Inter and Intra Observer Variability in Anthropometric Measurements

Dahiya YS [*], **Biswal P** [+], **Sharma MD** [#]

Abstract

Background: There are two aspects in measurement errors: the closeness of the measured value to the true value (accuracy) and the closeness of two repeated measurements (precision). Anthropometric data is unique because it is virtually impossible to measure accuracy as there exists no 'True Value' of measurement against which it can be compared. Therefore, there is never an absolute agreement of measurements and certain amount of uncertainty is inherent to the process. This difference in two measured values (which are expected to be same) can only be quantified in terms of 'Observer Variability'. This study was carried out to study the inter and intra observer variability in anthropometric measurement on the IAM anthropometric platform.

Methods: The study was conducted on 07 volunteers, who were part of an Anthropometry Workshop at IAM. The volunteers were divided into two groups. One of the volunteers acted as the subject. Rest of the six volunteers measured four parameters of the subject viz. Sitting Height, Leg Length, Thigh Length and Stature. This was done following the standardized methodology taught to them. The subject was measured by the same six observers the following day for all the four parameters. The inter observer variation was quantified in terms of 'agreement' which depends on the within subject Standard Deviation (SD). The intra observer variation was quantified in terms of 'repeatability' which depends on the between-observer SD. To estimate both these SD, one-way Analysis Of Variance (ANOVA) was used to model the data.

Results: The mean Standard Deviation for Intra & Inter Observer measurements was 4 mm & 2.4 mm respectively for Standing Height. The mean Standard Deviation for Intra & Inter Observer measurements was 1.1 mm & 1.3 mm respectively for Sitting Height. The mean Standard Deviation for Intra & Inter Observer measurements was 6 mm & 2.7 mm respectively for Thigh Length. The mean Standard Deviation for Intra & Intra Observer measurements was 0.2 mm & 2 mm respectively for Leg Length.

Conclusions: An awareness of the quantity and source of observer variability will assist in implementing the corrective actions like a structured 'training methods in anthropometry'. The results of the study have reinforced the fact that regular training of the personnel involved in performing anthropometry in IAF will certainly aid in limiting the variability within the expected range.

Introduction

No measured data are free from measurement errors. There are two aspects in *Measurement Errors* viz. closeness of the measured value to the true value (Accuracy) and the closeness of two repeated measurements (Precision). Anthropometric data is unique (in the sense that), it is virtually impossible to measure accuracy. This difficulty arises from the fact that the 'true value' of a measurement is unknown. There exists no Gold Standard of measurement against which other methods of anthropometric measurements can be compared. Therefore, there is never an absolute agreement of measurements and certain amount of uncertainty is inherent to the process. This difference in

[*] *Classified Specialist (Aerospace Medicine) & HoD, Department of Human Engineering, IAM, IAF.*

[+] *Classified Specialist (Aerospace Medicine), Department of Human Engineering, IAM, IAF.*

[#] *Classified Specialist (Aerospace Medicine), Department of Acceleration Physiology & Spatial Orientation, IAM, IAF.*

two measured values which are expected to be 'same' can only be quantified in terms of **Observer Variability**. This study was carried out to study the inter and intra observer variability in anthropometric measurements carried out on the IAM Anthropometry Platform.

Need for the Study

The Department of Human Engineering at Institute of Aerospace Medicine, Indian Air Force (IAF) is one of the major centres of the IAF where anthropometric measurements are carried out for the candidates and the aircrew of various streams of the IAF and the Indian Navy (IN).

Candidates once declared 'Unfit' for IAF or IN on any ground, including inability to meet the laid down anthropometry standards, have a right to appeal against the decision of the selection medical board. In such cases, these anthropometry parameters are measured once again by a different observer and on a different anthro platform. Differences do arise in such measurements, which could be due to observer variability. Hence, a need was felt, to quantify this difference and confidently attribute it to observer variation whether than on any procedural or instrument error.

A similar situation is created, when the pilot cadets undergo anthropometry before being assigned aircraft stream viz. fighter, helicopter or transport (called trifurcation board) at No. 2 Aeromedical Training Centre at Air Force Academy. An error in the observation at this stage may result in a scenario, where a pilot cadet is assigned to a particular aircraft stream based on his anthropometry, whereas, in actuality, he may be anthropometrically incompatible.

The values of inter and intra observer variability have been brought out in the multicentric IAF Anthropometry Survey (2013) report [1]. However, the methodology of determining such values is not highlighted in the project report.

Considering the methodology, equipment and the training to be uniform across all the centres, these differences in measured values need to be quantified. The sources of such differences need to be identified and remedial

measures to minimize the extent of such variation need to be deliberated upon. With this background, a study was designed, to evaluate this observer variability.

Aim

To quantify the inter and intra observer variability in measurement of the four critical anthropometry parameters for aircrew selection i.e. Sitting Height, Leg Length, Thigh Length and Stature using the standard methods on IAM Anthropometry Platform.

Objectives

The study had the following objectives:-

1. To measure the four critical anthropometry parameters on the same subject by six observers under standard conditions using standard methodology on the IAM Anthropometry Platform.

2. To measure four critical anthropometry parameters on the same subject measured on two different occasions by the same observer under standard conditions using standard methodology on the IAM Anthropometry Platform.

3. To identify deviations from standard protocols (if any).

4. To quantify inter and intra observer variation.

5. To evaluate the causes for the variation and suggest remedial measure for each of the four parameters.

Methodology

Study Design. The study was conducted with a repeated measure design.

Parameters recorded. Sitting Height, Leg Length, Thigh Length and Stature.

Subjects. The study was conducted on seven volunteers who were part of an Anthropometry Workshop at IAM. The composition of the volunteers was 04 Medical

Officers and 03 Medical Assistants posted at the 04 bases of the IAF involved in the anthropometric measurements of candidates/UT pilots/ aircrew. The observers and the measured subjects were chosen randomly from within this group.

Methodology. The study was conducted in the following manner:-

1. Training. The volunteers underwent two hours of didactic lectures followed by one hour of demonstration and hands-on practice on standard methods of anthropometric measurements as per Section II of IAP 4303 4th ed.

2. Inter observer Variation. The volunteers were divided into two groups. One of the volunteers acted as the subject. Rest of the six volunteers measured the four parameters of Sitting Height, Leg Length, Thigh Length and Stature of this subject. This was done following the standardized methodology taught to them. Thus, the subject had six readings for each of his four parameters taken by six different observers. These measurements were used for calculating the inter observer variation.

3. Intra observer Variation. The subject was measured by the same six observers the following day for all the four parameters. Thus the subject had two values for each parameter as measured by the same observer on two different days. These measurements were used to calculate the intra observer variation.

Analysis

The inter observer variation has been quantified in terms of '**agreement**' which depends on the 'within subject' Standard Deviation (SD). The intra observer variation has been quantified in terms of '**repeatability**' which depends on the 'between observer' SD [2]. To estimate both these SDs, one way Analysis Of Variance (ANOVA) was used to model the data. *ANOVA partitions variability in the data into that, which can be ascribed to differences between groups and that remaining to within groups. Using the ANOVA model, thus gives us estimates of between-subject (repeatability) and within-subject (agreement) SDs.* A further comparison was made with the available inter and intra observer variation given in the IAF Anthropometry Survey report (2013).

In addition, two most commonly used measures of precision, the **Technical Error of Measurement (TEM)** and the **Coefficient of Reliability (R)** were also used for analysis of intra observer variability. TEM and R can provide most of the information needed to determine whether a series of anthropometric measurements can be considered 'precise' or not [4].

The TEM is the most commonly used measure of precision, which is the square root of measurement error variance. TEM was calculated with the following formula, where $\sum d^2$ is the summation of deviations raised to the second power and N is the number of volunteers measured.

$$TEM = \sqrt{(\Sigma d^2/2N)}$$

Lower the TEM obtained, better is the precision of the appraisers to perform the measurement.

The Coefficient of Reliability (R) is calculated as percentage with the following equation, where SD^2 is the total intra-subject variance for the study, including measurement error.

$$R = 1 - \left(\frac{TEM^2}{SD^2}\right)$$

This coefficient shows the proportion of between subject variance free from measurement error. Scores can range from 0 to 1, where a value of 0 indicates that all between-subject variation was due to measurement error and a value of 1 indicates that no measurement error was present. Thus, higher R values indicate greater measurement precision.

Results

The Tables 1 to 4 present the ANOVA results depicting the inter and intra observer variation in measurement for each of the four parameters of Sitting Height, Leg Length, Thigh Length and Standing Height.

Table 1. Intra and Inter Observer Variation in Standing Height as per the current study and the IAF Anthropometry Survey (2013)

Source of Variation	Variance	df	Mean Variance	F	P-value	F crit	SD= $\sqrt{Variance}$	2013 Survey
					Standing height			
					ANOVA			
Intra Obs	0.213	1	0.213	3.535	0.089	4.964	**0.462**	**0±0.2**
Inter Obs	0.603	10	0.060				**0.246**	**-0.003±0.173**

Table 2. Intra and Inter Observer Variation in Sitting Height as per the current study and the IAF Anthropometry Survey (2013)

Source of Variation	Variance	df	Mean Variance	F	P-value	F crit	SD= $\sqrt{Variance}$	2013 Survey
					Sitting Height			
					ANOVA			
Intra Obs	0.013	1	0.013	0.727	0.414	4.965	**0.115**	**-0.027±0.479**
Inter Obs	0.183	10	0.018				**0.135**	**0.26±0.33**

Table 3. Intra and Inter Observer Variation in Thigh Length as per the current study and the IAF Anthropometry Survey (2013)

Source of Variation	Variance	df	Mean Variance	F	P-value	F crit	SD= $\sqrt{Variance}$	2013 Survey
					Thigh Length			
					ANOVA			
Intra Obs	0.368	1	0.368	5	0.049	4.964	**0.606**	**0.218±0.905**
Inter Obs	0.735	10	0.073				0.271	**-0.063±0.58**

Table 4. Intra and Inter Observer Variation in Leg Length as per the current study and the IAF Anthropometry Survey (2013)

Source of Variation	Variance	df	Mean Variance	F	P-value	F crit	SD= $\sqrt{Variance}$	2013 Survey
					Leg length			
					ANOVA			
Intra Obs	0.0008	1	0.0008	0.01	0.91	4.96	**0.029**	**0.272±0.558**
Inter Obs	0.6216	10	0.0621				0.249	**0.071±0.518**

Discussion

There are two aspects of measurement errors i.e. the closeness of the measured value to the 'true value' (Accuracy) and the closeness of two repeated measurements (Precision).

In anthropometry, accuracy is difficult to measure as there exists no Gold Standard of measurement which can give the 'true value' of any anthropometric parameter [3]. Each method of measurement in a dynamic biological system introduces its own error into the system. Thus, various methods are used to estimate the 'true value' in such a circumstance.

One of the simplest ways to measure the true value is to use the 'mean' value, out of a number of measurements. This method stands to statistical logic, as the measured value

would have some amount of error. If this is due to a type 2 error (methodology/standardization) and is random, then the error value will follow a normal distribution. Thus, some deviation from the true value would be negative and some would be positive. If sufficiently large number of readings are taken (i.e. 'n' is large enough), the plot of these errors would follow a normal curve, the mean error would approximate to zero and thus the mean, so calculated, would approximate to the 'true mean'. Using this method, a mean standard deviation and a confidence interval of deviation from the mean can be calculated. A study of the inter and intra observer variation given below each table of the IAF Anthropometry Survey (2013) mentions an error value and a confidence interval. However, the methodology does not mention how such values were arrived at. It is assumed that these values are mean SD and the 95% confidence interval of the SD [1].

Considering the above, it is a common practice, therefore, to approach measurement errors by quantifying 'precision'. In this case, the equipment and the method of measurement are standardized by using the best available practices. However, no attempt is made at identifying the true value. The focus of measurement is, to achieve repeatability of measurement such that there is minimal difference in values measured by one observer multiple times (intra observer) or by many observers (inter observer). The underlying principle or philosophy of measurement is that, Type 1 errors, if any, are assumed to be of a fixed magnitude and an inherent limitation of the system, therefore, these errors will be uniform irrespective of the time, place or person taking measurement. The aim is to minimize Type 2 errors by standardization and also to quantify the degree of variation that exists when repeated measurements are taken on the same person.

This study attempted to quantify the precision in anthropometric measurements by observers using the IAM Anthropometry Platform. The objective was to quantify the repeatability and the reproducibility of measurements made by this particular method.

1. Repeatability. This was measured by making at least two measurements on the same subject under identical conditions. This means the measurements were made on different occasions by the same observer on the same equipment using the same methodology at the same time of the day. Thus,

it was assumed that these errors exist due to the measurement process itself. This is also called **intra observer error**.

2. Reproducibility. This was measured by making at least two measurements on the same subject by different observers under identical conditions. It was assumed that these errors exist due to some differences in the way the two observers carry out measurements. This error is also called **inter observer error**.

In this study, ANOVA has been used to quantify the differences within the groups (same subject, same parameter, different observers) and between the groups (same subject, same parameter, same observer, different instance of measurement).

ANOVA is a robust tool for this type of analysis. Not only it gives a mean and SD for each of the within and between groups, it also gives a 'F' statistic which helps in evaluating whether between and within group variances are statistically same or not. This has implications on the interpretation as discussed later.

Intra-observer Variation. The Technical Error of Measurement (TEM) and the Coefficient of Reliability (R) were also used for analysis of the results. These tools provided the information to determine whether the results can be considered precise or not.

Table 5. Technical Error of Measurement (TEM) for four parameters

Parameter	Standing Height	Sitting Height	Thigh Length	Leg Length
Technical Error of Measurement (TEM)	0.29	0.14	0.29	0.08

Lower the TEM, better is the precision of the appraisers to perform the measurement.

The Coefficient of Reliability (R) was calculated as percentage where the scores range from **0 to 1, where 0 indicates that the between-subject variation was due to measurement error** and **1 indicates no measurement error.** Higher R values indicate greater measurement precision.

Table 6. Coefficient of Reliability (R) for four parameters

Parameter	Standing Height	Sitting Height	Thigh Length	Leg Length
Coefficient of Reliability (R)	0.54	0.63	0.29	0.97

The results show that measurement error is highest for 'Thigh Length' and lowest for 'Leg Length'. The cause lies in the measurement technique for these parameters and is discussed in succeeding paragraphs.

In the following paragraphs, findings for each of the four critical parameters are discussed separately in detail.

Standing Height [Table 1]

The mean Standard Deviation for **Intra Observer** measurements is **4 mm**. This implies that for standing height, the same observer taking two reading on the same subject is likely to have a difference of about 4 mm.

The mean Standard Deviation for **Inter Observer** measurements is **2.4 mm**. This implies that when two observers measure the same subject for stature, the difference in reading is likely to be about 2.4 mm.

The F statistic is lower than the F critical. It means that the inter and intra observer variation, as recorded in this study, are not statistically different from each other.

In measurement of a biological system involving soft tissues, the difference of 2.4 to 4 mm is unlikely to be due to procedural errors. In this study, the variation is between 0.2% to 0.1% of the measured value which is 174 cm. The Mean and SD as per the 2013 survey for standing height is also within this range.

The apparent paradox of the intra observer variation being higher than the inter observer variation in this study is probably due to the fact that the intra observer measurements were taken on two different days. The additional difference of less than 2 mm could be due to the time of day and the effects of erect posture on the intervertebral disc height. Although it is not statistically significant, a similar difference is seen in the 2013 survey data as well.

Sitting Height [Table 2]

The mean Standard Deviation for **Intra Observer** measurements is **1.1 mm.** This implies that for sitting height, the same observer taking two reading on the same subject is likely to have a difference of about 1 mm.

The mean Standard Deviation for **Inter Observer** measurements is **1.3 mm.** This implies that when observers measure the same subject for sitting height, the difference in reading is likely to be about 1.3 mm.

The F statistic is lower than the F critical. It means that the inter and intra observer variation recorded in this study is not statistically different from each other.

The variation in this parameter is approximately 1 mm. This difference cannot be deemed to have any significance and should be considered tending towards 'no variation' at all. In this study, the variation is about 0.1% of the measured value of 82 cm.

The Mean and SD as per the 2013 survey for sitting height is seen to be higher than that found in this study. The mean and SD for both inter and intra observer variation is close to 5 mm.

Within the biological system, a 5mm variation cannot be considered significant. However, considering the procedure for measurement to be correct, the following sources of subjectivity leading to variation of up to 5 mm can be considered for this parameter :-

1. The amount of inhalation before holding breath.

2. The positioning of head in the 'Frankfurt Plane', is determined visually. This changes the highest point on the cranium touching the datum probe. The alignment of the inferior margin of the orbit with the superior margin of the external auditory meatus can be achieved using a spirit level.

3. The height of the seat is adjusted visually to keep the thighs parallel to the floor. Identification of the greater trochanter of the femur may be difficult in many cases (due to overlying fat). Improper sitting can lead to a rotation of the hip, thereby bring variation.

4. Amount of pressure applied on the measuring arm of the short arm anthropometer can also cause variation in the readings.

Thigh Length [Table 3]

The mean Standard Deviation for **Intra Observer** measurements is **6 mm.** This implies that for thigh length, the same observer taking two reading on the same subject is likely to have a difference of about 6 mm.

The mean Standard Deviation for **Inter Observer** measurements is **2.7 mm.** This implies that when different observers measure the same subject for thigh length, the difference in reading is likely to be about 2.7 mm. This is approximately half of the intra observer variation.

The F statistic in this case is higher than the F critical, which implies that the inter and intra observer variations as recorded in this study are statistically different from each other. It is likely that even in the general population of observers, such a difference between inter and intra observer measurements would occur.

The variation in this parameter is the highest amongst the four parameters as evaluated in this study. The Mean and SD as per the 2013 survey for thigh length is seen to be even higher than that found in this study. The intra observer variation is more than 10 mm. It is the fact that even in the 2013 survey, the inter observer variation is about 5 mm which is half of the intra observer variation.

As seen in Table 6, the Coefficient of Reliability (R) is 0.29 for Thigh length. This indicate a high component of measurement error. The likely causes of this are discussed below.

Of all the four critical parameters, Thigh length/ buttock knee length has the maximum number of subjective variables to be ensured before taking a reading. The element of a number of observer 'judgement based' positionings add up to be the primary cause for the variation. The following sources of subjectivity leading to variation of up to 6 mm in this study and 10 mm in 2013 survey have been considered for this parameter :-

1. The positioning of the subject involves the adjustment of the seat height in such a manner that the thigh is parallel to the ground. This subjective assessment of the thigh positioning may lead to rotation of the hip joint, thereby, changing the part of the buttock touching the rear board of the anthropometer.

2. The amount of pressure exerted backwards may vary from measurement to measurement, thus leading to different degrees of compression of the gluteal fat. This can be avoided by the subject pressing himself maximally against the rear board.

3. The leg is to be kept vertical to the ground. If the knee is flexed or extended, then different parts of the knee touch the short anthropometer probe and lead to variation. This can be standardized by using the spirit level to align the head of fibula with the lateral malleolus.

4. The short anthropometer arm is to be kept parallel to the ground while taking this measurement. This is done by first slotting the lower end of the arm in one of the holes in the rear board of the platform. If the slot selection is improper, there is a chance of variation. Use of spirit level to align the arm horizontally or parallel to the ground can minimise this type of error.

All the above sources of variations, however, are unable to explain the higher intra observer variation as compared to inter observer variation both in this study as well as the 2013 survey. It may only be speculated that the procedure of finding inter observer variation entails repeated measurement of the same subject by different observers, usually in the same session. This repeated measurement may introduce a certain degree of standardization of position by the subject leading to a lower variation. When the subject is measured at a different point of time by the same observer, the standardization artificially introduced previously may be lost and thus show greater variation.

Leg Length [Table 4]

The mean Standard Deviation for **Intra Observer** measurements was **0.2 mm**. This implies that for leg length, the same observer taking two reading on the same subject is likely to have a difference of about 0.2 mm.

The mean Standard Deviation for **Inter Observer** measurements is **2 mm**. This implies that when observers measure the same subject for leg length, the difference in reading is likely to be about 2 mm.

The F statistic is lower than the F critical which implies that the inter and intra observer variation as recorded in this study are not statistically different from each other.

The inter and intra observer variation in leg length is found to be low. This is contrary to the expectations of most observers *(as brought out by them during this study)*. The notion that most of the observers had was, the difficulty in carrying out this measurement, particularly in subjects with high hamstrings muscle tone or low hamstring flexibility. In such cases, straightening the legs caused forward shifting of the buttocks thereby creating a gap between the buttocks and the rear board of the anthropometer, which is considered as a source of error and possible variation in the measurments by most observers.

The Mean and SD as per the 2013 survey for leg length is 5 and 7 mm respectively. This is greater than this study, but still it is low as compared to that of thigh length.

As seen in Table 6, the Coefficient of Reliability of this parameter is 0.9. This implies that there is very little measurement error for this parameter. The cause for this could be in the measurement protocol. In this, the amount of standardization using subjective estimates is 'nil'. There is no adjustment of seat, no estimation of vertical or horizontal limbs or body segments. Once the subject is seated with his legs partially flexed against the rear board of the anthropometry platform, the leg is straightened against resistance. Once the lower limbs are straight and the knees locked, the reading is taken with the resistance still being applied. Thus, the subject is pushed maximally against the rear board of the anthropometric platform. There can be no variation in the position/degree of flexion of the knee. The only variation that may occur is in terms of the 'gap' at the back. However, for a given subject, the amount of flexion that can occur at the hip with force being applied at the legs is fixed. Thus, irrespective of the observer, once the force is applied at the foot end, the gap at the back decreases to a fixed value (for that individual). This is possibly the reason for such low variation in inter and intra observer measurements and the high coefficient of reliability (R).

Summary

Anthropometry data is unique, as it is virtually impossible to measure accurately in the absence of the 'true value' of the measurement. Therefore, the closeness of repeat measurements (called 'precision') is a better parameter to ascertain measurement errors. This difference in two measured values (which are expected to be 'same') can be quantified in terms of *Observer Variability*. This study was carried out to study the inter and intra observer variability in four critical anthropometric parameter measurements carried out using the IAM Anthropometry Platform. Considering that the methodology, training and equipment are uniform, the observer variability was quantified and measures have been suggested to reduce it.

The inter and intra observer variability was maximum while measuring thigh length (2.7 and 6 mm respectively). This has been attributed to a series of observers' 'judgment-based' actions required for the measurement of thigh length. The best way to address this variability is training in the methods of anthropometry.

The measurement of leg length on the IAM Anthropometric Platform, which was considered to have maximum errors was negated during the study. The inter and intra observer variability was only 2.0 and 0.2 mm respectively.

Conclusion

An awareness in to the quantity and source of observer variability has assisted in implementing the corrective actions, in form of a structured 'training in methods of anthropometry'. The results of the study have reinforced the fact that regular training of the personnel involved in performing anthropometry in IAF will certainly aid in limiting the variability within the expected range.

The previous attempts in defining the limits for error also did not result in a policy change [5]. However, it is recommended that considering the fact that inter or intra observer variability can not be reduced to zero, limiting them to the range should be ensured.

References

1. IAF Aircrew anthropometric survey (2013) report: Institute of Aerospace Medicine, IAF, Bangalore; 2013.

2. Arroyo M, Freire M, Ansotegui L, Rocandio AM. Intraobserver error associated with anthropometric measurements made by dietitians. Nutr Hosp. 2010; 25(6):1053–6.

3. Kouchi M, Mochimaru M, Tsuzuki K, Yokoi T. Interobserver errors in anthropometry. J Hum Ergo. 1999; 28:15–24.

4. Zschokke S, Lüdin E. Accuracy : How much is necessary ? Bull Ecol Soc Am. 2001; Oct: 237–43.

5. Tripathi KK. Aeromedical decision making in anthropometry: defining limits of agreement between two observations and setting boundaries of uncertainty. Podium presentation at 55th Annual Conference of ISAM, 2015 Aug 28-30; Bangalore.