Ceylon J. Med. Sci. **31** (No. 2 December 1988) p. 25 — 34

OBSERVER VARIABILITY IN ANTHROPOMETRY

by

S. BALASURIYA

(*Faculty of Medicine, University of Peradeniya*)

SUMMARY. Sixteen trained field workers, working in eight teams, measured eight children to find the observer variability in the common anthropometric measurements. All measurements were found to be subjected to a siginficant observer error, stature being the one with least error and skinfold thickness the one with the highest error. It is recommended that the latter measurement should not be used unless absolutely indicated. The importance of recognising the high degree of observer error is discussed and certain precautions to reduce it are indicated.

INTRODUCTION

The commonly used anthropometric indices for the assessment of nutritional status are body weight, stature and skinfold thickness. These measurements are easy to obtain and are widely used for the identification of beneficiaries of food supplementation programmes, However, they suffer from the serious drawback in that they are subjected to errors of instruments and observers. Observer errors have been widely reported in the measurement of skinfold thickness in both adults (1, 2, 3, 4) and childern (5). However, we have not been able to trace any literature on the observer variability in other anthropometric measurements.

The aim of the present study, carried out in London, England, was to estimate the importance and magnitude of observer variability in the measurement of height, skinfold thickness, mid-arm circumference and head circumference.

MATERIAL AND METHODS

Eight teams of field workers with two in each, working in four areas were tested. These field workers had undergone an initial one week training programme in standard measurement techniques (2, 6). Working in pairs, they had been measuring pre-school children in the field for a period of over four months.

Eight children, four boys and four girls, aged 2—6 years were used as subjects.

The instruments used were a modified Holtain Infantometer to measure the height, Holtain skinfold calipers and Miniflex flexible steel tapes.

The investigation was carried out in two stages.

Stage 1 : The eight teams measured the height and length of the eight children using eight different sets of instruments. This suggested a latin square design where the subjects, instruments and teams are randomly allocated numbers or letters (Table 1).

TABLE 1. Latin square design for stage 1

(Teams given in the body of the table)

Child	1	2	3	4	5	6	7	8
Equipment								
1	A	F	D	G	C	H	B	E
2	E	B	H	C	G	D	F	A
3	B	E	C	H	D	G	A	F
4	F	A	G	D	H	C	E	B
5	G	D	F	A	E	B	H	C
6	C	H	B	E	A	F	D	G
7	H	C	E	B	F	A	G	D
8	D	G	A	F	B	E	C	H

Stage 2 : The other anthropometric measurements were taken the following day. Each observer, with the assistance of her team-mate, measured the eight children using their own equipment. This resulted in 16 measurements for each child taken by 8 teams using 8 sets of equipment.

RESULTS

The randomised block design in a Latin square employed in the study is expected to eliminate the systematic variation between subjects and observers. The technique of analysis of variance (ANOVA) was used in the analysis as it separates the total variation into components–between children and between observers and a residual. The between observers variation can be further subdivided into between team variation and between observers within team variation. The ratio of such explained variation to the unexplained variation (residual) gives a F value, the significance of which can be read off from a table.

Stature and length

The equipment variation was not significant for either measurement while the team variation was significant only for the measurements of length (Tables 2 and 3). Close scrutiny of the raw data revealed that one team, team F, obtained consistently low readings for length measurements. The mean length for the other teams was around 1042 mm while the mean for team F was 1034 mm, which was significantly different from the means of most other teams. When the measurement of team F are excluded the variability between teams was not significant at 5% level (Table 4).

TABLE 2. Analysis of variance (ANOVA) for height measurements

Source	DF	Sum of squares	Mean square	F	Significance
Children	7	574426.985	82060.99	8856	
Equipment	7	121.985	17.426	1.88	NS
Team	7	139.485	19.926	2.15	NS
Residual	41	379.905	9.266		
Total	63	575068.360			

NS = Not significant at 5% level

TABLE 3. ANOVA of Length measurements

Source	DF	Sum of squares	Mean square	F	Significance
Children	7	552462.235	78923.176	3110.176	
Equipment	7	167.235	23.891	0.9415	NS
Teams	7	863.235	119.462	4.7077	**
Error	41	1040.405	25.376		
Total	63	5545506.110			

NS = Not significant at 5% level

** $P<0.01$

TABLE 4. ANOVA of length measurements excluding Team F

Source	DF	Sum of squares	Mean square	F	Significance
Children	7	480842.214	68691.745	2862.4	
Equipment	7	178.869	25.553	1.065	NS
Teams	6	328.357	54.726	2.28	NS
Residual	35	839.916	23.998		
Total	55	482189.357			

NS = Not significant at 5% level

Head circumference

The observed variation between the observers was significant at 5% level. When the total observer sum of squares was subdivided into "between teams" and "within teams" the former was significant at 1% level and the latter not signiificant (Table 5). Close scrutiny of the data revealed that one worker was consistently reading lower than the others and this worker belonged to team H.

TABLE 5. ANOVA of measurement of head circumference

Source	DF	Sum of squares	Mean squares	F	P value
Between teams	7	3.9988	0.5713	3.24	< .01
Within teams between observers	8	1.3950	0.1744	<1	NS
Between observers	15	5.3938	0.3596	2.042	< .05
Between children	7	211.5113	30.2159	171.55	< .001
Residual (Children × observers)	105	18.4937	0.1761		
Total	127				

Mid-arm circumference

The variability between observers was again highly significant. As in the head circumference measurement, the between-team variation was significant ($p<0.001$) while the pooled effect within teams between observers was not significant (Table 6). In this instance too, one observer, no. 16 (Team F), was found to obtain consistently low readings.

TABLE 6. ANOVA of measurement of mid – arm circumference

Source	DF	SS	MS	F	P value
Between teams	7	11.5206	1.6458	4.27	< .001
Within teams between observers	8	2.6843	0.3353	< 1	NS
Between observers	5	14.2049	0.9470	2.458	< .01
Between children	7	139.5643	19.9378	51.76	< .001
Residual	105	40.4493	0.38523		
Total	127	165.8087			

Triceps skinfold thickness

The between observers variability was again very highly significant and so was the between teams variation. A significant variation was also observed ($p<0.05$) within teams between observers (Table 7). Examination of the raw data reveals that observer no. 12 (Team F) gave significantly low readings compared to others, while observer no. 10, too, gave readings different from the others. Exclusion of these two workers and reanalysis revealed that the significance level can be reduced to 5%. The team differences were also apparent. Team F gave measurements significantly diflerent from all others while teams E and C gave higher readings than most.

TABLE 7. ANOVA of measurement of triceps skinfold thickness

Source	DF	SS	MS	F	P value
Between teams	7	46.5225	6.6461	8.62	< 0.001
Within teams between observers	8	16.1587	2.0198	2.62	< 0.05
Between observers	15	62.6812	4.1787	5.42	< 0.001
Between children	7	232.9188	33.2741	43.16	< 0.001
Residual	105	80.9487	0.7709		
Total	127	376.5487			

Subscapular skinfold thickness

The observer and the team variability were both highly significant, while that between observers within teams was significant only at 5 % level (Table 8). However close scrutiny of the data revealed that there was a wide dispersion throughout the array of mean values and no observer could be picked out as being responsible for the observed difference.

TABLE 8. ANOVA of measurement of subscapular skinfold thickness

Source	DF	SS	MS	F	P value
Between teams	7	17.2662	2.4666	14.78	< 0.001
Within teams between observers	8	3.0368	0.3796	2.27	< 0.05
Between observers	15	21.3030	1.4202	8.506	< 0.001
Between children	7	102.1424	14.5918	87.394	< 0.001
Residual	105	17.5314	0.1670		
Total	127	140.9768			

TABLE 9. ANOVA of measurement of supra-iliac skinfold thickness

Source	DF	SS	MS	F	P value
Between teams	7	102.2192	14.602	13.9385	< 0.01
Within teams	8	18.2369	2.2796	2.170	< 0.05
Between observers	15	120.4561	8.0	7.6654	< 0.001
Between children	7	412.3848	58.9121	56.2333	< 0.001
Residual	104	641.7947	1.0476		
Total	126				

Supra-iliac skinfold thickness

There was one missing value and this was estimated and used in the analysis. ANOVA reveals that both the observers and the teams were significantly different at 1% level, while the pooled effect within teams between workers was significant only at 5% level (Table 9). For this measurement the team differences were striking and the teams could be split into three distinct groups. Team D gave a mean value significantly different from all others while team G, E and A were significantly different from C, F, B and H, although the differences between the above sub-groups were not significant.

Biceps skinfold thickness

The observer difference was very highly significant and so was the team difference. but the within-team between observer difference was significant only at 1% level (Table 10), Re-examination of data showed that observers 8, 5 and 7 were responsible for most of the differences. The observers no. 7 and 8 (Team D) had significantly higher values than all others.

TABLE 10. ANOVA of measurement of biceps skinfold thickness

Source teams	SS	DF	MS	F	P value
Between teams	72.1362	7	10.3052	12.413	< 0.001
Within teams between observers	23.0588	8	2.8823	3.472	< 0.01
Between observers	95.195	15	6.3463	7.644	< 0.001
Between children	71.9487	7	10.2784	13.38	< 0.001
Residual	87.1712	105	0.8302		
Total	254.315	127			

Coefficient of variation of skinfold thickness

The coefficient of variation was calculated for each child based on the measurements of 16 observers. When single skinfold thicknesses are considered the one with the lowest variability is the subscapular. The summation of all four thicknesses reduces the variability to some degree, but the best results are obtained by the summation of triceps and subscapular folds (Table 11).

TABLE 11. Coefficient of variation of skinfold measurements

Children	Triceps	Biceps	Sub-scapular	Supra-iliac	Sum of 4	Triceps and Subscp.
1	11.8	24.9	7.7	15.4	8.5	6.7
2	17.4	19.6	9.5	33.2	16.6	11.2
3	13.3	20.7	12.0	14.7	10.0	10.1
4	8.5	15.8	9.7	19.3	8.6	5.6
5	11.3	32.3	13.6	18.0	11.9	7.5
6	15.1	16.8	9.2	19.6	10.3	10.0
7	14.6	26.5	8.2	19.2	9.9	9.1
8	11.9	21.5	7.8	18.1	10.3	8.4

TABLE 12. Means of different anthropometric measurements on eight children by teams of observers

Team	Length mm	Head Circumference cm	Arm Circumference cm	Skinfold thickness mm			
				Triceps	Subscapular	Suprailiac	Biceps
A	1038.1	50.05	17.37	7.72	5.94	6.28	4.61
B	1039.9	49.90	17.34	7.51	5.00**	5.23	4.54
C	1044.4	50.18	17.35	8.73	6.03	5.32	5.74
D	1043.6	50.11	17.29*	8.30	4.97**	7.66**	6.87**
E	1045.0	50.05	17.07	8.72	5.48	6.45	5.05
F	1033.9**	50.14	16.67**	6.83**	5.39	5.31	4.54
G	1044.2	50.14	17.69	7.99	5.73	6.80	5.71
H	1041.5	49.60**	17.20	7.84	5.39	4.92	5.20
Difference between highest and lowest mean	11.9	0.58	1.02	1.9	1.06	2.74	2.33

* Significantly different from most readings ** Significantly different from all other readings

DISCUSSION

The anthropometric measurements used in the study are those usually employed in the assessment of nutritional status. Body weight, the most widely used measurement was not used as the errors involved in its measurement are well known. The results presented here reveal that all measurements (with the possibe exception of stature, where the observed variability is short of significance) are subjected to a significant observer variability, the worse offenders being the skinfold thicknesses. Even a very simple measurement like the mid-arm circumference is not free from this error. The variability observed within teams (between observers), though significant in most instances, was less marked than the between team variation. Examination of the mean values of each team shows that there is no consistency in the error, with some teams giving high values for some measurements and other giving high values for others (Table 12). This variation, though a random occurrence, cannot be ignored as the magnitude is very high, viz, a difference of 12 mm of the mean value for length measurement and over 10 mm for the mean value for the mid-arm crcumference measurement. The magnitude is even higher for the measurement of skinfold thickness.

The observers taking part in the study had followed the same training programme and had been working in the field for a minimum period of four months. At the end of the training programme, the author was satisfied that all the observers were using the same standard technique, but some deviations by some workers were observed during this experiment. This suggests that the field workers tend to modify the technique while working in the field without realising that they are doing so. The sixteen field workers were very carefully selected from a large group of applicants and they were all well educated and were found to be very responsible in other ways. The equipment used were the most uptodate ones with a high degree of accuracy but these precautions appear to be inadequate in eliminating measurement variability. The only other precaution one may take is to regularly supervise the field workers in the field. This can be done by a "Standard Observer" who can visits the field regularly to check the measurements and correct early any deviations by the observers from the standard technique.

In most countries a large number of national anthropometric surveys are undertaken by various organizations, both governmental and non-governmental. Sometimes they are used to identify beneficiaries of supplementation programmes and sometimes the findings are used to plan national policies. The observers used is such surveys in developing countries are less educated than the ones used in this study and the training given to them is at times inadequate. Field supervision in some instances is poor and the accuracy of the equipment used for measurements may be questionable. Regular standardisation is not possible as the numbers of the equipment used are high. While conducting anthropometric measurements of school children using good equipment and trained observers, the author on one occasion, detected that the weighing scale was reading nearly 4 kg lower. This error could be put right as the author was available at the site.

It is therefore imperative that the field workers used for anthropometric surveys should be very carefully selected, well trained and regularly supervised, if one were to use their findings in a meaningful way. Skinfold thickness should not be used as it is subjected to the highest variability, unless there is an absolute indication. In such instances the use of the combination of triceps and subscapular folds is recommended as it showed the minimum coefficient of variation.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Gregor Sutherland and Miss Elizabeth Thom, both of the London School of Hygiene and Tropical Medicine, for their help in planning the study. Thanks are also due to Miss P. Peiris for typing the manuscript.

Some of the results appeared in a M. Phil thesis presented to the University of London by the author.

REFERENCES

1. Durnin JVGA, Armstrong WH, Womersley J. An experimental study on the variability of measurements of skinfold thickness by three observers on twenty three young women and twenty seven young men. Proceedings of the Nutrition Society 1971; 30: 9A.

2. Jelliffe DB. The assessment of the nutritional status of the community. Geneva ; World Health Organization 1966.

3. Ruiz L, Colley JRT, Hamilton PJS. Measurement of triceps skinfold thickness : an investigation of sources of variation. British Journal of Preventive and Social Medicine 1971 ; 25: 165—167.

4. Womersley J, Durnin JVGA. An experimental study of variability of measurements of skinfold thickness in young adults. Human Biology 1973 ; 45: 281—292.

5. Edwards DAW, Hammond WH, Healy NJR, Tanner JM, Whitehouse RH. Design and accuracy of calipers for measuring sub-cutaneous tissue thickness. British Journal of Nutrition 1955 ; 9: 133—137.