J. Human Ergol., **28**: 15-24, 1999

INTEROBSERVER ERRORS IN ANTHROPOMETRY

MAKIKO KOUCHI, MASAAKI MOCHIMARU, KAZUYO TSUZUKI, AND TAKASHI YOKOI

National Institute of Bioscience and Human-Technology
Human-Environment System Department
Higashi 1-1, Tsukuba, Ibaraki 305-8566, Japan

To present basic information on the interobserver precision and accuracy of 32 selected anthropometric measurement items, six observers measured each of 37 subjects once in two days. The data were analyzed by using ANOVA, and mean absolute bias, standard deviation of bias, and mean absolute bias in standard deviation unit were used as measures of bias. By comparing the results of the two days, the effects of the practice on measurement errors were also investigated. Variance was overestimated by more than 10% in five measurements. Interobserver error variance and random error variance were highly correlated with each other. Measures of the bias were significantly correlated with interobserver and especially with random error variances. The interobserver errors were drastically reduced on the second day in the measurement items in which the causes of the interobserver errors could be specified. It was speculated that even when the definitions of the landmarks and measurement items were clear, the ambiguity in the practical procedures in locating landmarks, applying instruments, and so on, permitted each observer to develop his or her own measurement technique, and it in turn caused interobserver errors. To minimize interobserver and random errors, the standardization of measurement technique should be extended to the details of the practical procedures.

INTRODUCTION

No anthropometric data are free from measurement errors. Interobserver measurement errors have serious effects on the population comparisons and growth assessment (Heathcote, 1981; Cameron, 1984; Kouchi and Koizumi, 1985). Since we must use published anthropometric data for comparative purposes, information on the magnitude of measurement errors is essential in interpreting the results of statistical tests.

There are two aspects in measurement errors: the closeness of the measured value to the true value (accuracy) and the closeness of the two repeated measurements (precision). For 219 measurement items, we have presented the data on the precision when the repeated measurements are taken by the same observer (Kouchi et al., 1996). Since interobserver errors are greater than intraobserver errors, many studies have treated the precision when the repeated measurements are taken by different observers (Gordon et al., 1992; Jamison and Ward, 1993; Williamson et al., 1993).

On the other hand, it is difficult to assess the accuracy or the bias of the measurement taken by an observer because the true value is unknown. Mueller and Martorell (1988) suggested to assess the accuracy by comparing the values taken by a well-trained supervisor and the observers being evaluated. However, the experience does not necessarily guarantee the accuracy. An experienced observer may develop a measurement technique that is very precise, but biased (Utermohle et al., 1983).

Usually no information is available for the precision and accuracy of the measurements taken by particular observers participating in a particular research whose report we may cite. However, this information would be useful in interpreting the results of a population comparison and in determining the dimensions of industrial products based on human body dimensions.

Received for publication February 5, 1996

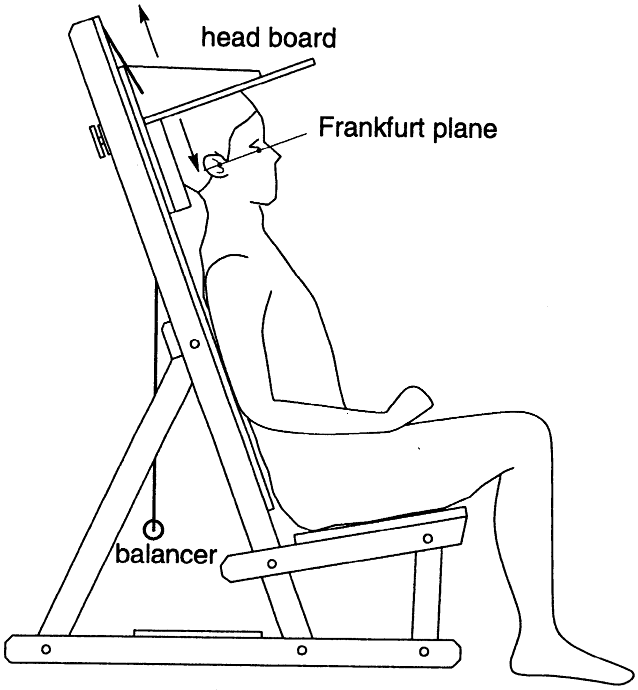

Fig. 1. Special chair for a head measurement apparatus

In the present paper, we present the information on the interobserver precision and the accuracy of 32 body dimensions, using the data taken on 37 subjects by six different observers. Another purpose of this study is to investigate the effective way to decrease interobserver measurement errors.

METHODS

Measurement item

Thirty-two measurement items used in the present study are shown in Table 2. They were selected so that the major dimensions of each part of the body were covered and major instruments were used. All measurements were taken in mm, and the right side was measured for bilateral measurements. For 18 measurement items taken according to Martin and Knussmann (1988), the item numbers they assigned are indicated in Table 2. Buttock-knee length, subscapular skinfold, and calf skinfold were measured according to Weiner and Lourie (1969), except that the right instead of the left side of the body was measured. The remaining 11 items are as follows:

6. Occiput to subnasale distance: the distance from the rearmost point of the occiput to subnasale measured parallel to the Frankfurt plane. It was measured with a head measurement apparatus that consists of a chair and a height gauge (Fig. 1).

7. Entocanthion to vertex height: the distance from entocanthion to the highest point of the head measured perpendicular to the Frankfurt plane. It was measured with a head measurement apparatus.

12. Crotch height: height from the floor to the lowest point of the ischial bone. The subject stands erect. Insert a hard thin plastic sheet between the legs against the inner surface of the right thigh, and lift it until its upper edge touches the lowest point of the right ischial bone. Confirm that the upper edge of the sheet is horizontal and the height from the floor to the upper edge is measured.

13. Sitting height: height from the sitting surface to vertex. The subject sits erect on a hard horizontal

Table 1. Subjects.

The first day (N=10)

	Mean	S.D.	Min.	Max.
Age (years)	25.4	3.37	19.1	30.3
Height (cm)	168.6	5.25	163.0	177.7
Weight (kg)	66.4	10.84	50.1	87.1

The second day (N=27)

	Mean	S.D.	Min.	Max.
Age (years)	23.4	3.74	18.5	30.3
Height (cm)	170.1	6.25	160.0	182.3
Weight (kg)	65.4	12.83	49.5	102.9

surface with thighs fully supported. Feet are supported so that the knees and ankles bend at right angles.

17. Chest depth: the maximum horizontal anteroposterior diameter of the chest at the nipple level. Measured at the end of normal expiration with a large sliding caliper.

18. Median chest depth: horizontal anteroposterior diameter of the chest at the nipple level in mid-sagittal plane. Measured at the end of normal expiration with a large spreading caliper.

19. Maximum lower leg breadth: breadth of the lower leg at the level of calf circumference. Measured perpendicular to the sagittal plane containing the foot axis.

21. Arm reach from back: the distance from the wall to the right dactylion. The subject stands erect with the back touching a vertical wall and arms extending forward. Measured with a section paper and a triangle.

23. Neck circumference: circumference of the neck measured perpendicular to the long axis of the neck at the level just below the Adam's apple.

24. Chest circumference: horizontal circumference of the trunk measured at the nipple level. Measured at the end of normal expiration.

25. Shoulder circumference: horizontal circumference of the trunk and arms measured at the same level with bideltoid breadth. The subject stands erect with arms hanging free.

Experiment

Six observers participated in the experiment. Four were well-trained anthropometrists with experience of more than 15 years, and one had experience of 3 years. One observer was a novice, who participated in the experiment after one month of training. Among the six observers, there were two trainer-trainee pairs.

Five of the six recorders had no experience in anthropometry. One observer and one recorder made a permanent pair during the two-day experiment.

The subjects are healthy Japanese male volunteers aged from 18 to 30 years. Their ages, heights and weights are shown in Table 1.

On the first day, all observers and recorders practiced together on two subjects for three hours. After a one-hour rest, each observer measured each of 10 subjects in a session of about 150 minutes. Each observer located the landmarks and made marks on them when necessary. After he or she took all the measurements on a subject, all marks were removed before the subject was measured by another observer. The measurements were analyzed for the bias.

On the next day, the observers were informed of the results on the bias of the first-day experiment. They discussed the causes of the bias, then practiced for 30 minutes. After that, each observer measured each of 27 subjects in three sessions. The subjects of the second day were different from those of the first day. About 10 were measured in each session of about 150 minutes. Between sessions, one-hour rest was allowed.

Statistical model and measures of errors

Means and standard deviations (variances) are the most frequently used statistics cited from papers. Therefore the effects of the measurement errors on these statistics are examined.

The measurement x_{ij} taken on subject i by observer j is the sum of the true mean (μ), the effect of the subject (s_i), the effect (bias) of the observer (o_j), and random error (e_{ij}) (formula 1).

$$x_{ij}=\mu+s_i+o_j+e_{ij} \qquad (1)$$

When more than one observer is involved, the expected values of mean is expressed as formula 2. The expected value of variance is the sum of intersubject variance (true variance), random error variance, and interobserver error variance (formula 3).

$$\mathrm{E}[x_{ij}]=\mu+\mathrm{E}[o_j] \qquad (2)$$

$$\mathrm{V}[x_{ij}]=\mathrm{V}[s_i]+\mathrm{V}[o_j]+\mathrm{V}[e_{ij}] \qquad (3)$$

The mean is biased by the mean of the biases of the observers, and variance is overestimated by intra- and interobserver error variances ($\mathrm{V}[e_{ij}]$ and $\mathrm{V}[o_j]$).

Some observers have positive and others have negative biases. When each of n observers measures each of m subjects once and when n is large enough, positive and negative biases are canceled out and $\mathrm{E}[o_j]$ would approximate to 0. That is, the grand mean (GM) would approximate to the true mean.

When GM is used instead of μ, the bias of observer j is expressed as M_j-GM, where M_j is the mean by the observer j. The mean absolute bias (MAB) is a measure of the bias expressed in the same unit as the measurement (formula 4). Standard deviation of the bias (SDB) is another measure of the bias, also expressed in the same unit as the measurement.

$$\mathrm{MAB}=[\sum_{j=1}^{n}(|\mathrm{M}_j-\mathrm{GM}|)]/n \qquad (4)$$

To make MAB unitless, MAB is expressed in the standard deviation unit of the measurement (MAB in SD unit). The intersubject variance was used for this calculation (formula 5).

$$\text{MAB in SD unit}=\mathrm{MAB}/\sqrt{\mathrm{V}[s_i]} \qquad (5)$$

As for the variance, the overestimation rate of the variance (OER) is calculated as formula 6 and reliability coefficient (R) as formula 7.

$$\mathrm{OER}=(\mathrm{V}[o_j]+\mathrm{V}[e_{ij}])/\mathrm{V}[s_i]=(\mathrm{V}[x_{ij}]-\mathrm{V}[s_i])/\mathrm{V}[s_i] \qquad (6)$$

$$\mathrm{R}=\mathrm{V}[s_i]/\mathrm{V}[x_{ij}] \qquad (7)$$

Statistical analyses

The data of the first day and of the second day were analyzed separately. An analysis of variance was conducted, and the total variance was divided into the three components: intersubject variance

Table 2. Grand mean and measures of the errors for the data of the second day (N=27)

	Measurement item	GM (mm)	F-value	V*o*	V*e*	%V*o*	%V*e*
1	Head length (C1)	189.6	13.2 **	1.8	0.8	3.8	1.7
2	Nose breadth (C3)	36.7	12.5 **	0.5	0.2	6.5	2.8
3	Interpupillary breadth (C12)	64.7	19.6 **	1.2	0.8	9.4	6.5
4	Morphologic face height (C18)	124.4	13.2 **	5.5	2.5	15.3	6.9
5	Bitragion chin arc (C52)	327.1	28.8 **	6.0	6.2	3.3	3.4
6	Occiput to subnasale distance	241.3	21.6 **	3.9	3.0	6.5	4.9
7	Entocanthion to vertex height	126.8	26.1 **	7.8	7.2	22.5	20.9
8	Height (S1)	1701.2	6.6 **	12.9	2.7	0.3	0.1
9	Iliospinale height (S13)	905.2	16.3 **	37.2	21.1	2.1	1.2
10	Trochanterion height (S14)	867.0	21.3 **	44.1	33.1	2.4	1.8
11	Cervicale height (S19)	1446.4	21.2 **	40.5	30.4	1.1	0.8
12	Crotch height	786.4	11.6 **	37.1	14.6	2.5	1.0
13	Sitting height	924.7	6.6 **	20.6	4.3	2.3	0.5
14	Biacromial breadth (S35)	399.1	40.4 **	11.4	16.7	3.5	5.1
15	Bideltoid breadth (S35b)	456.4	36.8 **	12.4	16.4	1.5	1.9
16	Hip breadth (S42a)	326.2	40.7 **	12.2	18.0	2.8	4.2
17	Chest depth	210.2	41.3 **	10.0	15.0	2.0	3.0
18	Median chest depth	191.8	17.1 **	12.9	7.7	3.2	1.9
19	Maximum lower leg breadth	120.0	21.5 **	3.2	2.4	3.4	2.6
20	Upper arm length (S47)	308.7	34.4 **	9.4	11.6	3.3	4.1
21	Arm reach from back	824.8	3.8 **	98.8	10.2	7.6	0.8
22	Bicondylar humerus (S52(3))	66.3	13.2 **	1.8	0.8	3.8	1.7
23	Neck circumference	364.4	14.9 **	17.9	9.2	2.6	1.3
24	Chest circumference	892.1	23.7 **	73.2	61.6	1.2	1.0
25	Shoulder circumference	1101.8	3.3 **	108.8	9.2	1.8	0.1
26	Minimum abdominal c. (S62)	756.0	4.7 **	61.6	8.5	0.6	0.1
27	Buttock c. (S64(1))	918.0	4.9 **	40.4	5.8	0.8	0.1
28	Upper arm c., flexed (S65(1))	297.2	18.3 **	9.8	6.3	1.0	0.6
29	Calf circumference (S69)	367.4	6.8 **	16.2	3.5	2.1	0.5
30	Buttock-knee length, sitting	569.4	5.7 **	24.7	4.3	3.0	0.5
31	Subscapular skinfold thickness	13.6	10.1 **	3.2	1.1	4.8	1.6
32	Medial calf skinfold thickness	8.3	11.7 **	0.4	0.2	2.7	1.1

(V*s*), interobserver error variance (V*o*), and random error variance (V*e*). The percentage of each error variance to the total variance was also calculated. They will be referred to as percentage interobserver error variance (%V*o*) and percentage random error variance (%V*e*) in the following text. The reliability coefficient (R) and overestimation rates of variance (OER) were also calculated. Assuming that the grand mean of the six observers (GM) is very close to the true mean, the mean absolute bias (MAB), MAB in standard deviation unit, and standard deviation of bias (SDB) were calculated. Excel 4.0 for Macintosh computers was used to conduct ANOVA.

Correlation coefficients between the measures of the error were calculated, in which the square root of variance instead of variance itself was used. By using GM to represent the measurement size, the relation between the measures of the error and the measurement size were also examined. For this purpose, five sets of measurement items were used. All 32 measurement items and measurement items belonging to the following four different size groups were included: 1) GM<200 mm (10 items),

Table 2. (continued)

	Measurement item	R	V*s*	OER	MAB	MAB in SD unit	SDB
1	Head length (C1)	0.945	44.6	5.8	0.66	0.098	0.93
2	Nose breadth (C3)	0.908	7.6	10.1	0.40	0.147	0.50
3	Interpupillary breadth (C12)	0.841	10.6	18.9	0.71	0.218	0.93
4	Morphologic face height (C18)	0.778	28.0	28.5	1.20	0.227	1.64
5	Bitragion chin arc (C52)	0.933	171.9	7.1	1.65	0.126	2.54
6	Occiput to subnasale distance	0.886	53.9	12.9	1.55	0.211	1.78
7	Entocanthion to vertex height	0.566	19.6	76.6	2.21	0.499	2.75
8	Height (S1)	0.996	3927.6	0.4	1.33	0.021	1.78
9	Iliospinale height (S13)	0.967	1685.0	3.5	3.79	0.092	4.74
10	Trochanterion height (S14)	0.958	1770.9	4.4	3.95	0.094	5.89
11	Cervicale height (S19)	0.981	3579.9	2.0	4.54	0.076	5.65
12	Crotch height	0.966	1449.2	3.6	2.60	0.068	3.99
13	Sitting height	0.973	882.8	2.8	1.89	0.064	2.25
14	Biacromial breadth (S35)	0.914	300.1	9.4	3.15	0.182	4.14
15	Bideltoid breadth (S35b)	0.966	816.4	3.5	3.45	0.121	4.11
16	Hip breadth (S42a)	0.930	403.5	7.5	3.34	0.166	4.30
17	Chest depth	0.951	480.8	5.2	3.22	0.147	3.92
18	Median chest depth	0.949	382.3	5.4	2.19	0.112	2.85
19	Maximum lower leg breadth	0.940	88.5	6.4	1.25	0.133	1.60
20	Upper arm length (S47)	0.926	262.2	8.0	2.55	0.157	3.46
21	Arm reach from back	0.916	1191.4	9.1	2.91	0.084	3.72
22	Bicondylar humerus (S52(3))	0.945	44.6	5.8	0.73	0.109	1.03
23	Neck circumference	0.961	661.8	4.1	2.51	0.098	3.14
24	Chest circumference	0.978	5985.0	2.3	6.63	0.086	8.02
25	Shoulder circumference	0.981	6072.0	1.9	2.96	0.038	3.63
26	Minimum abdominal c. (S62)	0.993	10701.9	0.7	2.68	0.026	3.28
27	Buttock c. (S64(1))	0.990	4778.5	1.0	1.80	0.026	2.70
28	Upper arm c., flexed (S65(1))	0.984	969.2	1.7	2.20	0.071	2.58
29	Calf circumference (S69)	0.974	737.0	2.7	1.40	0.052	2.02
30	Buttock-knee length, sitting	0.965	791.0	3.7	1.67	0.060	2.27
31	Subscapular skinfold thickness	0.935	62.6	6.9	0.85	0.107	1.10
32	Medial calf skinfold thickness	0.962	14.8	3.9	0.34	0.088	0.43

Numbers in the parentheses indicate those in Martin and Knussmann (1988).
C denotes cephalic and S somatic measurements. **: significant at the 1% level.
GM: grand mean; V*o*: interobserver error variance; V*e*: random error variance; %V*o*: percentage of interobserver error variance; %V*e*: percentage of random error variance; R: reliability coefficient; V*s*: intersubject variance; OER: overestimation rate of variance (%); MAB: mean absolute bias (mm); SDB: standard deviation of bias.

2) 200 mm≤GM<400 mm (9 items), 3) 400 mm≤GM<900 mm (7 items), and 4) GM≥900 mm (6 items).

RESULTS

Measures of the error

Table 2 shows the results of the second day's experiment. The results of the F-test were signifi-

Table 3. Correlation coefficients between grand mean and measures of the error for the data of the second day.

$\sqrt{Vo}$	0.722	**
$\sqrt{Ve}$	0.492	*
%V*o*	-0.422	*
%V*e*	-0.399	*
R	0.423	*
OER	-0.345	
MAB	0.514	**
MAB in SD unit	-0.507	**
SDB	0.530	**

*:p<0.05; **:p<0.01

cant at the 1% level for all measurement items: The means obtained by six observers were judged not to be equal. The four cephalic measurement items (interpupillary distance, morphologic face height, occiput to subnasale distance, and entocanthion to vertex height) have very low reliability coefficients. Their variances are overestimated by as much as 13% to 77%. Table 2 indicates that R is low in them because both %V*o* and %V*e* are large. They also have large MAB in the SD unit.

Table 3 presents the correlation coefficients between GM and measures of the error. The measurement size is positively correlated with $\sqrt{Vo}$, $\sqrt{Ve}$, MAB, SDB, and R and negatively correlated with %V*o*, %V*e*, and MAB in the SD unit. Larger measurements tend to have larger error variances and biases, but they tend to be more reliable and have relatively smaller biases. When the measurements were divided into size groups, this tendency disappeared in groups with the GM larger than 200 mm. Even in the size group with a GM smaller than 200 mm, the tendency disappeared in %V*o*, %V*e*, R, and MAB in the SD unit.

The correlation coefficient between MAB and SDB was as high as 0.99, and both represent nearly the same thing, the absolute magnitude of the bias. MAB is also significantly correlated with $\sqrt{Vo}$ (r=0.69) and especially with $\sqrt{Ve}$ (r=0.99). On the other hand, R, %V*o*, %V*e*, OER, and MAB in the SD unit are highly correlated with one another (r=0.88 to 0.99).

No tendency was observed that means were especially close in the two trainer-trainee pairs.

Differences between the first and the second day

Figure 2 shows the difference in R and MAB between the two days. In about half of the measurements, neither R nor MAB changed much. In the following nine measurements, total error variance decreased by more than 5% and R increased: interpupillary distance, morphologic face height, occiput to subnasale distance, entocanthion to vertex height, trochanterion height, crotch height, upper arm length, bicondylar humerus, and buttock-knee distance, sitting. It was due to the decreased %V*e* in interpupillary distance, morphologic face height, trochanterion height, upper arm length, and buttock-knee distance, sitting. In fact, %V*o* increased in interpupillary distance. In occiput to subnasale distance and entocanthion to vertex height, it was mainly due to the decreased %V*o*. In crotch height and bicondylar humerus, both error variances decreased and the decrease in %V*e* was more conspicuous.

When interobserver error variance decreased, it was expected that MAB would also decrease. In occiput to subnasale distance, entocanthion to vertex height, and bicondylar humerus in which %V*o* decreased by more than 5%, MAB in the SD unit did decrease on the second day.

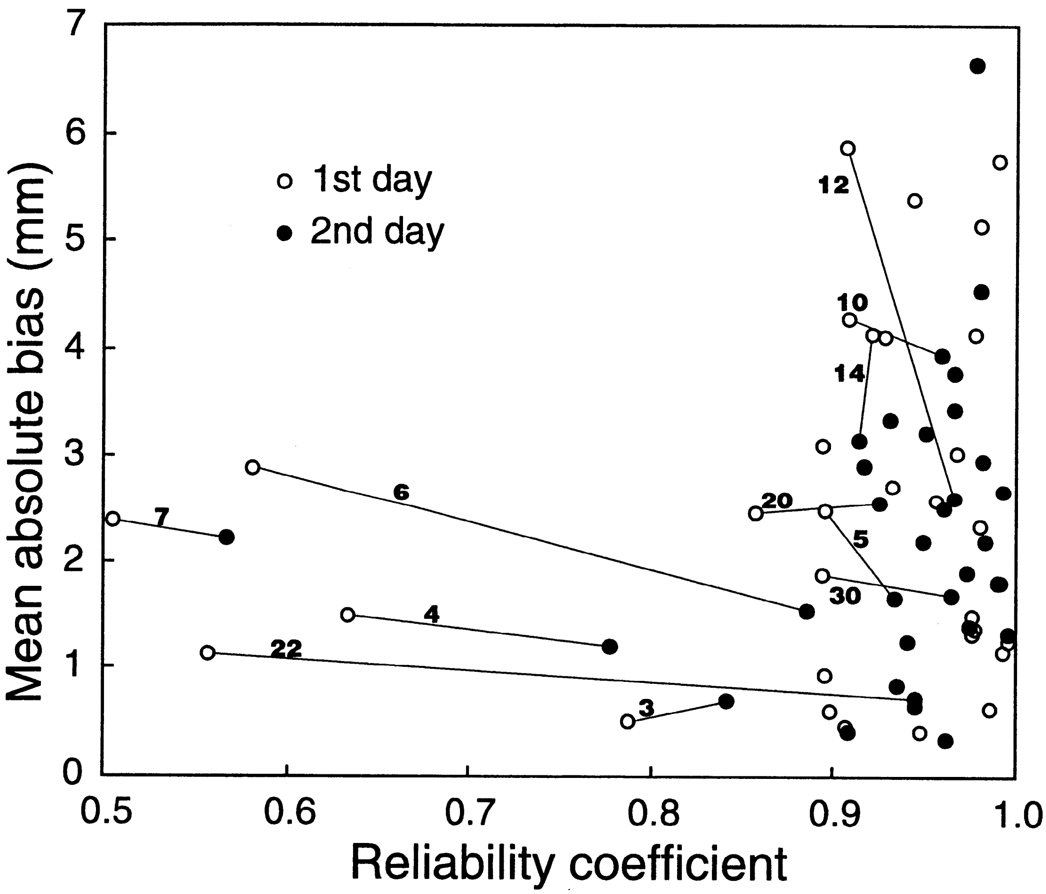

Fig. 2. Comparison of reliability coefficient and mean absolute bias between the two days. The numbers are the same as the item number in Table 2.

DISCUSSION

Factors of errors

When more than one observer was involved, variance was overestimated by 10% to 77% in five measurements, as mentioned in the previous section. They have large %V*o* (7 to 23%) and %V*e* (3 to 21%). Since %V*o* is highly correlated with %V*e* (r=0.89), it is reasonable to assume that the same factor may contribute to both error variances. It should also be noted that the measures of bias have a higher correlation with random error variance than with interobserver error variance.

The measurements with larger error variances are small measurements. In the intraobserver precision, the smaller measurements are less reliable in the size range of 0 to 200 mm (Kouchi et al., in press). This tendency was significant only when all measurements were used in the present analysis, probably because of small sample size. The small measurements are considered to be less reliable because about the same amount of ambiguity exists in locating the landmark irrespective of the measurement size. This ambiguity compels an observer to make his or her own judgment in taking the measurement. For example, the landmark is clearly defined as the midpoint of the pupil in interpupillary distance. However, many other factors that must be determined by the observer influence the actual measurements; instruction to the subject about which way to look; how to locate the tip of the sliding caliper on which part of the face without touching the landmark; and so on. To minimize both intra- and interobserver errors, the measurement technique should not depend on the judgment of each observer.

Cephalic measurements

Although cephalic measurements are generally small, some are especially unreliable. The possible causes are as follows: The tip of the instrument must not touch the landmark itself (interpupillary distance); the subject easily deforms because of the change in expressions and the pressure of the instrument (nose breadth); the landmark is extremely difficult to locate accurately (morphologic face height); and the measurement is seriously influenced by the orientation of the head, which depends not only on how to locate tragion and orbitale, but also on the technique of the observer to position the head of the subject (occiput to subnasale distance and entocanthion to vertex height). In these measurements, we cannot expect to drastically reduce the measurement errors as far as the traditional manual method is adopted. A possible solution is to introduce a different measurement method, such as photogrammetry or three-dimensional measurement without contact. However, even if the different measurement method is introduced, landmarks must be located and marked by a person. Detailed instructions in locating a landmark are necessary to minimize interobserver technique difference. Clear definition is not enough.

Training and practice

In some measurements, the reliability drastically improved on the second day. In these measurements, the cause of the interobserver difference could be specified through the discussions on the results of the first day's experiment.

In occiput to subnasale distance and entocanthion to vertex height, the difference in the technique of positioning the head in the Frankfurt plane caused large interobserver errors. In these measurements, the Frankfurt plane must be parallel to the head board of the head measurement instrument (Fig. 1). Two observers tended to position the subject's head tilting downward. In crotch height, a strong male observer lifted the hard thin plastic sheet much higher than two weak female observers did, and thus gave a higher crotch height. In bicondylar humerus, one observer with small hands had difficulties in touching the subject's medial and lateral condyles at the same time and tended to give different values. When the observers noticed such clear causes, the practice was more efficient and the interobserver errors decreased.

Most of the present observers were experienced anthropometrists. However, some of the measurement items (interpupillary distance, occiput to subnasale distance, entocanthion to vertex height, bicondylar humerus, trochanterion height, buttock-knee distance, sitting, and arm reach from back) and instruments (head measurement apparatus and section paper with a triangle) used in the present study were unfamiliar to them. Furthermore, except for the novice and the trainer pair, the observers practiced the measurements only three hours before the experiment. Practice is more useful than longtime experience in reducing the measurement errors (Gordon and Bradtmiller, 1992). The smaller random errors on the second day are probably due to the practice on the first day. After a longer period of practice, reliability and bias may both be further improved even in the experienced observers. In this sense, the data given in Table 2 are not the minimum errors of experienced anthropometrists.

CONCLUSIONS

Among 32 measurement items investigated in the present study, variance was overestimated by more than 10% in 5 items because of random or interobserver errors or both. Since interobserver and random error variances are highly correlated with each other, it is reasonable to consider that the same factors are responsible for both kinds of errors. Judging from the unreliable measurements, ambiguity in procedures in locating the landmark and taking measurements compels the observer to make his or her own judgment, and this may cause the errors.

To minimize the errors in some cephalic measurements in which landmarks must not be touched or the part deforms because of the pressure applied by the instrument, the introduction of a new measurement method that does not require touching the subject may be useful.

When the cause of the interobserver errors could be specified, the interobserver errors were drastically reduced. A specification of the practical causes is essential in reducing interobserver errors, but it guarantees the standardization only among the members of a particular team. Many measurement items have measurement errors that cannot be ignored even when the definitions of the landmark and the measurement item are clear. To minimize both interobserver and random measurement errors, the standardization of measurement technique should be extended to the details, such as what to instruct the subject, how to locate the landmark, and how to handle the instrument.

REFERENCES

CAMERON, N. (1984) The Measurement of Human Growth. Croom Helm, Sydney.

GORDON, C. and BRADTMILLER, B. (1992) Interobserver error in a large scale anthropometric survey. *Am. J. Hum. Biol.*, **4**:253-263.

HEATHCOTE, G. M. (1981) The magnitude and consequences of measurement error in human craniometry. *Canadian Review of Physical Anthropology*, **3**(1):18-40.

JAMISON, P. L. and WARD, R. E. (1993) Measurement size, precision, and reliability in craniofacial anthropometry: bigger is better. *Am. J. Phys. Anthrop.*, **90**:495-500.

KOUCHI, M. and KOIZUMI, K. (1985) An analysis errors in craniometry. *J. Anthropological Society of Nippon*, **93**(4):409-424.

KOUCHI, M., MOCHIMARU, M., TSUZUKI, K., and YOKOI, T. (1996) Random errors in anthropometry. *J. Human Ergol.*, **25**:155-166.

MARTIN, R. and KNUSSMANN, R. (1988) Anthropologie. Handbuch der vergleichenden Biologie des Menschen, Band I. Gustav Fischer, Stuttgart (*in German*).

MUELLER, W. H. and MARTORELL, R. (1988) Reliability and accuracy of measurement. *In* Anthropometric Standardization Reference Manual, ed. by LOHMAN, T.G., ROCHE, A. F., and MARTORELL, R., Human Kinetic Books, Champaign, Illinois, pp.83-86.

UTERMOHLE, C., ZEGURA, S. L., and HEATHCOTE, G. M. (1983) Multiple observers, humidity, and choice of precision statistics: factors influencing craniometric data quality. *Am. J. Phys. Anthrop.*, **61**:85-95.

WEINER, J. S. and LOURIE, J. A. (1969) Human Biology- A Guide to Field Methods. Blackwell, Oxford.

WILLIAMSON, D. F., KAHN, H. S., WORTHMAN, C. M., BURNETTE, J. C., and RUSSEL, C. M. (1993) Precision of recumbent anthropometry. *Am. J. Hum. Biol.*, **15**:159-167.