Accuracy and reliability of measurements obtained from Computed Tomography 3D volume rendered images

Kyra E. Stull[a], Meredith L. Tise[b], Zabiullah Ali[c], David R. Fowler[c]

[a] Department of Anatomy, University of Pretoria
Private Bag x323, Arcadia, 0084, South Africa
kstullster@gmail.com

[b] Department of Anthropology, University of South Florida
4202 E. Fowler Avenue, SOC 107, Tampa, FL 33620, United States
mtise@mail.usf.edu

[c] Office of the Chief Medical Examiner, State of Maryland
900 W. Baltimore Street, Baltimore, MD 21223, United States
aliz@ocmemd.org; fowlerd@ocmemd.org

Corresponding Author:
Kyra Stull, PhD
Department of Anatomy
University of Pretoria
Private Bag x323, Arcadia, 0084, South Africa
(US) +1 (864) 230-2301
kstullster@gmail.com

Abstract

Forensic pathologists commonly use computed tomography (CT) images to assist in determining the cause and manner of death as well as for mass disaster operations. Even though the design of the CT machine does not inherently produce distortion, most techniques within anthropology rely on metric variables, thus concern exists regarding the accuracy of CT images reflecting an object's true dimensions. Numerous researchers have attempted to validate the use of CT images, however the comparisons have only been conducted on limited elements and/or comparisons were between measurements taken from a dry element and measurements taken from the 3D-CT image of the same dry element.

A full-body CT scan was performed prior to autopsy at the Office of the Chief Medical Examiner for the State of Maryland. Following autopsy, the remains were processed to remove all soft tissues and the skeletal elements were subject to an

additional CT scan. Percent differences and Bland-Altman plots were used to assess the accuracy between osteometric variables obtained from the dry skeletal elements and from CT images with and without soft tissues. An additional seven crania were scanned, measured by three observers, and the reliability was evaluated by technical error of measurement (TEM) and relative technical error of measurement (%TEM).

Average percent differences between the measurements obtained from the three data sources ranged from 1.4% to 2.9%. Bland-Altman plots illustrated the two sets of measurements were generally within 2mm for each comparison between data sources. Intra-observer TEM and %TEM for three observers and all craniometric variables ranged between 0.46 mm and 0.77 mm and 0.56% and 1.06%, respectively. The three-way inter-observer TEM and %TEM for craniometric variables was 2.6 mm and 2.26%, respectively. Variables that yielded high error rates were orbital height, orbital breadth, inter-orbital breadth and parietal chord. Overall, minimal differences were found among the data sources and high accuracy was noted between the observers, which prove CT images are an acceptable source to collect osteometric variables.

1. Introduction

Multiple research studies have suggested the use of three-dimensional (3D) reconstructed computed tomography (CT) scans to provide a means of accurate data collection from the human body, allowing the anthropologist to bypass the need to remove soft tissues [1–4], a process that is time consuming and may conflict with religious beliefs. Furthermore, when skeletal samples are not available to create population specific formula, anthropologists may need to utilize a different source to acquire suitable data. The application of CT has also gained popularity in the forensic pathology community to assist in determining cause and manner of death and in preparation for a mass disaster situation [5–7]. In mass disaster operations and disaster victim identification investigations, forensic investigators frequently need to conduct extensive preparation and processing of remains to obtain data that assists with the

identification process, such as age, ancestry, stature and sex [6,8,9]. The application of CT rather than conventional X-ray allows for better contrast resolution that results in more detailed images of bones and soft tissues and offers a rapid processing time [6]. Furthermore, CT images do not present with distortion as compared to images generated from conventional X-ray machines because of the physics involved with the design of CT scanners.

While the technology implies that measurements collected from skeletal remains on a 3D-CT image with soft tissues would be accurate, a study has yet to be conducted that validates the claim for the entire skeleton. The majority of published studies demonstrate high accuracy between measurements taken from a dry element and measurements taken from the 3D-CT image of the same dry element [3,10–18]. However, the exclusion of soft tissues when imaged inherently produces a CT image that is slightly smaller in three dimensions because of the influence of partial volume effects (PVE) during volume rendering (VR), which occurs when the CT scanner is unable to differentiate between materials with varying Hounsfield units (i.e. air and bone). Several studies compared the measurements obtained from bones within the soft tissues and measurements of the same bones following the removal of soft tissues [1,2,19]. Both Decker et al. [1] and Robinson et al. [2] argue that the virtual models are highly accurate and measurements obtained from CT images can be used in forensic anthropological applications. However, the 95% confidence interval for the measurement difference between CT and dry bone measurements of the lower limb and foot was approximately +/- 5 mm [2], which is considerably larger than the generally accepted level in anthropology. The range of measurement error obtained by Verhoff et al. [19] is between 1 and 2 mm; a result that is more acceptable within forensic applications. Besides validating the measurement error between dry and CT images for the areas that have previously been evaluated, measurement errors need to be recorded for the entire skeleton as almost all elements are used in estimation techniques for ancestry, sex, stature and age.

Measurement error, both because of uncertainty in landmark identification/location and alterations in the objects true dimension as a consequence of imaging, has the potential to drastically affect the interpretation of results, thus this error should be considered in research design [20,21]. Furthermore, the *Daubert* guidelines

emphasize the importance of precision, accuracy and reliability in forensic science research [20,22–24]. Two broad categories are associated with measurement error. The first includes reliability and precision, terms associated with the variation in repeated measures, and the second includes validity and accuracy, terms associated with the extent to which the true value of the object is obtained with the measurement [25–28]. In the current study, the main interest lies in the evaluation of the accuracy of skeletal measures obtained from CT images when soft tissues are present and the reliability of osteometric data collected from CT images. The potential impact of the current study includes the enablement of identification in mass disaster situations because of the elimination of maceration of skeletal remains and the ability to collect metric data from images and the increase in modern, comparative reference collections to create and/or validate current anthropological techniques.

2. Materials and Methods

A full-body CT scan of a deceased individual was performed with a General Electric (GE) Light Speed RT-16 multi-detector scanner prior to autopsy at the Office of the Chief Medical Examiner for the State of Maryland (OCME). Following the OCME protocol, the skull was scanned with a slice thickness of 0.625 and the postcrania was scanned with a slice thickness of 1.25 mm. The acquired images were reconstructed in a contiguous fashion using the GE Advanced CT Workstation (AW-2) (Version: aws-2.0-5.5). A formal consent was obtained from the next of kin through the State Anatomy Board and the remains were processed to extract specific skeletal elements, including the cranium, mandible, left clavicle (the right was damaged during autopsy) and the right and left scapulae, humerii, ulnae, radii, femora, tibiae, and fibulae. Following the removal of all soft tissues, the skeletal remains were processed and the dry elements were re-scanned using the previously described settings. Thus, measurements were obtained from three sources: the dry skeletal elements (dry), the CT images with soft tissues prior to autopsy (CT), and the CT images of the dry skeletal elements after soft tissue removal (dryCT).

An additional seven crania were scanned following the same CT settings in order to evaluate reliability in repeated measurements. Demographics associated with the crania were trivial considering the purpose of the paper was for measurement error and

not estimation of a biological parameter.

Linear cranial (n=35) and postcranial measurements (n = 61) were collected following measurement definitions from Buikstra and Ubelaker [29] and Urcid [30] (Tables 1 and 2). Maxillo-alveolar length, mastoid height, mandibular measurements that required a mandibulometer (i.e. maximum ramus height, mandibular length, and mandibular angle), circumference measurements on the long bones, and pelvic measurements were excluded from data collection. Measurements were excluded because of the difficulty in identifying the landmarks that define the measurement or because of highly unreliable measurements [27,31]. Measurements were taken from both left and right sides to increase the number of available measurement comparisons. Each bone was isolated from all other elements to allow for full visibility of features and landmarks and the measurements were collected using the AW-2 program. Figures 1 and 2 demonstrate cranial and postcranial measurements collected from CT images.

During creation of a VR image, an opacity curve determines the opacity and transparency of various tissues. Landmarks were identified using a preset 3D filter displaying bone in color with a wide opacity ramp set at window/level operation (W/L) of 594/41. This single step was sufficient to identify most landmarks, but was not reliable in identifying the exact location of all anatomical landmarks of interest because of differences in skull thickness or other degenerative changes, such as low bone density (i.e. osteoporosis). This was especially true in identifying landmarks in the ocular and temporal regions. Therefore, the opacity ramp was manually lowered for better visualization purposes (Figure 3). Although the thickness of the bone changes when the opacity is adjusted, the distance between landmarks is unaffected.

The measurements associated with the dry skeletal remains were collected using sliding and spreading calipers and an osteometric board.

Accuracy

One author (MLT) collected all of the cranial and postcranial measurements from the three data sources to reduce error. Percent differences were used as a means to compare the measurements obtained from the dry, CT, and dryCT because the calculation takes the size of the measurement into account. For example, a 2 mm error is drastically

Table 1 – Cranial measurements collected from the three data sources.

Maximum cranial length (GOL)	Biorbital breadth (EKB)
Maximum cranial breadth (XCB)	Interorbital breadth (DKB)
Bizygomatic breadth (ZYB)	Frontal chord (FRC)
Basion-bregma height (BBH)	Parietal chord (PAC)
Cranial base length (BNL)	Occipital chord (OCC)
Basion-prosthion length (BPL)	Foramen magnum length (FOL)
Maxillo-alveolar length (MAB)	Foramen magnum breadth (FOB)
Biauricular breadth (AUB)	Biasterion breadth (ASB)
Upper facial height (UFHT)	Zygomaxillary breadth (ZMB)
Minimum frontal breadth (WFB)	Chin height (GNI)
Upper facial breadth (UFBR)	Body height (HMF)
Nasal height (NLH)	Body thickness (TMF)
Nasal breadth (NLB)	Bigonial diameter (GOG)
Orbital breadth (OBB)	Bicondylar breadth (CDL)
Orbital height (OBH)	Minimum ramus breadth (WRB)

Table 2 - Postcranial measurements collected form the three data sources. AP = Anterior - Posterior; ML = Medio - lateral; SI = Superior - Inferior

Clavicle	Maximum length AP midshaft SI midshaft	**Femur**	Maximum length Bicondylar length AP Subtrochanteric ML Subtrochanteric Vertical diameter head Epicondylar breadth AP midshaft ML midshaft
Scapula	Maximum height Maximum breadth		
Humerus	Maximum length Maximum diameter midshaft Minimum diameter midshaft Head diameter Epicondylar breadth	**Tibia**	Condyllo-malleolar length Proximal epiphyseal breadth Distal epiphyseal breadth AP at the nutrient foramen ML at the nutrient foramen
Radius	Maximum length AP midshaft ML midshaft		
Ulna	Maximum length Physiological length AP midshaft ML midhaft	**Fibula**	Maximum length Diameter at midshaft

Table 3 – Average percent differences between the data sources and the three measurement subsets. Abbreviations: Dry = dry skeletal elements, CT = CT images with soft tissue, and dryCT = CT images without soft tissue.

Source Comparison	Mean Percent Differences		
	Combined	*Cranial*	*Postcranial*
dry – dryCT	1.4%	0.9%	1.7%
dry – CT	1.5%	0.6%	2.0%
dryCT – CT	2.9%	1.5%	3.7%

Fig. 1.
Four standard cranial measurements collected on a CT image.
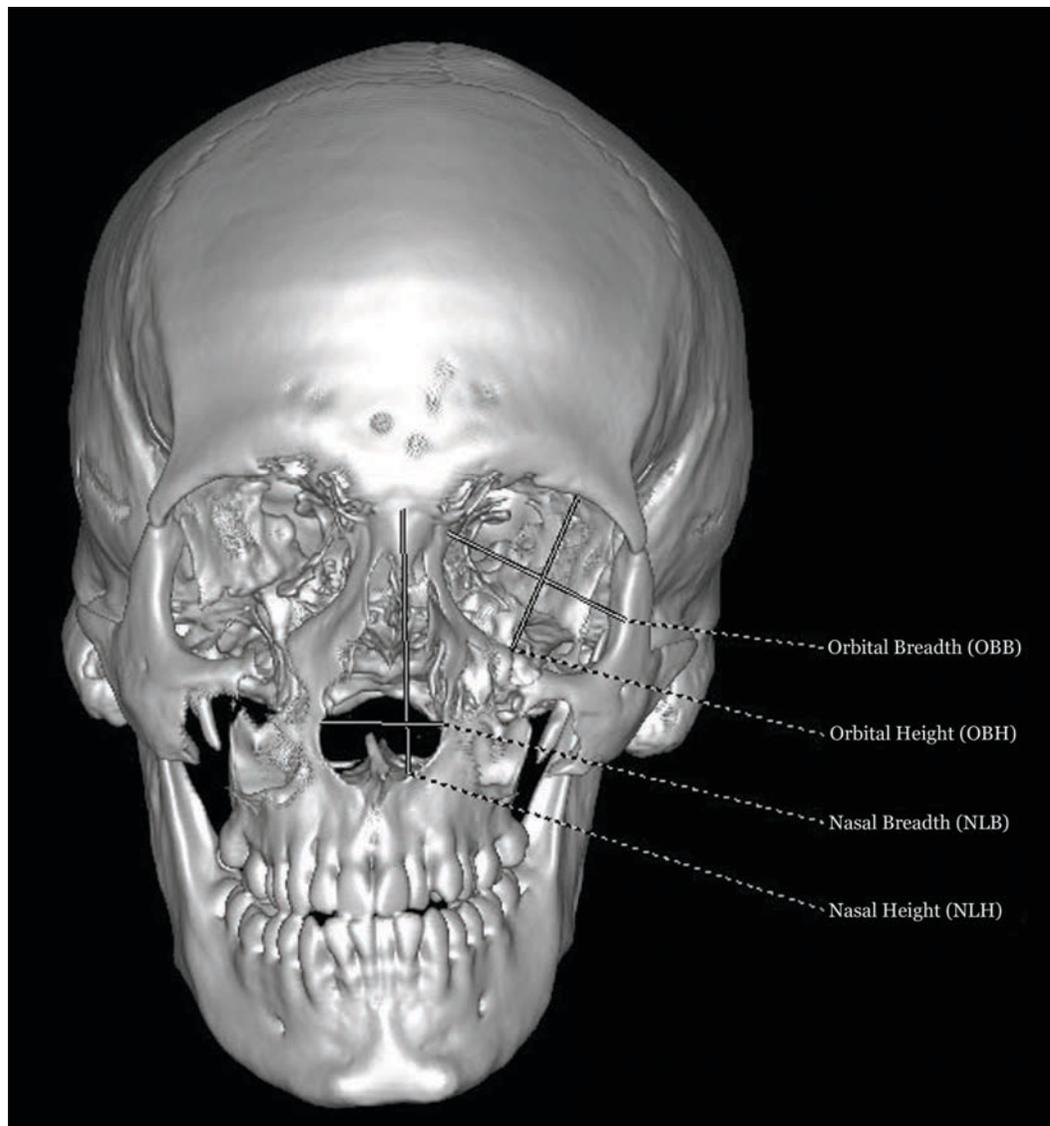
Fig. 2.
Two standard humeral measurements
collected on a CT image.

Epicondylar breadth

Maximum Length

Fig. 3.

The same crania imaged at two different opacity ramps, demonstrating the increased clarity of the zygomaticofrontal suture. The image on the left is at a window/level operation (W/L) of 594/41 and the image on the right is at a W/L operation of 310/268.
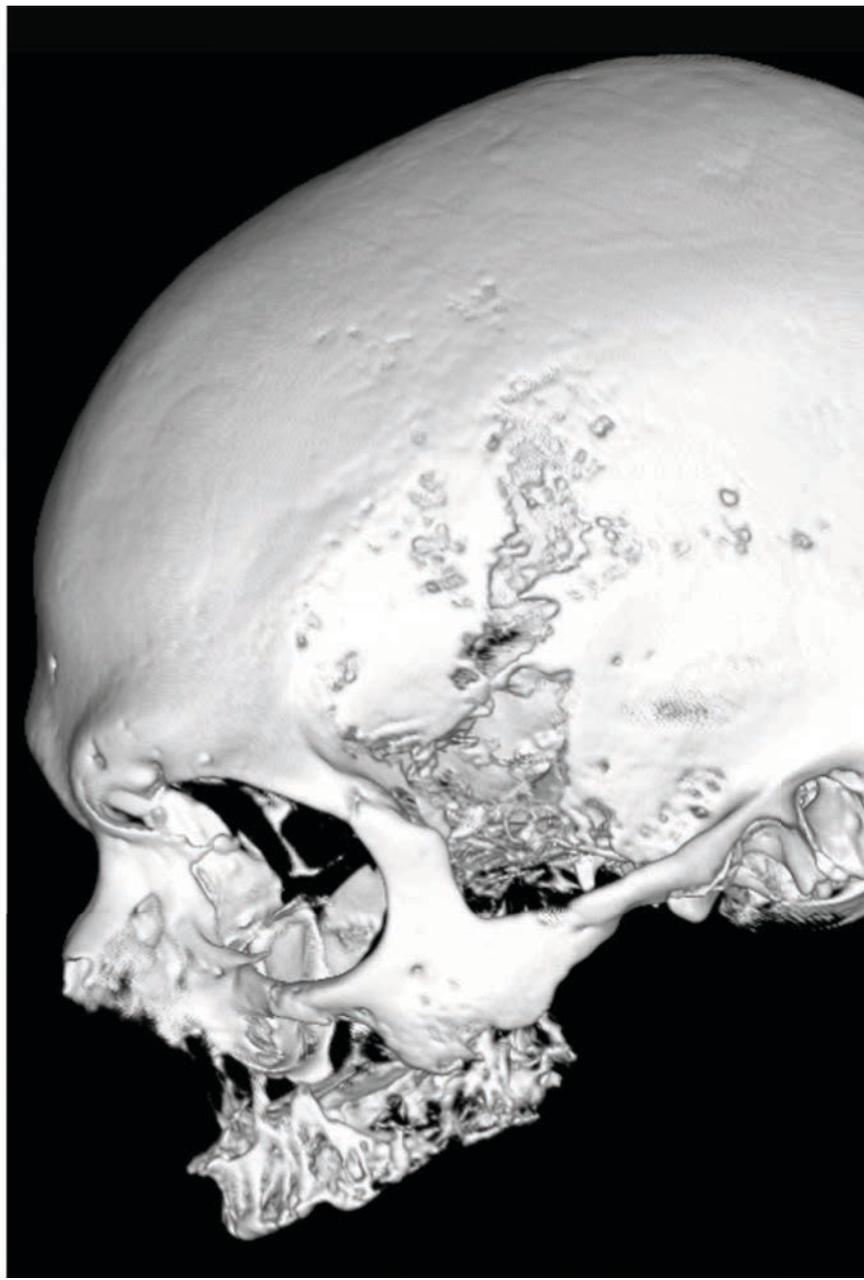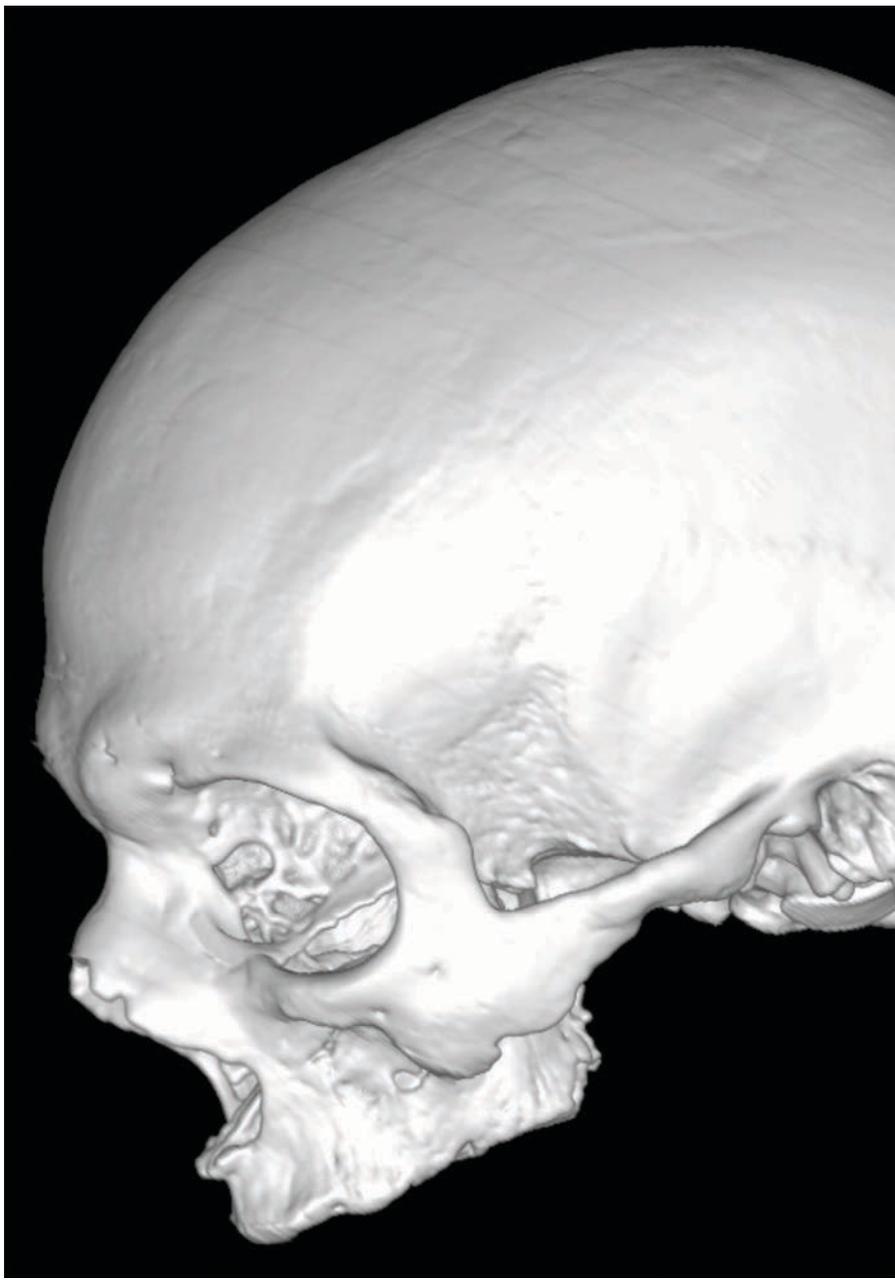
different on a 20 mm measurement versus a 100 mm measurement. Percent differences were calculated between the three data sources and comparisons were made for all measurements combined and also separated by cranial and postcranial measurements.

$$Percent\ differences = \frac{x_1 - x_2}{\left(\frac{x_1 + x_2}{2}\right)} * 100$$

A Bland-Altman plot was employed to visualize the amount of agreement between the measurements obtained from each data source [32]. The plot reveals the overall trends in the agreement of measurements and identifies any systematic biases and outliers by plotting the means of the repeated measures along the x-axis and the differences between the corresponding measurement pairs on y-axis [33–35]. The limits of agreement, both positive and negative, are the reference interval that is based on the mean and standard deviation and provide insight into the amount of random variation that is present [32,34]. If the two sets of measurements tend to agree, the plot shows a random scatter of differences around a mean of zero; if the two sets of measurements tend to disagree, the scatter will increase causing the limits of agreement to widen [32,34].

Reliability

Three observers, two biological anthropologists and one medical examiner with training in radiology, measured the seven dry crania and the VR images of the dry crania. Because reliability refers to the consistency in measures, the technical error of measurement (TEM) was utilized to assess inter- and intra-observer error for each measurement. The equation for intra-observer error TEM is

$$TEM = \sqrt{\frac{(\Sigma D^2)}{2N}}$$

where D is the difference between the measurements and N is the number of individuals measured [26]. The equation for inter-observer TEM differs when there are more than two observers and is as follows

$$\text{TEM} = \sqrt{(\Sigma_1^N((\Sigma_1^K M^2) - ((\Sigma_1^K M)^2/K)))/N(K-1)}$$

where N is the number of measurements, K is the number of observers, and M is the measurement. TEM retains the same unit of the measurement and is directly related to the measurement size. For example, a large mean value will have a large TEM and thus comparison of measurements of different sizes cannot be assessed [26]. To surmount this, TEM can be converted to relative TEM (%TEM), which is the error expressed as a percentage that corresponds to the total average of the variable analyzed (see below) [36]. The converted percentage has no units and allows for direct comparisons of all measurement sizes [26].

$$\% \text{ TEM} = \left(\frac{TEM}{mean}\right) * 100$$

3. Results

Accuracy

The average percent differences for all measurements combined were 1.4% for the dry-dryCT comparison, 1.5% for the dry-CT comparison, and 2.9% for the dryCT-CT comparison (Table 3). Cranial measurements resulted in lower percent differences in comparison to the postcranial measurements. The overall average smallest percent difference of 0.6% was between the cranial measurements obtained from the dry and CT images while the overall average highest percent difference of 3.7% was between the postcranial measurements obtained from the dryCT and CT images. The majority of measurements fell within the upper and lower agreement levels in the Bland-Altman plots, which was approximately 2 mm (Figures 4 – 6).

Reliability

The average intra-observer TEM and %TEM for all craniometric variables and three observers was between 0.46 mm and 0.71 mm and 0.56% and 1.06%, respectively. Variables that yielded the highest error rates were orbital height (OBH), inter-orbital

Fig. 4.
Bland–Altman plot depicting the differences in osteometric variables collected from the dry skeletal elements and from a CT image of the same bones in situ with soft tissue (dry–CT).
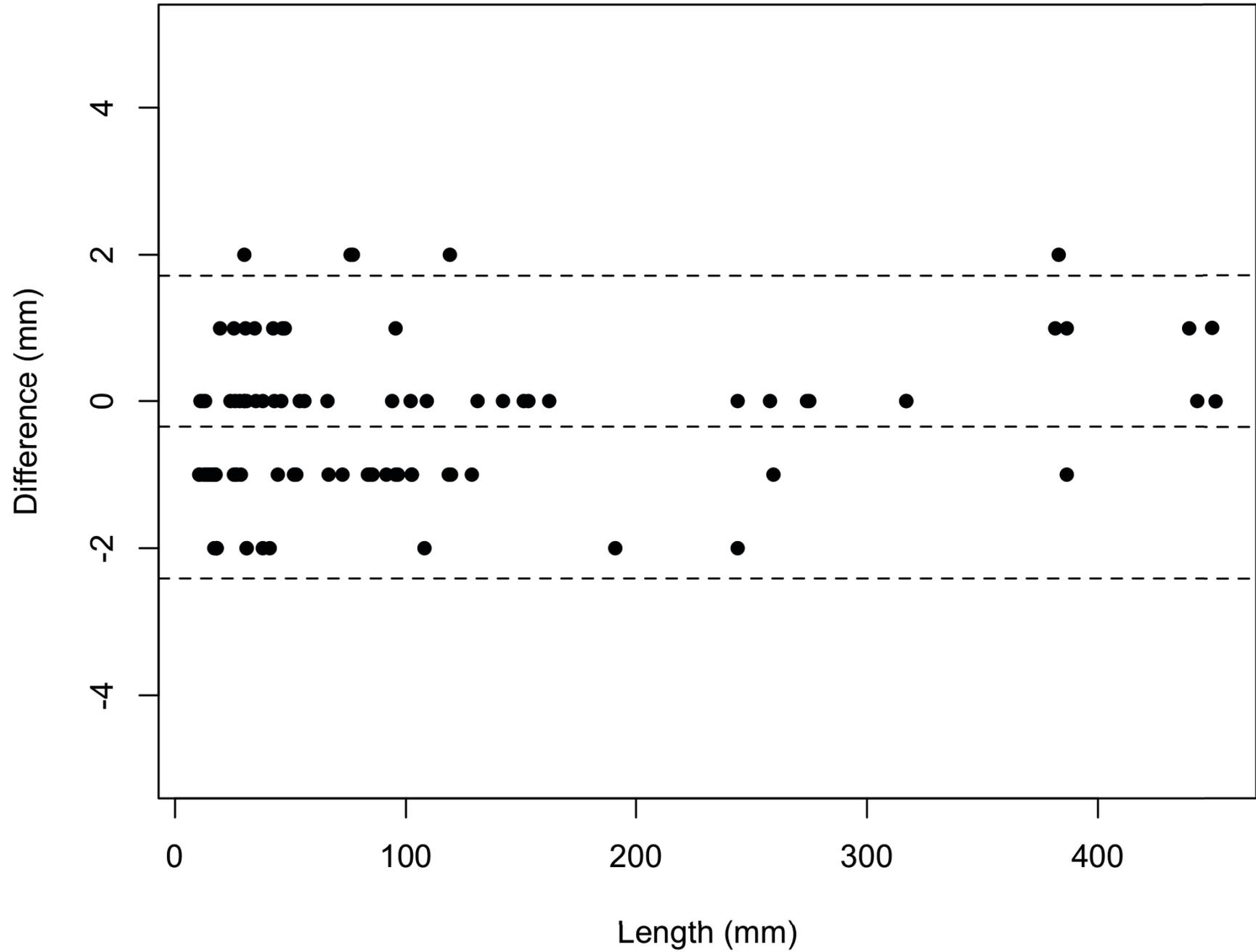
Fig. 5.
Bland–Altman plot depicting the differences in osteometric variables collected from dry skeletal elements and from a CT image of the dry skeletal elements (dry–dryCT).
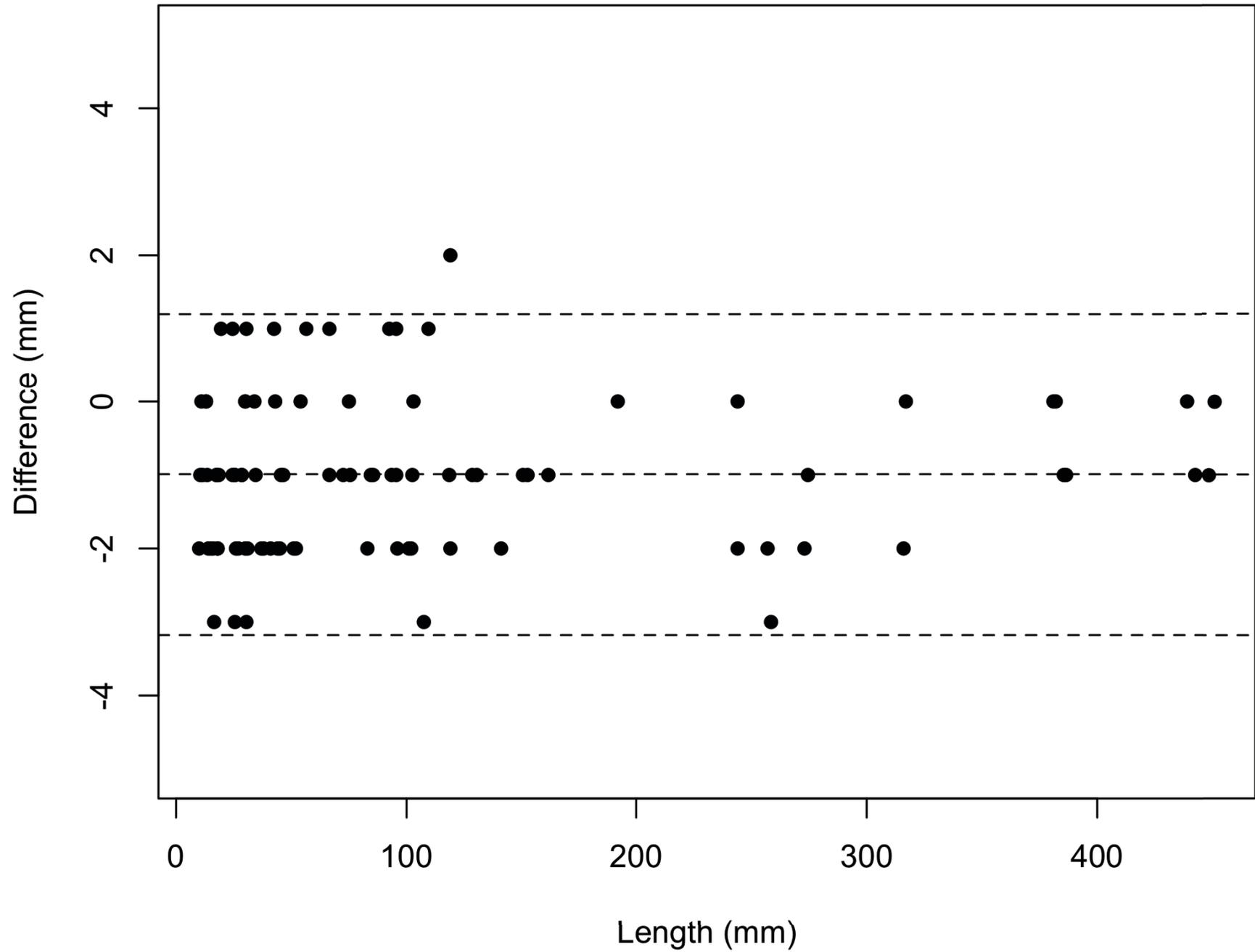
Fig. 6.
Bland–Altman plot depicting the differences in osteometric variables collected from a CT image of the dry skeletal elements and a CT image of the skeletal elements in situ with soft tissue (dryCT–CT).
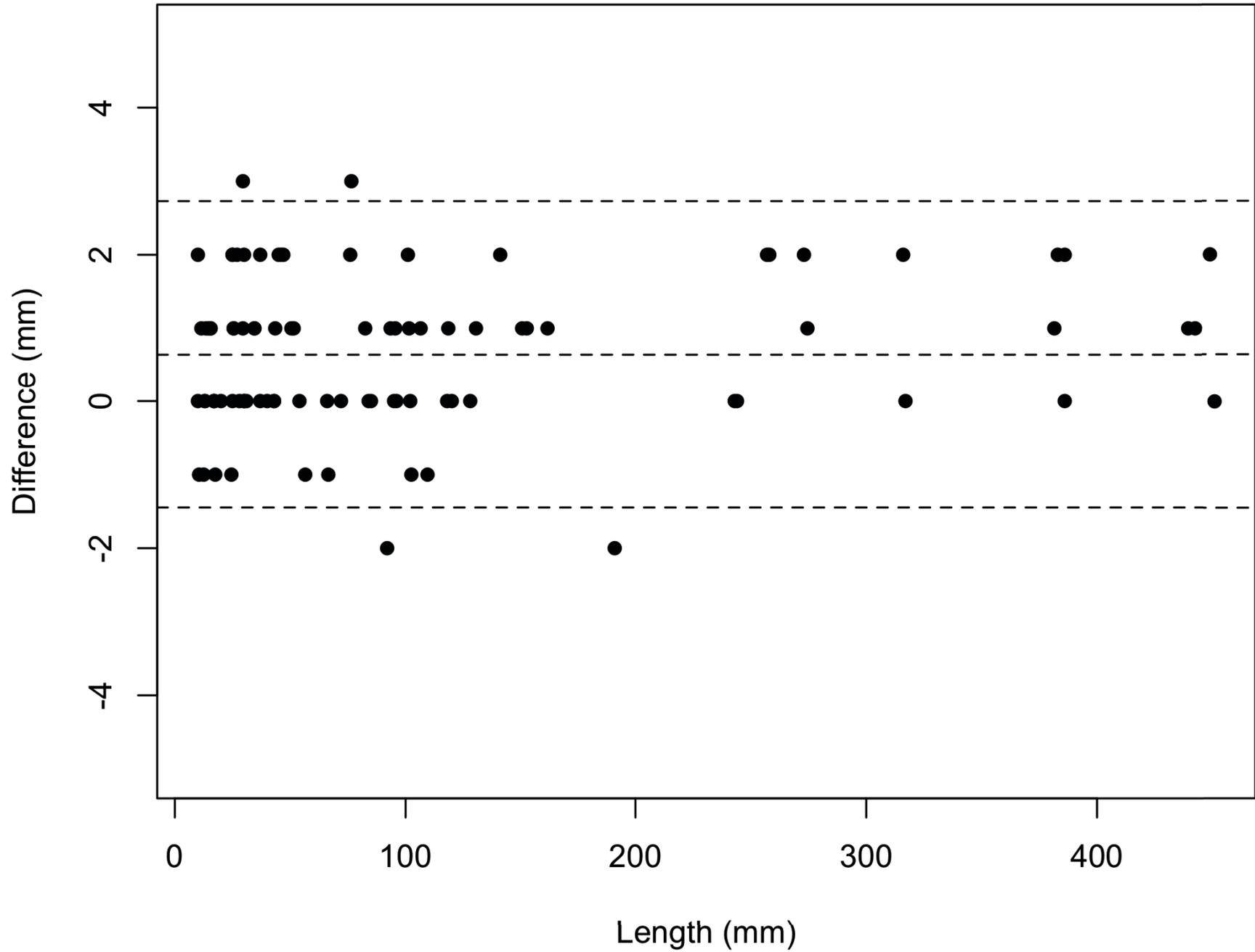
breadth (DKB) and orbital breadth (OBB). The three-way inter-observer TEM and %TEM, averaged across all craniometric variables, was 2.6 mm and 2.26%, respectively, slightly higher than the intra-observer error. The measurements that presented with the most error among the three observers were DKB, OBB and parietal chord (PAC).

4. Discussion

The primary concern with utilization of CT images is if the CT image reflects the same size and dimensions as the original object. Of primary importance is if the comparison between measurements collected from dry skeletal remains are accurate to measurements collected on CT images. Within the current study, the dry-CT comparison demonstrated an average percent difference of 1.5%, which is comparable to average percent differences of measurements obtained on dry bone and on Lodox Statscan-generated radiographic images of the dry bone [37,38]. Furthermore, prospective longitudinal growth studies that collected metric data from radiographic images with controlled settings to generate images with the least distortion possible, note distortion between 1% and 3% [39–44]. The average percent differences were higher in the postcrania than crania, which is likely because of the increased number of smaller measurements and decreased number of Type I landmarks (see below). As illustrated by Figures 4 – 6, the majority of measurement differences were within 2 mm, which is considered an acceptable amount of error in forensic anthropology. Overall, measurement differences in the current study were similar to most published studies, and in some instances the differences were much smaller. For example, a study conducted on the lower limb revealed errors as wide as 7 mm [2].

The largest percent difference was noted in the dryCT – CT source comparison (2.9%). The increased percent difference is related to the differential Hounsfield units of air and soft tissue. Although CT images accurately represent scanned objects, the imaging process is susceptible to certain artifacts, such as PVE. Essentially, while CT scans of dry material are beneficial for morphological observations only CT scans inclusive of soft tissues are recommended for metric data collection if the goal is to create an applicable anthropological technique.

Though studies vary in design, measurement tools and objects, the evaluation of published intra- and inter-observer TEM and %TEM values demonstrate the values acquired in the current study were comparable to previous reports [26,36,37,45–47]. The small percent differences and the acceptable levels of repeatability of measurements, not only between dry bone and CT images but within CT images, suggests the measurement error in the data sources is more a consequence of measurement repeatability and less because of artifacts associated with CT generated images.

Landmarks were historically chosen and subsequently defined to be repeatedly located with high accuracy and high precision on each object within and between populations [49]. Type I landmarks are based on biologically unique patterns on the form, Type II landmarks are defined by geometric criteria (e.g. point of maximum curvature) and Type III landmarks are dependent on the location of other landmarks [12,13,50]. Because Type II and III landmarks present with lower precision compared to Type I landmarks [21,51,52], measurements inclusive of Type II and III landmarks were expected to result in lower repeatability. The measurements that presented with the highest intra-observer error were OBH defined by Type III landmarks, OBB defined by Type I and III landmarks, and DKB defined by Type I landmarks. Although these measurements had the largest error, the difference between the original and secondary measurement for the three variables was only 1 mm. The measurements that displayed the most error in the three-way inter-observer error were DKB, OBB, and PAC. Similar to DKB, PAC is defined by Type I landmarks.

DKB has long been recognized as a variable with high levels of error between observers on dry skeletal elements [53]. The high error associated with DKB and OBB is related to the location of dacryon. The ability to identify dacryon with high precision is likely to increase with a smaller slice thickness, as this would increase clarity in the images and increase reliability of Type I landmarks. For the current study, the authors chose to validate a retrospective data source. The slice thickness in the current research was smaller or comparable (0.625 mm and 1.25 mm) to the majority of recent publications (0.75 mm to 1.25 mm) that investigated the accuracy of measurements collected on CT images [2,3,48]. However, a smaller slice thickness should be considered if designing a prospective study.

Similarly to the results of the current study, Utermohle and Zegura [53] and Utermohle et al. [54] identified PAC to be a measurement with increased levels of TEM even though both bregma and lambda are Type I landmarks. Suture obliteration can cause the landmark location to be estimated and thus higher error between observers. By adjusting the opacity level of the CT images, as previously described, the intersection of the sutures can be observed (Figure 7). The measurement error noted in PAC of current study emphasizes the unreliable and inaccurate placement of estimated landmarks. Obliterated sutures often reduces the number of measurements that can be used in anthropological analyses; however, the use of CT images permits a larger number of variables included in analyses as the opacity ramp permits visibility of obliterated sutures.

Midshaft measurements from the humerus, ulna, clavicle, radius, and fibula presented with the highest percent differences in the postcrania, likely because midshaft measurements are small in size (i.e. a 1 mm difference can account for upwards of 10% of the error) and because midshaft measurements are not defined by anatomic landmarks [21,50]. When measuring the skeletal elements on CT, the most difficult aspect is not being able to handle the remains. For example, the bones were isolated within the CT image and each element was sectioned at midshaft (clavicle, humerus, radius, and fibula), the greatest development of the crest (ulna), or at the nutrient foramen (tibia). Through trial and error the measurements were obtained while simultaneously trying to orientate the element in anatomical position. The tibia was especially difficult, as the nutrient foramen appeared as a continuous groove as opposed to a groove that terminates at a foramen. Similar to comments regarding the location of dacryon, scanning at thinner slices (i.e. 0.5 mm) would generate a more detailed visualization of anatomical landmarks (i.e. foramina), which would result in higher measurement accuracy. Even with the acknowledgement of the difficulties in obtaining some postcranial measurements, all measurements were within 2 mm and the majority within the upper and lower agreement levels. Because the measurements in the current study that presented with the highest error are also measurements that have been identified as unreliable on dry skeletal remains, results suggest that the largest source of measurement error is actually human error and not associated with imaging.

Fig. 7.
Although the sutures appeared obliterated on the dry cranium with the standard opacity ramp (left), adjustment of the opacity ramp allows for bregma to be observed (right).
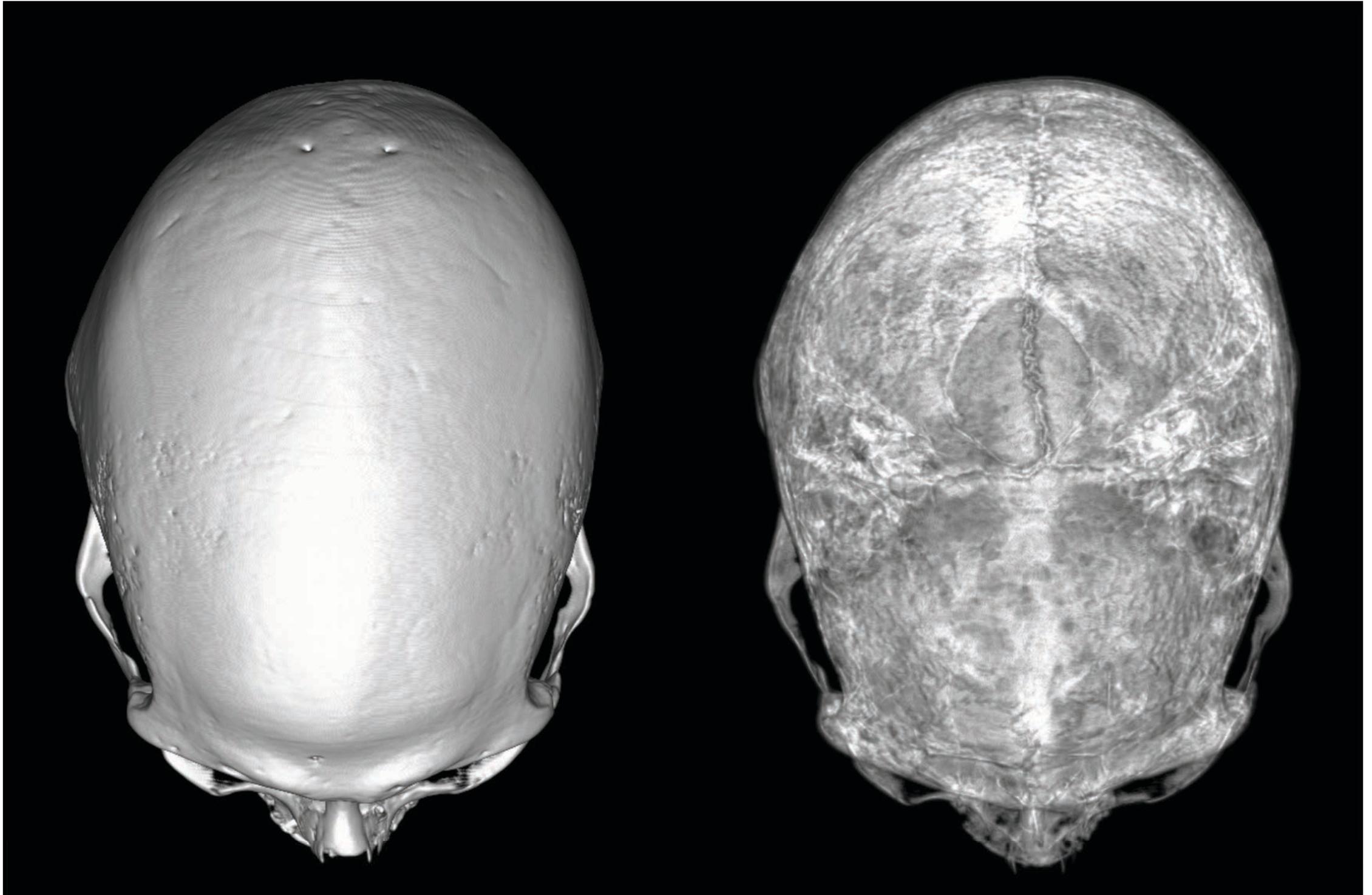
Forensic Applications

Both forensic pathologists and forensic anthropologists can use CT scans during mass disaster response operations in which the practitioners work together to assist in identification of human remains and estimate a minimum number of individuals. In particular, forensic anthropologists can collect cranial and postcranial data directly from the CT image [55]. Additionally, CT data are stored in PACS (Picture Archiving and Communication System) and transmitted using DICOM (Digital Imaging and Communication in Medicine), which is a global information technology standard that is used in virtually all hospitals worldwide and is designed to produce, manage, and distribute images. Use of PACS allows for collaboration and/or consultation even if an anthropologist cannot be physically present at the scene of a mass disaster [2]. Utilization of CT images allow for a more efficient and less invasive way to obtain data that ultimately assists in victim identification. Besides being an efficient tool to facilitate victim identification, the use of CT images also offers a means to accommodate humanitarian and religious beliefs. Additionally, as the CT images are stored permanently in PACS, they can be reviewed if additional data needs to be collected or a second opinion is necessary. Although the process of obtaining measurements on CT images takes slightly longer than directly from dry skeletal elements and there is a learning curve for the software, use of CT images is nevertheless quicker than the time required to process a complete body.

Researchers in forensic anthropology are encouraged to develop population specific techniques as each population differs in size and shape as well as experiences different extrinsic factors that affect the skeleton's biomechanical adaptation. However, many locations where population specific methods are needed do not have suitable skeletal collections to develop or validate standards. Therefore, the use of 3D-CT images offers a data source that one can collect metric and morphological variables and ultimately facilitate the creation of standards throughout the world.

5. Conclusions

The high consistency of measurements to be within the acceptable measurement range for anthropologists (~2 mm) validates the claim that CT images are accurate

representations of the true objects dimensions. Furthermore, the small percent differences between the data sources, comparable TEM and %TEM for the inter- and intra-observer error, and the measurements noted as unreliable in the current study being the same measurements consistently recognized as unreliable on studies of dry bones suggests that the measurement error is because of human error rather than CT imaging (i.e. distortion). If a 3D reconstruction of CT images is available, the time consuming procedure of soft tissue removal is unnecessary to obtain metric variables, as seen in the current study, or morphological variables, as noted in previous publications [56]. The VR process allows the anthropologist to view elements from different angles and take measurements that are comparable to measurements obtained on dry bones. Furthermore, consultations can be made from anywhere in the world through the use of DICOM. However, anthropologists must first be competent in measurement techniques and then perfect the manipulation of CT images prior to data collection. The results of this study prove that measurements obtained from CT images can be considered accurate and reliable, and subsequently, CT images can be treated as a practical option for anthropologists to utilize during the development or validation of forensic anthropological techniques.

Acknowledgements

6. References

[1]   S. Decker, J. Ford, E. Hoegstrom, D. Hilbelink, Virtual anatomy: three-dimensional computer modeling and measurement of human cranial anatomy, in: Colorado Springs, CO, 2008: p. 312.

[2] C. Robinson, R. Eisma, B. Morgan, A. Jeffery, E.A.M. Graham, S. Black, et al., Anthropological measurement of lower limb and foot bones using multi-detector computed tomography, J. Forensic Sci. 53 (2008) 1289–1295.

[3] A.L. Brough, J. Bennett, B. Morgan, S. Black, G.N. Rutty, Anthropological measurement of the juvenile clavicle using multi-detector computed tomography--affirming reliability, J. Forensic Sci. 58 (2013) 946–951.

[4] S. Decker, The Human in 3D: Advanced Morphometric Analysis of High-Resolution Anatomically Accurate Computed Models, University of South Florida, 2010.

[5] C. O'Donnell, N. Woodford, Post-mortem radiology—a new sub-speciality?, Clinical Radiology. 63 (2008) 1189–1194.

[6] C. O'Donnell, M. Iino, K. Mansharan, J. Leditscke, N. Woodford, Contribution of postmortem multidetector CT scanning to identification of the deceased in a mass disaster: Experience gained from the 2009 Victorian bushfires, Forensic Science International. 205 (2011) 15–28.

[7] B. Daly, S. Abboud, Z. Ali, C. Sliker, D. Fowler, Comparison of whole-body post mortem 3D CT and autopsy evaluation in accidental blunt force traumatic death using the abbreviated injury scale classification, Forensic Science International. 225 (2013) 20–26.

[8] M. Sidler, C. Jackowski, R. Dirnhofer, P. Vock, M. Thali, Use of multislice computed tomography in disaster victim identification—Advantages and limitations, Forensic Science International. 169 (2007) 118–128.

[9] S. Blau, C.A. Briggs, The role of forensic anthropology in Disaster Victim Identification (DVI), Forensic Science International. 205 (2011) 29–35.

[10] C.F. Hildebolt, M.W. Vannier, R.H. Knapp, Validation study of skull three-dimensional computerized tomography measurements, American Journal of Physical Anthropology. 82 (1990) 283–294.

[11] A.A. Waitzman, J.C. Posnick, D.C. Armstrong, G.E. Pron, Craniofacial skeletal measurements based on computed tomography: Part I. Accuracy and reproducibility, Cleft Palate Craniofac. J. 29 (1992) 112–117.

[12] J.T. Richtsmeier, C.H. Paik, P.C. Elfert, T.M. Cole III, H.R. Dahlman, Precision, repeatability, and validation of the localization of cranial landmarks using computed tomography scans, The Cleft Palate-Craniofacial Journal. 32 (1995) 217–227.

[13] C.J. Valeri, T.M. Cole 3rd, S. Lele, J.T. Richtsmeier, Capturing data from three-dimensional surfaces using fuzzy landmarks, Am. J. Phys. Anthropol. 107 (1998) 113–124.

[14] M.J. Citardi, B. Herrmann, C.S. Hollenbeak, B.C. Stack, M. Cooper, R.D. Bucholz, Comparison of Scientific Calipers and Computer-Enabled CT Review for the Measurement of Skull Base and Craniomaxillofacial Dimensions, Skull Base. 11 (2001) 5–11.

[15] F.L. Williams, J.T. Richtsmeier, Comparison of mandibular landmarks from computed tomography and 3D digitizer data, Clin Anat. 16 (2003) 494–500.

[16] B. Swift, G.N. Rutty, Recent Advances in Postmortem Forensic Radiology, in: M. Toskas (Ed.), Forensic Pathology Reviews, Humana Press, 2006: pp. 355–404.

[17] L. Lou, M.O. Lagravere, S. Compton, P.W. Major, C. Flores-Mir, Accuracy of measurements and reliability of landmark identification with computed tomography (CT) techniques in the maxillofacial area: a systematic review, Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 104 (2007) 402–411.

[18] P.M.L. Lopes, C.R. Moreira, A. Perrella, J.L. Antunes, M.G.P. Cavalcanti, 3-D volume rendering maxillofacial analysis of angular measurements by multislice CT, Oral Surg Oral Med Oral Pathol Oral Radiol Endod. 105 (2008) 224–230.

[19] M.A. Verhoff, F. Ramsthaler, J. Krähahn, U. Deml, R.J. Gille, S. Grabherr, et al., Digital forensic osteology--possibilities in cooperation with the Virtopsy project, Forensic Sci. Int. 174 (2008) 152–156.

[20] P. Guyomarc'h, F. Santos, B. Dutailly, P. Desbarats, C. Bou, H. Coqueugniot, Three-dimensional computer-assisted craniometrics: A comparison of the uncertainty in measurement induced by surface reconstruction performed by two computer programs, Forensic Science International. 219 (2012) 221–227.

[21] S.B. Sholts, L. Flores, P.L. Walker, S.K.T.S. Wärmländer, Comparison of coordinate measurement precision of different landmark types on human crania using a 3D laser scanner and a 3D digitiser: Implications for applications of digital morphometrics, International Journal of Osteoarchaeology. 21 (2011) 535–543.

[22] Daubert vs. Merrell Dow Pharmaceuticals, 113 Supreme Court 2786, 1993.

[23] Dennis C. Dirkmaat, Luis L. Cabo, Stephen D. Ousley, S.A. Symes, New Perspectives in Forensic Anthropology, Yearbook of Physical Anthropology. 51 (2008) 33–52.

[24] R.L. Melnick, A Daubert motion: a legal strategy to exclude essential scientific evidence in toxic tort litigation, Am J Public Health. 95 Suppl 1 (2005) S30–34.

[25] S. Ousley, Should We Estimate Biological or Forensic Stature?, J Forensic Sci. 40 (1995) 768–773.

[26] S. Ulijaszek, D. Kerr, Anthropometric Measurement Error and the Assessment of Nutritional Status, Br J Nutr. 82 (1999) 165–177.

[27] B.J. Adams, J.E. Byrd, Interobserver Variation of Selected Postcranial Skeletal Measurements, Journal Forensic Sciences. 47 (2002) 1–10.

[28] R. Goto, N. Mascie-Taylor, Precision of Measurement as a Component of Human Variation, J Physio Anthropol. 26 (2007) 253–256.

[29] J.E. Buikstra, D.H. Ubelaker, Standards for Data Collection from Human Skeletal Remains: Proceedings of a Seminar at the Field Museum of Natural History, Arkansas Archaeological Research Series, Fayetteville, 1994.

[30] J. Urcid, Manual for Post-Cranial Measurements, Smithsonian Institution's National Museum of Natural History, 1992.

[31] S. Ousley, R. Jantz, FORDISC 3.1, The University of Tennessee, Knoxville, 2005.

[32] M. Bland, D. Altman, Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement., Lancet. (1986) 307–310.

[33] P. Rothwell, Analysis of Agreement Between Measurements of Continuous Variables: General Principles and Lessons from Studies of Imaging of Carotid Stenosis, J Neurol. 415 (2000) 825–834.

[34] A. Geeta, H. Jamaiyah, M. Safiza, G. Khor, C. Kee, A. Ahmad, et al., Reliability, technical error of measurements and validity of instruments for nutritional status assessment of adults in Malaysia, Singapore Med J. 50 (2009) 1013–18.

[35] E.F. Harris, R.N. Smith, Accounting for measurement error: A critical but often overlooked process, Archives of Oral Biology. 54, Supplement 1 (2009) S107–S117.

[36] T. Perini, G. de Oliveira, J. dos Santos Ornellas, F. de Oliveira, Technical Error of Measurement in Anthropometry, Rev Bras Med Esporte. 11 (2005).

[37] K. Stull, An Osteometric Evaluation of Age and Sex Differences in the Long Bones of South African Children from the Western Cape, Dissertation, University of Pretoria, 2013.

[38] K.E. Stull, E.N. L'abbé, S. Steiner, Measuring distortion of skeletal elements in Lodox Statscan-generated images, Clin Anat. 26 (2013) 780–786.

[39] W. Green, G. Wyatt, M. Anderson, Orthoroentgenography as a method of measuring the bones of the lower extremities, J Bone Joint Surg Am. 28 (1946) 60–65.

[40] M. Maresh, Linear growth of long bones of extremities from infancy through adolesence, American Journal of Diseases of Children. 89 (1955) 725–742.

[41] M. Anderson, M. Messner, W. Green, Distribution Lengths of the Normal Femur and Tibia in Children from One to Eighteen Years of Age, J Bone Jt Surg. 46A (1964) 1197–1202.

[42] P. Buschang, Differential long bone growth of children between two months and eleven years, American Journal of Physical Anthropology. 58 (1982) 291–295.

[43] S. Smith, P. Buschang, Variation in Longitudinal Diaphyseal Long Bone Growth in Children Three to Ten Years of Age, American Journal of Human Biology. 16 (2004) 648–657.

[44] S. Smith, Stature Estimation of 3-10 year-old Children from Long Bone Lengths, J Forensic Sci. 52 (2007) 538–46.

[45] WHO Multicentre Growth Reference Study Group, Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study, Acta Paediatr. Suppl 450 (2006) 38–46.

[46] M. Sicotte, M. Ledoux, M.-V. Zunzunegui, S.A. Aboubacrine, V.-K. Nguyen, Reliability of anthropometric measures in a longitudinal cohort of patients initiating ART in West Africa, BMC Medical Research Methodology. 10 (2010) 102.

[47] H.F.V. Cardoso, J. Abrantes, L.T. Humphrey, Age estimation of immature human skeletal remains from the diaphyseal length of the long bones in the postnatal period, Int. J. Legal Med. (2013).

[48] S.J. Decker, S.L. Davy-Jow, J.M. Ford, D.R. Hilbelink, Virtual determination of sex: metric and nonmetric traits of the adult pelvis from 3D computed tomography models, J. Forensic Sci. 56 (2011) 1107–1114.

[49] I.L. Dryden, K.V. Mardia, Statistical Shape Analysis, in: John WIley & Sons, New York, 1998: p. Chapters 1–3.

[50] F. Bookstein, Morphometric Tools for Landmark Data: Geometry and Biology, Cambridge University Press, Cambridge, 1991.

[51] D. Slice, C. Untereggre, K. Schaefer, F. Bookstein, Modeling the precision of landmark location data, Am. J. Phys. Anthropol. 123 (Supp 36) (2004) 183.

[52] C. Simonis-Sueur, M. Friess, F. Detroit, Skull shapes, maps, and microscribes, American Journal of Physical Anthropology. suppl. 48 (2009).

[53] C. Utermohle, S. Zegura, Intra- and Interobserver Error in Craniometry: A Cautionary Tale, Am J Phys Anthropol. 57 (1982) 303–310.

[54] C.J. Utermohle, S.L. Zegura, G.M. Heathcote, Multiple observers, humidity, and choice of precision statistics: factors influencing craniometric data quality, Am. J. Phys. Anthropol. 61 (1983) 85–95.

[55] E.A. Gursky, M.F. Fierro, Death in Large Numbers: The Science, Policy, and Management of Mass Fatality Events, American Medical Association, 2012.

[56] F. Ramsthaler, M. Kettner, A. Gehl, M.A. Verhoff, Digital forensic osteology: Morphological sexing of skeletal remains using volume-rendered cranial CT scans, Forensic Science International. 195 (2010) 148–152.