The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title:	**Evaluation of Osteometric Measurements in Forensic Anthropology**
Author(s):	**Natalie R. Langley, Ph.D., Lee Meadows Jantz, Ph.D.**
Document Number:	**250465**
Date Received:	**December 2016**
Award Number:	**2013-DN-BX-K038**

FINAL SUMMARY REPORT

for award number 2013-DN-BX-K038

Evaluation of Osteometric Measurements in Forensic Anthropology

OMB No. 1121-0329
Submitted April 16, 2016

Principal Investigators:

Natalie R. Langley, Ph.D., Principal Investigator
Associate Professor
Lincoln Memorial University
DeBusk College of Osteopathic Medicine
6965 Cumberland Gap Parkway, 430 Math and Science Building
Harrogate, TN 37752
Telephone: (423) 869-6363
Fax: (423) 869-7174
natalie.langley@lmunet.edu
Adjunct Assistant Professor, University of Tennessee
Department of Anthropology

Lee Meadows Jantz, Ph.D., Co-Principal Investigator
Adjunct Associate Professor and Associate Director of the Forensic Anthropology Center
Department of Anthropology
University of Tennessee
250 South Stadium Hall
Knoxville, TN 37996
Telephone: (865) 974-4408
ljantz@utk.edu

Keywords: forensic science, forensic anthropology, osteometric, craniometric, observer error, measurement error, NAS Report, *Daubert*

Table of Contents

Purpose of the Project 1

Implications for US Criminal Justice Policy and Practice 2

Project Design 3

Materials and Methods 3

Results 7

Dissemination of DCP 2.0 8

References Cited 10

"Anthropologists have followed the trend among scientists in general in placing increasing emphasis on measurement data as the fuel for their analytical fire. One of the reasons for this emphasis is the alleged reliability of measurements and the results of the mathematical analyses of the same. We have formally evaluated the efficacy of this belief on the part of biological anthropologists by measuring the variability in selected skeletal measurements due to interobserver error. Where this variation is found to be inordinately high, then our faith is misplaced." (Adams and Byrd 2002)

Purpose of the Project

This research evaluates the reliability and repeatability of osteometric data and provides the forensic anthropology community with error rates for standard skeletal measurements. The investigators have delivered a new and considerably revised open access online edition of a widely-used laboratory manual (*Data Collection Procedures for Forensic Skeletal Material, Moore-Jansen et al. 1994*) that was previously available only in hard copy for purchase. An instructional video was made to accompany the manual and placed on YouTube to facilitate training opportunities for national and international audiences.

The use of metric data in forensic case analyses is increasingly common, particularly in light of the judicial atmosphere since *Daubert* (1993). Methods employing metric data are considered more objective than nonmetric techniques that require visual assessments of skeletal form. Metric data also form the basis of the Fordisc (Jantz and Ousley 2005) computer program used by forensic anthropologists in the United States and provide a straightforward means to quantify error. Many components of a forensic anthropology case report are derived from osteometric data (e.g. sex, ancestry, and stature). Error associated with any method that utilizes osteometric data is further

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

compounded by the error inherent in a given measurement or set of measurements, whether a function of the observer, the instrumentation, or both. Measurement error can be minimized by using appropriate instrumentation, understanding the measurement definition, and by using highly reliable and repeatable measurements. Knowing the reliability of a given measurement provides a foundation from which to proceed with metric estimations of sex, ancestry, and stature, as well as method development.

Implications for US Criminal Justice Policy and Practice

On a broad scale, the results of this work provide foundational knowledge for forensic case analyses, research, data collection, and method development. The measurements in the *Data Collection Procedures 2.0* (Langley et al. 2016) interface with the Fordisc (Jantz and Ousley 2005) computer software used to derive a biological profile of unidentified remains (sex, ancestry, stature). The DCP 2.0 is also the vehicle through which the managers of the Forensic Data Bank (FDB, a database of osteometric data from modern skeletons) obtain osteometric data for the Fordisc reference database. In the spirit of non-proprietary data, FDB data is available to anyone who wishes to use it for research purposes, and the DCP 2.0 is now also available online at no cost. The manual will be versioned going forward (2.0, 3.0, etc.), and new versions will be uploaded to the University of Tennessee Forensic Anthropology Center web page and distributed to the forensic anthropology community at conferences and on social media.

The first edition of the *Data Collection Procedures* manual was released in 1988 as the product of NIJ Grant No. 85-IJ-CX-0021 with the goal of standardizing recording procedures to facilitate osteometric data collection for the Forensic Data Bank. The second edition was released in 1990, and the FDB had 850 cases. The third edition was released in 1994; the FDB had 1,200 cases. The FDB now has 4,000 cases, and the Fordisc software has undergone revisions and new versions, but the *Data Collection Procedures* has not been revised. This revision (*DCP 2.0*) introduces essential reference data on measurement accuracy and precision into forensic anthropology laboratory manuals. The updated measurements will be incorporated into the Fordisc software. By working with the Fordisc developers, the principal investigators ensure that the results of this grant work will reach forensic practitioners. In accordance with the recommendations of the NAS report to provide known error rates and promote consistent practices that can be integrated into standard operating procedures, this research will strengthen forensic anthropology as it moves forward as a scientific discipline. Ultimately, the impact on the criminal justice system will be an improved and more accurate reference database and methods for identifying unknown decedents.

Project Design

Materials and Methods

Osteometric data was 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 and Observer 4 having the least experience. Ninety-nine measurements were taken on each skeleton using spreading calipers, digital sliding calipers, a tape measure, an osteometric board, and a mandibulometer. Instruments were calibrated with calibration rods before each measuring session. 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* (Moore Jansen et al. 1994) and *Cranial Variation in Man* (Howells 1978); the latter describes how to locate cranial landmarks in the event that sutures are obliterated, Wormian or apical bones are present, etc. **Figure 1** provides a schematic of the data collection design for each measurement evaluated in this project.

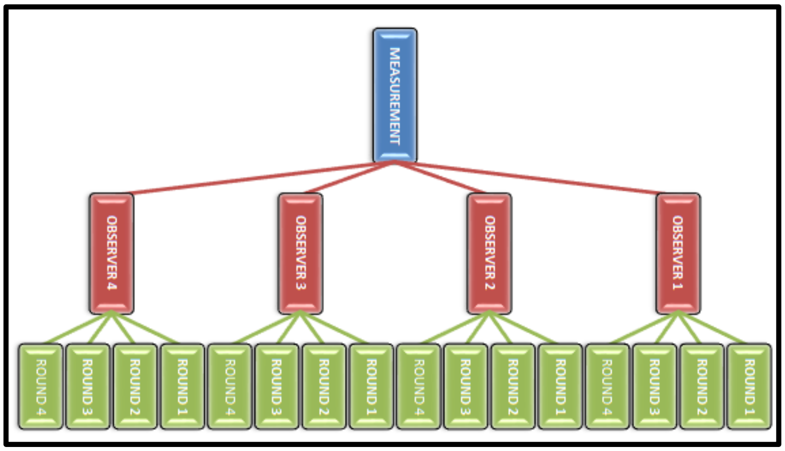

Figure 1. Schematic representation of data collection procedure for each measurement tested in this study.

In total, 78 standard measurements from *Data Collection Procedures* (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 evaluated in this project, as well, for a total of 99 measurements. Three of the 21 additional measurements are craniometric measurements incorporated into the Fordisc 3 program [biasterionic breadth (ASB), mid-orbital width (MOW), zygomaxillary breadth (ZMB)], and one measurement is an alternative method of measuring mastoid height (MDH). The remaining additional measurements are postcranial measurements chosen because of their potential to reduce the subjectivity of the currently used standard measurements or because they capture information about highly dimorphic joint dimensions. Maximum and minimum midshaft diameters of the clavicle (Shirley 2009) 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 in informal analyses. Maximum and minimum midshaft diameters of the radius, ulna, and tibia were also evaluated as alternatives to measures defined by anatomical planes.

The data were analyzed to assess measurement reliability and repeatability. Unacceptable values were inspected to ascertain if the problem was the measurement definition, observer experience, or the specific landmark(s) used to define the measurement. Repeated measures ANOVAs were run in NCSS 10 (Hinze 2016) to examine intraobserver (within-subjects) and interobserver (between-subjects) variability simultaneously. Mauchly's test of compound symmetry was used to determine if the sample covariance matrix violated the assumptions of the repeated measures ANOVA. Repeated measures ANOVAs can elucidate at which level most of the variation is occurring (interobserver or intraobserver) and give insight into what processes may be causing this variation (experience level, problematic measurement, etc.). If the ANOVA detected significant effects, Tukey-Kramer post hoc tests were be used to examine the significant differences.

The relative technical error of measurement (TEM) was calculated to assess reliability between observers. TEM represents the variability encountered between observers when the same measurements are taken multiple times. This commonly used measure of precision or imprecision is unaffected by scale or sample size and allows for the direct comparison of measurements of different scales. TEM is calculated as

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

where N is the sample size, K is the number of observers, M is the measurement, and M(n) is the nth repetition of the measurement. Relative TEM is calculated by dividing TEM by the mean and multiplying by 100. Acceptable ranges for the relative, or percent, TEM in anthropometry are <1.5% for intra-examiner error and <2% for inter-examiner error (Perini et al. 2005).

Scaled Error Index (SEI) was calculated to examine intraobserver error and determine if measurement variability was constant for each observer for the four rounds of measurement. Like the TEM, SEI also permits comparison of measurements regardless of scale. This statistic was used by Adams and Byrd (2002) in their evaluation of osteometric measurement error in a select number of postcranial measurements. To calculate SEI, the absolute value of the difference between a single measurement and the median of the four repeated measurements is divided by the median. This value is multiplied by 100 to convert it to the percent error from the median.

Results

Two general trends in the data significantly impacted the revisions to the *DCP*: maximum lengths and breadths have the lowest error across the board, and maximum and minimum diameters at midshaft are more reliable than their positionally-dependent counterparts (i.e. sagittal, vertical, transverse, dorso-volar). The repeated measures ANOVAs indicated problems with mastoid height, anterior sacral breadth, transverse diameter of the first sacral segment, ischium length, pubis length, distal tibial breadth, and the olecrenon-coronoid length of the ulna. In addition, several mandibular measurements (mandibular angle, maximum ramus height, mandibular body breadth, and maximum ramus breadth), mid-orbital width, interorbital breadth, and sixteen postcranial measurements had relative TEM>2.0%, indicating greater than acceptable interobserver error in these measurements. All of these measurements were scrutinized closely for issues with definitions and landmarks. In some cases (e.g. distal tibial breadth, transverse diameter of the first sacral segment, anterior sacral breadth) the measurement definition was unclear, and we were able to correct this and eliminate the subjectivity. In other cases (e.g. pubis length, ischium length, maximum ramus breadth) the landmarks used to define the measurement are highly variable and inconsistent; these measurements were removed from the DCP.

All positionally-dependent measurements were abandoned, and maxima and minima were substituted except for the antero-posterior and medial-lateral subtrochanteric diameters of the femur. These were retained because of their discriminatory utility for some analyses, but a note was added about low inter-rater agreement. Most of the new measurements of articular dimensions were added to the DCP 2.0. These measurements were highly reliable and have potential for accurate discrimination of sex (e.g. maximum radial head diameter, maximum olecranon breadth, and maximum antero-posterior diameters of the femoral condyles). In addition, nine new reliable measurements of the pelvis were added. All of the amendments in the DCP 2.0 relative to the 1994 edition are listed in the preface, and the error data for each measurement (inter- and intraobserver error) is included in appendices for easy reference.

Dissemination of DCP 2.0

Once data collection and analyses were completed, the observers met to discuss measurements. Measurements with significant differences in the repeated measures ANOVA, high TEM values, or high SEI values were were either eliminated or revised. Observers discussed the results and revised definitions with Drs. Richard Jantz and Steve Ousley, authors of the Fordisc software, to ensure that *DCP 2.0* is compatible with the latest version of the software. The Principal Investigator is fluent in German and checked all existing translations of all Martin and Knussman (1988) definitions and translated the new inclusions, as well.

Though not in the proposed scope of work, the Principal Investigators also met with Dr. George Milner to acquire the new age estimation standards that will be included in the next release of the Fordisc software. The PIs felt that this section should be included in *DCP 2.0* so that the final product is compatible with Fordisc and, ultimately, with the data that will be collected for the Forensic Anthropology Data Bank (FDB). During the course of the grant, the PIs met regularly with a graphic artist to revise the measurement definitions and figures and format the new age estimation materials. The graphic artist redesigned the manual into a two-column format, re-drew all of the existing line drawings, and created new drawings for new measurements.

The DCP 2.0 was officially released at the American Academy of Forensic Sciences 2016 Annual Meetings in Las Vegas Nevada. An oral presentation was presented at the NIJ Grantees Meeting, and two posters were presented at the AAFS meetings (Langley et al. 2016; Ousley et al. 2016). The PIs distributed 100 free printed copies of the manual to practitioners who have consistently contributed data to the FDB since its inception. QR codes that linked to a pdf of the DCP 2.0 and to the instructional video were placed on each poster and distributed on small business cards. The QR code was also placed on the University of Tennessee Forensic Anthropology Center web page and Facebook site. Since the February release the manual has received 123 scans, and the video has received 50 scans. Colleagues have also distributed the manual in Algeria and Nepal. Currently, we are arranging for Spanish and French translations of the video and hope to be able to do the same with the manual. **Figure 2** contains the QR codes for the DCP 2.0 and instructional video. We intend to continue to make these materials freely available to practitioners and researchers to encourage and facilitate osteometric data collection worldwide and contributions to the Forensic Data Bank.

Figure 2. DCP 2.0 manual in pdf format (right) and instructional video (left).

References Cited

Adams BJ, and Byrd JE. 2002. Interobserver variation of selected postcranial skeletal measurements. J Forensic Sci 47(6):1193-1202.

Daubert v. Merrell Dow Pharmaceuticals, Inc. 1993. In: Court USS, editor: United States Supreme Court.

Hintze, J. 2016. NCSS 10. NCSS, LLC. Kaysville, Utah.

Howells WW. 1973. Cranial Variation in Man. Papers of the Peabody Museum of Archaeology and Ethnology, 67. Harvard University, Cambridge Massachusetts.

Jantz RL and Ousley SD. 2005. FORDISC 3.0: Computerized Forensic Discriminant Functions. Knoxville: University of Tennessee.

Langley NR, Meadows Jantz L, Ousley SD, Jantz RL, Milner G. 2016. Data Collection Procedures for Forensic Skeletal Material 2.0. Knoxville: University of Tennessee.

Langley NR, Meadows Jantz L, McNulty S, Maijanen H, Ousley SD, Jantz RL. (2016). DCP 2.0: Changes in Data Collection Procedures for Forensic Skeletal Material. Proceedings of the American Academy of Forensic Sciences Annual Meetings in Las Vegas, NV 22: 95.

Martin R and R Knussmann. 1988. Anthropologie: Handbuch der vergleichenden Biologie des Menschen. Stuttgart: Gustav Fischer.

Moore-Jansen PH, Ousley SD, and RL Jantz. 1994. Data Collection Procedures for Forensic Skeletal Material. Knoxville, TN: University of Tennessee.

Ousley SD, Jantz RL, Langley NR, Spradley MK, Dudzik B. (2016). Thirty Years of the Forensic Data Bank and Data Collection Procedures (DCP) 2.0: Continuity and Transformation. Proceedings of the American Academy of Forensic Sciences Annual Meetings in Las Vegas, NV 22: 157.

Perini TA, de Oliveira GL, Ornellas JS, de Oliveira FP. 2005. Technical error of measurement in anthropometry. Rev Bras Med Esporte 11: 86-90.

Shirley, NR. (2009). Age and sex estimation from the human clavicle: An investigation of traditional and novel methods. University of Tennessee, Knoxville.