Title: Error Quantification of Osteometric Data in Forensic Anthropology

Authors: Natalie R. Langley, Lee Meadows Jantz, Shauna McNulty, Heli Maijanen, Stephen D. Ousley, Richard L. Jantz

PII: S0379-0738(18)30159-2
DOI: https://doi.org/10.1016/j.forsciint.2018.04.004
Reference: FSI 9243

To appear in: *FSI*

Received date: 20-2-2018
Revised date: 27-3-2018
Accepted date: 3-4-2018

Error Quantification of Osteometric Data in Forensic Anthropology

Natalie R. Langley[a] (**corresponding author**)
Email: langley.natalie@mayo.edu

Lee Meadows Jantz[b]
Email: ljantz@utk.edu

Shauna McNulty[c]
Email: shauna.mcnulty@umpqua.edu

Heli Maijanen[d]
Email: heli.maijanen@oulu.fi

Stephen D. Ousley[e]
Email: sousley@mercyhurst.edu

Richard L. Jantz[b]
Email: rjantz@utk.edu

[a]Department of Anatomy, Mayo Clinic College of Medicine and Science, Mayo Clinic School of Medicine Arizona Campus, 13400 E. Shea Blvd., Scottsdale, AZ, USA 85259
[b]Department of Anthropology, University of Tennessee, 502 Strong Hall, 1621 Cumberland Ave., Knoxville, TN, USA, 37996
[c]Science Department, Umpqua Community College, 1140 Umpqua College Road, Roseburg, OR 97470
[d] University of Oulu, Archaeology, P.O. Box 1000, 90014 University of Oulu, Finland
[e]Department of Mathematics and Information Technology, Mercyhurst University, 501 E. 38th Street, Erie, PA, USA, 16546

Highlights

- The reliability of osteometric data used in forensic case analyses is evaluated.
- Twenty-two measurements had unacceptable technical error of measurement values.
- A new manual, *Data Collection Procedures 2.0*, was produced based on these results.
- The manual and an accompanying instructional video are freely available online.

Abstract

This study evaluates the reliability of osteometric data commonly used in forensic case analyses, with specific reference to the measurements in *Data Collection Procedures 2.0* (DCP 2.0). Four observers took a set of 99 measurements four times on a sample of 50 skeletons (each measurement was taken 200 times by each observer).

Two-way mixed ANOVAs and repeated measures ANOVAs with pairwise comparisons were used to examine interobserver (between-subjects) and intraobserver (within-subjects) variability. Relative technical error of measurement (TEM) was calculated for measurements with significant ANOVA results to examine the error among a single observer repeating a measurement multiple times (e.g. repeatability or intraobserver error), as well as the variability between multiple observers (interobserver error). Two general trends emerged from these analyses: (1) maximum lengths and breadths have the lowest error across the board (TEM < 0.5), and (2) maximum and minimum diameters at midshaft are more reliable than their positionally-dependent counterparts (i.e. sagittal, vertical, transverse, dorso-volar). Therefore, maxima and minima are specified for all midshaft measurements in DCP 2.0. Twenty-two measurements were flagged for excessive variability (either interobserver, intraobserver, or both); 15 of these measurements were part of the standard set of measurements in *Data Collection Procedures for Forensic Skeletal Material, 3rd edition*. Each measurement was examined carefully to determine the likely source of the error (e.g. data input, instrumentation, observer's method, or measurement definition). For several measurements (e.g. anterior sacral breadth, distal epiphyseal breadth of the tibia) only one observer differed significantly from the remaining observers, indicating a likely problem with the measurement definition as interpreted by that observer; these definitions were clarified in DCP 2.0 to eliminate this confusion. Other measurements were taken from landmarks that are difficult to locate consistently (e.g. pubis length, ischium length); these measurements were omitted from DCP 2.0. This manual is available for free download online (https://fac.utk.edu/wp-content/uploads/2016/03/DCP20_webversion.pdf), along with an accompanying instructional video (https://www.youtube.com/watch?v=BtkLFl3vim4).

Keywords: forensic science, forensic anthropology, observer variation, osteometric, Analysis of Variance, calipers, methods

1. Introduction

The *Daubert v. Merrell Dow Pharmaceuticals* [1] ruling set the precedent for relevant, reliable, and scientifically sound expert testimony more than two decades ago. As a result of this and several subsequent rulings, research in the forensic sciences has been driven by the need to critically evaluate, validate, and establish error rates for methodologies used in forensic case analyses. The prevailing atmosphere of scientific validation was reiterated in the National Academy of Sciences (NAS) Report *Strengthening Forensic Science in the United States: A Path Forward* [2], where the authors stressed the need for "rigorous systematic research to validate the discipline's basic premises and techniques" (p. 22). Many techniques in forensic anthropology employ osteometric data, although little work has been done to investigate the intrinsic error in these measurements. Forensic practitioners must be aware of the reliability of osteometric data to make informed decisions for case analyses. This study investigates the reliability and validity of osteometric data used in forensic methods (NAS Report Recommendation 3) and aims to support the effort to establish valid and reliable methods and protocols for proficiency testing, training, and certification (NAS Report

Recommendation 6). The primary objectives are to provide the forensic anthropology community with error rates for standard osteometric measurements, evaluate the efficacy of these measurements, and investigate alternatives for problem measurements. These data will provide a foundation for forensic case analyses, research, training, and method development.

Adams and Byrd [3] investigated interobserver error in 22 postcranial measurements and made several recommendations that were used to formulate the research design and goals of the current study:

(1) Procedures using skeletal measurements should favor measurements that are relatively easy to take.

(2) Clear definitions of the measurements should be provided in any publications.

(3) Problematic measurements such as pubis length are invalid due to the problem of locating a particular landmark (i.e. the junction of the pubis, ischium, and ilium in the acetabulum); these measurements should not be used in analyses.

(4) On account of the significant implications that the results of these metric analyses hold (e.g. the identification or exclusion of an unknown individual), it is of utmost importance that measurements used by forensic anthropologists can be accurately and reliably taken and that they are replicable between observers.

(5) Significant interobserver measurement variation could compromise pooled datasets compiled from multiple researchers and, in turn, bias research based on these data. Interobserver error in reference data (e.g. the Forensic Data Bank described below)

will increase the standard error and introduce potential bias in models to estimate stature, sex, or ancestry.

(6) University training in osteometrics promotes continuity in data collection. Beyond the university, forensic laboratories should include detailed measurement descriptions in their standard operating procedures and provide osteometric training to new staff.

Many components of forensic anthropology case reports are derived from osteometric data (e.g. sex, ancestry, and stature). Methods employing metric data are considered more objective than nonmetric techniques requiring visual assessments of skeletal form. Nonetheless, error associated with any method based on osteometric data is compounded by the error inherent in a given measurement or set of measurements, whether a function of the observer, instrumentation, or both. Measurement error can be minimized by using appropriate instrumentation, carefully reading the instrumentation, understanding the measurement definition, adequate training, and by using highly reliable and repeatable measurements. Knowing measurement reliability provides a foundation from which to proceed with metric estimations of sex, ancestry, and stature, as well as method development.

Osteometric data also form the basis of the Fordisc [4] computer program used by forensic anthropologists mainly in the United States to estimate sex, stature, and ancestry. Fordisc uses the Forensic Data Bank (FDB) as its reference database. Measurement data is provided to the FDB curators by forensic practitioners and researchers who use a reference text such as *Data Collection Procedures for Forensic*

Skeletal Material [5-7]or *Standards for Data Collection from Human Skeletal Remains* [8] to take the measurements. The *Data Collection Procedures* manual was designed specifically to interface with the FDB and Fordisc. The first edition of *Data Collection Procedures* [6] was released in 1986 with the goal of standardizing recording procedures and establishing a means to amass a centralized database of skeletal data on modern humans (the FDB) that would keep pace with the changing US population and preserve data that would otherwise be lost when remains are returned. The second edition was released in 1990, and the FDB had 850 entries, 60% of which were documented forensic cases [5]. The third edition was released in 1994; the FDB had 1,200 cases. As of this writing the FDB has over 4,000 cases, and the *Data Collection Procedures* manual was revised in 2016 (*Data Collection Procedures 2.0*) [9].

It is essential that forensic anthropologists continue to compile reference databases of modern skeletal data so methods and software packages can keep pace with changing populations worldwide. However, the reliability of these data must be determined, and standardized data collection protocols employed by all practitioners to ensure best practice. This paper reports the technical error of measurement of osteometric data commonly used in forensic case analyses, with specific reference to the measurements in the revised *Data Collection Procedures 2.0* (DCP 2.0) [9]. This manual is available for free download online (https://fac.utk.edu/wp-content/uploads/2016/03/DCP20_webversion.pdf), along with an accompanying instructional video (https://www.youtube.com/watch?v=BtkLFl3vim4).

2. Materials and Methods

2.1 Data Collection

A priori power analyses were conducted to estimate the sample size necessary to achieve adequate power for several interobserver (between-subject) and intraobserver (within-subject) analyses. A probability level of 0.05 and statistical power of 0.80 was used to estimate *n* (sample size necessary to achieve power). Although there are no formal standards for power, 0.80 is commonly used as a standard for adequacy, as it implies a four-to-one trade-off between the risk of a Type II error and a Type I error. Data from the FDB were used to calculate means and standard deviations of several measurements, and these values were used to approximate effect size. Cohen [10] classifies effect sizes into small, medium, and large values, depending on the type of statistical analysis employed. F values of 0.1, 0.25, and 0.4 represent small, medium, and large effect sizes [10]. The smaller the effect size, the more difficult it is to detect the degree of deviation of the null hypothesis; consequently Cohen [11] recommends a medium effect size because it can approximate the average effect size in various fields. Estimated effect size values for several osteometric measurements ranged between small and medium, so a medium effect size (0.25) was used in the power analyses. The analysis assumed that the four observers would take each measurement four times on each skeleton. The analyses were conducted using the freeware GPower (latest version available at http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3). The total sample size necessary for power=.80 with a medium effect size was n=24 for a within-factors repeated measures ANOVA and n=116 for a between-factors repeated measures ANOVA. Therefore, each observer took a set of 99 measurements four times on a sample of 50 skeletons, ensuring that each metric was taken 200 times by each observer.

Figure 1 provides a schematic of the data collection design for each measurement evaluated in this project. The osteometric data were collected on a random sample of William M. Bass Donated Collection skeletons (n=50). Four observers measured the left elements of 50 skeletons. The observers were assigned numbers based on experience level, with Observer 1 having the most experience (27 years) and Observer 4 having the least experience (3 years). Ninety-nine measurements were taken on each skeleton using the instrument specified in the measurement definition in *Data Collection Procedures,* 3rd edition [7] (e.g. spreading calipers, digital sliding calipers, tape measure, osteometric board, mandibulometer). Once all 50 skeletons were measured, the process was repeated for a total of four rounds. Observers were provided copies of *Data Collection Procedures for Forensic Skeletal Material* [7] and *Cranial Variation in Man* [12]; the latter describes how to locate cranial landmarks if sutures are obliterated, Wormian or apical bones are present, etc. Observers calibrated their instruments with calibration rods before each measuring session, and the following conditions were modeled to establish the repeatability of the measurements according to the National Institute of Standards and Technology's *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results* [13]:

1. The measurement procedure was performed the same each time.

2. The same observer performed each measurement with the same measuring instrument.

In total, 78 measurements from *Data Collection Procedures* [7] (34 cranial and 44 postcranial) were recorded for each skeleton from the following elements: cranium, mandible, clavicle, scapula, humerus, radius, ulna, femur, tibia, fibula, os coxa, sacrum, and calcaneus. Twenty-one additional measurements were also evaluated, for a total of 99 measurements. The additional measurements tested in this study are defined in Table 1. Three of the 21 additional measurements are craniometric measurements incorporated into the Fordisc 3 software [biasterionic breadth (ASB), mid-orbital width/zygoorbitale breadth (MOW/ZOB), zygomaxillary/bimaxillary breadth (ZMB)]. An alternative method of measuring mastoid length (MDH) using the landmarks porion and mastoidale (see Table 1 for definition) as opposed to visually sighting the tip of the mastoid process [14] is also evaluated. The remaining additional measurements are postcranial measurements chosen because of their potential to reduce subjectivity of existing measurements or because they capture information about highly dimorphic joint dimensions. Maximum and minimum midshaft diameters of the clavicle were included to evaluate alternatives to sagittal and vertical diameters at midshaft, as the latter two measurements have been found to be subject to considerable observer error [15]. Maximum and minimum midshaft diameters of the radius, ulna, and tibia were also evaluated as alternatives to measures defined by anatomical planes. For example, maximum midshaft diameter [16] and minimum midshaft diameter of the ulna were included as a means of eliminating hypothesized error associated with determining the maximum expansion of the crest when measuring dorso-volar and transverse diameters.

2.2 Statistical Analysis

Box and whisker plots and scatterplot matrices were constructed to screen the data for extreme outliers (e.g. data input errors), and Q-Q plots were used to examine normality. Two-way mixed ANOVAs and repeated measures ANOVAs were run in SPSS 23 [17] to examine intraobserver (within-subjects; factor=repeated measurements) and interobserver (between-subjects; factor=observer) variability [18]. Pairwise comparisons were examined for variables with significant main effects. Simple main effects were run for variables with significant interactions between the between- and within-subjects factors, and Greenhouse-Geisser corrections were used for variables that failed Mauchly's test of sphericity.

Absolute and relative technical error of measurement (TEM) was calculated for measurements with significant ANOVA results to determine the degree of the error. TEM was calculated to examine the variability among a single observer repeating a measurement multiple times (e.g. repeatability or intraobserver error), as well as the variability between multiple observers (interobserver error). Absolute TEM is calculated as

$$\sqrt{\frac{\sum_1^N \left[\sum_1^K M(n)^2 - \frac{\left(\sum_1^K M(n)\right)^2}{K}\right]}{N(K-1)}}$$

where N is the sample size (N=50 skeletons), K is the number of observers or the number of repeated rounds per observer (K=4), M is the measurement, and M(n) is the

n^{th} repetition of the measurement [19]. Relative TEM is calculated by dividing absolute TEM by the mean and multiplying by 100. Relative TEM is a measure of precision (or imprecision) unaffected by scale or sample size that allows for the direct comparison of measurements of different scales [19, 20]. Acceptable ranges for the relative, or percent, TEM in anthropometry are <1.5% for intra-examiner error and <2% for inter-examiner error [20]. To calculate intraobserver relative TEM, the relative TEM was calculated for the four measurement rounds on one skeleton, and the average of the 50 relative TEM values was used as the relative TEM. To calculate interobserver relative TEM, relative TEM was calculated for each measurement round using the data from all four observers; the mean relative TEM from all four rounds was used as the relative TEM.

Upon completion of the statistical analyses, observers met to discuss the method they used to obtain the measurements. Measurement definitions in DCP 2.0 were revised or omitted after the statistical results and observer discussions were reconciled.

3. Results and Discussion

The figures and tables in this section highlight measurements flagged as problematic by the analytical procedures, as well as the new measurements included in DCP 2.0. Measurements with insignificant ANOVA results and acceptable relative TEM values (< 2.0) are not reported here; however, this information is available in Appendix C of *Data Collection Procedures 2.0* [9]. Table 2 provides a key to the measurement abbreviations, and Table 3 presents the absolute and relative TEM values for measurements with significant ANOVA results. All variables in Table 3 had significant interobserver error (α=0.05), as also reflected by relative TEM values > 2.0. Twenty-two measurements have greater than acceptable interobserver error (relative TEM>2.0), and several measurements also had significant intraobserver variability among one or more observers. Fifteen of these measurements were included in the standard set of measurements in *Data Collection Procedures for Forensic Skeletal Material, 3rd edition* [7]; seven were part of the alternative measurements investigated in this study (see Table 1).

Pairwise comparisons of observers and of repeated measurements among observers elucidated patterns among some of the measurements. For several measurements only one observer differed significantly from the remaining observers, indicating a likely problem with the measurement definition as interpreted by that observer: glenoid cavity breadth and height, anterior sacral breadth, antero-posterior diameter of the first sacral segment, antero-posterior diameter of the medial condyle of the femur, interorbital breadth, and both mastoid measurements. Often the discrepancy was due to the inclusion of osteophytes in a joint surface measurement or incorrect placement of the calipers. These discrepancies became obvious in the post-analysis conversations among the observers, and the measurement definitions were clarified in DCP 2.0 to eliminate this confusion. The issue with minimum circumference of the distal ulna was only present in one observer's data and may have been due to misreading of the tape measure.

Other measurements displayed errors likely due to ambiguous measurement definitions or landmarks that were difficult to locate consistently: pubis length, ischium length, innominate breadth, minimum breadth of the olecranon, olecranon to coronoid process length, distal epiphyseal breadth of the tibia, transverse and dorso-volar diameters of the ulna, vertical and sagittal diameters of the clavicle. In some cases, a definition correction rectified the issue (e.g. distal epiphyseal breadth of the tibia and innominate breadth—see discussion below), but in others the measurement was omitted altogether from DCP 2.0.

Two general trends emerged from these analyses: (1) maximum lengths and breadths have the lowest error across the board (TEM < 0.5), and (2) maximum and minimum diameters at midshaft are more reliable than their positionally-dependent counterparts (i.e. sagittal, vertical, transverse, dorso-volar) (Fig. 2). As a result, all diameter measurements of long bone shafts were changed to maximum and minimum measures at midshaft in DCP 2.0 (clavicle, radius, ulna, femur, and tibia; humerus and fibula diameters were already specified as such in previous editions). Innominate height and breadth were also specified as maxima. Several new articular dimension measurements were added, with definitions clarified as needed according to the results of this analysis: glenoid cavity breadth and height, maximum breadth of the olecranon, antero-posterior diameter of the first sacral segment, and maximum antero-posterior length of the medial and lateral femoral condyles. "Maximum distal epiphyseal breadth of the tibia" was changed to "distal epiphyseal breadth" because, as defined, the measurement is not a maximum measurement of the distal epiphysis. Pubis length and ischium length were omitted because the landmark denoting the junction of the pubis, ischium, and ilium in the acetabulum cannot be located with any degree of consistency. The definition of dacryon was clarified, and the landmarks porion and mastoidale were added to take mastoid length from two physical landmarks (MDH_{TIP}, as is done with digitizers) rather than visual sighting (MDH_{SIGHT}). Biasterionic breadth (ASB), bimaxillary breadth (ZMB), and mid-orbital width/zygoorbitale breadth (MOW/ZOB—the latter term is preferred) were also added to DCP 2.0 since these measurements are used in Fordisc and were found to be reliable among and between observers. Other changes in DCP 2.0 beyond the scope and relevance of this discussion are explained in the preface [9].

Observer experience also played a role in the ability to consistently reproduce measurements. Average intraobserver relative TEM values of the measurements in Table 3 from lowest to highest were 2.31 (Observer 2), 3.25 (Observer 1), 3.36 (Observer 3), and 3.41 (Observer 4). Observer 2 had the lowest TEM for most measurements, and Observer 4 had the highest TEM most frequently. While Observer 1 had the most experience in number of years (27 years), Observer 2 had more technical training than any other observers. Observer 2 had 14 years of experience, but had measured approximately 900 skeletons (more than any other observer) during this time. Observer 3 had 10 years of experience, and Observer 4 had three years of experience.

Conclusions

This systematic investigation of the reliability of osteometric data documented significant observer error in 15 measurements used in forensic casework and recorded in the Forensic Data Bank. Interobserver error was the primary source of variability; however, some measurements also exhibited significant intraobserver error. A revised edition of *Data Collection Procedures for Forensic Skeletal Material* [7] was created to incorporate these results: *Data Collection Procedures 2.0* [9]. Some measurements were omitted altogether (e.g. pubis length and ischium length), while others were revised to eliminate confusion with the definition or landmarks (e.g. distal epiphyseal breadth of the tibia and transverse diameter of the first sacral segment). Diameter measurements of long bone shafts were changed to maximum and minimum measures at midshaft, as measures of maxima and minima have the lowest TEM values of all measurements investigated. Several articular dimension measurements were also added in DCP 2.0 due to their discriminatory power for sex estimation. Other changes are detailed in the Preface of DCP 2.0, which will be versioned as additional changes in skeletal data collection procedures arise. The manual is available for free download online (https://fac.utk.edu/wp-content/uploads/2016/03/DCP20_webversion.pdf), along with an accompanying instructional video (https://www.youtube.com/watch?v=BtkLFl3vim4).

Data Collection Procedures 2.0 [9] is a first step in incorporating essential reference data on measurement accuracy and precision into forensic anthropology laboratory manuals. Similar in-depth studies are needed to quantify error associated with landmark data used to obtain measurements with a digitizer. In accordance with the recommendations of the NAS report to provide known error rates and promote consistent practices to be integrated into standard operating procedures, this research will establish an accurate reference database of osteometric data and bolster the foundations upon which forensic anthropology methods, research and applications are constructed.

Disclosure of Interests

NL, LMJ, RJ, and SO are authors of the Data Collection Procedures 2.0 Laboratory Manual but have no financial interest and do not stand to benefit from financial gain, as this is a lab manual freely available to the public for download. RL and SO created the Fordisc computer program mentioned herein but do not stand to benefit from financial gain and collect no proceeds from the sale of this software. The University of Tennessee owns the license to this software.

Contribution to Authorship

The study was designed by NL, LMJ, and RLJ. NL, LMJ, SM, and HM collected the data. NL and SM analyzed the data and interpreted results with help from LMJ, HM, RLJ, and SDO. NL wrote the article and all authors commented on and contributed to the manuscript.

Funding

This work was supported by the National Institute of Justice (grant number 2013-DN-BX-K038). The views and opinions in expressed this article are the authors' own and do not reflect the view of the National Institute of Justice, the US Department of Justice, or the US government.

Mendeley Data

The raw data used in this analysis has been published on Mendeley Data: **DOI:** 10.17632/6xwhzs2w38.1

ACKNOWLEDGEMENTS

We would like to thank Charlene Weaver in the University of Tennessee Anthropology Department and Carolyn Gulley and Melissa Miracle in the Lincoln Memorial University grants office for their assistance with administrating the grant that made this research effort possible. We also recognize Neil Ward, the graphic artist who did line drawings from bones and bone images and designed the layout of the DCP 2.0 manual. We also thank the unnamed individuals who donate their remains to the Forensic Anthropology Center and to the dedicated practitioners who submit data to the Forensic Data Bank.

References

[1] Daubert v. Merrell Dow Pharmaceuticals, Inc., in: U.S.S. Court (Ed.) United States Supreme Court, 1993.

[2] N.R. Council, Strengthening Forensic Science in the United States: A Path Forward, The National Academies Press, Washington, DC, 2009.

[3] B.J. Adams, J.E. Byrd, Interobserver variation of selected postcranial skeletal measurements, J Forensic Sci 47(6) (2002) 1193-202.

[4] R. Jantz, S. Ousley, FORDISC 3.0: Computerized Forensic Discriminant Functions, University of Tennessee, Knoxville, 2005.

[5] P. Moore-Jansen, R. Jantz, Data collection procedures for forensic skeletal material, 2nd ed., The University of Tennessee Department of Anthropology and Forensic Anthropology Center, Knoxville, 1990.

[6] P. Moore-Jansen, R. Jantz, Data collection procedures for forensic skeletal material, 1st ed., The University of Tennessee Department of Anthropology and Forensic Anthropology Center, Knoxville, 1986.

[7] P.M. Moore-Jansen, S.D. Ousley, R.L. Jantz, Data collection procedures for forensic skeletal material, 3rd ed., The University of Tennessee Department of Anthropology and Forensic Anthropology Center, Knoxville, 1994.

[8] J.E. Buikstra, D.H. Ubelaker, D. Aftandilian, Standards for data collection from human skeletal remains : proceedings of a seminar at the Field Museum of Natural History, organized by Jonathan Haas, Arkansas Archeological Survey, Fayetteville, Ar, 1994.

[9] N.R. Langley, L. Meadows Jantz, S.D. Ousley, R.L. Jantz, G.R. Milner, Data collection procedures for forensic skeletal material 2.0, 3rd ed., The University of Tennessee Department of Anthropology and Forensic Anthropology Center, Knoxville, 2016.

[10] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, Lawrence Earlbaum Associates, Hillsdale, NJ, 1988.

[11] J. Cohen, Quantitative methods in psychology: A power primer, Psychol Bull 112(1) (1992) 155-159.

[12] W.W. Howells, Cranial Variation in Man: A Study by Multivariate Analysis of Patterns of Difference Among Recent Human Populations, Harvard University Press, Cambridge, MA, 1978.

[13] B.N. Taylor, C.E. Kuyatt, Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. National Institute of Standards and Technology Technical Note 1297, Gaithersburg, MD, 1994.

[14] P. Moore-Jansen, S.D. Ousley, R. Jantz, Data collection procedures for forensic skeletal material, 3rd ed., The University of Tennessee Department of Anthropology and Forensic Anthropology Center, Knoxville, 1994.

[15] N.R. Shirley, Age and sex estimation from the human clavicle: an investigation of traditional and novel methods, Anthropology, University of Tennessee, Knoxville, 2009, p. 142.

[16] H.M. McHenry, R.S. Corruccini, F.C. Howell, Analysis of an early hominid ulna from Omo Basin, Ethiopia, Am J Phys Anthropol 44(2) (1976) 295-304.

[17] I. Corp., IBM SPSS Statistics for Windows, IBM Corp., Armonk, NY, 2014.

[18] E.F. Harris, R.N. Smith, Accounting for measurement error: a critical but often overlooked process, Arch Oral Biol 54 (Suppl 1) (2009) S107-17.

[19] P.K. Mony, S. Swaminathan, J.K. Gajendran, M. Vaz, Quality Assurance for Accuracy of Anthropometric Measurements in Clinical and Epidemiological Studies: [Errare humanum est = to err is human], Indian J Community Med 41(2) (2016) 98-102.

[20] T.A. Perini, G.L. de Oliveira, J.S. Ornellas, F.P. de Oliveira, Technical error of measurement in anthropometry, Rev Bras Med Esporte 11 (2005) 86-90.

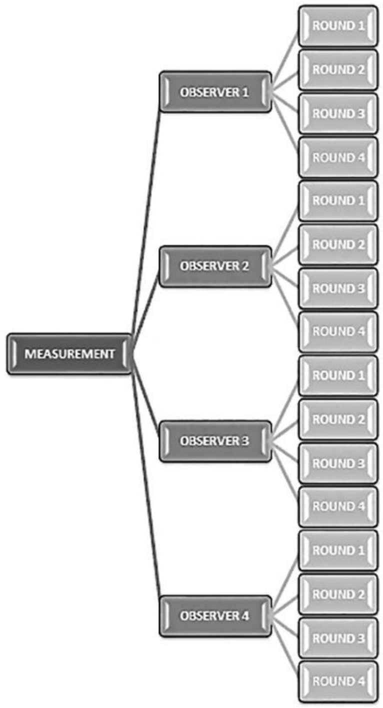

Fig. 1. Schematic representation of data collection procedure for each measurement. This procedure was executed on 50 skeletons; 99 measurements were taken on each skeleton.

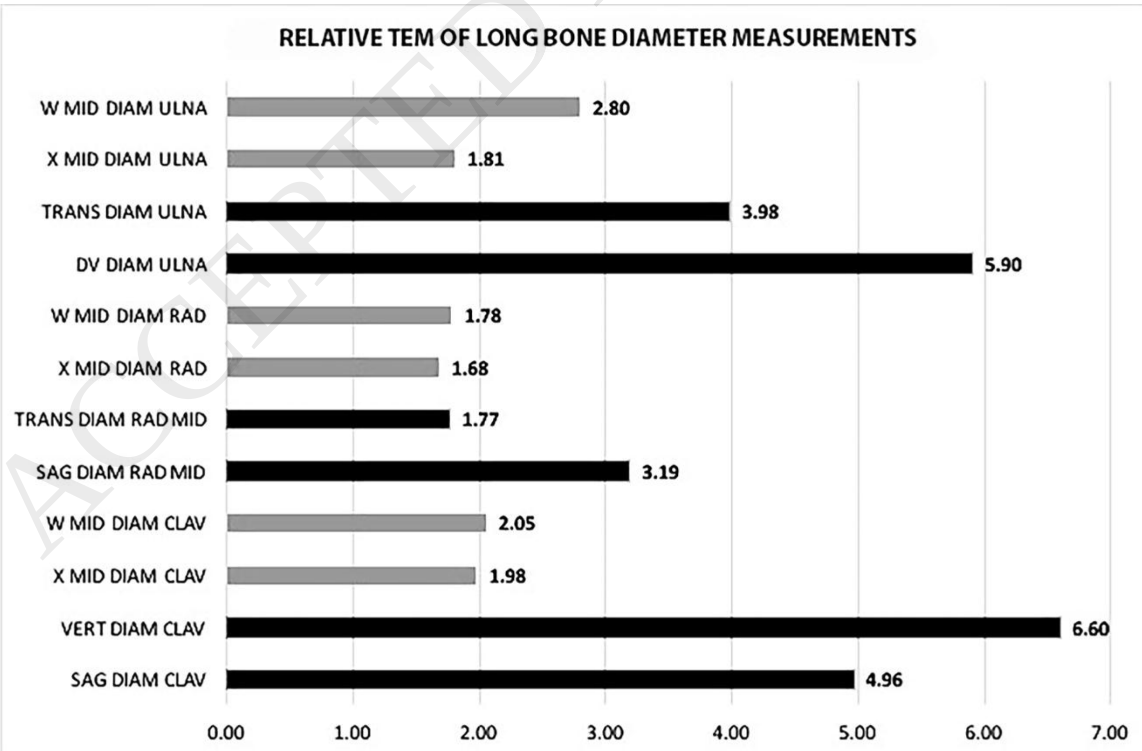

Fig. 2. Relative TEM of Long Bone Diameter Measurements. Intraobserver error of maxima and minima (grey bars) is uniformly less than positionally-dependent diameter measurements (black bars).

Table 1. Additional Measurements. These measurements are not included in the 3rd edition of *Data Collection Procedures* but were evaluated in this study as alternatives to potentially error-prone measurements and/or useful additions to current standards.

CRANIUM

1. **Biasterion breadth (ASB)**: the distance between right and left asterion. Asterion is a landmark at the junction of lambdoidal, parietomastoid, and occipitomastoid sutures. (Howells 1973; Martin and Knussmann 1988)

2. **Zygoorbitale breadth (ZOB)**: the distance between right and left zygoorbitale. Zygoorbitale is a landmark at the junction of the zygomatic bone and the maxilla (i.e. the zygomatico-maxillary suture) at the orbital border. (Jantz and Ousley 2005)

3. **Bimaxillary breadth (ZMB)**: The breadth across the maxillae, from the left to right zygomaxillare anterior (zma). The endpoints of the measurement are located on the facial surface and not on the inferior aspect of the zygomaxillary suture. (Howells 1973)

4. **Mastoid height (MDH)**: The direct distance between porion and mastoidale. Place the fixed arm of the caliper on porion and move the movable arm until it touches mastoidale. This may be most easily accomplished by holding the calipers in a coronal plane. (Jantz and Ousley 2005)

CLAVICLE

5. **Maximum diameter at midshaft**: *Instrument*: sliding caliper. Find midshaft using an osteometric board, mark with a pencil, and rotate calipers around the shaft until the maximum is reached. This is frequently found in the sagittal dimension. (Shirley 2009)
6. **Minimum diameter at midshaft**: *Instrument*: sliding caliper. Find midshaft using an osteometric board, mark with a pencil, and rotate calipers around the shaft until the minimum is found. (Shirley 2009)

SCAPULA

7. **Glenoid Cavity Breadth**: *Instrument:* sliding caliper. Taken at a point just below the constriction of the ventral border. Measured across the breadth of the glenoid cavity from the ventral to the dorsal margin. (Corruccini and Ciochon 1976)
8. **Glenoid Cavity Height**: *Instrument:* sliding caliper. Taken from the superior to the inferior margin of the glenoid cavity, being sure that the measurement is taken perpendicular to glenoid cavity breadth. (Corruccini and Ciochon 1976)

RADIUS

9. **Maximum diameter at midshaft**: *Instrument*: sliding caliper. Find midshaft using an osteometric board, mark with a pencil, and rotate calipers around the shaft until the maximum is found.
10. **Minimum diameter at midshaft**: *Instrument*: sliding caliper. Find midshaft using an osteometric board, mark with a pencil, and rotate calipers around the shaft until the maximum is found.

11. **Maximum Diameter of the Head**: *Instrument:* sliding caliper. Taken from a point on the edge of the articular surface of the bone across to the opposite side. The bone is rotated until the maximum distance is obtained. (Trotter and Gleser 1952)

ULNA

12. **Maximum Diameter at Midshaft**: *Instrument*: sliding caliper. Find midshaft using an osteometric board, mark with a pencil, and rotate calipers around the shaft until the maximum is found.
13. **Minimum Diameter at Midshaft**: *Instrument*: sliding caliper. Find midshaft using an osteometric board, mark with a pencil, and rotate calipers around the shaft until the minimum is found.
14. **Maximum Breadth of the Olecranon Process**: *Instrument:* sliding caliper. Measured from the medial and lateral margins of the olecrenon process' articular surface at its greatest breadth. (McHenry et al. 1976)
15. **Minimum Breadth of the Olecranon Process**: *Instrument:* sliding caliper. Measured from the medial and lateral margins of the olecrenon process' articular surface where the constriction on the medial margin becomes apparent. (Zobeck 1983)
16. **Olecranon Process to Coronoid Process Length**: *Instrument:* sliding caliper. From the most anteriorly projecting point on the olecrenon process to the peak of the coronoid process. (McHenry et al. 1976)

FEMUR

17. **Anterior-Posterior Diameter of the Lateral Condyle**: *Instrument:* sliding caliper. The projected distance between the most posterior point on the lateral condyle and lip of the patellar surface taken perpendicular to the axis on the shaft. (Montagu 1960)
18. **Anterior-Posterior Diameter of the Medial Condyle**: *Instrument:* sliding caliper. The projected distance between the most anterior point on the joint surface and the most posterior point on the medial condyle and the lip of the patellar surface taken perpendicular to the axis of the shaft. (Montagu 1960)

SACRUM

19. **Anterior-Posterior Diameter of S1**: *Instrument:* sliding caliper. Maximum possible diameter of the first sacral vertebra measured by taking one point on the antero-superior border and the other point on the postero-superior border. (Mishra et al. 2003)

TIBIA

20. **Maximum Diameter at Midshaft**: *Instrument*: sliding caliper. The maximum diameter of the tibial shaft at midshaft. This measurement is instrumentally determined but usually located in the anterior-posterior orientation.
21. **Minimum Diameter at Midshaft**: *Instrument*: sliding caliper. The minimum diameter of the tibial shaft at midshaft. This measurement is instrumentally determined but usually located in the medial-lateral orientation.

Table 2. Measurement Abbreviation Key. Abbreviations for measurements discussed in the figures, tables, and results. °definition revised in DCP 2.0 [9] relative to the 3rd edition of *Data Collection Procedures* [7], *new measurement, $^{--}$measurement omitted.

ABBREVIATION	MEASUREMENT
DKB°	interorbital breadth (dacryon-dacryon)
$\text{MDH}_{\text{SIGHT}}^{--}$	mastoid length (after Howells, 1973)
MDH_{TIP}*	mastoid length (porion-mastoidale)
SAG DIAM CLAV^{--}	sagittal diameter of clavicle at midshaft
VERT DIAM CLAV^{--}	vertical diameter of clavicle at midshaft
X MID DIAM CLAV*	maximum diameter of clavicle at midshaft
W MID DIAM CLAV*	minimum diameter of clavicle at midshaft
GLEN CAV BR*	glenoid cavity breadth
GLEN CAV HT*	glenoid cavity height
SAG DIAM RAD MID^{--}	sagittal diameter of radius at midshaft
TRANS DIAM RAD MID^{--}	transverse diameter of radius at midshaft
X DIAM MID RAD*	maximum diameter of radius at midshaft
W DIAM MID RAD*	minimum diameter of radius at midshaft
DV DIAM ULNA^{--}	dorso-volar diameter of the ulna
TRANS DIAM ULNA^{--}	transverse diameter of the ulna
X MID DIAM ULNA*	maximum midshaft diameter of the ulna
W MID DIAM ULNA*	minimum midshaft diameter of the ulna
W CIRCUM ULNA	minimum circumference of the ulna
X BR OLEC ULNA*	maximum breadth of the olecranon process
W BR OLEC ULNA*$^{--}$	minimum breadth of the olecranon process
OLEC-CORON L ULNA*$^{--}$	olecranon to coronoid process length
ANT BR SAC°	anterior breadth of sacrum
TRANS DIAM S1°	transverse diameter of sacral segment 1
AP DIAM S1*	antero-posterior diameter of sacral segment 1
PUBIS L^{--}	pubis length
ISCHIUM L^{--}	ischium length
AP SUBTROCH DIAM FEM	antero-posterior subtrochanteric diameter of femur
TRANS SUBTROCH DIAM FEM	transverse subtrochanteric diameter of femur
AP DIAM LAT COND FEM*	antero-posterior diameter of lateral femoral condyle
AP DIAM MED COND FEM*	antero-posterior diameter of medial femoral condyle
EPICOND BR DIST TIB°	distal epiphyseal breadth of tibia

Table 3. TEM Values for Inter- and Intraobserver Error. Relative and absolute TEM values are provided, the latter in parentheses. Relative (%) TEM is unitless; absolute TEM is in cm. °definition revised in DCP 2.0 [9] relative to the 3rd edition of *Data Collection Procedures* [7], *new measurement, --measurement omitted.

CRANIAL/MANDIBULAR MEASUREMENTS	**INTEROBS TEM**	**INTRAOBS 1 TEM**	**INTRAOBS 2 TEM**	**INTRAOBS 3 TEM**	**INTRAOBS 4 TEM**
DKB°	4.70 (0.97)	3.21 (0.69)	1.39 (0.27)	4.46 (0.92)	5.31 (1.10)
MDH$_{TIP}$*	3.34 (1.03)	3.10 (0.96)	2.86 (0.90)	2.49 (0.76)	3.10 (0.95)
MDH$_{SIGHT}$--	4.35 (1.29)	3.13 (0.93)	2.62 (0.79)	2.41 (0.72)	3.17 (0.89)
POSTCRANIAL MEASUREMENTS					
SAG DIAM CLAV--	4.96 (0.60)	4.21 (0.53)	2.17 (0.26)	3.86 (0.46)	3.89 (0.47)
VERT DIAM CLAV--	6.60 (0.71)	7.40 (0.78)	3.07 (0.34)	5.01 (0.54)	4.84 (0.51)
GLEN CAV BR*	3.48 (0.99)	2.25 (0.65)	1.78 (0.49)	2.32 (0.65)	3.80 (1.10)
GLEN CAV HT*	2.97 (1.12)	2.01 (0.75)	2.09 (0.77)	1.98 (0.74)	1.80 (0.70)
SAG DIAM RAD MID--	3.19 (0.38)	5.95 (0.75)	1.13 (0.14)	2.88 (0.35)	4.03 (0.50)
DV DIAM ULNA--	5.90 (0.79)	4.15 (0.57)	4.08 (0.53)	6.31 (0.86)	2.21 (0.28)
TRANS DIAM ULNA--	3.98 (0.62)	2.61 (0.44)	1.93 (0.32)	3.16 (0.53)	2.31 (0.40)
W CIRCUM ULNA	3.25 (1.12)	4.21 (1.54)	1.99 (0.70)	2.50 (0.88)	1.78 (0.60)
W BR OLEC ULNA*--	14.55 (1.76)	8.45 (1.64)	6.04 (1.07)	8.83 (1.30)	9.57 (1.69)
OLEC-CORON L ULNA*--	8.41 (1.73)	2.45 (0.61)	2.11 (0.58)	1.82 (0.45)	3.44 (0.84)
ANT BR SAC°	5.70 (3.79)	2.08 (2.12)	1.61 (1.67)	2.03 (2.10)	1.75 (1.90)
TRANS DIAM S1°	5.97 (2.94)	1.83 (1.00)	1.50 (0.77)	1.85 (0.91)	5.97 (3.14)
AP DIAM S1*	2.06 (0.68)	1.75 (0.58)	0.79 (0.26)	1.43 (0.46)	1.09 (0.36)
PUBIS L--	6.95 (6.19)	3.02 (2.81)	2.88 (2.42)	4.47 (3.83)	5.02 (4.09)
ISCHIUM L--	6.67 (6.20)	2.41 (2.21)	3.73 (3.19)	5.86 (4.77)	2.06 (1.89)
AP SUBTROCH DIAM FEM	3.90 (2.01)	2.01 (0.55)	2.33 (0.66)	3.57 (1.02)	3.89 (1.08)
TRANS SUBTROCH DIAM FEM	3.67 (1.78)	2.08 (0.63)	2.81 (0.85)	3.03 (0.94)	3.80 (1.19)
AP DIAM MED COND FEM*	2.36 (1.31)	1.27 (0.81)	0.91 (0.57)	1.24 (0.79)	1.12 (0.73)
EPICOND BR DIST TIB°	4.83 (2.13)	1.92 (1.00)	1.04 (0.51)	2.51 (1.30)	0.96 (0.47)