of the Northern Region in Ghana. A community-based cohort study was conducted on a convenience sample of 120 community health nurses and nursing students from selected communities in the Northern region of Ghana. Each CHN measured MUAC for four unique children on three consecutive days to assess intra-observer variability. To assess inter-observer variability, 50 independent nursing students took MUAC measures of four children on the same day. Bland Altman plots showed a high degree of agreement of MUAC measures taken repeatedly for three days by the same observers and ICC= 0.993. MUAC measures among observers (inter-observer) and ICC =0.042 showed a considerable level of variability among different observers. This study showed that MUAC is reliable when repeated measures are taken on children by the same observers, but using different observers for the same children showed a high level of variability.

Key words: Mid-upper arm circumference, Reliability,
Inter and intra observer variability, Community health nurses.

INTRODUCTION

Malnutrition remains an important risk factor for childhood mortality especially in developing countries. Severe acute malnutrition (SAM) is a fatal condition which requires urgent nutritional and medical support (Myatt, Khara, & Collins, 2006). It accounts for 35% of childhood deaths and 10-40% of hospital admissions in children under 5 years in sub-Saharan Africa (Black *et al.*, 2008). To enable the

treatment and management of malnutrition, effective identification of cases is very important.

Weight-for-Height (W/H) and Mid Upper Arm Circumference (MUAC) are the two anthropometric indicators most often used to identify SAM cases. They are two proxies of the nutritional status of an individual, but the concept underlying their use as a screening tool for SAM differs. For so many years MUAC has been used as a pre-screening tool

to identify wasted children and W/H is used as a confirmatory test. In the MUAC approach, a single cut-off of <110 mm is applied for defining SAM in children between 6 and 60 months of age, as MUAC is assumed to be relatively stable in this age bracket. The 110 mm cut-off was chosen because studies have demonstrated that the risk of dying sharply increases below this measure (Briend & Zimicki, 1986). Recently a cut-off of 115mm has been recommended to gain sensitivity in detecting severely wasted children

MUAC-based diagnosis of SAM is seen as very attractive because of its assumed simplicity of measuring arm circumference and of applying a single cut-off, without having to compute an index and compare it to a reference population. MUAC was originally used in emergency settings, where huge populations had to be screened within a short time. However, in the past few years, the development of therapeutic care of acute malnutrition at the community level implied a need for a diagnostic tool. MUAC was proposed because its interpretation is assumed to be easier than W/H, because it is cheaper and easy to measure (a simple arm strip). This has been endorsed by experts in WHO, UNICEF, WFP and SCN (statement). The new definition of SAM is W/H Z-score <-3SD or MUAC < 115 mm or bilateral oedema (WHO, 2009).

However, using MUAC because of its assumed simplicity and convenience raises some concerns. A convenient tool must be able measure the arm circumference with a fair level of precision and accuracy (Myatt *et al.*, 2006) Like other anthropometric measurements, MUAC is a non-invasive, and a suitable method for evaluating the nutritional status in population studies, and hence could be a valuable tool for use by community health workers (CHWs) for early detection of acute malnutrition in infants at the community level. However, reliability of MUAC measurements among children is unknown. This situation is worsened by the rising number of self-trained anthropometrists, which may contribute to lower quality of findings (Farkas, 1996).

Furthermore, although the right arm is usually used for MUAC measurements, a published

rationale behind the use of the right arm is difficult to locate (Jamaiyah *et al.*, 2008).

The aim of the study was to show the inter- and intra-observer variations that occur in taking MUAC measurements of children. In this study, MUAC measurements taken by community health nurses in selected communities in the Northern Region of Ghana were analysed for inter- and intra-observer variability. This study focused on MUAC as a single indicator.

MATERIALS AND METHODS

This study was a community-based cohort study in which convenience sampling was used to recruit seventy (70) community health nurses (CHN) and fifty (50) nursing students. The nurses were recruited from the health posts in communities from the northern region. The communities were: Yapei, Yipala, Junayili, Ntereso, Chaangni, Tolon, Jenikura, Choggu, Yendi, Nyaani, Limafoong and Kusugu. Data was collected between July and August, 2011.

Study procedure

At the community health centres, the study was explained to the community health nurses. Those CHN's who showed interests in partaking were selected. The nurses talked to mothers in the community and selected children under five years old who were apparently healthy, with no signs of kwashiorkor, marasmus or any other disease. CHN's who could not be available to take repeated measurements were excluded from the study. Children who were not going to be available for subsequent measurements were excluded from the study. The measurements were taken on children under 5 years within the following subgroups: 1-2 years, 2-3 years, 3-4 years, and 4-5 years for both males and females. Informed consent was obtained from observers and mothers who were willing to take part after the study procedure was explained to them.

MUAC measurement

Non-stretch MUAC tapes, known as Shaker's tape, were used to measure MUAC. The child's right arm and left arm were flexed to 90 degrees at the elbow. The midpoint of the arm

(between the lateral acromion and distal olecranon) was identified and marked. The arm was then relaxed and the measurement taken and recorded to the nearest 0.1 cm.

Intra-observer assessment

70 nurses participated in the intra-observer variability testing. Each CHN recruited 4 children under the age of five and measured the children on three consecutive days. Children who met the eligibility criteria were selected from among parents who showed interest in participating in the study.

Inter-observer assessment

The observers for the inter-observer assessment were recruited only from the community health nursing school in Tamale. Students were trained prior to the study as part of the curriculum to take anthropometry measurements. 5 children of ages 26, 37, 42, 54 and 59 months respectively whose parents had been selected to be measured participated in inter-observer assessment of MUAC. Each observer measured a child once. The children were seated far apart from one another and the students measured once at a time without disclosing the measurement to one another. There were two supervisors overseeing that students measured independently.

Ethical Considerations

Ethical approval for the study was sought from the Institutional Review Board of the Noguchi Memorial Institute for Medical Research, Ghana.

Data was double-entered in EPIDATA and exported to Stata 10 (Statacorp, USA). Bland Altman plots and Pitman's test were used to assess intra-observer. 70 observers were recruited of which data sheets from 56 of them was used for the analysis due to missing data on the remaining sheets.

To assess inter-observer variability, the differences between the average of all the measurements and the individual measurements were determined and further explored using distribution plots (normal histogram).

Intra-class correlation was estimated using a random intercept model from an Xtre command

in STATA. The same procedure was used to explore the association between average left hand MUAC measurements on day 1 and 3 and covariates such as gender and age. Results were presented as regression coefficients with the associated 95% confidence interval and P-values. (See Table 1 and 2).

RESULTS

Assessment of Intra-observer variability

Analysis was mainly focused on difference in the MUAC measurements taken on day 1 and day 3. Figure 1 shows the Bland Altman plot for intra-observer measurements assessing the agreement between day 1 and 3. The scatter around the bias line seemed to get smaller as the average increases. The 95% limit lines are about plus or minus 0.9 with few differences outside or quite close to the limit lines (limit of agreement: -0.90cm to 0.83cm). The outlying measurements however were within the 5% allowable error that could probably have occurred by chance. The estimate of the bias is -0.04 cm with a 95% confidence interval that does include 0 (95% CI: -0.09 to 0.02). In addition, the pitman's correlation coefficient between the difference and the average of day1 and day3 is quite small and might have probably occurred by chance (pitman's correlation coefficient (r) = -0.02, p = 0.77). Comparing measures of day 1 and day 3 as reported from Bland Altman test did not show much variation.

Assessment of Intra-observer variability between left and right MUAC

The bland Altman plot for assessing the intra-observer variability between left and right MUAC on day 1 is presented in Figure 2. The 95% limit of agreement varied between -0.249 to 0.228. The bias is quite small and is estimated to be -0.010cm (95%CI -0.026 to 0.005). Besides, the correlation between the two measurement is not significantly different from 0 at alpha level of 0.05 (pitman's correlation coefficient (r):-0.02; p =0.8

Exploratory analysis of the data used for assessment of intra- observer variability

A two-way scatter plot for the repeated left hand MUAC measurements on day 1 and 3 is presented in Figure 3. The plot revealed that measurements taken from the same person

were correlated (i.e. clustered together). Thus for each child, it appeared that the mean of the two observations are normally scattered (like the normal distribution) around the overall mean (i.e. the red circles shown in the Figure 3).

The estimated regression coefficients for the random intercept model constructed for the average left hand MUAC measurements between day 1 and 3 is shown in Table 1. At the subject specific level, a 10 month increase in age increases significantly the average left hand MUAC measurement between day 1 and 3 by 0.8 cm when the effect of gender is controlled for. (95% CI: 0.07 to 0.09; $P < 0.001$).

In contrast to females, sub population of males showed on average a non-significant decrease of 0.08 cm in left hand MUAC measurements between day 1 and 3 after controlling for the effect of age (95% CI: -0.40 to 0.23; $P = 0.60$).

The estimate of the overall mean is 12.174 and the estimate of the standard deviation of the random intercept of children is 1.212 (i.e. σ_u). Hence we expect from 95% of children who are females to have MUAC left measurement that ranges from 10 to 15 $\{12.174 \pm 2 \times 1.212\}$ cm.

The Intra class correlation coefficient (ICC), which reflects left hand MUAC measurement at

different days for the same child, was very close to 1, (i.e. 0.993) indicating that the left hand MUAC measurement for the intra-observer data is very reliable. In addition, this indicates that the repeated observations within children are highly correlated and the proportion of the total variance that is between children is very large.

The estimated within-subject standard deviation (σ_u) is 0.0995. Hence we expect 95% of the repeated observations on a child to fall within 0.2 (= approximately 0.099×2) cm from the subject-specific mean.

Compared to the intra-observer data, the ICC for the inter-observer data, reflecting the left hand MUAC of 4 different children measured once by each of the 48 observers was very close to 0 (i.e. 0.042) as shown in Table 2 indicating quite a low reliability of this measure.

Inter-observer variability assessment

The frequency histogram for the MUAC difference between the average of the overall measurement of the 4 children and the individual measurement is shown in Figure 4. The histogram does approximate a standard normal or Gaussian distribution. Thus assuming we have a difference in left MUAC of 1 cm in this population with mean 0 and standard deviation of 0.5, this difference

Table. 1: Estimates from regression model (intra-observer variability)

| Mean left MUAC | coefficient | 95% | P values |
|----------------|-------------|-----------------|----------|
| age | 0.800 | 0.069 to 0.091 | 0.000 |
| male | -0.084 | -0.403 to 0.234 | 0.604 |
| rho | 0.993 | | |

CI = confidence interval

Table. 2: Estimates from regression model (inter-observer variability)

| Mean left MUAC | Coefficient | 95% CI | P values |
|----------------|-------------|-----------------|----------|
| age | -0.011 | -0.028 to 0.007 | 0.232 |
| rho | 0.042 | | |

CI = confidence interval

of 1cm is calculated to have a z-score of 2. For a normal distribution, a z-score of 2 (i.e. value-mean)/sd = 0.5), = 1/0.5}, always corresponds to the 95th percentile. Thus we can assume that 1cm difference in left MUAC is the 95th percentile in the empirical data, meaning that 95% of the scores lie at or below 1cm.

Assuming we have 1cm difference in left MUAC for this child. This difference is calculated to have a z score of 2 and implies that 95% of the scores for this child fall at or below 1cm difference in MUAC left. The implications of a 0.5 cm error in MUAC are higher than a 0.5 cm error in height. If for the same children, the difference in MUAC measures taken by different observers' results in values with such a huge variation as shown in Figure 5, it implies that the reliability of MUAC is poor. In figure 5, the difference in MUAC measures of one child showed a wide range. This can result in false positives and false

Figure 5 displays the left MUAC difference between the average of overall measurements for this child and the individual measurement. This variable does approximate a Gaussian distribution.

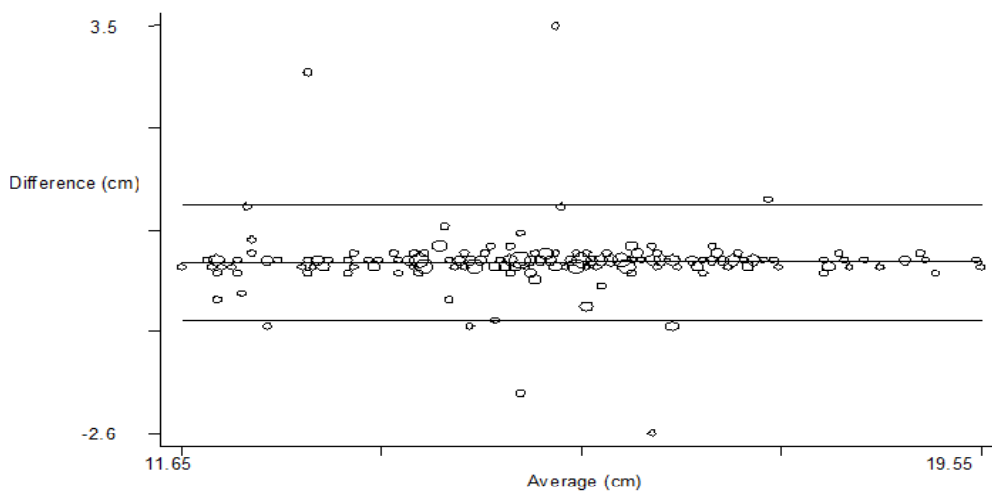


Fig. 1: Bland Altman plot for intra-observer measurements of MUAC between day 1 and day 3

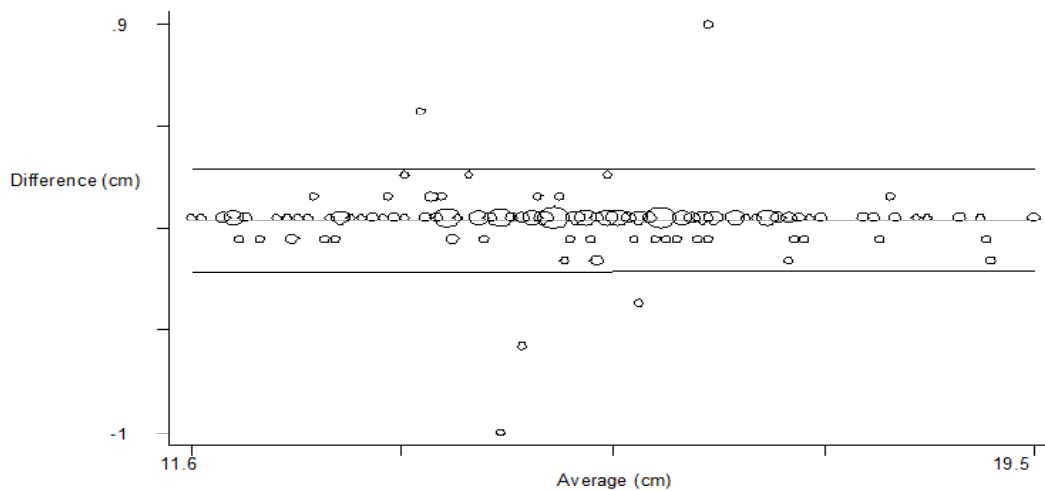


Fig. 2: Bland Altman plot for intra-observer assessment between left and right MUAC for day 1

negatives misdiagnosis. Assuming this child has to be screened using the MUAC cut –off for severe malnutrition of MUAC< 115 mm, the child could be misclassified either to be adequately nourished or malnourished. The clinical implication for screening is that malnourished children, who should be referred for specialised treatment, could be missed and this situation can reduce the chances of their survival. Additionally, false positive misdiagnosis can have negative consequences for the child and the mother.

DISCUSSION

MUAC was initially proposed for use in emergency settings, but for the past few years it has been applied in other settings. These include estimating the prevalence of wasting in children with a risk of severe acute malnutrition at the community level.

The aim of the study was to investigate some test characteristics of MUAC. Previous reliability studies on MUAC aimed at comparing it with other indices. In these studies, different methods were used to recruit and train observers, and to analyze data in order to assess reliability. We assessed the intra and inter-observer variability in using MUAC

at the community level. It is rational to assume that inter-observer error should be greater than intra-observer error. However, more often than not, inter- and intra-observer R measures of weight, height, length arm and waist circumference are similar to each other (Ulijaszek & Kerr, 1999; Ulijaszek & Lourie, 1994). We showed that reproducibility within community health nurses (intra –observer) in taking MUAC among children 12- 60 months was high, while reproducibility between observers (inter-observer) was low. Our findings imply that measurement error for repeated MUAC measures between two or more observers is higher than error within an observer.

For intra-observer reproducibility of MUAC, Bland Altman graphs plots showed a very high degree of agreement between day 1 and 3. The Intra class-correlation coefficient (ICC) was very close to 1 (i.e. 0.993) indicating that the MUAC left measurement for the intra-observer data is very reliable. This agrees with findings from a study among minimally trained health workers in which fewer errors were made for arm circumference and a Guatemalan study in which health promoters were trained to measure weight, height and MUAC (Collins, 1996; Velzeboer, Selwyn, Sargent, Pollitt, & Delgado, 1983).

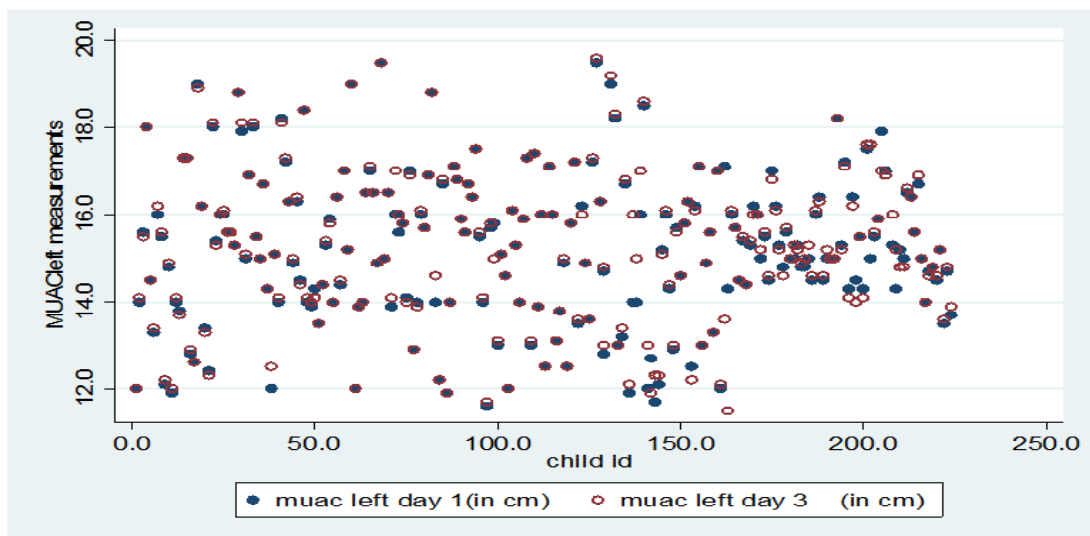


Fig. 3: Plot showing the repeated left hand MUAC measure for each subject and the overall mean left hand MUAC between day 1 and 3

Additionally, our study compared MUAC taken on the left arm to that of the right. The bias between the two arms was quite small and estimated at -0.010 (95%CI -0.026 to 0.005). Also the correlation between the two arms was not statistically different (P=0.8). This implies that, among children 12-60 months, using the left or right makes no difference. Hand dominance is probably not expressed in children up until they are older.

may be lower with these straps than the numbered straps. A look at the raw data showed that, most of the measures from the observers were rounded up to the nearest whole number (e.g., 15.0 cm, 16.0 cm instead of 15.2, 16.7). Reporting data in this manner will result in the loss of a lot of data. This is because measures between 0.1 and 0.9 will not be accounted for. It is worth noting that a difference of about 0.5 cm (50mm) will make a huge difference in terms of classification of children for malnutrition.

A possible explanation for our high reproducibility could be the use of the Shakir's tape. It is generally assumed by field workers that errors

The findings from Inter-observer variability assessed with histograms showed wide variation

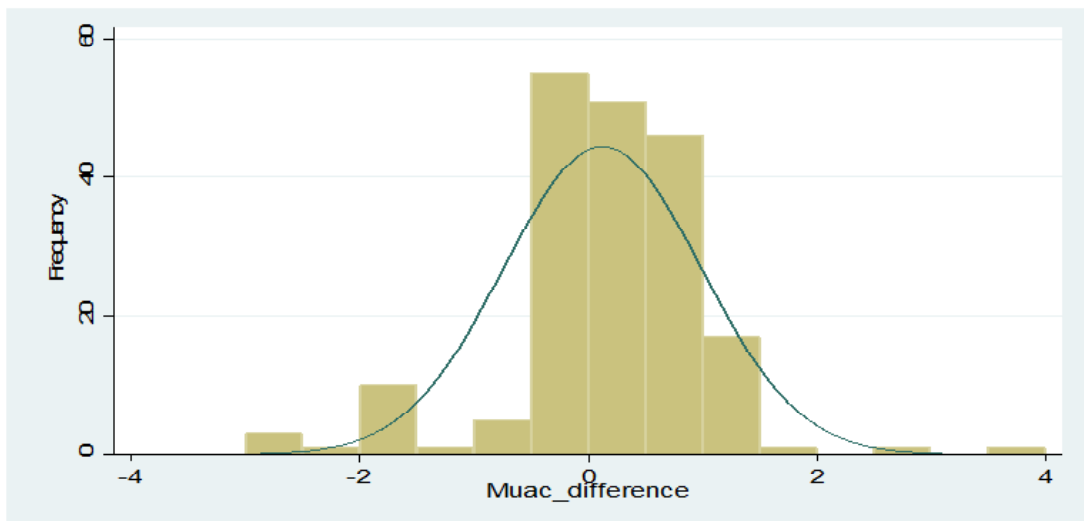


Fig. 4: The histogram of left MUAC difference of the 4 children

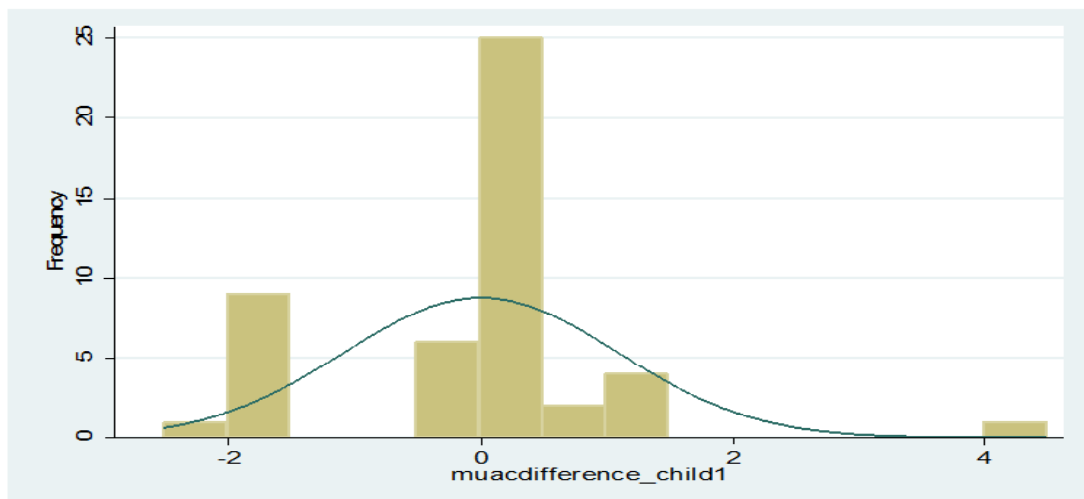


Fig. 5: The histogram of differences in measures of left MUAC of child 1 (age: 26 months)

between observers. This finding is similar to the results of a previous study in which investigations showed that the errors could have been from the different ways the MUAC strips were pulled to facilitate admission into programs (Myatt *et al.*, 2006). The ICC for the inter-observer data, reflecting the left MUAC measurements of 4 different children measured once by each of the 48 observers was very close to 0 (i.e. 0.042) indicating quite a low reliability of this measure compared to the intra-observer data.

The differences in how tightly the strips are pulled, the non-uniformity in how the arm should be placed (i.e. whether the arm is bent or left hanging) differs from one observer to the other. Also there is variation in the estimation of the midpoint of the arm among observers. These differences could explain why the variation between observers is more than within observers. It also shows that having a large number of observers is not always an advantage.

The Spanish AVENA study for adolescents and healthy lifestyle also measured intra-observer reliability. Agreement in measurement was good following a training of observers (Moreno *et al.*, 2003). The study emphasizes the need to train observers in order to reduce measurement errors in anthropometry.

The high inter-observer variability reported in our study among community health nursing students may have some clinical and diagnostic implications. If children who are apparently healthy are misclassified as malnourished, it will put an extra burden on health services. This problem is more serious in communities where resources are limited. With a high number of false positives the quality of care given to the true positives will be affected. Misclassification can also result in correct diagnosis misses, delayed diagnoses, suffering and for children at the border line for malnutrition; this could reduce their chances of survival.

At the subject specific level, a 10 month increase in age increases significantly the average MUAC left measurement between day 1 and 3 by 0.8 cm when the effect of gender is controlled for (95% CI: 0.07 to 0.09; $P < 0.001$). This association found between age and MUAC measures is in contrast to a

study (Stomfai *et al.*) in which age did not influence reliability results. It is however similar to other studies (Ayele *et al.*, 2012) in which reliability was greater for smaller children compared with larger children.

Additionally, in contrast to females, the sub-population of males showed on average a non-significant decrease of 0.08 cm in MUAC left measurements between day 1 and 3 after controlling for the effect of age (95% CI: -0.40 to 0.23; $P = 0.60$). The association found between gender and MUAC measures is similar to a study (Ayele B, 2012), in which the arm circumference of the males was more than that of the females and the differences between the two sexes were statistically insignificant (P values ranged between 0.1 to 0.9).

Different studies in which inter- and intra-observer errors have been reported show wide variability (Chumlea, Guo, Kuczmarski, Johnson, & Leahy, 1990). Some studies have also reported extremely high TEM values for intra-observer error for measurements of triceps, and skin folds (Ferrario, Carpenter, & Chambless, 1995), and weight, waist and hip circumference (Rimm *et al.*, 1990). The inconsistencies in the findings of studies on accuracy and reliability generally could be related to the study methods and sample selection which varies in each case.

Evidence to show high measurement error has been identified in some circumstances. These include large epidemiological studies in which anthropometry forms part of a complex study design and in self-reported measurements taken by anthropometrists with limited experience (Ulijaszek & Kerr, 1999).

This study assessed variability in MUAC measurement among a convenient sample of community health nurses and nursing students. The results from the intra-observer testing could be explained by the fact that recruited observers are the community health nurses who are very meticulous in taking measurements. Further studies are needed among community health nurses that are randomly selected.

Assessment of inter- and intra-observer variability for a master's thesis may have provided

an artificial environment for the community health nurses and could thereby not reflect what exactly happens in the field. It is also not surprising that measurements taken on the three consecutive days for the intra-observer assessment did not vary much, because each nurse measured only four children. There is a tendency to remember previous measurement. Future studies should use more children per community health nurse.

The findings of this study may not be applicable to all community health nurses doing a rapid assessment for malnutrition at the community level.

ACKNOWLEDGMENT

Data collection was funded by the Ghent university and this research was supervised by Professor Patrick Kolsteren, we acknowledge, with gratitude their role in this research. The authors are also grateful to Professor Carl Lachat, Dr. Lieven F. Huybregts, Kimberley Bouckaert, Ir. Anne-Marie Remaut-De Winter and Priscilla Uba for their various roles in this research.

REFERENCES

1. Ayele B, A. A., Gebre T, Tadesse Z, Stoller NE, reliability of measurements performed by community drawn anthropometrist from rural Ethiopia. *PLoS ONE*, **7**(1) (2012).
2. Ayele, B., Aemere, A., Gebre, T., Tadesse, Z., Stoller, N. E., See, C. W., . . . Porco, T. C. Reliability of measurements performed by community-drawn anthropometrists from rural Ethiopia. *PLoS one*, **7**(1), e30345 (2012).
3. Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., De Onis, M., Ezzati, M., . . . Rivera, J. Maternal and child undernutrition: global and regional exposures and health consequences. *The Lancet*, **371**(9608), 243-260 (2008).
4. Briend, A., & Zimicki, S. Validation of arm circumference as an indicator of risk of death in one to four year old children. *Nutrition research*, **6**(3), 249-261 (1986).
5. Chumlea, W. C., Guo, S., Kuczmarski, R. J., Johnson, C. L., & Leahy, C. K. Reliability for anthropometric measurements in the Hispanic Health and Nutrition Examination Survey (HHANES 1982-1984). *Am J Clin Nutr*, **51**(5), 902S-907S (1990).
6. Collins, S. Using middle upper arm circumference to assess severe adult malnutrition during famine. *JAMA: The Journal of the American Medical Association*, **276**(5), 391-395 (1996).
7. Farkas, L. G. Accuracy of anthropometric measurements: past, present, and future. *The Cleft palate-craniofacial journal*, **33**(1), 10-22 (1996).
8. Ferrario, M., Carpenter, M. A., & Chambless, L. E. Reliability of body fat distribution measurements. The ARIC Study baseline cohort results. Atherosclerosis Risk in Communities Study. *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, **19**(7), 449 (1995).
9. Jamaayah, H., Jr., Geeta, A., Safiza, M. N., Wong, N. F., Kee, C. C., Ahmad, A. Z., . . . Faudzi, A. Reliability and Technical Error of Calf Circumference and Mid-half Arm Span Measurements for Nutritional Status Assessment of Elderly Persons in Malaysia. *Malaysian Journal of Nutrition*, **14**(2), 137-150 (2008).
10. Moreno, L. A., Joyanes, M., Mesana, M. I., González-Gross, M., Gil, C. M., Sarría, A., . . . Bueno, M. Harmonization of anthropometric measurements for a multicenter nutrition survey in Spanish adolescents* 1. *Nutrition*, **19**(6), 481-486 (2003).
11. Myatt, M., Khara, T., & Collins, S. A review of methods to detect cases of severely malnourished children in the community for their. *Putting the management of severe malnutrition back on the international health agenda*, **27**(3), 7 (2006).
12. Rimm, E. B., Stampfer, M. J., Colditz, G. A., Chute, C. G., Litin, L. B., & Willett, W. C. Validity

- of self-reported waist and hip circumferences in men and women. *Epidemiology*, **1**(6), 466-473 (1990).
13. statement, W. U. W. S. J. . Community- based management of severe acute malnutrition: Geneva, New York, Rome, (2007).
 14. Stomfai, S., Ahrens, W., Bammann, K., Kovacs, E., Marild, S., Moreno, L. A., . . . Veidebaum, T. Intra-and inter-observer reliability in anthropometric measurements in children.
 15. Ulijaszek, S. J., & Kerr, D. A. Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition*, **82**, 165-177 (1999).
 16. Ulijaszek, S. J., & Lourie, J. A. Intra-and inter-observer error in anthropometric measurement. *Anthropometry: the Individual and the Population*, 30-55 (1994).
 17. Velzeboer, M. I., Selwyn, B. J., Sargent, F., Pollitt, E., & Delgado, H. The use of arm circumference in simplified screening for acute malnutrition by minimally trained health workers. *Journal of Tropical Pediatrics*, **29**(3), 159 (1983).
 18. WHO. *WHO Child Growth Standards and the Identification of Severe Acute Malnutrition in Infants and Children: A Joint Statement by the World Health Organization and the United Nations Children's Fund*: World health organization (WHO) (2009).