Inter- and Intra-observer Error in Forensic Anthropology: Based upon Langley et al. 2016 *Data Collection Procedures for Forensic Skeletal Material 2.0* for the Cranium and Femur

Maggie M. Klemm
University of Nebraska-Lincoln, mklemm2@unl.edu

Dakota L. Taylor
University of Nebraska-Lincoln, dakota.taylor@unl.edu

Nebraska Anthropologist (2024) Volume 30

Inter- and Intra-observer Error in Forensic Anthropology: Based upon Langley et al. 2016 *Data Collection Procedures for Forensic Skeletal Material 2.0* for the Cranium and Femur

Maggie M. Klemm & Dakota L. Taylor

Abstract

Following the Daubert v. Merrell Dow Pharmaceuticals (1993) ruling three decades ago, guidelines were established for determining the reliability and validity of expert testimony in court. These guidelines include considerations of whether the theory and methodologies employed have been tested, the establishment of potential or known error rates, and the existence and maintenance of the standards (Christensen, 2004). Testifying in court as an expert witness has become a critical role as a forensic anthropologist. As such, the methodologies, techniques, and standards utilized by forensic anthropologists must be carefully considered and adhere to the Daubert guidelines within the court system. One of the established standard manuals is the *Data Collection Procedures for Forensic Skeletal Material 2.0* (Langley et al. 2016). This manual provides the most up-to-date data collection procedures and instructions for taking osteometric measurements of skeletal landmarks, providing the basis for estimates of biological sex, population affinity, and stature of unknown human remains. Within forensic anthropology, the use of osteometric data is frequent, yet little research has explored both interobserver and intraobserver variability for various measurements, as well as testing the reliability and accuracy of different measurement devices.

In this study, two observers took repeated measurements of the femora and crania of two skeletal casts. The observers adhered to the instructions provided in the *Data Collection Procedures for Forensic Skeletal Material 2.0* (Langley et al. 2016) when taking these measurements. The aim of this study is to quantify the intra- and interobserver variability expressed in these measurements and assess the thoroughness of the instructions and descriptions provided in the data collections manual. Overall, it was observed that the crania measurements reflected higher rates of inter- and intraobserver variability, whereas the femora measurements showed the least

variability. Discrepancies were also observed between the units of measurements for certain tools (i.e., sliding and spreading calipers measure in centimeters while the osteometric board measures in millimeters). This data is important as it sparks conversation concerning the variability and accuracy of osteometric measurements in connection with the directions provided in the standard manual. This research will be expanded in the future to include more skeletal casts and observers to quantify the inter- and intraobserver error rates and inter-device reliability with a larger sample dataset.

Introduction

In forensic anthropology, one of the standard data collection manuals that anthropologists utilize is the *Data Collection Procedures for Forensic Skeletal Material 2.0* written by Langley et al. (2016). The latest edition of this manual was written to appropriately line up with databases, such as FORDISC 3.1 (Langley et al. 2016). FORDISC is a program frequently used in forensic casework to estimate biological sex, population affinity, and stature through discriminant function analysis of measurements. The latest edition of the *Data Collections Procedures* manual was written to update the skeletal analysis methods and provide more detailed descriptions of the different measurements. The stated goals of this version are to provide more detailed instructions and descriptions of the measurements and skeletal landmarks, simplify the measurement process, and remove measurements that are deemed problematic due to the location of the landmarks. Further, the authors indicate the manual should assist in teaching standardized measurements to individuals of varying experience levels, make the measurements easily replicable, and account for any interobserver error that might affect error ranges or affect biological profile estimates (Langley et al. 2016).

Due to the changes made to this version, the goal of this research is in the quantitation of the interobserver and intraobserver variability of measurements taken by two observers. The measurements taken were of the crania and right femora of a European Female (denoted as “1” in graphs) and an Asian Female (denoted as “2” in graphs). Intraobserver variability refers to the difference between multiple measurements taken by one individual, whereas interobserver variability refers to the difference between multiple individuals performing the same measurement (Henson et al. 2020). Despite both observers having similar experience and education levels, it is hypothesized there will be differences observed between the measurements taken by the two observers, and there will be discrepancies between the observers' first and second attempts at taking the measurements. This research will also focus on the inter-device variability of measuring devices frequently utilized by forensic anthropologists. Inter-device variability refers to the difference between multiple measurement devices performing the same measurement (Henson et al., 2020). It is hypothesized there will be discrepancies observed in the measurements collected by the different devices. This is due to the limited malleability of some devices and the fact that certain measurements require the tool to be maneuvered and will pose difficulties due to the overall shape of the bone. Overall, this research will demonstrate if Langley et al. (2016) effectively updated their manual, making it easier to use by individuals who are not fully trained, or if there needs to be another revision that explains in simpler terms or provides more detailed photographs on how to perform these measurements.

Background

Under the Daubert Standard (1993), guidelines were established to determine the validity of expert testimony and the employed scientific techniques. It is important that forensic anthropologists are mindful of the Daubert guidelines to ensure the utilized methodologies are admissible in the instance the techniques are considered in court (Christensen, 2004). Researchers within the forensic science community have focused on instituting ways to evaluate and validate research, while establishing known error rates for different methodologies within the field. Within forensic anthropology, the use of osteometric data is frequent, yet little research has explored the variability and error rates, both interobserver and intraobserver, for various measurements, as well as testing the reliability and accuracy of measurement devices. As stated in Langley et al. (2018), measurement error can be reduced by utilizing the appropriate instrumentation, carefully reading and understanding instructions for specific measurements, sufficient training, and using reliable and repeatable measurements. Further, understanding measurement reliability provides the foundation for metric estimations of aspects for building the biological profile of unidentified individuals, including biological sex, population affinity, and stature (Langley et al., 2018, 184).

Data Collection Procedures for Forensic Skeletal Material 2.0 (2016) is a manual providing the most up-to-date data collection procedures and instructions for taking measurements of skeletal landmarks which are the basis for estimates of biological sex, population affinity, and stature of unknown remains (Langley et al., 2016). An existing gap in the literature pertains to the accuracy and reliability of different measurement instruments, including digital calipers, spreading calipers, and manual sliding calipers. The manual produced by Langley et al. (2016) specifies the instrument that should be utilized to take the different measurements. However, some measurements indicate the potential use of two instruments, allowing the observer to determine the most appropriate instrument to utilize.

Data and Analytic Methods

For this study, two observers measured two crania and right femora to test the interobserver and intraobserver variability. The crania and right femora were casted by Bone Clones (Bone Clones, Inc) and represent a European Female and Asian Female. When conducting the measurements, it was decided by the observers to use the right femora for consistency. The casts and measurement tools were in the forensic lab on the University of Nebraska- Lincoln Campus. The measurements taken were those established in Langley et al. (2016) *Data Collection Procedures for Forensic Skeletal Material 2.0*, the standard most forensic anthropologists use for casework.

For the crania, twenty-eight different measurements were recommended to be taken, according to the *Data Collection Procedures for Forensic Skeletal Material 2.0* (Langley et al.

2016). For each measurement, descriptions of the landmarks and recommendations of the measuring device are provided. While most descriptions have one option, sliding or spreading caliper, some measurements state that either tool is acceptable (Langley et al. 2016). In this scenario, the observers decided to use both the digital and manual sliding calipers, as well as spreading calipers to determine if the measurements differed depending on which tool was used. For consistency in this study, the same set of tools were used by both observers, and measurements of the mandible were not taken in this study. The sliding calipers and spreading calipers were made by GPM Swiss made, while the digital caliper is made by ProDent USA. The sliding calipers are designated by the letter “S”, the digital calipers are designated by the letter “D”, and the spreading calipers are designated by the word “Spread” in the graphs.

For the femora, eleven different measurements were recommended to be taken, according to the *Data Collection Procedures for Forensic Skeletal Material 2.0* (Langley et al. 2016). The same set of sliding calipers, digital and manual, that were used for the crania were also used for the femur along with an osteometric board with a standardized ruler used by the observers. Again, the sliding calipers and spreading calipers were made by GPM Swiss made, while the digital caliper is made by ProDent USA. The osteometric board is custom-made by Restoration Fine Woodworking and Upholstery in Hawaii and is made of wood and contains a NIST traceable ruler for Dr. William Belcher.

The measurements were taken over three days in the span of a week, and each observer performed the measurements on their own. Both observers had similar levels of experience in the field with some casework and research experience, as well as both having Bachelor of Science degrees in Anthropology with concentrations in Forensic Anthropology from Radford University (Observer A), and Western Carolina University, (Observer B).

These measurements were then brought into RStudio to create graphs that represent the inter- and intraobserver variability for each bone. The first and second measurements collected by Observer A and Observer B are represented in the bar plots below (Figures 1 and 2). The x-axis represents the titles of the different measurements that were taken, while the y-axis reflects the measurements in millimeters, as that is the standard measurement parameter in forensic anthropology. The code used for the creation of the graphs can be seen below in Appendix A (Figure 11). Bar plots representing the interobserver rates for each observer's two attempts were averaged and then graphed in RStudio. These bar plots reflect the averages of each observer for cranium one (Figure 12), cranium two (Figure 13), right femur one (Figure 14), and right femur two (Figure 15).

Results

Interobserver Variability of Cranium One

The measurements taken by the observers in the two trials from the cranium one was fairly consistent, within 2 millimeters of each other. Four cranial measurements were observed to have variability higher than 2 mm, falling outside the acceptable range of error as established by the American Board of Forensic Anthropology (ABFA). The measurement with the highest variability from cranium one is the digital caliper measurement of the Biauricular Breadth, or AUB-D. The calculated spread between the highest and lowest observed measurement of AUB-D was 9.88 mm. Other measurements observed with high interobserver variability are MAB (Maxillo- Alveolar Breadth), PAC-S (Parietal Chord), and ASB-S (Biasterionic Breadth). The measurement MAB is taken with the spreading calipers and PAC-S and ASB-S are both taken with the manual sliding calipers. The calculated spread between the highest and lowest observed measurements of the MAB, PAC-S, and ASB-S are 4 mm, each. Observed in this graph are noticeable gaps in the data for measurement ZYB-D (Bizygomatic Breadth), FRC-D (Frontal Chord), PAC-D (Parietal Chord), and ASB-D (Biasterionic Breadth). These gaps are attributed to the digital calipers measuring a maximum distance of 104.5 mm, and these specific measurements were longer than this.

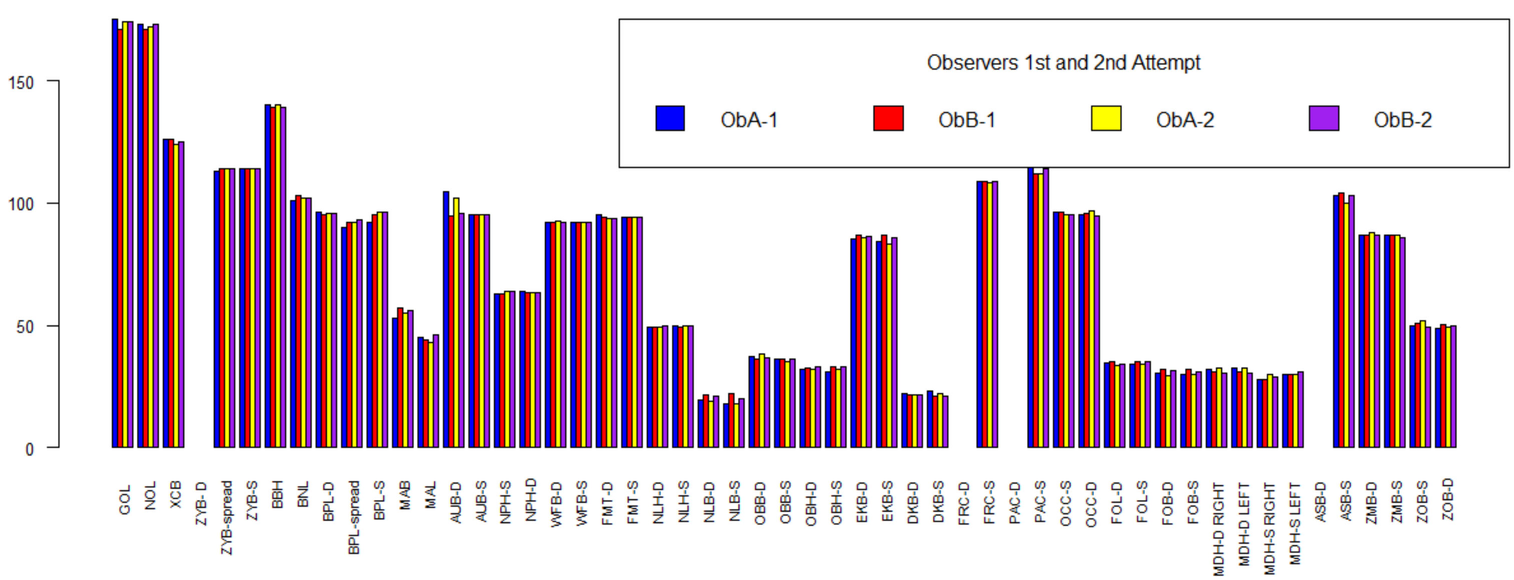

Figure 1: Observer A versus Observer B Cranium 1 Measurements (Interobserver Error)

Interobserver Variability of Cranium Two

Similar to the measurements taken from cranium one, the measurements for cranium two were consistent. Five measurements reflected significant variability, falling outside the acceptable range of variability established by the ABFA. The measurement with the highest observable variability was the Parietal Chord, or PAC-S, taken with the manual sliding calipers. The calculated variability between the highest and lowest measurements of the PAC-S was 7 mm. Another set of measurements with high variability was the Occipital Chord, or the OCC-S and OCC-D, measured with the manual sliding calipers and digital calipers, respectively. The spread for OCC-S was 5 mm and 4.46 mm for the OCC-D. The high variability in these measurements can largely be attributed to measurement descriptions and the landmarks on the cranium itself (Figures 9 and 10). Another measurement observed with higher variability from cranium two was the Basion- Prosthion Length, shortened to BPL- D and measured with the manual digital calipers. The calculated difference between the highest and lowest observed measurements of BPL-D was 4.27 mm. The final measurement with higher variability was the Biauricular Breadth, shortened to AUB- S and measured with the manual sliding calipers. The spread from the highest to the lowest measurement taken for AUB-S was 4 mm. It is worth noting that the observers were not able to measure AUB (Biauricular Breadth), ZYB (Bizygomatic Breadth), and PAC (Parietal Chord) using the digital calipers because the digital calipers measure a maximum distance of 104.5 mm.

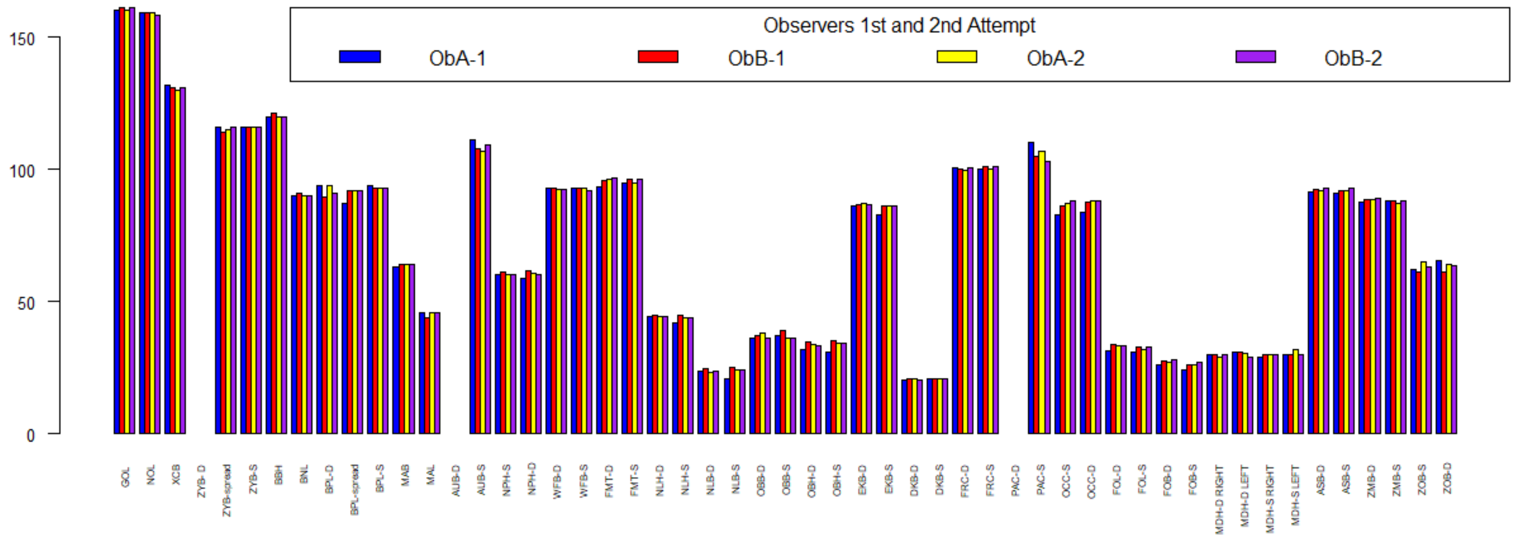

Figure 2: Observer A versus Observer B Cranium 2 Measurements (Interobserver Error)

Interobserver Variability for Right Femora One

The interobserver variability for the right femora one between both observers' first and second attempts at measurements was less compared to the variability of the crania (Figure 3). Only one femora measurement reflected significant variability, falling outside the acceptable range of variability. The measurement with the highest spread was the transverse subtrochanteric diameter of the femur, measured with the manual sliding calipers, with a range of 4 mm. Three femora measurements were at the acceptable range of variation of 2 mm. These measurements were the anterior- posterior subtrochanteric diameter of the femur, the maximum antero-posterior length of the lateral condyle, and maximum antero-posterior length of the medial condyle. The rest of the measurements taken from femora one by the two observers were within the acceptable range of variation.

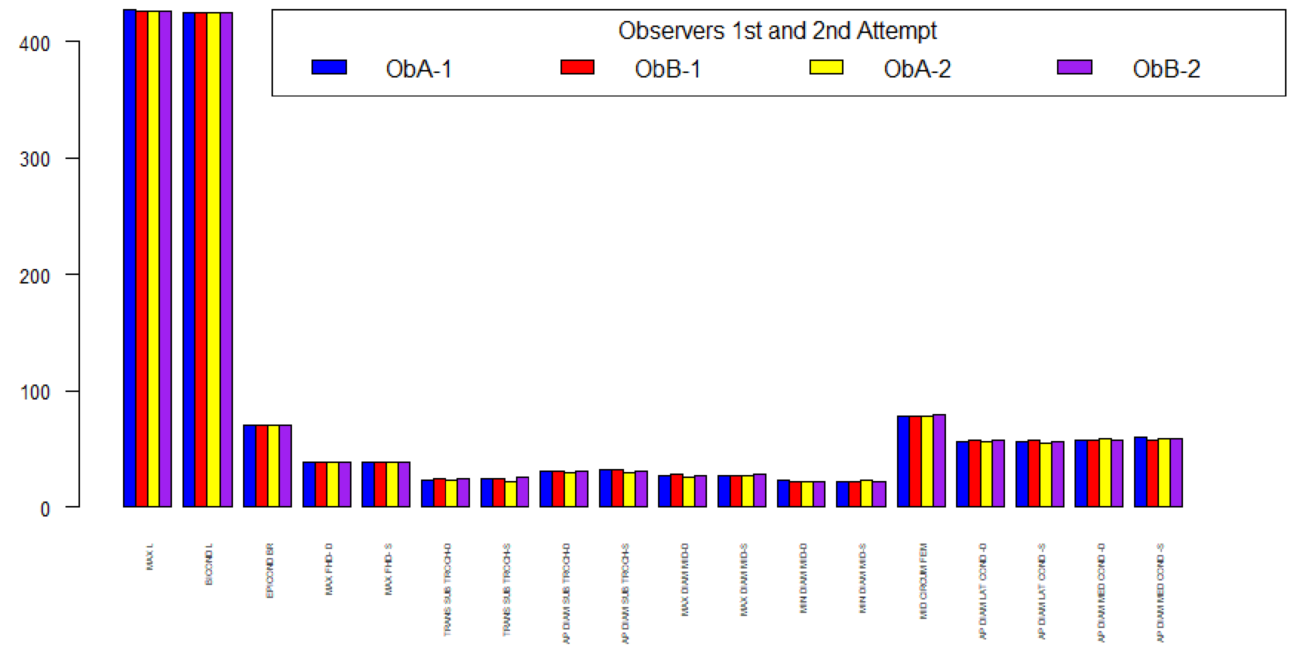

Figure 3: Observer A versus Observer B Femur 1 Measurements (Right Side) (Interobserver Error)

Interobserver Variability for Right Femora Two

The interobserver variability for the right femora two was consistent with the variability observed in the measurements from the right femora one (Figure 4). Again, the transverse subtrochanteric diameter of the femur measurement fell outside the acceptable range of variation, but this time, measured with the digital calipers. The calculated spread between the highest and lowest measurements for the transverse subtrochanteric diameter was 2.83 mm. Measurements of the epicondylar breadth, transverse subtrochanteric diameter of the femur, circumference of femur at the midshaft, and the maximum antero-posterior length of the lateral body were at the acceptable range of variation at 2 mm. The remaining measurements taken from the right femur two were within the acceptable range of variation.

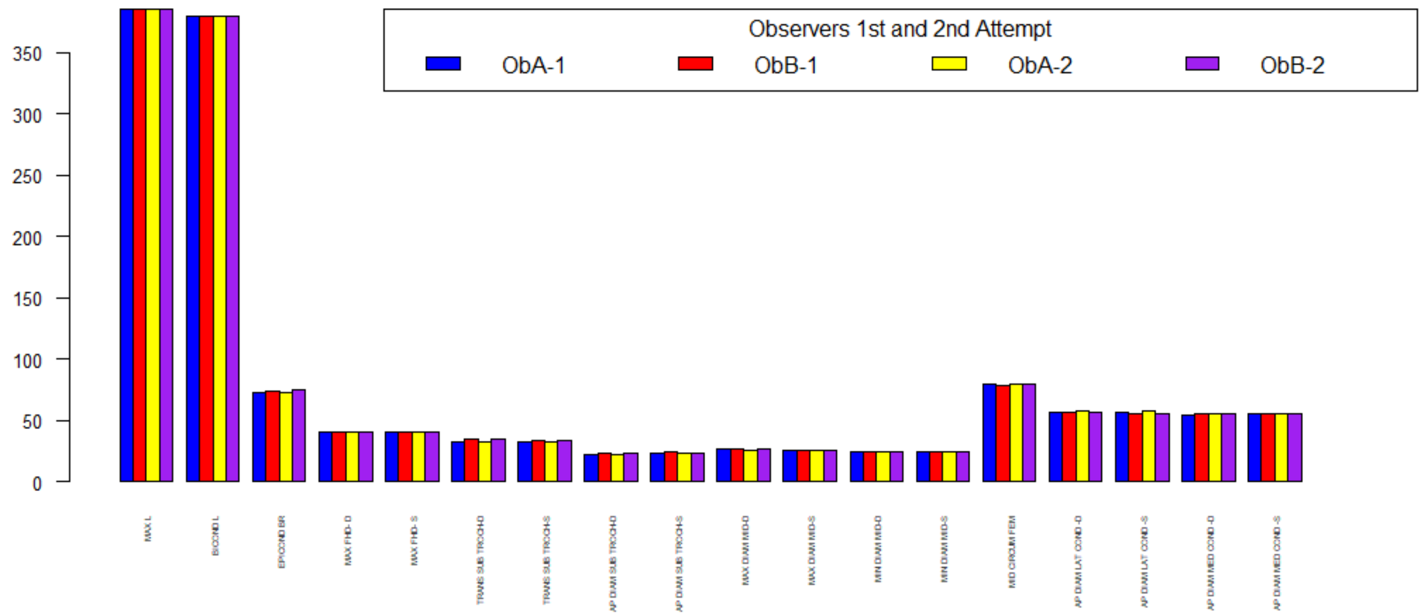

Figure 4: Observer A versus Observer B Femur 2 Measurements (Right Side) (Interobserver Error)

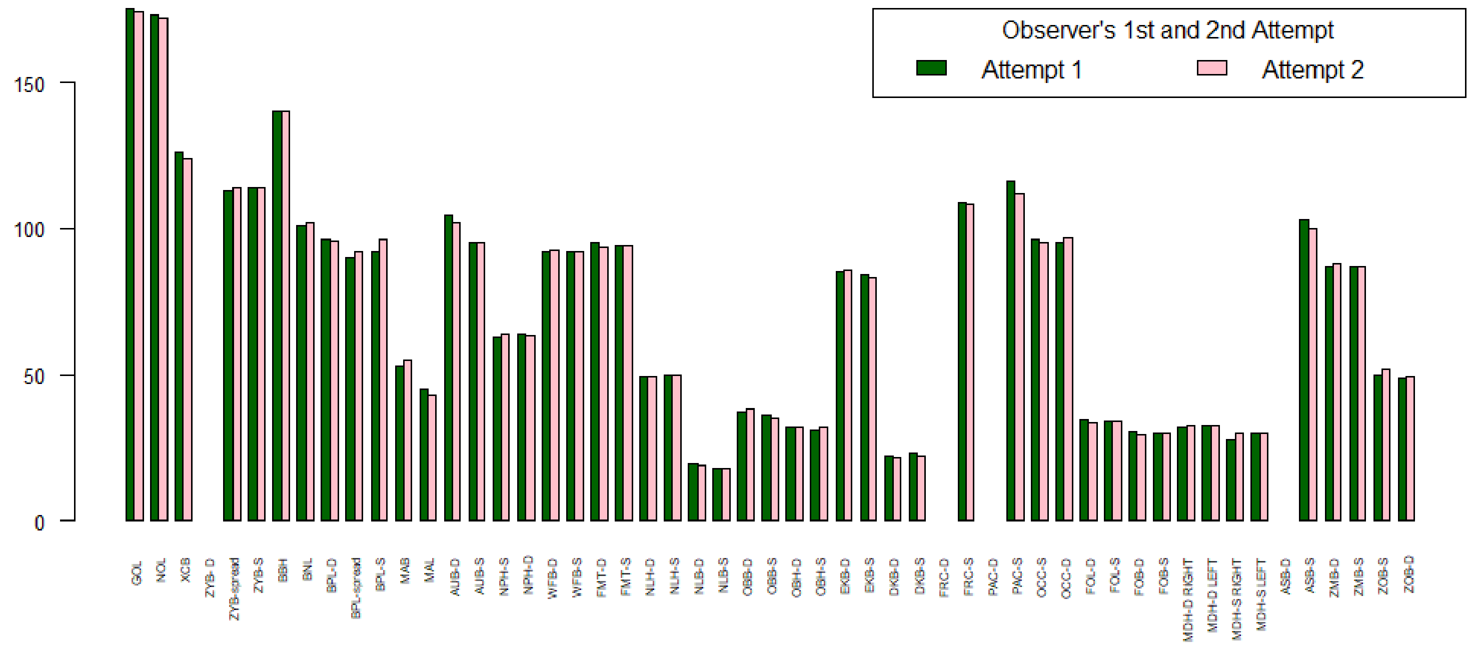

Figure 5: Observer A Cranium 1 Measurements (Interobserver Error)

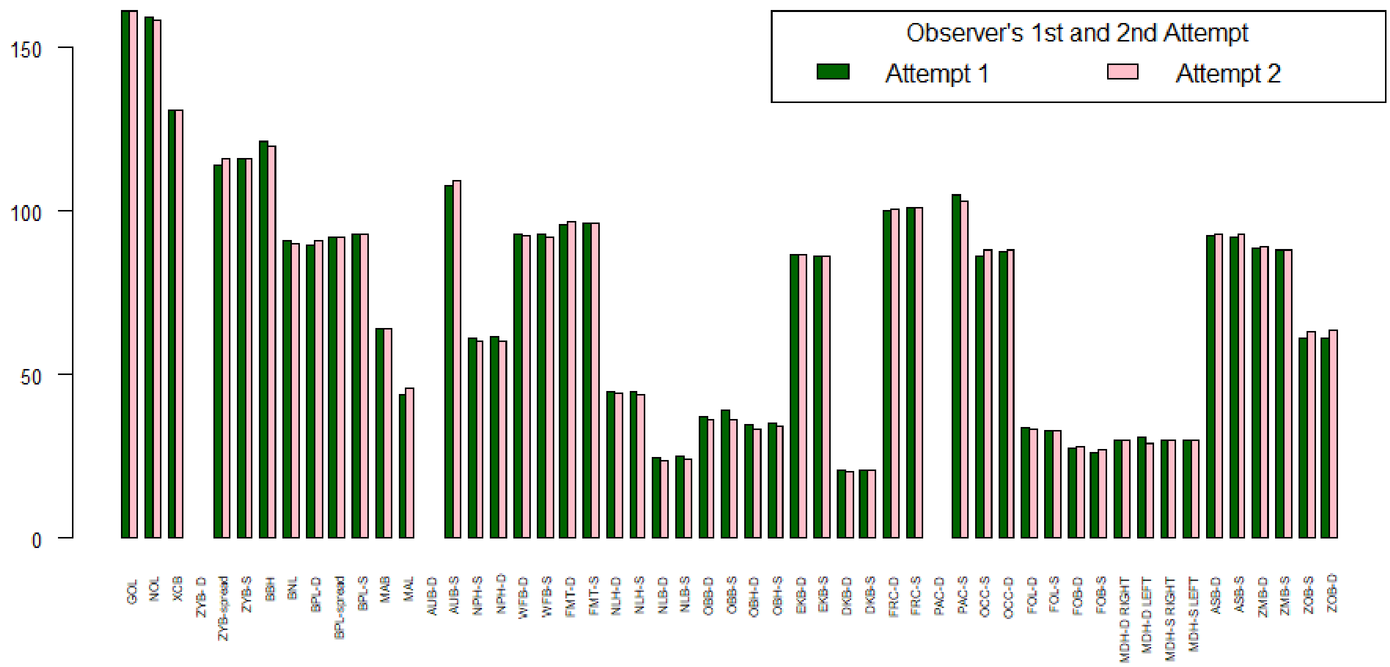

Figure 6: Observer B Cranium 2 Measurements (Intraobserver Error)

Intraobserver Variability for the Crania

Figures 5 and 6 represent the intraobserver variability, referring to the comparison of a single observer's first and second attempts for each bone (the other graphs are in Appendix B, for the cranium, and Appendix C, for the femur). For the cranial measurements, the highest intraobserver variability were measurements of the Parietal Chord (PAC-S), Zygoorbitale breadth (ZOB-S), and the Occipital Chord (OCC- S), all measured with manual sliding calipers. The OCC-S had a spread of 4 mm, and the PAC-S and ZOB-S had a spread of 3 mm, respectively. The PAC-S and OCC-S were measurements taken from the lambda cranial landmark on the lambdoid suture and measured to another cranial feature, as specified in the manual. Cranial sutures have high variability as they can change or be completely obliterated depending on the individual, and interpretation of the suture lines can vary by the individual taking the measurements, thus increasing the potential variability for these specific measurements. With PAC-S, the shape of the cranium can also make the measurement harder to take because the sliding calipers cannot bend around some of the natural variations in the shape of the cranium. The other measurements with variability are within the 2 mm range accepted by the ABFA as natural observer variation.

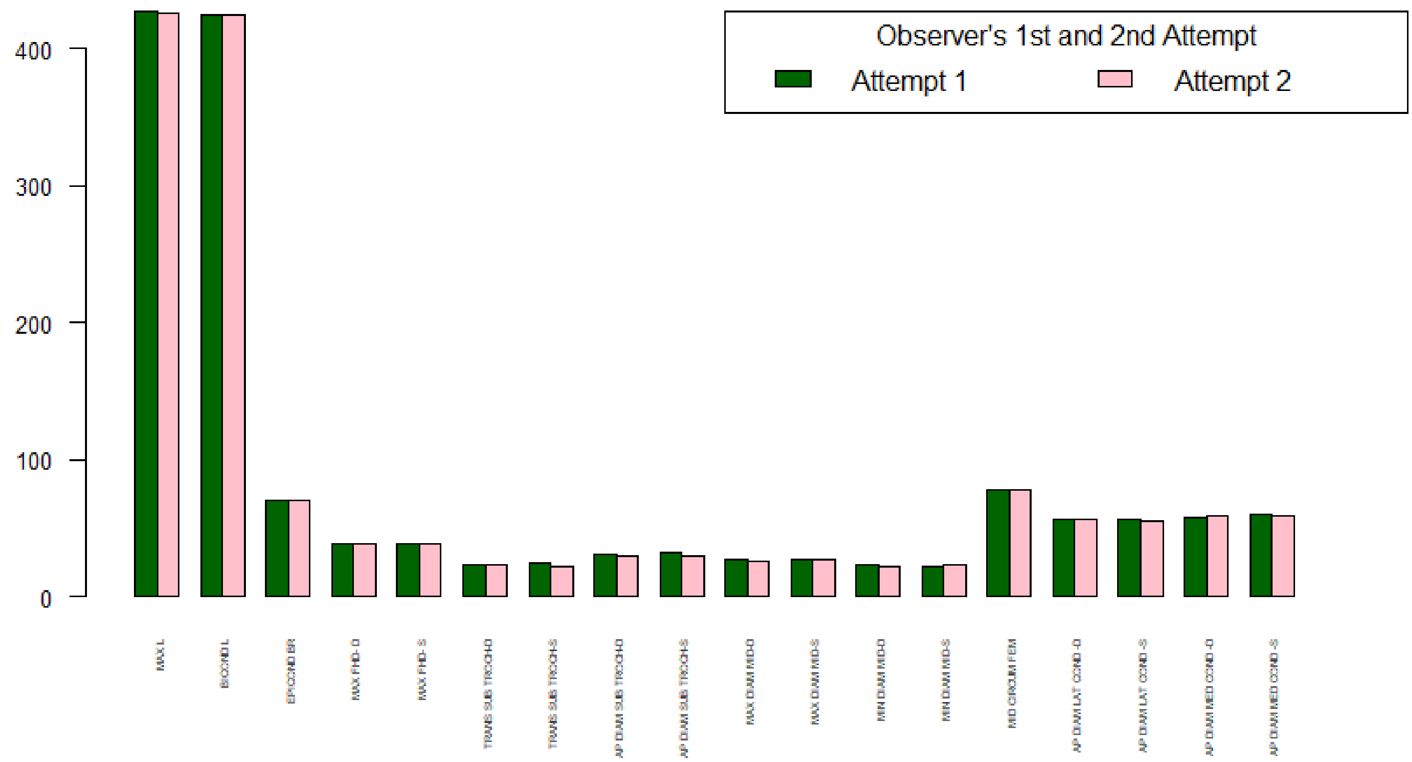

Figure 7: Observer A Femur 1 Measurements (Right Side) (Intraobserver Error)

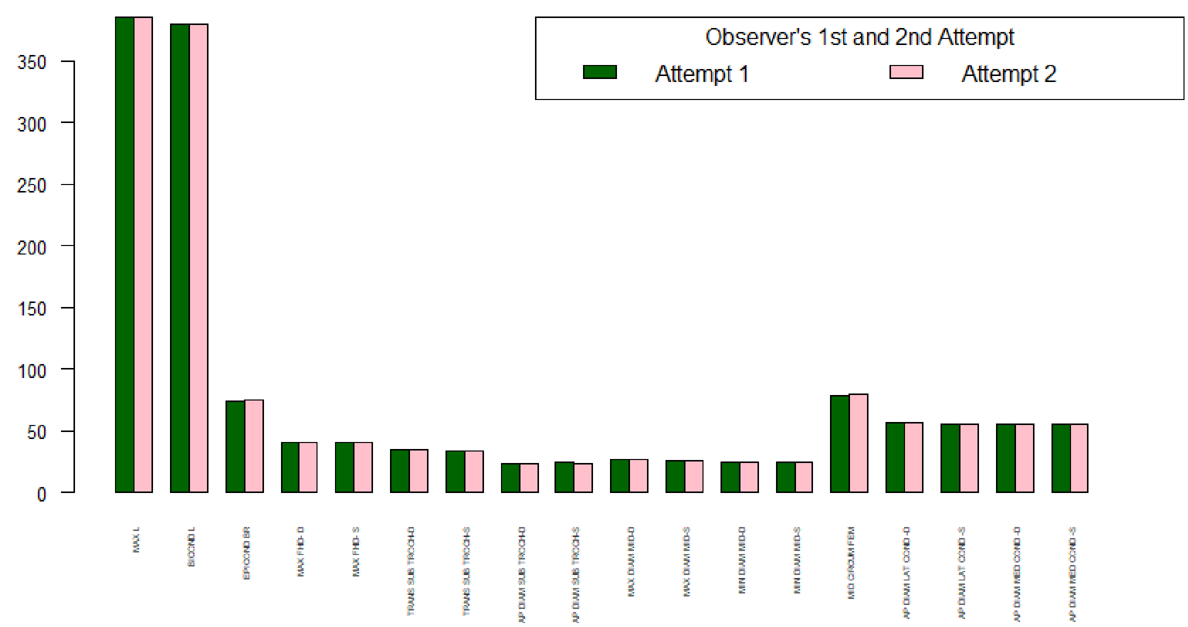

Figure 8: Observer B Femur 2 Measurements (Right Side) (Intraobserver Error)

Intraobserver Variability for the Femora

Both observers' intraobserver results for the femora were consistent (Figures 7 & 8). Some of the differences between the observer's first and second attempts were of the transverse and anterior-posterior subtrochanteric measurements. This is due to unclear descriptions of where to take these measurements from. Only one measurement, the transverse subtrochanteric diameter of the femur, had a significant difference of 3 mm. The rest of the femora measurements did not differ significantly as the variability was under 2 mm, an error range considered acceptable by the ABFA. Overall, the femora measurements taken by the observers reflected a low amount of intraobserver variability.

Discussion

The cranial measurements showed higher variability compared to the femora measurements. The cranial landmarks express some variability in their location, especially the cranial suture landmarks. Further, the measurement devices had limited malleability to be flexible and reach precisely designated cranial landmarks. These factors potentially contribute to higher rates of measurement and observer interpretation variability. The observers documented ambiguous and vague directions provided by Langley et al. (2016) for some of the cranial and femora measurements. Also, the photos in the manual occasionally did not line up with their subsequent description, presenting another contributor of variability (Figures 9 & 10).

There were a few documented issues resulting from the standardized measurement devices. Of note, the digital calipers posed challenges due to their limited length (maximum length of 104.5 mm). However, the digital calipers yield more detailed measurements by providing the measurement to the nearest hundredth. Further, the digital calipers provide the observer with a digitized measurement, reducing the observation error as each observer must manually read and interpret the measurement from the other measuring devices. Special attention was required when taking the measurements as the spreading and sliding calipers were measured in centimeters, meaning the observer must convert those measurements to millimeters. The two resisting points on the osteometric board make it a more stable measurement device compared to handheld devices, resulting in less varied results. This research shows that students of similar experience levels can exhibit discrepancies in their skeletal measurements, both interobserver and intraobserver. These discrepancies have the potential to contribute to inaccuracies when estimating the biological sex, age, and stature of an unidentified individual. Similar to findings from Langley et al. (2016), the observer inconsistencies may be the result of landmark identification and location. Measurements involving suture landmarks are especially problematic due to their unpredictable location and curvature. Despite Langley et al. (2016) deleting problematic measurements, they did not provide clarity in the descriptions for the remaining measurements.

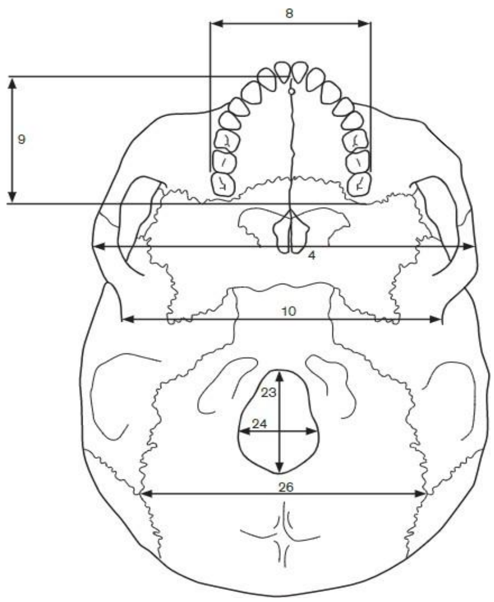

Figure 9: An example of cranium measurements from Langley et al. (2016) (pg.67)

8. **Maxillo-Alveolar Breadth (ecm-ecm, MAB):** The maximum breadth across the alveolar borders of the maxilla measured on the lateral surfaces at the location of the second maxillary molars (Figure 3.5).

   Instrument: spreading caliper

   Comments: The points of measurement (ecm) are not found on the alveolar processes when reactive alveolar bone is present; in these cases ecm is located on the bony segment superior to the second maxillary molars. Apply both arms of the caliper to the alveolar borders above the tooth row from an anterior position (Howells 1973:176; Martin and Knussmann 1988: 182 #61).

9. **Maxillo-Alveolar Length (pr-alv, MAL):** The distance from prosthion (pr) to alveolon (alv) (Figure 3.5).

   Instrument: Spreading or sliding caliper. A sliding caliper is only useful if the central incisors have been lost.

   Comments: Place the skull so the base faces up. Apply a thin wire, rubber band, or other similar implement to the posterior borders of the alveolar arch and measure the distance from prosthion (pr) to the middle of the wire/band in the mid-sagittal plane (alv). (Martin and Knussmann 1988: 182 #60).

10. **Biauricular Breadth (ra-ra, AUB):** The least exterior breadth across the roots of the zygomatic processes (Figure 3.5).

    Instrument: sliding caliper

    Comments: With the skull resting on the occiput and with the base toward the observer measure to the outside of the roots of the zygomatic processes at their deepest incurvature, generally slightly anterior to the external auditory meatus, with the sharp points of the caliper (Howells 1973:173; Martin and Knussmann 1988: 170, #11b).

Figure 10: Descriptions from Langley et al. (2016) (pg. 66) also demonstrate some of the measurements do not appear on the same page with the subsequent diagram showing readers how to take those measurements.

Conclusions

Osteometric measurements are utilized frequently in forensic anthropological casework involving unidentified human remains. The accuracy of these measurements is critical to the identification of these individuals. The present research is important because it raises questions about current standards and directions for taking osteometric measurements through quantification of the inter- and intraobserver variability. Overall, when looking at the collected data and subsequent results, the directions for taking osteometric measurements provided by Langley et al. (2016) should provide more clarity to ensure individuals of differing experience levels can accurately take these measurements. The cranial measurements showed higher inter- and intraobserver variability, compared to the variability in femora measurements. This difference in variability is likely contributed to the differences in surface area and variability in location of skeletal features on the crania. The crania have more surface area and more landmarks that are affected by individual variation, thus impacting the ability of observers to collect measurement data. These findings demonstrate Langley et al. (2016) may need to update the direction in their manual to be more encompassing of individuals of differing experience levels and to succeed in achieving the goals that are expressed at the beginning of their manual.

The next steps for this research are to expand the data being analyzed and incorporate measurements taken by other students within the School of Global Integrative Studies at the University of Nebraska- Lincoln. These recorded measurements were documented by Dr. William Belcher and this data will be analyzed and collaborated with the present findings, as it will increase the number of observers and sample size. The interobserver reliability and error rates will be calculated from this collected data. Further, the inter-device reliability and variability will be calculated from the India-made calipers and the Swiss-made calipers. The expected findings of this research will help guide recommendations for using the current manual and give rise to conversation concerning updates to the manual.

Limitations

This study includes limitations such as the small sample and observer size. Due to these limitations, future researchers should expand the sample and observer size to draw more substantial conclusions regarding inter and intra-observer reliability and error rates.

Acknowledgments

We would like to thank Dr. William Belcher and Dr. Elizabeth Clausing for their guidance throughout this research and their encouragement to continue with more data, provided by Dr. William Belcher, for a future publication.

References

Christensen A. M. (2004). The impact of Daubert: implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *Journal of forensic sciences*, *49*(3), 427–430.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993)

Henson K., Harding T., Starcher K., Smith A., Seccurro D. (2020). Intraobserver and Interobserver Error in Osteological Analysis as an Indicator for Non-Expert Skeletal Analysis. Journal of Forensic Science and Criminology. 8(1), 1-7. ISSN:2348-9804.

Human Anatomy - bone clones, inc. - osteological reproductions. (n.d.). https://boneclones.com/category/human-anatomy

Langley N.R., L.M. Jantz, S.D. Ousley, R.L. Jantz, G.S. Milner. (2016). Data Collection Procedures for Forensic Skeletal Material 2.0. The University of Tennessee, Knoxville

Langley, N. R., Meadows Jantz, L., McNulty, S., Maijanen, H., Ousley, S. D., & Jantz, R. L. (2018). Error quantification of osteometric data in forensic anthropology. *Forensic science international*, *287*, 183–189. https://doi.org/10.1016/j.forsciint.2018.04.004

Appendix A – RStudio code and Averages

```r
na.omit(Cranium1,na.rm=TRUE)
Cranium1$Measurements
dat <- `dimnames<-`(t(as.matrix(Cranium1[c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2")])),
                    list(c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2"), Cranium1$Measurements))
barplot(dat, beside=TRUE, col=c("blue", "red", "yellow", "purple"), las=2,cex.axis = 0.8,cex.names = 0.7,
        )
legend("topright", inset=.00, title="Observers 1st and 2nd Attempt",
       c("ObA-1","ObB-1","ObA-2","ObB-2"), fill=c("blue","red","yellow","purple"), horiz=TRUE)
Femur1$Measurements
dat1 <- `dimnames<-`(t(as.matrix(Femur1[c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2")])),
                    list(c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2"), Femur1$Measurements))
barplot(dat1,beside=TRUE, col=c("blue", "red","yellow","purple"), las=2,cex.axis = 0.8,cex.names = 0.4,
        )
legend("topright", inset=.00, title="Observers 1st and 2nd Attempt",
       c("ObA-1","ObB-1","ObA-2","ObB-2"), fill=c("blue","red","yellow","purple"), horiz=TRUE)
na.omit(Cranium2,na.rm=TRUE)
dat2 <- `dimnames<-`(t(as.matrix(Cranium2[c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2")])),
                     list(c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2"), Cranium2$Measurements))
barplot(dat2,beside=TRUE, col=c("blue", "red","yellow","purple"), las=2,cex.axis = 0.8,cex.names = 0.5,
        )
legend("topright", inset=.00, title="Observers 1st and 2nd Attempt",
       c("ObA-1","ObB-1","ObA-2","ObB-2"), fill=c("blue","red","yellow","purple"), horiz=TRUE)
dat3 <- `dimnames<-`(t(as.matrix(Femur2[c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2")])),
                     list(c("Dakota.1", "Maggie.1", "Dakota.2", "Maggie.2"), Femur2$Measurements))
barplot(dat3,beside=TRUE, col=c("blue", "red","yellow","purple"), las=2,cex.axis = 0.8,cex.names = 0.4,
        )
legend("topright", inset=.00, title="Observers 1st and 2nd Attempt",
       c("ObA-1","ObB-1","ObA-2","ObB-2"), fill=c("blue","red","yellow","purple"), horiz=TRUE)
na.omit(Cranium1.AVG,na.rm=TRUE)
dat4 <- `dimnames<-`(t(as.matrix(Cranium1.AVG[c("Dakota.AVG", "Maggie.AVG")])),
                    list(c("Dakota.AVG", "Maggie.AVG"), Cranium1.AVG$Measurements))
barplot(dat4,beside=TRUE, col=c("blue", "red"), las=2,cex.axis = 0.8,cex.names = 0.5,
        )
legend("topright", inset=.00, title="Observers Averages",
       c("ObA","ObB"), fill=c("blue","red"), horiz=TRUE)
```

Figure 11: Code used to create graphs in RStudio

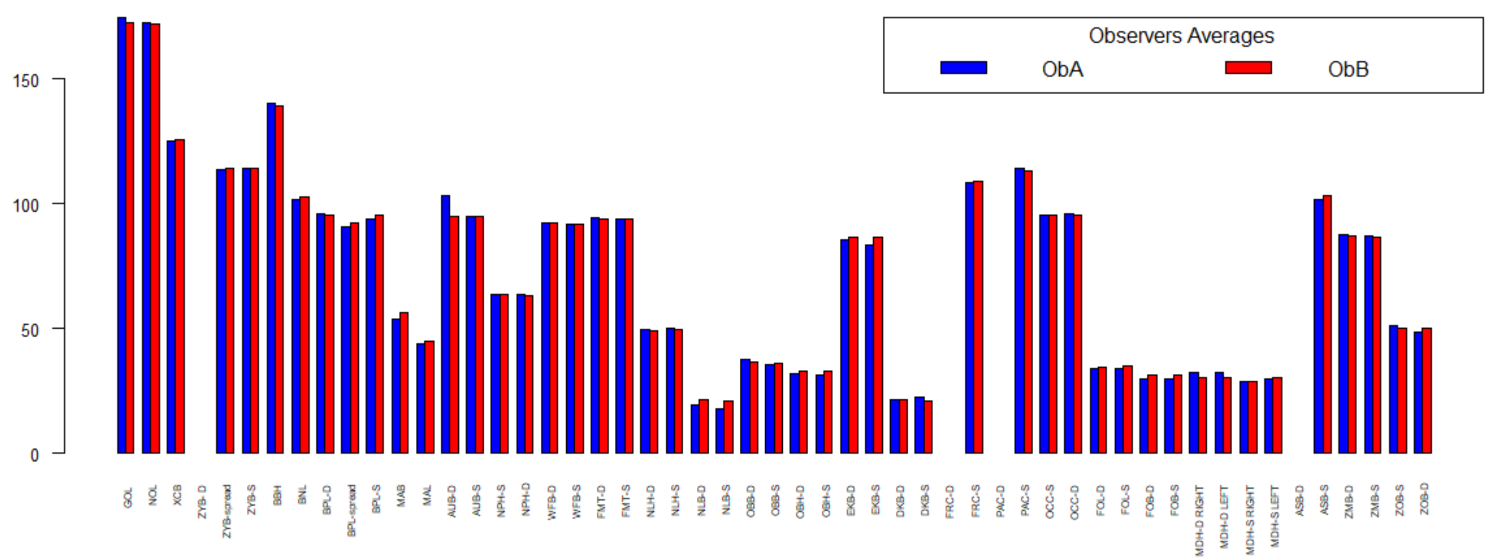

Figure 12: Averages of both Observer's measurements for Cranium 1 (Interobserver Error)

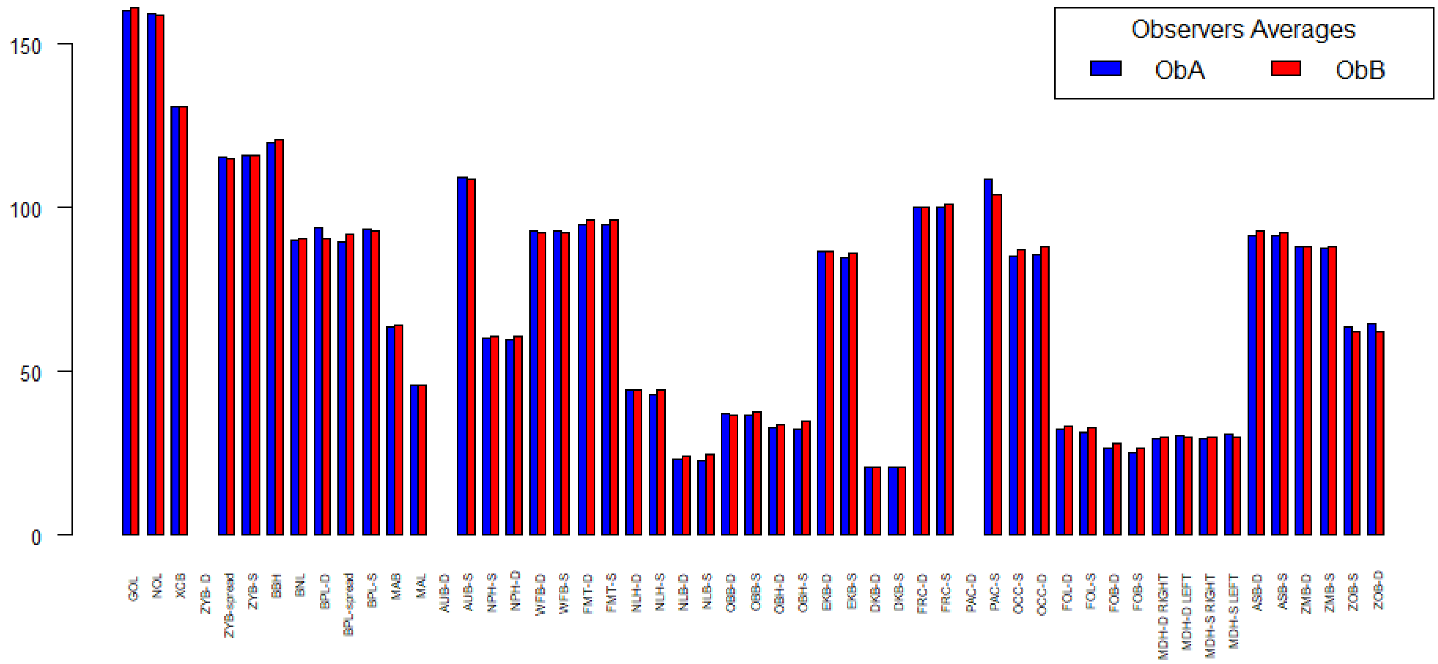

Figure 13: Averages for both Observer's measurements Cranium 2 (Interobserver Error)

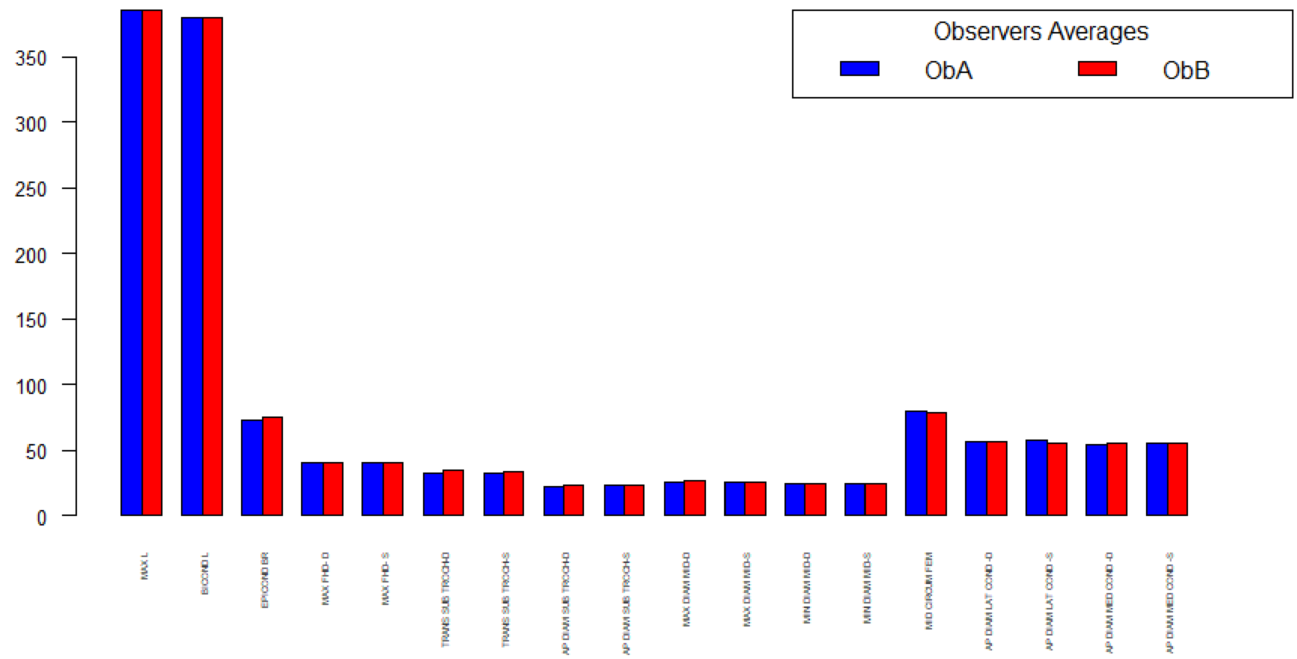

Figure 14: Averages for both Observer's measurements Femur 1 (Interobserver Error)

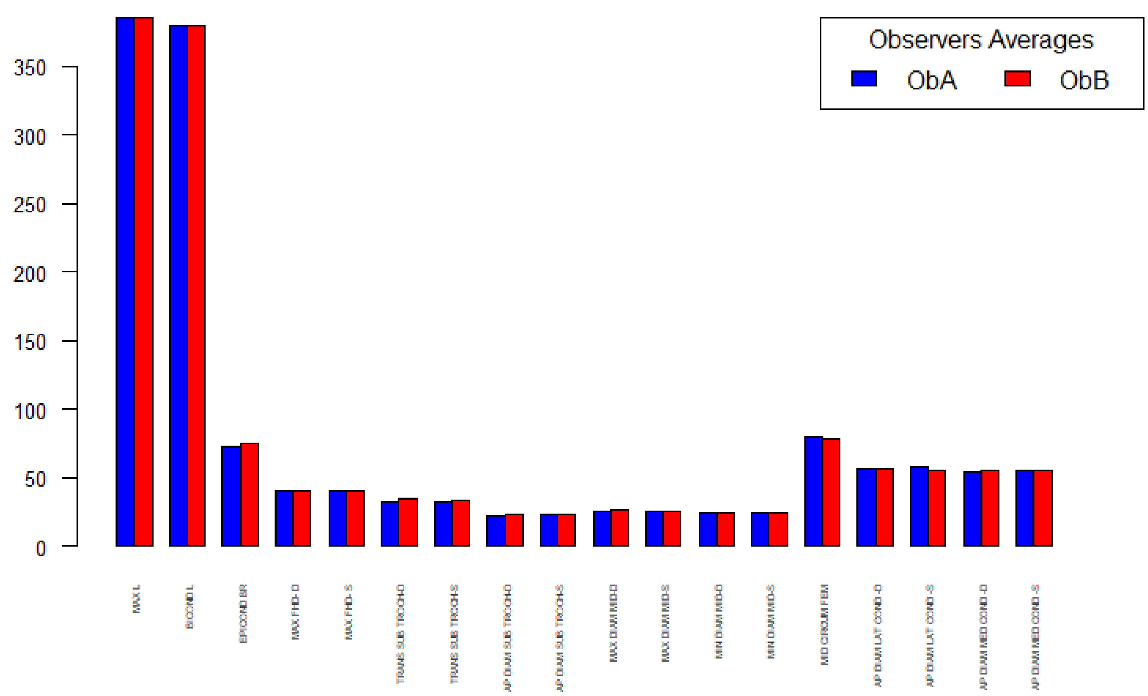

Figure 15: Averages for both Observer's measurements Femur 2 (Interobserver Error)

Appendix B – Individual Cranium Measurements

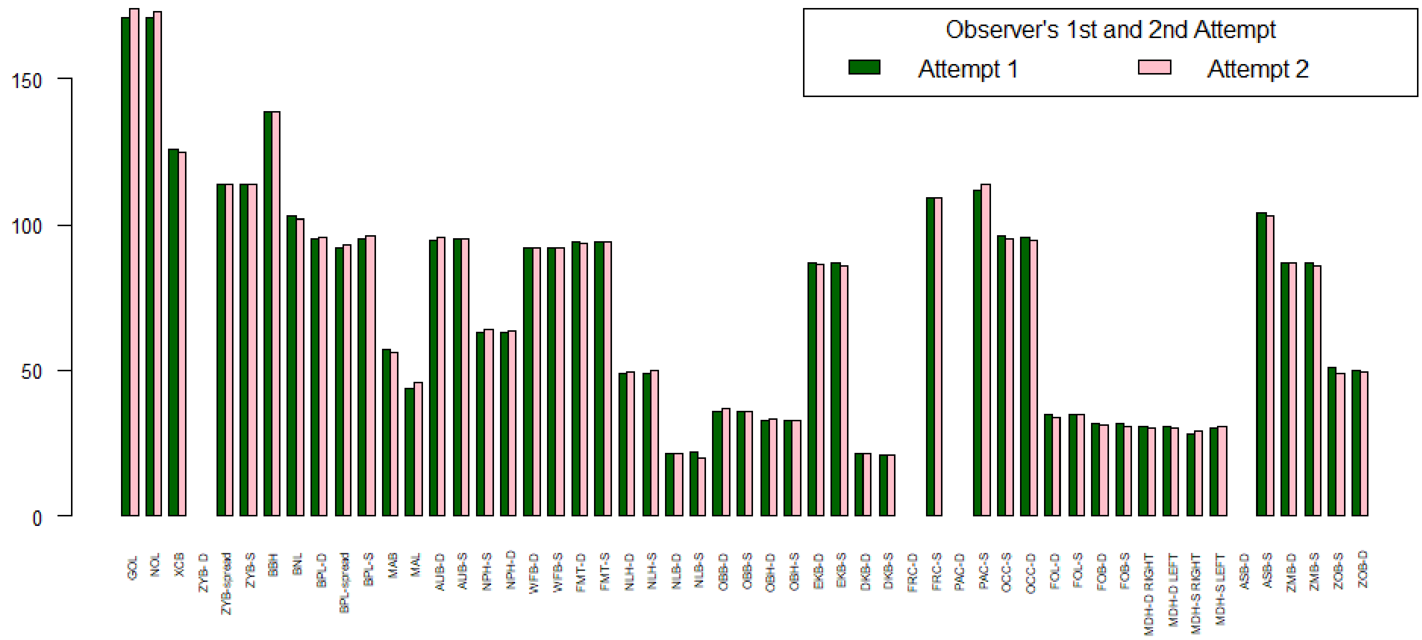

Figure 16: Observer B's Measurements for Cranium 1 (Intraobserver Error)

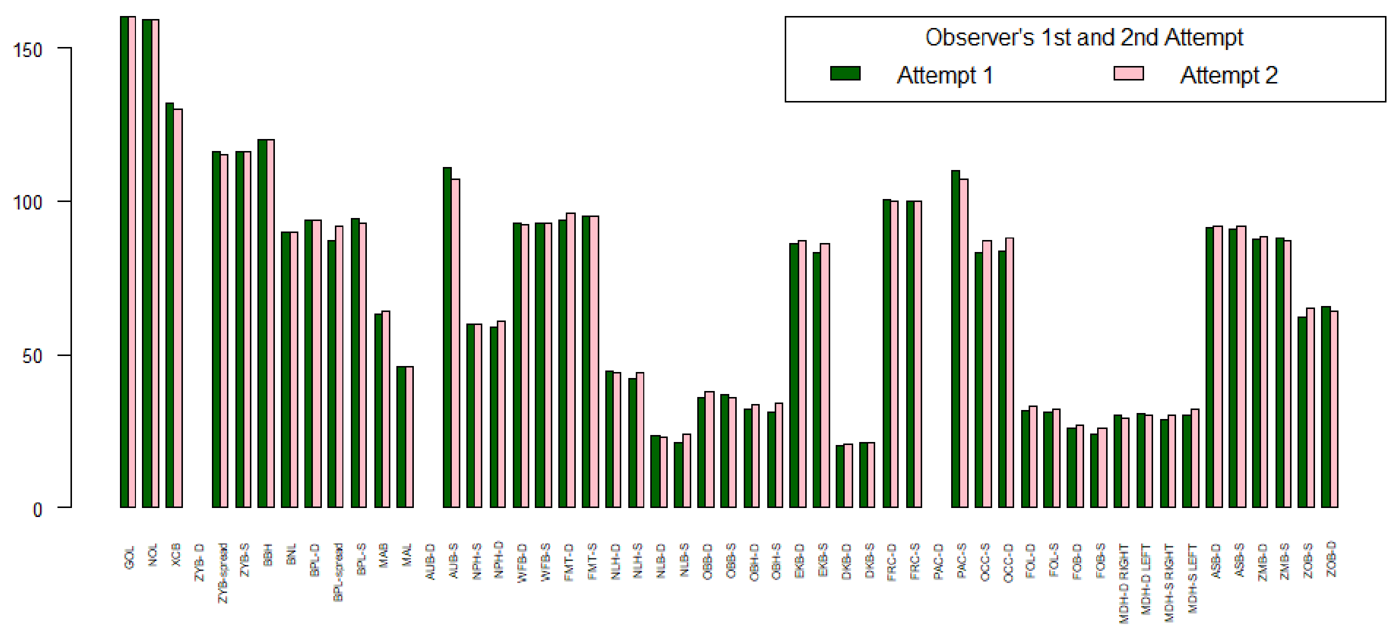

Figure 17: Observer A's Measurements for Cranium 2 (Intraobserver Error)

Appendix C – Individual Femur Measurements

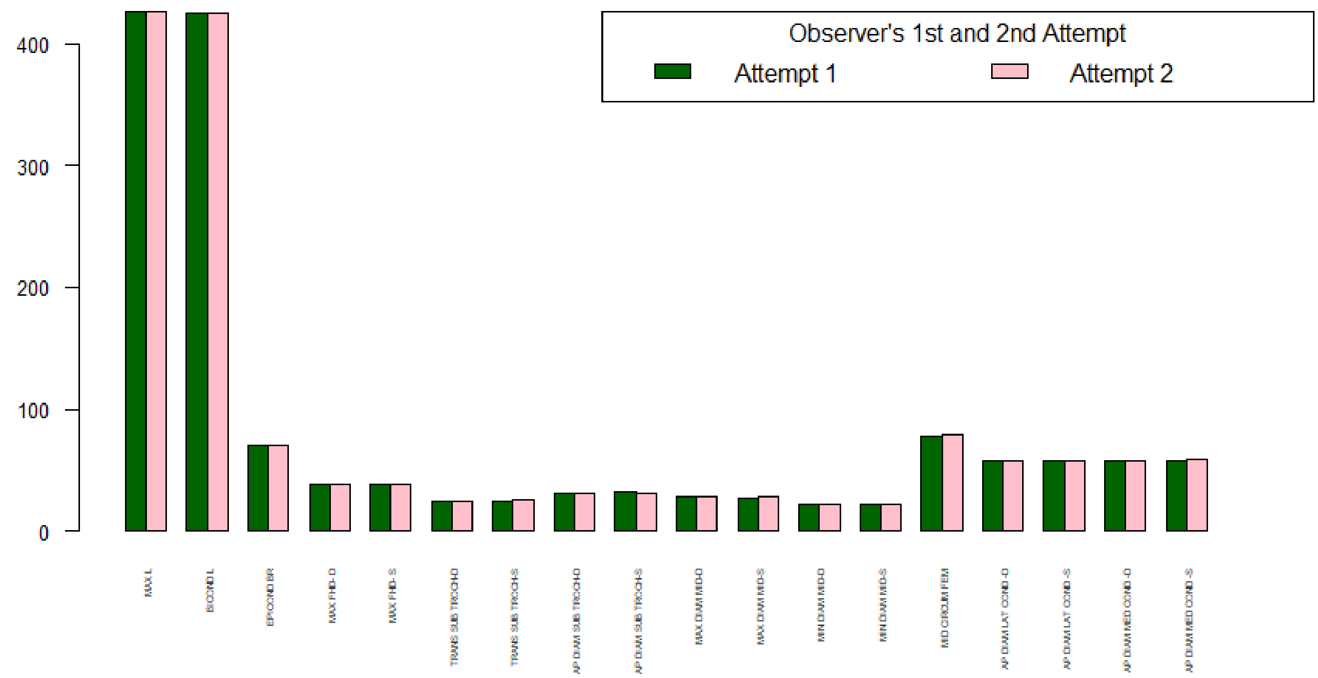

Figure 18: Observer B's Measurements for Femur 1 (Intraobserver Error)

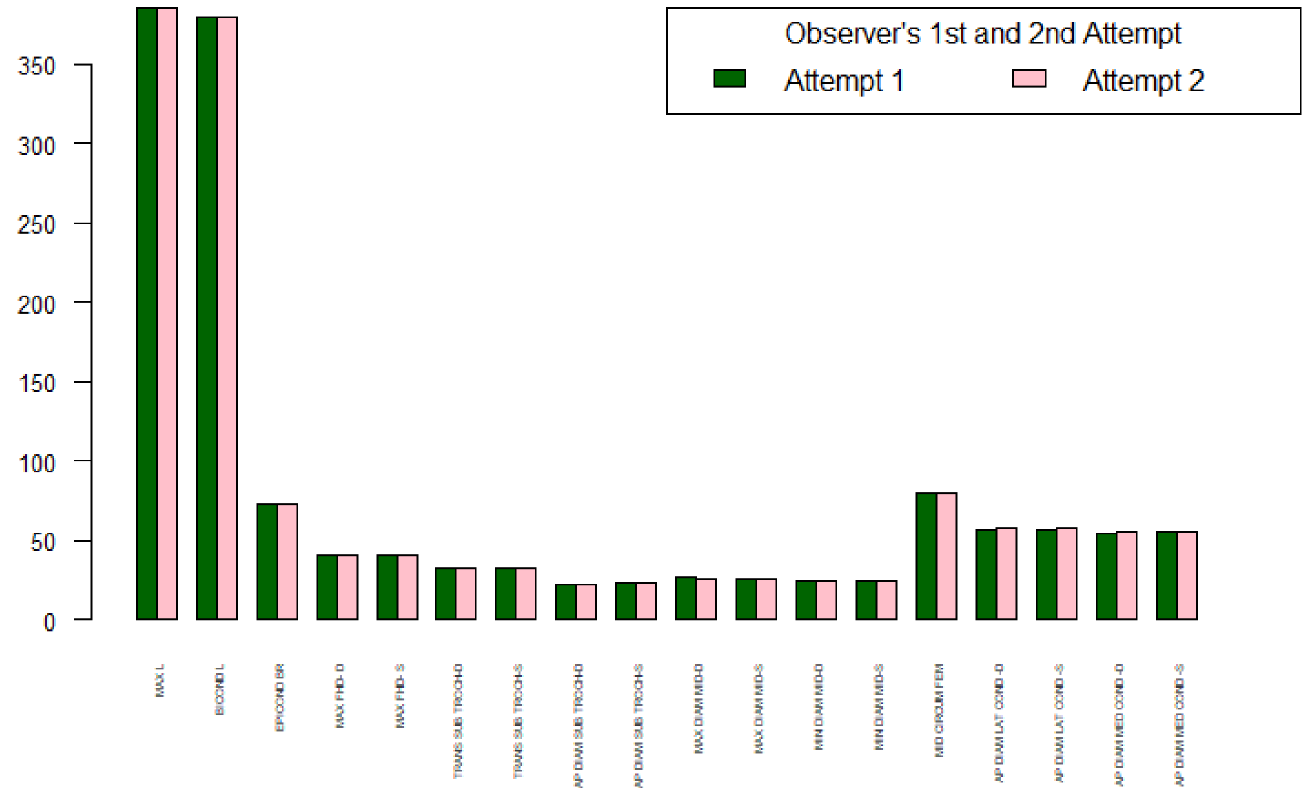

Figure 19: Observer A's Measurements for Cranium 2 (Intraobserver Error)

Appendix D – Definitions of Measurements

Abbreviations on Graphs (Cranium)	Name in Langley (2016)
GOL	Maximum Cranial Length
NOL	Nasio-occipital length
XCB	Maximum Cranial Breadth
ZYB- D	Bizygomatic Breadth (Digital Caliper)
ZYB-spread	Bizygomatic Breadth (Spreading Caliper)
ZYB-S	Bizygomatic Breadth (Sliding Caliper)
BBH	Basion-Bregma Height
BNL	Cranial Base Length
BPL-D	Basion-Prosthion Length (Digital Caliper)
BPL-spread	Basion-Prosthion Length (Spreading Caliper)
BPL-S	Basion-Prosthion Length (Sliding Caliper)
MAB	Maxillo-Alveolar Breadth
MAL	Maxillo-Alveolar Length
AUB-D	Biauricular Breadth (Digital Caliper)
AUB-S	Biauricular Breadth (Sliding Caliper)
NPH-S	Nasion-Prosthion Height (Sliding Caliper)
NPH-D	Nasion-Prosthion Height (Digital Caliper)
WFB-D	Minimum Frontal Breadth (Digital Caliper)

WFB-S	Minimum Frontal Breadth (Sliding Caliper)
FMT-D	Upper Facial Breadth (Digital Caliper)
FMT-S	Upper Facial Breadth (Sliding Caliper)
NLH-D	Nasal Height (Digital Caliper)
NLH-S	Nasal Height (Sliding Caliper)
NLB-D	Nasal Breadth (Digital Caliper)
NLB-S	Nasal Breadth (Sliding Caliper)
OBB-D	Orbital Breadth (Digital Caliper)
OBB-S	Orbital Breadth (Sliding Caliper)
OBH-D	Orbital Height (Digital Caliper)
OBH-S	Orbital Height (Sliding Caliper)
EKB-D	Biorbital Breadth (Digital Caliper)
EKB-S	Biorbital Breadth (Sliding Caliper)
DKB-D	lnterorbital Breadth (Digital Caliper)
DKB-S	lnterorbital Breadth (Sliding Caliper)
FRC-D	Frontal Chord (Digital Caliper)
FRC-S	Frontal Chord (Sliding Caliper)
PAC-D	Parietal Chord (Digital Caliper)
PAC-S	Parietal Chord (Sliding Caliper)
OCC-S	Occipital Chord (Sliding Caliper)
OCC-D	Occipital Chord (Digital Caliper)
FOL-D	Foramen Magnum Length (Digital Caliper)
FOL-S	Foramen Magnum Length (Sliding Caliper)
FOB-D	Foramen Magnum Breadth (Digital Caliper)
FOB-S	Foramen Magnum Breadth (Sliding Caliper)
MDH-D RIGHT	Mastoid Height (Digital Caliper)
MDH-D LEFT	Mastoid Height (Digital Caliper)
MDH-S RIGHT	Mastoid Height (Sliding Caliper)
MDH-S LEFT	Mastoid Height (Sliding Caliper)
ASB-D	Biasterionic Breadth (Digital Caliper)
ASB-S	Biasterionic Breadth (Sliding Caliper)
ZMB-D	Bimaxillary breadth (Digital Caliper)
ZMB-S	Bimaxillary breadth (Sliding Caliper)
ZOB-S	Zygoorbitale breadth (Sliding Caliper)
ZOB-D	Zygoorbitale breadth (Digital Caliper)

Abbreviations on Graphs (Femur)	**Name in Langley (2016)**
MAX L	Maximum Length of the Femur
BICOND L	Bicondylar Length of the Femur
EPICOND BR	Epicondylar Breadth of the Femur
MAX FHD- D	Maximum Diameter of the Femur Head (Digital Caliper)
MAX FHD- S	Maximum Diameter of the Femur Head (Sliding Caliper)
TRANS SUB TROCH-D	Transverse Subtrochanteric Diameter of the Femur (Digital Caliper)
TRANS SUB TROCH-S	Transverse Subtrochanteric Diameter of the Femur (Sliding Caliper)
AP DIAM SUB TROCH-D	Anterio-posterior Subtrochanteric Diameter of the Femur (Digital

	Caliper)
AP DIAM SUB TROCH-S	Anterio-posterior Subtrochanteric Diameter of the Femur (Sliding Caliper)
MAX DIAM MID-D	Maximum Midshaft Diameter of the Femur (Digital Caliper)
MAX DIAM MID-S	Maximum Midshaft Diameter of the Femur (Sliding Caliper)
MIN DIAM MID-D	Minimum Midshaft Diameter of the Femur (Digital Caliper)
MIN DIAM MID-S	Minimum Midshaft Diameter of the Femur (Sliding Caliper)
MID CIRCUM FEM	Circumference of the Femur at Midshaft
AP DIAM LAT COND -D	Maximum Antero-posterior Length of the Lateral Condyle (Digital Caliper)
AP DIAM LAT COND -S	Maximum Antero-posterior Length of the Lateral Condyle (Sliding Caliper)
AP DIAM MED COND -D	Maximum Antero-posterior Length of the Medial Condyle (Digital Caliper)
AP DIAM MED COND -S	Maximum Antero-posterior Length of the Medial Condyle (Sliding Caliper)