Institute of Experimental Morphology, Pathology and Anthropology with Museum
Bulgarian Anatomical Society

Acta morphologica et anthropologica, 23
Sofia ● 2016

Intra- and Interobserver Measurement Error of Linear Measurements on Three-dimensional Computed Tomography Models of Dry Mandibles

D. Toneva[1], S. Nikolova[1], I. Georgiev[2,3], A. Tchorbadjieff[3]

[1] Institute of Experimental Morphology, Pathology and Anthropology with Museum, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
[2] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
[3] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria

The study aimed to establish the precision of linear measurements taken on 3D models of human mandibles created by computed tomography and to compare the measurement error with the one obtained for the corresponding measurements taken directly on the dry mandibles. Ten mandibles were scanned through computed tomography. The polygonal models in STL format were generated using VG Studio Max 2.2 software. Ten linear measurements were taken on both dry mandibles and 3D models. The conventional measurements of the mandibles were taken with a digital caliper and the digital measurements were accomplished on the 3D models using the software Geomagic Verify Viewer. All parameters were measured twice by two observers. The intra- and interobserver measurement error was estimated using the technical error of measurement and the reliability of the mandibular measurements was assessed with the coefficient of reliability. All digital measurements showed acceptable measurement error. According to the coefficients of reliability, most of the digital measurements had values above 0.95, indicating high reliability.

Key words: 3D models, mandible, computed tomography, technical error of measurement, coefficient of reliability.

Introduction

The integration of three-dimensional (3D) technologies in different anthropological fields has led to an increasing utilization of the 3D models in anthropometric studies. The steadily increasing usage of digital measurements inevitably puts the question for their precision and reliability. Precision is a measure of the closeness of repeated measurements of the same quantity [5], and reliability is the degree to which within-subject variability is due to factors other than measurement error [14]. Since measurements are not free of error due to instrument imprecision and human inconsistencies [5], the determination of the measurement error appears to be an important component of the anthropometric studies [15].

The most commonly used measures of precision are the technical error of measurement (TEM) and the coefficient of reliability (R). The use of these two errors estimates contributes sufficiently to the determination of the precision of a series of anthropometric measurements [15]. The TEM is the variability encountered between dimensions when the same specimens are measured multiple times by the same observer (intraobserver TEM) or by two or more observers (interobserver TEM) [5, 14]. The lower variability between repeated measurements indicates greater precision [3]. Furthermore, the coefficient of reliability (R) reveals the proportion of between-subject variance, which is free from measurement error [15].

This study aimed to establish the precision of linear measurements taken on 3D models of human mandibles created by computed tomography (CT) and to compare the measurement error with the one obtained for the corresponding measurements taken directly on the dry mandibles.

Materials and Methods

The study was conducted on a sample of ten mandibles from the Military Mausoleum with Ossuary, National Museum of Military History, Bulgaria. The bones belonged to adult male individuals, who served and died during the wars from the beginning of the 20th century.

To obtain 3D models, the mandibles were scanned through computed tomography. The CT scanning was performed on a Nikon Metrology XT H 225 system with reflection head and a voltage of 85 kV with a power of 8.1 W and 95 μA tube current. To generate a 3D CT volume, a series of sequential 2D X-ray images (projections) were captured as the object was rotated through 360°. For each scan 3000 projections were registered, where each projection was taken with an exposure time of 500 ms. The images were then reconstructed to generate a 3D volumetric representation of the object with voxel size of 86 μm. The polygonal models in STL format were generated from voxel data by surface determination and surface extraction using VG Studio Max 2.2 software.

Ten linear measurements between definite landmarks (**Fig. 1**), described according to Martin and Saller [8], were taken on both dry mandibles and 3D models (**Table 1**). The

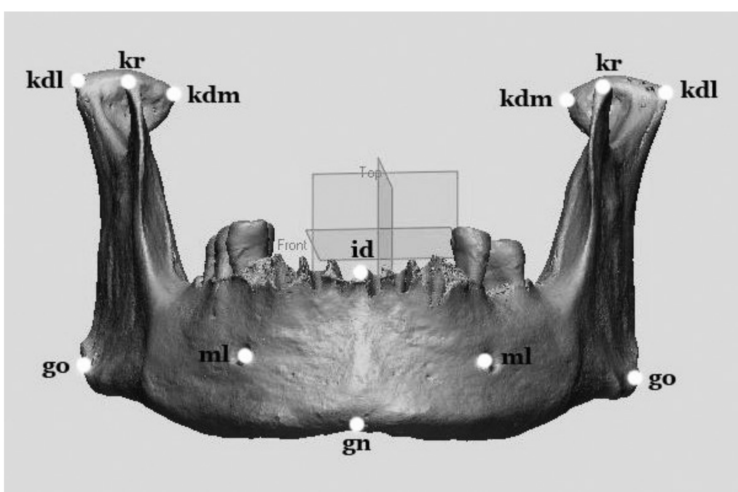

Fig. 1. Location of the landmarks used in the study

Table 1. Mandibular measurements

Measurements		Definition
M1	kdl-kdl	The direct distance between the left and right kondilion laterale (kdl)*.
M2	kdm-kdm	The direct distance between the left and right kondilion mediale (kdm).
M3	kr-kr	The direct distance between the left and right koronion (kr).
M4	go-go	The direct distance between the left and right gonion (go).
M5	ml-ml	The direct distance between both landmarks mentale (ml).
M6	id-gn	The direct distance from infradentale (id) to gnation (gn).
M7	kdl-kdm (L)	The direct distance between left kondilion laterale and left kondilion mediale.
M8	id-kr (L)	The direct distance from infradentale to the left koronion.
M9	id-kdl (L)	The direct distance from infradentale to the left kondilion laterale.
M10	id-kdm (L)	The direct distance from infradentale to the left kondilion mediale.

*Landmarks' abbreviations are written according to Martin and Saller [8].
(L) – left

conventional measurements of the mandibles were taken with a digital caliper (Würth, Germany) with a measuring unit of 0.01 mm and accuracy to 0.03 mm. The digital measurements were accomplished on the 3D models using the free software Geomagic Verify Viewer (3D Systems, Inc).

All parameters were measured twice by two observers. The 1st and 2nd measurements of all samples were taken on separate days.

Statistics

The intra- and interobserver measurement error was estimated using TEM and the reliability of the mandibular measurements was assessed with R. The intraobserver measurement error was computed for each parameter using the duplicate measurements taken by each observer. The interobserver measurement error was calculated based on the values of repeated measurements of both observers.

The absolute TEM was calculated by the following equitation [15]:

$$TEM = \sqrt{\left(\sum D^2\right)/2N} \ ,$$

where $\sum D^2$ is the sum of the deviations raised to the second degree and N is the number of the measured specimens.

However, the absolute TEM depends on the magnitude of the measurements and thus, the different parameters are not comparable. Therefore, the absolute TEM was transformed into relative TEM (%TEM or rTEM) [14]:

$$\%TEM = (TEM/mean) \times 100 \ .$$

For comparison of different studies with more than one observer involved, the total TEM was computed in the following way:

$$total\ TEM = \sqrt{(((TEM_{(intra_1)})^2 + TEM_{(intra_2)}{}^2)/2) + TEM_{(inter)}{}^2)},$$

where TEM(intra$_1$) is the intraobserver TEM for the first observer, TEM(intra$_2$) is the intraobserver TEM for the second observer, and TEM(inter) is the interobserver TEM between the two of them [14]. The relative total TEM (% total TEM) was obtained using the equation: % total TEM = ((total TEM)/mean) × 100.

According to the calculations of the total TEM, the measurement error of the digital and direct mandibular measurements was classified as acceptable and non-acceptable. A threshold of 5% was chosen as an admissible cut-off in accordance with the osteometric studies of Richard et al. [11], Franklin et al. [2], Lottering et al. [7].

The coefficient of reliability (R) was calculated from the total TEM [14]:

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

where SD^2 is the total between subject variance, including measurement error. The R values can range from 0 (all between-subject variation was due to measurement error) to 1 (no measurement error). Values of R greater than 0.95 indicate high reliability of the measurements with more than 95% of the variance due to factors other than measurement error [14].

Comparisons between the TEMs of both observers for digital and direct measurements as well as between the TEMs of the digital and conventional measurements of each observer were performed using a paired t-test.

Results

The basic descriptive statistics of the measurements of both observers are presented in **Table 2**. The values obtained for the intra- and interobserver absolute TEM, %TEM, total TEM and R are given in **Tables 3** and **4**.

Intraobserver TEM and %TEM

The intraobserver absolute TEMs for the **digital measurements** ranged within 0.26–0.60 mm for the first observer and 0.26-0.59 for the second one. The values of the %TEM ranged up to 2.18% and 1.34%, respectively. As can be seen from **Table 3**, the most affected by measurement error was the measurement with smallest magnitude (M7). The mean %TEMs for the digital measurements were 0.80% for the first and 0.70% for the second observer.

The TEMs observed for the **direct measurements** were lower than these for the digital ones (**Table 4**). The intraobserver absolute TEMs for both observers ranged within 0.17-0.53 mm. The largest values of the %TEM reached up to 1.04% for the first and 0.95% for the second observer. The most precise measurement was M1 and the least one – M7. The mean %TEMs for the direct measurements of both observers were very close – 0.48% and 0.47%, respectively.

Table 2. *Means* and *SD* of the two measurement sets of both observers. Deviations (*D*) between the measurements of each observer or both observers

Measurements	Observer I			Observer II			Interobserver *D*
	Trial 1	Trial 2	$D^2 \pm SD$	Trial 1	Trial 2	$D^2 \pm SD$	$D^2 \pm SD$
	Mean ± SD	*Mean ± SD*		*Mean ± SD*	*Mean ± SD*		
Digital measurements							
M1	115.84±7.57	116.06±7.46	0.13±0.18	116.52±7.23	116.21±7.36	0.35±0.52	0.31±0.35
M2	80.54±6.08	80.43±6.11	0.39±0.60	80.11±6.08	80.24±5.88	0.26±0.36	0.17±0.23
M3	92.90±3.95	92.72±4.12	0.32±0.27	93.44±3.93	93.29±3.71	0.61±0.79	0.84±0.88
M4	99.42±5.06	99.72±5.00	0.41±0.37	99.28±4.62	99.08±4.32	0.70±1.01	1.02±1.42
M5	44.29±2.93	44.65±2.64	0.60±0.85	44.98±2.53	44.98±2.92	0.45±0.48	0.38±0.50
M6	31.06±2.27	31.01±2.20	0.24±0.32	30.85±2.14	30.87±2.29	0.27±0.38	0.36±0.31
M7	18.52±2.26	18.49±2.37	0.33±0.29	19.06±2.16	19.09±2.36	0.13±0.22	0.44±0.43
M8	82.13±3.31	81.99±4.00	0.68±0.69	82.97±3.77	82.96±3.62	0.50±0.52	1.16±1.21
M9	107.06±2.66	107.29±2.34	0.71±0.72	107.93±2.84	107.96±2.59	0.52±1.21	1.38±1.66
M10	103.15±2.42	103.08±2.31	0.69±0.98	103.24±1.98	103.22±2.07	0.20±0.28	0.72±0.87
Direct measurements							
M1	116.82±7.33	116.84±7.42	0.06±0.05	117.06±7.36	117.14±7.40	0.06±0.08	0.11±0.14
M2	79.41±6.22	79.54±6.20	0.39±0.43	79.85±6.09	80.02±5.89	0.37±0.38	0.39±0.48
M3	92.51±4.54	92.54±4.43	0.15±0.13	92.37±4.66	92.38±4.47	0.55±0.68	0.29±0.33
M4	100.56±4.72	100.40±4.86	0.18±0.27	100.50±4.86	100.46±4.86	0.07±0.10	0.10±0.18
M5	44.98±2.66	44.85±2.59	0.12±0.14	44.82±2,59	44.89±2.71	0.07±0.10	0.11±0.10
M6	32.71±2.01	32.74±1,90	0.11±0.13	32.62±1,77	32.53±1.88	0.10±0.11	0.17±0.41
M7	19.41±2.57	19.50±2.43	0.08±0.17	19.58±2.48	19.63±2.24	0.07±0.10	0.11±0.18
M8	82.66±4.29	82.94±3.91	0.33±0.53	82.52±4.00	82.56±4.34	0.42±0.41	0.49±0.76
M9	107.59±2.97	107.57±2.87	0.29±0.27	107.38±3.16	107.42±3.00	0.11±0.19	0.25±0.28
M10	103.52±2.52	103.59±2.35	0.29±0.33	103.47±2.33	103.47±2.39	0.34±0.51	0.32±0.42

Table 3. TEMs and R of the digital mandibular measurements

Measurements	Intraobserver TEM				Interobserver TEM		Total TEM		R
	Observer I		Observer II						
	Absolute	%	Absolute	%	Absolute	%	Absolute	%	
M1	0.26	0.22	0.42	0.36	0.39	0.34	0.37	0.32	0.997
M2	0.44	0.55	0.36	0.45	0.29	0.36	0.35	0.44	0.996
M3	0.40	0.43	0.55	0.59	0.65	0.69	0.57	0.61	0.978
M4	0.45	0.45	0.59	0.60	0.71	0.72	0.63	0.63	0.982
M5	0.55	1.24	0.47	1.05	0.44	0.98	0.48	1.06	0.968
M6	0.34	1.11	0.37	1.20	0.42	1.36	0.39	1.26	0.967
M7	0.40	2.18	0.26	1.34	0.47	2.51	0.41	2.18	0.966
M8	0.59	0.71	0.50	0.60	0.76	0.92	0.66	0.80	0.966
M9	0.60	0.56	0.51	0.47	0.83	0.77	0.71	0.66	0.923
M10	0.59	0.57	0.32	0.31	0.60	0.58	0.54	0.52	0.935

Table 4. TEMs and R of the direct mandibular measurements

Measurements	Intraobserver TEM				Interobserver TEM		Total TEM		R
	Observer I		Observer II						
	Absolute	%	Absolute	%	Absolute	%	Absolute	%	
M1	0.17	0.15	0.18	0.15	0.24	0.20	0.32	0.27	0.998
M2	0.44	0.56	0.43	0.54	0.44	0.56	0.69	0.87	0.986
M3	0.28	0.30	0.53	0.57	0.38	0.41	0.60	0.65	0.981
M4	0.30	0.30	0.18	0.18	0.23	0.23	0.40	0.40	0.993
M5	0.24	0.54	0.19	0.41	0.24	0.53	0.36	0.81	0.980
M6	0.23	0.70	0.22	0.68	0.29	0.90	0.40	1.24	0.951
M7	0.20	1.04	0.19	0.95	0.23	1.18	0.33	1.70	0.980
M8	0.41	0.49	0.46	0.55	0.49	0.60	0.72	0.87	0.968
M9	0.38	0.35	0.24	0.22	0.36	0.33	0.55	0.51	0.965
M10	0.38	0.37	0.41	0.40	0.40	0.39	0.62	0.60	0.927

Interobserver TEM and %TEM

The interobserver TEMs for the **digital measurements** were similar or slightly greater than the intraobserver TEMs, varying from 0.29 mm to 0.83 mm (**Table 3**). The %TEMs ranged from 0.34% to 2.51%. The largest interobserver %TEM was also observed for the M7. The mean interobserver %TEM obtained by all 10 digital measurements was 0.92%.

The interobserver TEM for the **direct measurements** varied from 0.23 mm to 0.49 mm (**Table 4**). The %TEM values were up to 1.18%. The mean %TEM for the direct measurements was 0.53%, which is approximately twice less than the mean %TEM established for the digital ones.

The mean interobserver TEMs did not exceed substantially the intraobserver TEMs. The intra-method differences between intra- and interobserver TEM and %TEM values were 0.11 mm and 0.17% for the digital measurements and 0.03 mm and 0.05% for the direct ones (calculations based on the averaged mean intraobserver TEM and %TEM of both observers).

Total TEM

All digital and direct measurements showed total TEMs below the 5% threshold, so the measurement errors were classified as acceptable. The largest total %TEM concerned the M7 and amounted to 2.18% for the digital and 1.70% for the direct measuring method.

R of digital and direct measurements

The coefficients of reliability showed that almost all **digital measurements** had values above 0.95, except for two measurements (M9 and M10), whose R values were below this threshold (**Table 3**). This result was indicative for low imprecision and demonstrated good repeatability of the mandibular measurements, especially the most used ones (M1-M7). On the other hand, only one of the **direct measurements** (M10) showed excessive variability ($R < 0.95$).

The most reliable measurements ($R > 0.99$) obtained on 3D models were M1 and M2. Concerning the direct measuring method, the most precise ones were M1 and M4.

Interobserver and inter-method comparisons of TEMs

The paired t-test did not establish statistically significant differences between the TEMs obtained by both observers ($p > 0.05$), which means that measurements produced by both observers were consistent. However, statistically significant differences were observed between the measurement errors of both measuring methods ($p < 0.05$). This was most likely due to the more consistent direct measurements performed by the observers compared to the digital ones.

Discussion

The precision of the digital measurements acquired from 3D CT bone models and the comparison of the obtained measurement error with the corresponding one of the direct bone measurements are very important topics for all anthropological subfields dealing with 3D models, especially forensic anthropology and paleoanthropology. Anthropometric measurement error is unavoidable, but could be minimized considering all aspects of the data collection process [14]. According to Harris and Smith [5], the landmark location is the major source of variability, since it depends mostly on the human judgement.

In previous studies using intraclass correlation coefficient for assessment of the reliability [12, 13], the mandibular measurements obtained by cone beam CT (CBCT) imaging were revealed to be highly reliable and reproducible. In our study, the digital measurements were assessed as highly precise based on the TEM widely used in anthropometry. According to the assumed 5% threshold of the total TEM, the measurement errors of all mandibular measurements from both measuring techniques were classified as acceptable. However, in some anthropometric studies, the acceptable levels for intra- and interobserver %TEM of the linear measurements were set at 1.5% and 2% for beginning anthropometrists and at 1% and 1.5% for skillful anthropometrists [9]. Thus, taking into account the acceptable level for intraobserver measurement error, the three digital measurements with smallest magnitude (M5, M6, and M7) were above the limit of 1% for both observers. According to the direct measurements, the M7 from the measurements of the first observer was the only one with %TEM over 1%, which illustrated again the least precision observed for this measurement. The interobserver %TEM, respectively, was above 1.5% for M6 and M7 from the digital measurements. Considering the conventional measuring method, there was no one measurement to pass the limit of 1.5% for interobserver measurement error. The interobserver absolute TEMs and %TEMs ranged within or slightly exceeded the intraobserver values. However, it has been noticed that the interobserver TEM was more susceptible to error [3].

The results of R showed that most of the digital and direct measurements had $R > 0.95$. Such values above 0.95 have been reported as indicative of good quality control [4]. However, in some studies, it has also been used a cut-off of 0.90 [2] or 0.75 [6, 10]. In case of accepting these thresholds, all digital and direct measurements in our study passed them.

Comparing the %TEM of both measuring techniques, we established lower %TEM values for direct bone measurements compared to the digital ones. Similar results were obtained by Franklin et al. [2]. However, Berco et al. [1] established that the CBCT-derived measurements tended to be more precise than the direct anthropometric standard. Although the direct measurements in our study were more precise, as a whole, the mean differences between both methods were very small (Observer I: TEM – 0.16 mm, %TEM – 0.32%; Observer II: TEM – 0.14 mm, %TEM – 0.23%).

Conclusion

The digital linear measurements taken on 3D models of human mandibles appeared to be highly precise (total TEM < 5%) with minimal measurement error and could be used for anthropometric research purposes. Taking into account the coefficients of reliability, almost all of the measurements indicated high reliability ($R > 0.95$).

Acknowledgement: This study was supported by the "Program for Career Development of Young Scientists, BAS", research grant DFNP – 75/27.04.2016.

The authors would like to acknowledge the kind assistance given by the staff of the National Museum of Military History, Bulgaria.

References

1. **Berco, M., P. H. Rigali, Jr, M. R. Miner, S. DeLuca, N. K. Anderson, L. A. Will.** Accuracy and reliability of linear cephalometric measurements from cone-beam computed tomography scans of a dry human skull. – Am. J. Orthod. Dentofac. Orthop., **136**, 2009, 17.e1-9.
2. **Franklin, D., A. Cardini, A. Flavel, A. Kuliukas, M. K. Marks, R. Hart, C. Oxnard, P. O'Higgins.** Concordance of traditional osteometric and volume-rendered MSCT interlandmark cranial measurements. – Int. J. Legal. Med., **127**, 2013, 505-520.

3. **Gómez-Cabello, A., G. Vicente-Rodríguez, U. Albers, E. Mata, J. A. Rodriguez-Marroyo, P. R. Olivares, N. Gusi, G. Villa, S. Aznar, M. Gonzalez-Gross, J. A. Casajús, I. Ara.** Harmonization process and reliability assessment of anthropometric measurements in the elderly EXERNET multi-centre study. – PLoS ONE, **7**(7), 2012, e41752.
4. **Goto, R., C. G. N. Mascie-Taylor.** Precision of measurement as a component of human variation. – J. Physiol. Anthropol., **26**, 2007, 253-256.
5. **Harris, E. F., R. N. Smith.** Accounting for measurement error: a critical but often overlooked process. – Arch. Oral. Biol., **54**(Suppl 1), 2009, S107-S117.
6. **Hemy, N.** Stature and sex estimation using anthropometry of feet and footprints in a Western Australian population. – Thesis, Perth, University of Western Australia, 2012.
7. **Lottering, N., D. M. MacGregor, M. D. Barry, M. S. Reynolds, L. S. Gregory.** Introducing standardized protocols for anthropological measurement of virtual subadult crania using computed tomography. – Journal of Forensic Radiology and Imaging, **2**, 2014, 34-38.
8. **Martin, R., K. Saller.** Kraniometrische Technik. – In: Lehrbuch der Anthropologie (Eds R. Martin, K. Saller), Band I, Stuttgart, Gustav Fischer, 1957.
9. **Perini, T. A., G. L. de Oliveira, J. S. Ornelia, F. P. de Oliveira.** Technical error of measurement in anthropometry. – Rev. Bras. Med. Esporte, **11**, 2005, 86-90.
10. **Reynolds, M. E., D. Franklin, M. A. Raymond, I. Dadour.** Bloodstain measurement using computer-fitted theoretical ellipses. – Journal of Forensic Identification, **58**(4), 2008, 469-484.
11. **Richard, A. H., C. L. Parks, K. L. Monson.** Accuracy of Standard Craniometric Measurements Using Multiple Data Formats. – Forensic Sci. Int., **242**, 2014, 177-185.
12. **Tarazona-Álvarez, P., J. Romero-Millán, D. Peñarrocha-Oltra, M. Á. Fuster-Torres, B. Tarazona, M. Peñarrocha-Diago.** Comparative study of mandibular linear measurements obtained by cone beam computed tomography and digital calipers. – J. Clin. Exp. Dent., **6**(3), 2014, e271-e274.
13. **Tomasi, C., E. Bressan, B. Corazza, S. Mazzoleni, E. Stellini, A. Lith.** Reliability and reproducibility of linear mandible measurements with the use of a cone-beam computed tomography and two object inclinations. – Dentomaxillofacial Radiology, **40**, 2011, 244-250.
14. **Ulijaszek, S. J., D. A. Kerr.** Anthropometric measurement error and the assessment of nutritional status. – Br. J. Nutr., **82**, 1999, 165-77.
15. **Ulijaszek, S. J., J. A. Lourie.** Intra- and inter-observer error in anthropometric measurement. – In: Anthropometry: The individual and the population. (Eds S. J. Ulijaszek, C. G. N. Mascie-Taylor), Cambridge, Cambridge University Press, 1994.