A Novel Method of Using Computed Tomography Data to Sort Commingled Skeletal Human Remains through Pair Matching

By

Shannon Christine Velasquez

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Biology

Middle Tennessee State University
June, 2023

Thesis Committee:

Dr. Yangseung Jeong, Chair

Dr. Sarah Bergemann

Dr. Frank C. Bailey

ACKNOWLEDGEMENTS

First and most importantly, thank you to all of the individuals and their families who gave the incredible gift of their bodies for scientific research. Without them, none of this work would be possible. Thank you to Dr. Dawnie Wolfe Steadman and the University of Tennessee Forensic Anthropology Center (FAC) for the use of their CT data from their donated skeletal collection. Thank you also to Dr. Benjamin Auerbach for your hospitality during our visit to the FAC and for the use of your lab space and equipment.

Thank you, Dr. Yangseung Jeong, for being an incredibly patient and thorough thesis advisor. I have grown immensely as a scientist from studying with you. Thank you to my other thesis committee members, Dr. Frank C. Bailey and Dr. Sarah Bergemann, for your time and commitment to ensuring I produced the best product possible. Also, thank you to Dr. Ying Jin for your time in assisting me with my R code.

Thank you to all of the undergraduate research assistants in Dr. Jeong's lab who assisted with the tedious task of doing over 1,600 mesh-to-mesh comparisons including Mirna Hanna, Sosna Kumssa, Meirola Endraws, Sandra Shafeek, Marcus Luciano, Haley Stiehler, and Casey Tomlin.

Finally, thank you to my family: my husband, Jonathan, and my daughter, Anna, for holding down the fort while I spent countless hours finishing my degree program. You inspire me to be my best self.

ABSTRACT

In situations of commingled human skeletal remains, the first task of a forensic anthropologist is to accurately sort the remains and determine which bones belong to the same individual. Pair matching bilateral elements is one of the techniques often employed for this purpose. Current pair-matching methods are complicated by issues of subjectivity, bones of similarly sized individuals being incorrectly sorted, dimension reduction, need for highly trained practitioners, or the need for specific and expensive software applications. This research presents a novel and objective method of pair matching using 3D images rendered from computed tomography (CT) scans. Left and right bilateral elements were compared by mirroring left elements in Meshmixer and overlaying the mirrored left with a corresponding right element in CloudCompare. Next, the amount of discrepancy between the 3D images (meshes) was calculated in CloudCompare and summarized using two statistics, the mean of the distances between the vertices of one mesh to the closest vertices of the other (D_{mean}) and the standard deviation of those distances (D_{sd}). These statistics were used to build logistic regression models to be utilized in future predictive analysis to determine the probability of bones belonging to the same individual. The methods presented overcame some of the limitations of current pair matching methods mentioned while providing accurate, objective, and easily learned techniques to contribute to successful resolution of commingled human skeletal remains.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: BACKGROUND

Natural disasters, wars, and human rights violations are just some of the scenarios that can cause mass fatality events. These scenarios can lead to intermixing of human remains from multiple individuals, known as commingling (Adams & Byrd, 2014). Multiple victims of genocide are often buried in mass graves, sites where two or more deceased victims are co-located (Jessee & Skinner, 2005). The bombing of Pearl Harbor in 1941 (Defense POW/MIA Accounting Agency, 2022), the Serbian war crimes against Kosovan Albanians in 1999 (Tuller & Hofmeister, 2014), and the terrorist attack on the World Trade Center in 2001 (Mundorff et al., 2014) are just a few examples of tragedies that caused commingled human skeletal remains in the last several decades. In scenarios such as these, forensic anthropologists are tasked with determining how many individuals are at the grave site and sorting the commingled remains into their respective individuals.

Determining the number of individuals at a disaster or burial site and identifying them can become quite complex, particularly if there were a large number of victims, if there were repeated burials in the same locale, or if the site had been disturbed by natural or human forces (Christensen et al., 2019). The Minimum Number of Individuals (MNI) is determined by sorting bones by element and side (left and right) and using the number of the group with the most elements (Adams & Konigsberg, 2008). However, since the MNI is often an underestimation, the Most Likely Number of Individuals (MLNI) was developed to provide a more accurate estimation (Adams & Konigsberg, 2004). Calculating the MLNI considers the number of elements by side like MNI. However, it also considers paired elements, using the equation $MLNI = \frac{(L+1)(R+1)}{(P+1)} - 1$, where L is the number of left elements, R the right elements, and P is the number of pairs of the

elements. Calculating the MNLI requires accurate pair matching, which refers to a sorting technique to determine if bilateral bones belong to the same individual (Fancourt et al., 2021). The underlying assumption of pair matching is that bones belonging to the same individual will be more similarly sized and shaped than bones belonging to different individuals (Adams & Byrd, 2008).

Although the MNLI can provide a more accurate estimate of the number of individuals at a mass grave site, the accuracy of the estimation is dependent on the accuracy of pair matching (Adams & Konigsberg, 2008). Because determining an accurate number of individuals and sorting them is the first step before biological profiles of individuals can be made, it is critical to use reliable methods for pair matching and sorting. Thus, these first steps are critical to ensuring accurate identification of remains and returning them to the next of kin.

1.1 Methods of Sorting and Pair Matching Commingled Remains

There are many methods of sorting commingled remains into their respective individuals, and often different approaches are used in combination depending on the context of the burial site to best fit each specific scenario (Christensen et al., 2019). Forensic analysis is often conducted using either molecular markers or anthropological methods for sorting skeletal remains.

DNA analysis can include the use of either nucleic (nDNA) or mitochondrial DNA (mtDNA). nDNA provides the advantage of high individuation. Though variation of DNA between humans is only 0.1 percent, there are still (primarily non-coding) regions of variability that result in a one in 594.1 trillion chance that two different individuals will share the exact same nDNA (except in cases of identical twins) (Keerti &

Ninave, 2022). However, DNA samples in skeletal remains are often degraded from a variety of factors, including time, temperature, and presence of water and chemicals to name just a few (Gaudio et al., 2019), and mtDNA can provide an alternate to nDNA in cases of degradation (Court, 2021). Although mtDNA is smaller in size than nDNA, it occurs in much higher copy numbers (Linacre & Ottens, 2016) throughout human nucleated cells (Court, 2021), and universal primers of highly conserved regions allow for ease of amplification and subsequent comparison of three hypervariable regions (Kowalczyk et al., 2021). However, mtDNA is inherited only maternally and, thus, could be the same for all descendants coming from the same mother barring mutations (Latham & Miller, 2018). Additionally, some ethnic groups have low mitogenomic diversity (Andersen & Balding, 2018), and mtDNA has higher mutation rates compared to nDNA (Court, 2021).

There are many methodologies of comparing nDNA for identification purposes, though the use of short tandem repeats (STRs) is widely used for nDNA throughout jurisdictions in the United States (Latham & Miller, 2018), whereas single nucleotide polymorphisms (SNPs) are useful for mtDNA analysis (Damann & Edson, 2008). For both types of DNA analysis, a DNA sample from the victim's remains must first be obtained by DNA extraction after pulverization of the skeletal remains. The sample is compared to a sample from the victim's personal items, a stored medical sample, or DNA samples obtained from a family member to determine whether the STR profile of the victim is consistent with the sample or a family member (Mundorff et al., 2014). Obtaining DNA from bones requires destructive sampling, which includes decontamination of the surface of the bones by physical removal, bleaching, or radiation

followed by pulverization of the sample prior to DNA extraction (Latham & Miller, 2018).

In cases of commingling, a DNA profile can conflict with a profile built with anthropological data, indicating that recollection of the DNA sample is necessary (Yazedjian & Kešetović, 2008). Additionally, due to insults from natural elements, DNA samples are often degraded in skeletal remains and yield insufficient data for DNA profiles, as exemplified by the World Trade Center disaster of 2001 where 60% of the remains did not have sufficient DNA for a complete short tandem repeat (STR) profile (Mundorff et al., 2014). Even when a sufficient STR profile can be obtained, most of the sequences are in non-coding regions and would, thus, not be indicative of phenotype (Keerti & Ninave, 2022). Solely using DNA analysis would require sampling each individual piece of remains in a commingled assemblage. Considering each adult skeleton has 206 bones, identifying remains solely through DNA results in extensive damage to skeletal remains and time for analyses, especially in a large assemblage. For these reasons, many other anthropological methods that rely on visual examinations or statistical testing can complement DNA analysis.

Anthropological methods of sorting commingled human skeletal remains consider morphological features of bones (Adams & Byrd, 2008) that can be used in conjunction with each other or in combination with DNA analysis to build a biological profile of the individual (Finlayson et al., 2017). These methods assess such features as articulation between elements (Buikstra & Gordon, 1980), robusticity, and how similar bilateral elements are to each other to reassociate remains (Adams & Konigsberg, 2008). Anthropologists may assess bone morphology visually (with the unaided eye) or

osteometrically, taking specific bone measurements and utilizing statistical models to determine whether the bones likely came from the same individual (Adams & Byrd, 2008). There are a variety of anthropological methods for determining the sex of remains as well, including assessing the height and depth of the sciatic notch on the hip bone (Shukla et al., 2022) and measuring the subpubic angle (Jeong et al., 2021). Anthropological methods can also be used to determine age by assessing features like epiphyseal unions (Schaefer, 2008), cranial suture closures (Ruengdit et al., 2020), and bone degeneration (Ubelaker & Khosrowshahi, 2019).

Pair matching refers to a variety of visual and ostemetric techniques to sort bilateral elements into their respective individuals. Since visual pair matching compares the size and unique features between paired bones using unaided eyes and hands, the procedure itself is simple and quick; however, the accuracy of this method fluctuates between 84–99% depending on the practitioner's level of experience and training (LeGarde, 2019). Moreover, this method is time-consuming when it comes to a large assemblage of skeletons (Bertsatos & Chovalopoulou, 2020; LeGarde, 2019).

Osteometric pair matching was developed to ensure objectivity using a series of statistical analyses on multiple dimensions of the bones (Byrd & Adams, 2003). The null hypothesis of this test is that the paired bones belong to the same individual, and a rejection of the null hypothesis concludes that the data is inconsistent with bones belonging to the same individual. Thus, this method is regarded as an effective way to exclude a pair of bones and form a condensed list of possible matches in a large skeletal assemblage (LeGarde, 2019). However, some issues with osteometric pair matching were proposed. For one, due to the nature of hypothesis testing, failing to reject the null

hypothesis (bones belong to the same individual) does not equate to acceptance of the null. While null hypothesis rejection lends to use to exclusion, this does not allow for confirmation that the bones definitely belong to the same individual. Additionally, Vickers et al. (2015) found that false rejections of a true pair using the method developed by Byrd and Adams (2003) can occur at rates as high as 22%. Lastly, dimension reduction is regarded as a limitation of the osteometric pair matching as its conclusions are based on only a few measurable dimensions of a three dimensional bone. To overcome the limitations of osteometric pair matching, different statistical approaches and automation were proposed (Bertsatos & Chovalopoulou, 2020).

1.2 Using 3D Data to Sort Commingled Remains

In recent times, there have been many methods developed to increase the number of dimensions used in sorting commingled skeletal remains. Photographs and statistical analysis of 2D surfaces of heel and ankle bones (calcanei and tali) were analyzed using a modified Hausdorff distance to identify true pairs (Lynch, 2018). 3D data can be compared using Geometric Morphometrics (GMM), using Cartesian coordinates for surface spatial information, and with Cross Sectional Geometry (CSG), which uses different measurements of skeletal remains (e.g. cross-sectional areas and dihedral angles) to conduct pair-matching (Fancourt et al., 2021). Recently, Bertsatos and Chovalopoulou (2020) developed a method to use CSG data to sort femora and tibiae with sensitivities reported as high as 1 for femora and 0.997 for tibiae. Their method used five cross sections from 3D image bone renderings to calculate 46 measurements for comparison, including cross section area and angles between planes. These measurements were used to compare probabilities of matches using z score analysis.

While their methods show the promise of using 3D data to sort commingled remains, identification of skeletal remains by 3D analysis of CT scans requires further testing and validation. Namely, there is a dimension reduction issue as 46 measurement variables were used for comparison instead of the entire 3D morphology. Also, the study only assessed the methodology on two long, intact bones. Finally, ancestry was not a consideration in the sample collection.

Mesh-to-mesh comparison, which uses imaging to compare full 3D bone scans, is another method that pairs bones with high accuracy and decreases the dimension reduction issue. Though 3D data can be collected with a 3D scanner, many methods are now shifting to computed tomography (CT) as these scans include both outer surface features (like 3D scanners) as well as information on internal morphology of the bone. Advantages of using CT data is that sampling is non-destructive and does not require a physical skeletal collection (Jeong et al., 2021).

Karell et al. (2016) presented a method of comparing an entire CT 3D mesh to another with software that computes a mesh-to-mesh value (MVC) to quantify how similar two bones are to each other. This method used a combination of CT and 3D surface scans of 45 humeri that originated from a variety of time periods and locations. 3D renderings of right bones were mirrored to the left, and the two renderings were aligned for comparison to calculate the MVC with lower MVCs indicating a more likely match. The manual version of this method had 100% sensitivity and specificity (Karell et al., 2016), demonstrating the power of using CT and 3D scan alignments for pair matching. However, Karell et al. (2016) noted important limitations of the study were that only intact humeri were used, ancestry was not considered, and non-open-source

software was used. Thus, the method needs to be tested on a variety of ancestries, bones, software to explore whether this method is broadly applicable.

Preliminary studies that this paper builds upon were conducted by Yangseung Jeong over the last few years (personal communication, May 16, 2023). Jeong's preliminary research used CT scan data of limb bones from 106 Korean individuals. Like Karell et al. (2016), Jeong used CT scans for mesh-to-mesh comparisons. Jeong's method used open-source software for file conversion, alignment, and mirroring. Instead of using one MVC value for the comparison, Jeong collected two descriptive statistics of bilateral asymmetry, D_{mean} (mean of distances in mm) and D_{sd} (standard deviation of distances in mm), for each comparison of two bones. These estimates were used to form three logistic regression models based on D_{mean} as a predictor, D_{sd} as a predictor, and both D_{mean} and D_{sd} as predictors. Jeong also analyzed males and females separately to account for sexual dimorphism in asymmetries. The models achieved 90.3–100% correct classification ratios (CCRs) depending on the bone and predictor combination. Jeong's preliminary research also included validation studies on both Korean and Chinese populations, revealing CCRs of up to 15.9% lower on Chinese populations. This is consistent with previous research that asymmetries are population specific (Auerbach & Ruff, 2006) and suggests that population-specific regression models will have more predictive power than general models applied to the entire population.

1.3 Thesis Objectives

The objective of this thesis is to establish a novel method of pair matching that is quick, cost effective, and accurate. It establishes a repeatable, quantifiable method to assess similarity between two bones, offering an objective alternative to subjective visual

pair matching methods. The method also addresses the dimension reduction issues of previous 3D methods by utilizing mesh-to-mesh comparison to compare the entire 3D structure of pairs. This thesis provides logistic regression models to overcome limitations of hypothesis testing in establishing a probability of a true match, while it is also the first digital pair-matching study of its kind to compare seven long bones in a U.S. population.

CHAPTER TWO: METHODS AND MATERIALS

2.1 Obtaining Sample Set from the Human Skeleton CT Scans

CT scans were obtained from the Forensic Anthropology Center (FAC) at the University of Tennessee. The FAC's Donated Skeletal Collection has over 1,800 individuals (University of Tennessee, 2022), approximately 800 of which have been CT scanned with the data stored as Digital Imaging and Communications in Medicine (DICOM) files. Each of these scans contains some combination of the cranium, mandible, clavicles, scapulae, humeri, radii, ulnae, femora, tibiae, fibulae, patellae, vertebrae, os coxae, sacrum, tali, calcanei, and metatarsals. The elements were previously CT scanned using specially designed cardboard boxes with markings and Styrofoam to ensure standardization of placement (Figure 1). Individuals for this study were selected that had ancestry data available. White American males and females were selected for analysis as there were insufficient samples to include collections from other ancestral groups. Of the 39 females included, age was known for 24 of the individuals and averaged 62.1 years with a range of 24–93 years (sd = 16.3) (Table 1). Ages were known for 55 of the 56 males and averaged 57.0 years with a range of 16–89 years and a standard deviation of 13.6 years (Table 1).

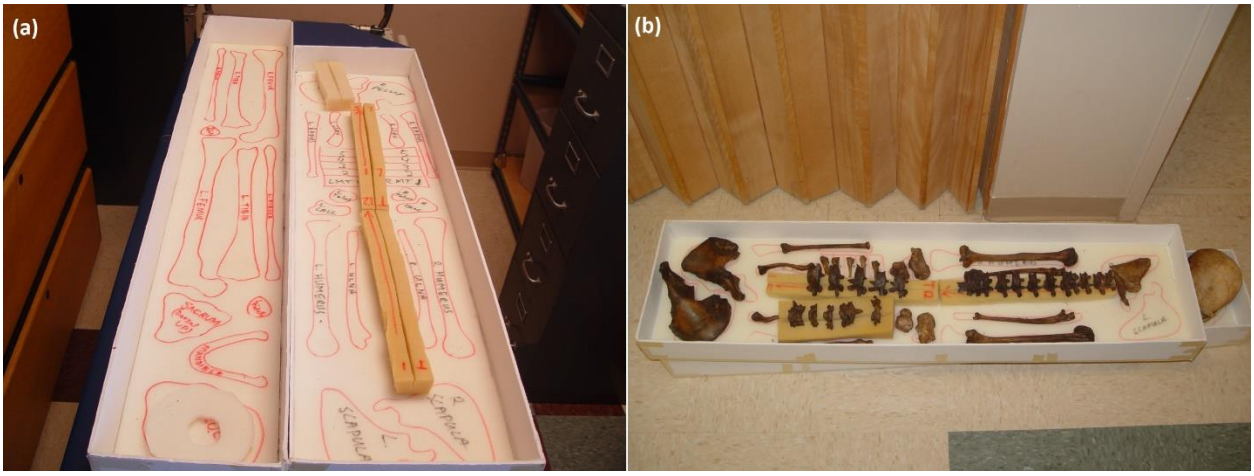

Figure 1

Preparation for CT scanning of skeletons at the Forensic Anthropology Center of the University of Tennessee

Note. CT scans were completed by previous researchers prior to this study. (a) The cardboard boxes included pre-cut Styrofoam and drawings for proper placement of. (b) The boxes were stacked and scanned together.

Table 1

Demographic information of age (years) for white males and females included in this study

	n	% age known	mean age of known	min. age	max. age	std. dev.
female	39	61.5%	62.1	24	93	16.3
Male	56	98.2%	57	16	89	13.6

Seven long bones were chosen for analysis: the clavicles, humeri, radii, ulnae, femora, tibiae, and fibulae. Individual skeletal elements were selected that were free from distortions due to scanning or rendering errors and completely intact with one

exception (one right tibia of a female that had an anterior excision for DNA extraction, about 3% of the tibiae used).

2.2 Segmentation of Human Skeletons in 3D Slicer

The DICOM files of the skeletons were imported and volume rendered in 3D Slicer (v5.0.1), an open-source software for displaying and analyzing CT scans (Brigham and Women's Hospital and 3D Slicer contributors, 2022; Fedorov et al., 2012). The lower image thresholds were increased to –500 to eliminate surrounding foam and other materials that had held the skeletal elements during the CT scan. Each individual long bone element of each skeleton was exported as its own stereolithography (.stl) file, a 3D image format (Figure 2).

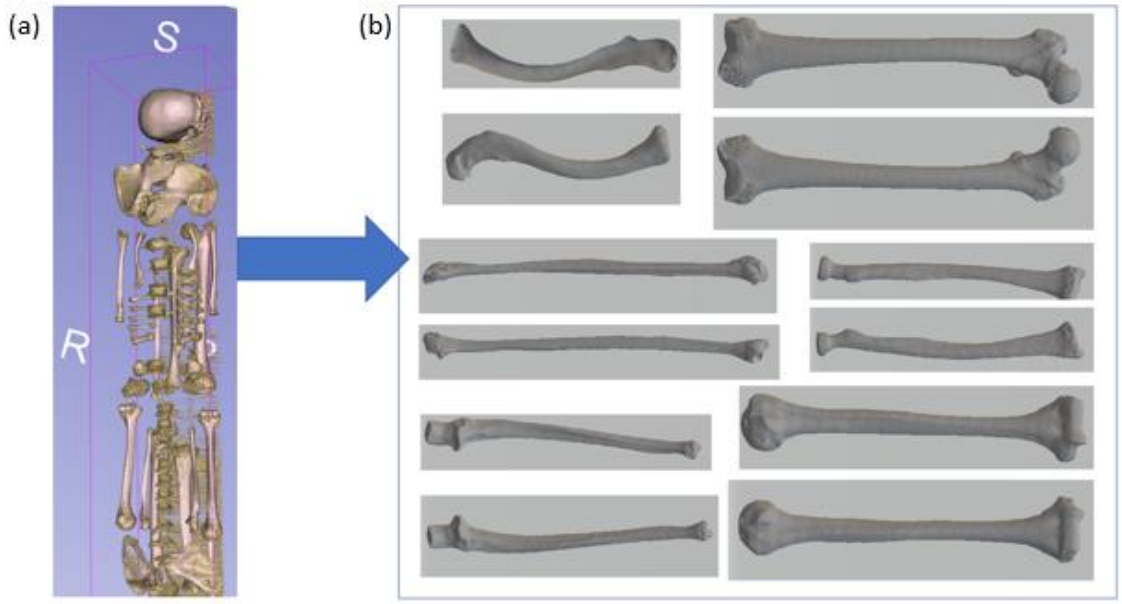

Figure 2

Segmentation of skeletons

Note. (a) The DICOM stacks were displayed in 3D slicer using the Volume Rendering module, and (b) individual 3D image files of the long bones used in this study were exported.

2.3 Mirroring of Left Bilateral Elements in MeshMixer

The files of the left bilateral skeletal elements exported from 3D slicer were mirrored to project a bilateral image in Autodesk Meshmixer (v3.5), an open-source software for creating and editing 3D scans and images (Autodesk, Inc., 2020) (Figure 3). The original bone was deleted from the bilateral image to leave only the mirror. The remaining mirror was exported as its own 3D .stil file.

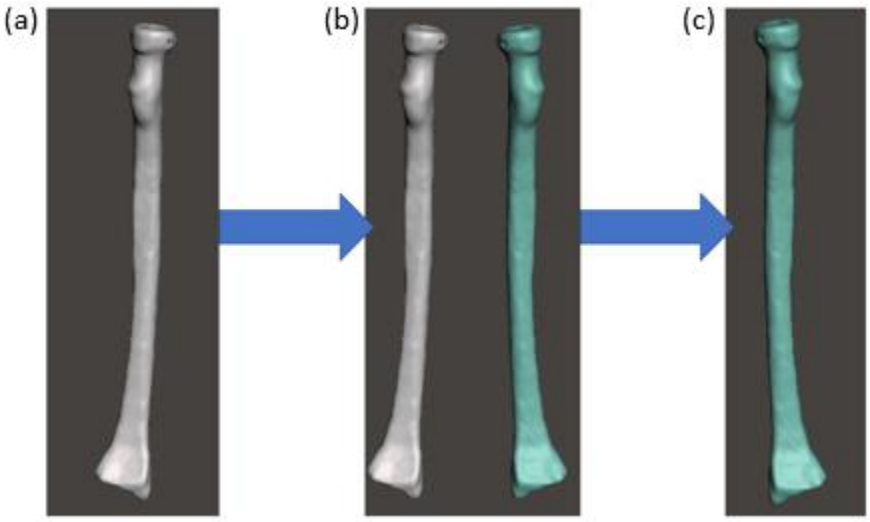

Figure 3

Mirroring of left bones

Note. Figure shows (a) an original left radius .stl file opened in MeshMixer; (b) the image after mirroring; and (c) the mirrored radius that was exported.

2.4 Aligning Right and Mirrored Left Bilateral Elements in CloudCompare

3D meshes of both the right long bone image and its corresponding left mirror image were analyzed in CloudCompare, an open-source software for analyzing 3D meshes (Girardeau-Montaut, 2023). A rough overlap of the bones was first performed by selecting three to four corresponding points on each of the meshes and aligning those. Then, the bones were finely aligned to allow for the maximum amount of overlap possible.

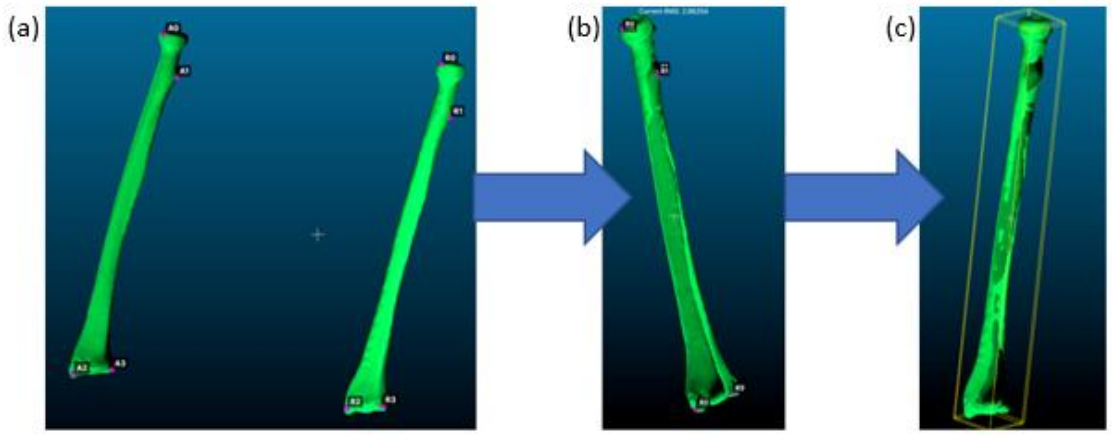

Figure 4

Alignment of right radius and left mirrored radius in CloudCompare

The image at left shows (a) Two radii (a right and a mirrored left image); (b) rough alignment of bones; (c) fine alignment.

2.5 Calculating D_{mean} and D_{sd} in CloudCompare

CloudCompare defines point clouds as the unorganized points that comprise a 3D image that can be organized into a "mesh" of triangles (CloudCompare, 2019). Thus, a mesh is a 3D rendering comprised of many vertices. The CloudCompare "Compute Cloud/Mesh Distance" tool assigns one of the bone element meshes as the "compared" and the other as the "reference". The distance was calculated between each vertex of the compared bone and the closest vertex on the reference bone. These distances were averaged to output the D_{mean}, and the standard deviation of these distances was output as the D_{sd}. By default, Cloud Compare will use a positive or negative sign depending on if the closest point was on the inside or outside of the reference. Thus, signed distances were removed to use the absolute value of the distances in the average.

This process was repeated to swap which bone was assigned as the "compared" and which was assigned as the "reference," and the D_{mean} and D_{sd} were recalculated. The D_{mean} and D_{sd} with the lowest sum were recorded. Repeating the calculation process twice with swapping which bone was the "compared" and which was the "reference" ensured the bone with the larger number of vertices was used as a reference per CloudCompare procedures (CloudCompare, 2019).

These methods of aligning a right bilateral element with its mirrored left were repeated for all elements within the same individual (i.e., match). Then, the methods were repeated with each right bilateral element and a corresponding left mirrored element from different individuals within the same sex (i.e., non-match). Since estimating non-match data was not feasible for all left and right elements, skeletal elements of non-matches were randomly selected for comparison. Table 2 outlines the number of comparisons made.

Table 2

Number of comparisons made for match and non-match for use in logistic regression models.

	female			male		
	match	**non-match**	**total**	**match**	**non-match**	**total**
clavicle	27	66	93	51	99	150
humerus	27	72	99	53	108	161
radius	35	71	106	52	96	148
ulna	35	72	107	51	102	153
femur	30	59	89	46	78	124
tibia	29	54	83	49	58	107
fibula	31	57	88	50	98	148

Note. Match represents comparisons made using bilateral elements of the same individual while non-match represents comparisons made where bilateral elements were from different individuals.

2.6 Statistical Analysis

For each long bone for each sex, means of D_{mean} and D_{sd} were calculated for the same individual (match) comparisons and for different individuals (non-match). For match and non-match datasets, z scores were calculated for each individual D_{mean} and D_{sd} data point. High leverage points (z scores +/–3) were checked to ensure that accurate measurements were obtained and were included in the analysis when bones exhibited no abnormalities or differences in mesh size due to scanning errors.

Independent t-tests were conducted with $\alpha = 0.0036$ (Bonferonni correction, $\alpha = 0.05/14$) to compare the means of D_{mean} and D_{sd} between the match and non-match groups separately within each sex. Additionally, D_{mean} and D_{sd} mean comparisons were conducted to compare match groups between the sexes and the non-match groups between sexes. Welch t-tests were used when the homogeneity of variances assumption

was not met. When the homogeneity of variance or normality assumptions were not met, a Mann-Whitney U test was also performed to compare the median ranks at $\alpha = 0.0036$. Additionally, the predictor data was transformed using the common logarithm to attempt to satisfy the normality and homogeneity of variances assumptions. Both the t-test and Mann-Whitney U tests were completed again for the transformed data with $\alpha = 0.0036$ (Appendix 1).

For each long bone of each sex, the match and non-match comparisons were then all combined into one data set with D_{mean} and D_{sd} as the predictors of the outcome variable "match", with 1 representing match comparisons and 0 representing non-match comparisons. Correlation was assessed between D_{mean} and D_{sd} in each data set to check for multicollinearity. Three logistic regression models were analyzed: i) D_{mean} was used as the predictor; ii) D_{sd} used as the predictor, and; iii) both D_{mean} and D_{sd} as the predictors. Classification tables were obtained for each of the models using a cutoff of 0.5. K-fold cross validation was performed on all models using 5 folds. Receiver operator characteristic (ROC) curves were generated for each model, and the area under the curve (AUC) was calculated for each ROC curve. Tables were generated of coordinates (data points) on the ROC curve and their corresponding to true positive (i.e., sensitivity) and false positive (i.e., 1 – specificity) rates for each regression model to assess the highest cutoff where the false positive rate (FPR, or 1 – specificity) was below 0.05.

Statistical analyses were conducted in R (v4.2.3; R Core Team, 2022), utilizing the following packages: caret (v6.0-93; Kuhn, 2008) and pROC (v1.18.0; Robin et al.,

2011). The R code used is available in the Appendix. Generation of tables of ROC coordinates and ROC plots were generated in SPSS, Version 27.0.

CHAPTER THREE: RESULTS

3.1 Collection of D_{mean} and D_{sd}

A total of 1,656 D_{mean}s and D_{sd}s were collected using CloudCompare from comparisons of right bone images to a mirrored left. Table 3 gives the means and standard deviations of the D_{mean} and D_{sd} values collected by bone and sex. Additionally, z scores were assessed for each data point in each set to identify high leverage points, those points having z scores +/–3, to ensure that there were no errors as a result of distortion or corruption of mesh sizes in the image, in the comparison process, or in recording the result and that there was no distortion or corruption of the image files. (Table 3).

The number of match comparisons made was 27–35 per bone for females and 46–53 per bone for males. For non-matches, there were 54–72 comparisons per bone for females and 78–108 comparisons per bone for males. For matches, the means of D_{mean} ranged from 0.48 mm (female ulna)–0.90 mm (female clavicle), and the means of D_{sd} were between 0.41 mm (female ulna)–0.79 mm (female clavicle). Meanwhile, for non-matches, means of D_{mean} ranged from 1.00mm (female radius)–2.35mm (male femur), and means of D_{sd} ranged from 0.90 mm (female radius)–2.31mm (female femur). In match groups, the male tibia had two D_{mean}s comparisons with a Z score greater than 3 or less than –3, representing the highest prevalence of high leverage points considering all

bones, sexes, and predictors. In non-match groups, the male ulna D_{mean}s, male clavicle D_{sd}s, and male radius D_{sd}s all had 3 high leverage points with z scores +/–3.

Table 3

Estimates of D_{mean} and D_{sd} for match and non-match comparisons

	match							non-match					
		D_{mean}			D_{sd}				D_{mean}			D_{sd}	
	n	$\bar{x}$	**s**	**z**	$\bar{x}$	**s**	**z**	**n**	$\bar{x}$	**s**	**z**	$\bar{x}$	**s**	**z**
							female						
cla	27	0.90	0.42	0	0.79	0.44	0	66	1.44	0.49	1	1.31	0.53	1
hum	27	0.64	0.11	1	0.57	0.15	1	72	1.50	0.46	1	1.46	0.64	1
rad	35	0.55	0.20	1	0.49	0.17	1	71	1.00	0.31	1	0.90	0.33	1
uln	35	0.48	0.12	1	0.41	0.08	1	72	1.03	0.29	1	0.91	0.35	1
fem	30	0.81	0.18	0	0.74	0.17	0	59	2.28	0.86	2	2.31	1.08	1
tib	29	0.71	0.20	0	0.68	0.23	0	54	1.55	0.47	2	1.42	0.52	1
fib	31	0.67	0.19	0	0.58	0.18	0	57	1.36	0.36	0	1.26	0.43	0
							male						
cla	51	0.86	0.18	0	0.71	0.18	0	99	1.36	0.31	1	1.19	0.34	3
hum	53	0.79	0.24	1	0.68	0.25	1	108	1.74	0.51	0	1.65	0.61	1
rad	52	0.56	0.21	1	0.49	0.18	1	96	1.24	0.38	2	1.11	0.47	3
uln	51	0.57	0.12	0	0.47	0.10	0	102	1.29	0.43	3	1.09	0.43	2
fem	46	0.83	0.20	0	0.73	0.22	0	78	2.35	2.41	1	0.98	1.17	0
tib	49	0.76	0.16	2	0.70	0.21	0	58	1.70	0.51	1	1.63	0.54	1
fib	50	0.69	0.17	1	0.58	0.16	1	98	1.37	0.38	0	1.29	0.49	1

Note. Table shows estimated means ($\bar{x}$) and standard deviations (s) of each of the data sets in mm. z indicates the number of data points with z scores that were either greater than 3 or less than -3 that were retained in the sample set. Clav – clavicle; hum – humerus; rad – radius; uln – ulna; fem – femur; tib – tibia; fib – fibula.

3.2 Mean Comparisons

Means were compared for females and males to assess if there were statistical differences between the match and non-match comparison groups. They were also compared between male and female match or non-match groups. It was concluded after

performing the Shapiro-Wilks and Barlett tests on these data sets that the assumptions of normality and homogeneity of variances were not met (Appendix 1). Density plots showed right skewness, and a log10 transformation was done on all data sets. This transformation allowed for the assumptions to be satisfied for some of the data sets. An independent t-test was performed for the data sets that met both assumptions after transformation, and the Welch t-test was substituted in cases where only the normality assumption was satisfied. For the remaining sets that still did not meet the assumptions after transformation, the original, untransformed data sets were used with a Mann Whitney U test. A Bonferroni corrected α of 0.0036 was used (= 0.05 / 14) to account for the total of fourteen hypothesis tests that were conducted. The tests indicated that for both sexes and for all bones, D_{mean} and D_{sd} were significantly different between match versus non-match (Table 4).

The same statistical process was completed for comparisons between male and female match groups and between male and female non-match groups (Table 5). The tests indicated that D_{mean} was similar between males and females for all bones except the humerus and ulna, and D_{sd} was similar between males and females for all bones except the humerus. For non-matches, both D_{mean} and D_{sd} were similar for all lower limb bones and clavicles. Because some D_{mean}s and D_{sd}s were similar between the sexes while others were not, logistic regression models were built separately for the sexes for consistency.

Table 4

Mean comparisons between match and non-match groups

	D_{mean}			D_{sd}		
	trans.	**test**	**result**	**trans.**	**test**	**result**
			female			
clavicle	n	MWU	2.8e–7*	n	MWU	5.5e–7*
humerus	n	MWU	1.0e–13*	n	MWU	4.8e–13*
radius	y	t-test	< 2e–16*	n	MWU	1.9e–11*
ulna	n	MWU	2.2e–15*	n	MWU	1.4e–15*
femur	n	MWU	1.9e–14*	n	MWU	2.0e–14*
tibia	n	MWU	1.4e–12*	n	MWU	2.7e–11*
fibula	n	MWU	1.2e–12*	n	MWU	1.5e–12*
			male			
clavicle	y	t-test	< 2e–16*	n	MWU	< 2e–16*
humerus	n	MWU	< 2e–16*	n	MWU	< 2e–16*
radius	n	MWU	< 2e–16*	n	MWU	< 2e–16*
ulna	n	MWU	< 2e–16*	n	MWU	< 2e–16*
femur	n	MWU	< 2e–16*	n	MWU	< 2e–16*
tibia	n	MWU	< 2e–16*	n	MWU	< 2e–16*
fibula	y	t-test	< 2e–16*	y	Welch	< 2e–16*

Note. trans. – indicates whether or not a transformation was used on the data sets with "y" indicating that a log10 transformation set was used and "n" indicating the original data set was used; MWU – Mann Whitney U; * indicates significance. $\alpha = 0.0036$.

Table 5

Mean comparison between male and female data sets of either match or non-match groups

	D_{mean}			D_{sd}		
	trans.	**test**	**result**	**trans.**	**test**	**result**
			match			
clavicle	n	MWU	0.16	n	MWU	0.34
humerus	n	MWU	2.6e–4*	n	MWU	3.2e–3*
radius	n	MWU	0.44	n	MWU	0.52
ulna	n	MWU	1.2e–4*	y	t-test	5.6e–3
femur	n	MWU	0.97	n	MWU	0.24
tibia	n	MWU	0.029	n	MWU	0.48
fibula	n	MWU	0.41	n	MWU	0.62
			non-match			
clavicle	n	MWU	0.83	n	MWU	0.57
humerus	n	MWU	1.11e–3*	n	MWU	0.017*
radius	n	MWU	1.6e–6*	n	MWU	4.0e–4*
ulna	n	MWU	1.6e–6*	n	MWU	3.6e–4*
femur	n	MWU	0.87	n	MWU	0.66
tibia	y	t-test	0.035	y	t-test	0.020
fibula	y	t-test	0.89	y	t-test	0.87

Note. trans. – indicates whether or not a transformation was used on the data sets with "y" indicating that a log10 transformation set was used and "n" indicating the original data set was used; MWU – Mann Whitney U; * – indicates significance. $\alpha = 0.0036$.

3.3 Generation of Logistic Regression Models

A total of 42 logistic regression models were generated, with coefficients to plug into the following equation:

$$[\boldsymbol{P(match)}] = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\beta_0 + \beta_1 \cdot D_{mean} + \beta_2 \cdot D_{sd})}}$$

P(match) represents the probability of a match (Tables 9–11). For each bone within each sex, three models were generated: i) D_{mean} as the predictor; ii) D_{sd} as the predictor; and

iii) both D_{mean} and D_{sd} as the predictors. D_{mean} and D_{sd} were correlated ($r > 0.7$) with Pearson's correlation coefficients in all sets of data ranging from 0.94 to 0.99 (Table 8). Though this indicates that the two predictor models are in violation of the assumption of logistic regression that the predictors are not correlated, the two predictor models were still included as the goal was to generate the highest accuracy possible.

Table 6

Pearson's correlation coefficient between D_{mean} and D_{sd}

	Female	**Male**
clavicle	0.97	0.95
humerus	0.94	0.96
radius	0.95	0.94
ulna	0.95	0.94
femur	0.98	0.99
tibia	0.97	0.98
fibula	0.95	0.94

A cutoff value of 0.5 was used to maximize correct classification ratios (CCR) and serve as a baseline for assessing the model's accuracy, which ranged from 0.86 to 1 for the female models and from 0.87 to 0.98 for the male models (Tables 7–9). For both male and female models, the model using both predictors either had the highest CCR or tied with a one predictor model for highest CCR for all bones (Table 9).

ROC AUC ranged from 0.84–1.00 for female models and from 0.89–0.99 for male models (Figures 5–8). Similarly to CCR, for all models of both sexes, the model using both predictors had the highest AUC or tied with at least one of the one predictor models for highest AUC.

K-fold cross validation with five folds was utilized to train and evaluate the models. Thus, the data set (containing both matches and non-matches) for each bone for each sex was divided into five nearly equal sets, where four of the data sets were used to train the data, while one of the data sets was used to test the data. This continued four more times until all of the sets had been used as a test set. This method allowed for using relatively small data sets to assess the future predictive power of the models without having to collect new data. Cross-validation accuracies ranged from 85–98%.

Table 7

Regression models with D_{mean} as a predictor

$$[\boldsymbol{P(match)}] = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\boldsymbol{\beta_0} + \boldsymbol{\beta_1} \cdot \boldsymbol{D_{mean}})}}$$

	β_0	β_1	CCR (0.5)	ROC AUC	FPR< 0.05 Cutoff	Sens	FPR	CV
					female			
clavicle	3.10	–3.58	0.88	0.84	0.48	0.67	0.045	0.87
humerus	11.89	–13.57	0.98	0.99	0.14	0.96	0.042	0.98
radius	6.45	–9.77	0.88	0.92	0.56	0.74	0.042	0.88
ulna	10.19	–15.96	0.96	0.97	0.32	0.94	0.042	0.95
femur	48.54	–37.58	0.97	1.00	0.19	1.00	0.034	0.97
tibia	10.14	–10.35	0.94	0.97	0.51	0.90	0.037	0.94
fibula	9.51	–10.77	0.91	0.96	0.51	0.87	0.035	0.93
					Male			
clavicle	10.28	–10.31	0.89	0.94	0.68	0.69	0.040	0.89
humerus	10.18	–9.80	0.96	0.98	0.47	0.94	0.046	0.96
radius	8.38	–10.75	0.95	0.97	0.35	0.94	0.042	0.96
ulna	17.43	–22.00	0.97	0.99	0.19	0.96	0.049	0.97
femur	11.76	–9.70	0.94	0.99	0.54	0.91	0.038	0.92
tibia	11.07	–10.31	0.93	0.97	0.73	0.92	0.034	0.93
fibula	10.18	–11.47	0.90	0.97	0.59	0.81	0.042	0.90

Note. CCR – correct classification ratio; ROC AUC – receiver operator characteristic area under the curve; FPR< 0.05 cutoff – cutoff value on ROC curve where false positive rate was less than 0.05; Sens – sensitivity, FPR – false positive rate; CV – cross validation accuracy using 5 folds.

Table 8

Regression models with D_{sd} as a predictor

$$[\boldsymbol{P(match)}] = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\beta_0 + \beta_1 \cdot D_{sd})}}$$

	β_0	β_1	CCR (0.5)	ROC AUC	FPR< 0.05 Cutoff	Sens	FPR	CV
				female				
clavicle	1.94	–2.85	0.86	0.83	0.46	0.63	0.045	0.86
humerus	9.00	–12.15	0.98	0.97	0.42	0.96	0.042	0.98
radius	5.17	–9.37	0.85	0.90	0.56	0.74	0.042	0.85
ulna	12.20	–22.91	0.96	0.98	0.43	0.94	0.042	0.96
femur	34.85	–30.76	0.97	1.00	0.27	1.0	0.034	0.96
tibia	6.47	–7.51	0.88	0.95	0.64	0.83	0.037	0.88
fibula	7.18	–9.49	0.92	0.96	0.73	0.74	0.035	0.87
				male				
clavicle	7.99	–9.71	0.88	0.92	0.61	0.75	0.040	0.88
humerus	7.76	–8.82	0.94	0.97	0.57	0.94	0.046	0.94
radius	8.23	–12.72	0.94	0.97	0.48	0.90	0.042	0.94
ulna	16.67	–25.80	0.95	0.99	0.27	0.96	0.049	0.94
femur	9.09	–8.16	0.92	0.99	0.65	0.89	0.038	0.92
tibia	8.12	–7.99	0.87	0.97	0.67	0.86	0.034	0.87
fibula	8.17	–11.08	0.90	0.96	0.69	0.71	0.042	0.90

Note. CCR – correct classification ratio; ROC AUC – receiver operator characteristic area under the curve; FPR< 0.05 cutoff – cutoff value on ROC curve where false positive rate was less than 0.05; Sens – sensitivity, FPR – false positive rate; CV – cross validation accuracy using 5 folds.

Table 9

Regression models with both D_{mean} and D_{sd} as predictors

$$[\boldsymbol{P(match)}] = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{-(\beta_0 + \beta_1 \cdot D_{mean} + \beta_2 \cdot D_{sd})}}$$

	β_0	β_1	β_2	CCR (0.5)	ROC AUC	FPR < 0.05 cutoff	Sens	FPR	CV
					female				
clavicle	3.57	–6.96	3.36	0.89	0.84	0.49	0.74	0.045	0.89
humerus	13.15	–19.47	5.01	0.98	0.99	0.091	1.0	0.042	0.98
radius	6.50	–12.25	2.81	0.89	0.92	0.53	0.77	0.042	0.88
ulna	11.09	–9.28	–9.70	0.96	0.98	0.38	0.94	0.042	0.95
femur	1776.9	–923.5	–489.8	1.00	1.00	5.9e–6	1.0	0.034	0.98
tibia	15.76	–35.07	21.06	0.96	0.99	0.27	0.97	0.037	0.92
fibula	9.80	–13.70	2.99	0.93	0.96	0.45	0.87	0.035	0.92
					male				
clavicle	10.35	–11.54	1.39	0.89	0.94	0.70	0.69	0.040	0.88
humerus	10.50	–11.22	1.27	0.96	0.98	0.45	0.94	0.046	0.94
radius	8.31	–15.29	5.49	0.96	0.97	0.33	0.94	0.042	0.96
ulna	17.92	–16.42	-7.55	0.97	0.99	0.21	0.96	0.049	0.96
femur	12.11	–12.08	2.14	0.94	0.99	0.56	0.91	0.038	0.92
tibia	11.26	–14.35	4.03	0.94	0.99	0.72	0.94	0.034	0.94
fibula	10.50	–10.36	–1.67	0.91	0.97	0.63	0.83	0.042	0.89

Note. CCR – correct classification ratio; ROC AUC – receiver operator characteristic area under the curve; FPR< 0.05 cutoff – cutoff value on ROC curve where false positive rate was less than 0.05; Sens – sensitivity, FPR – false positive rate; CV – cross validation accuracy using 5 folds.

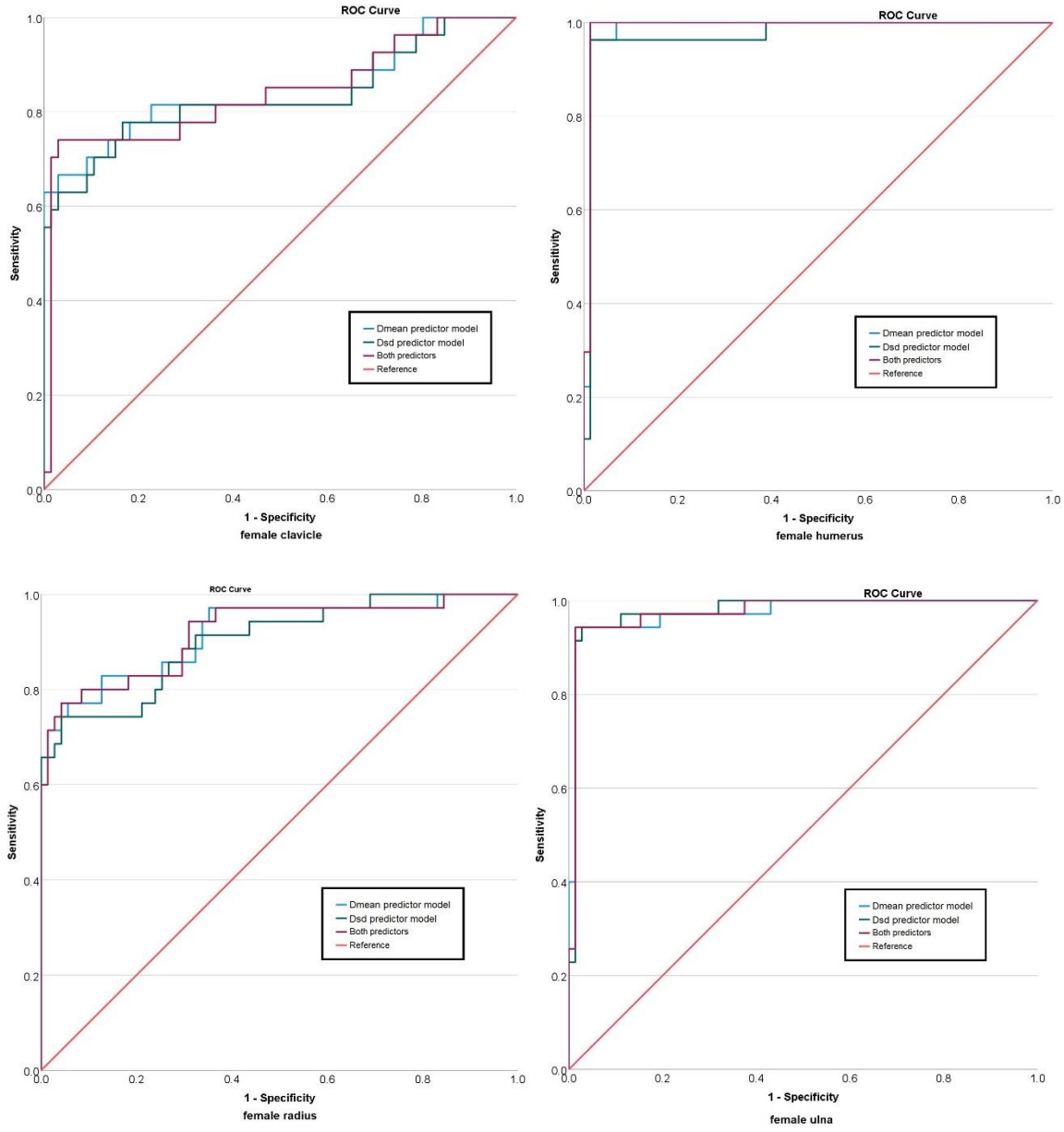

Figure 5

ROC curves for female models for upper body.

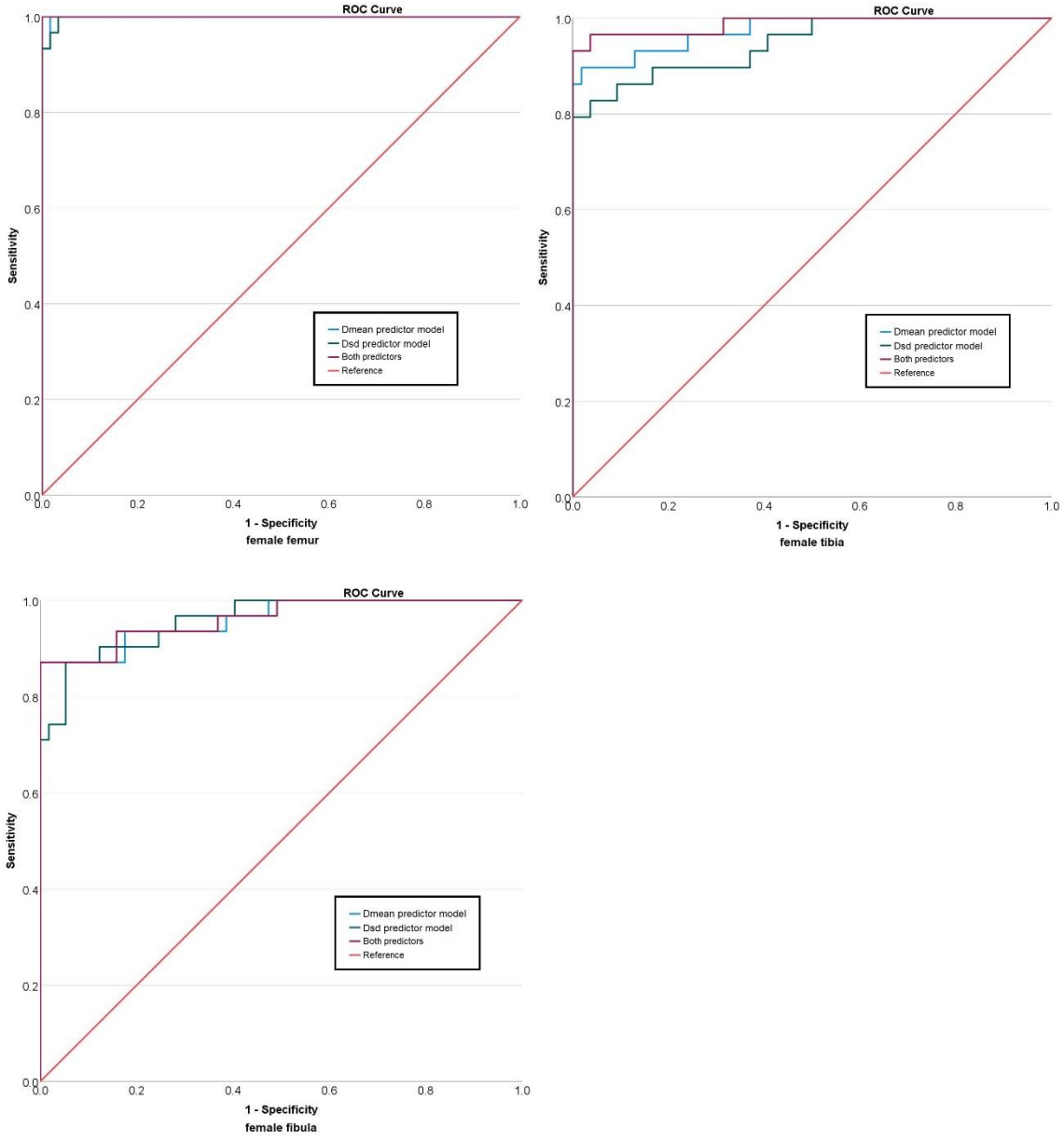

Figure 6

ROC curves for female models for lower body.

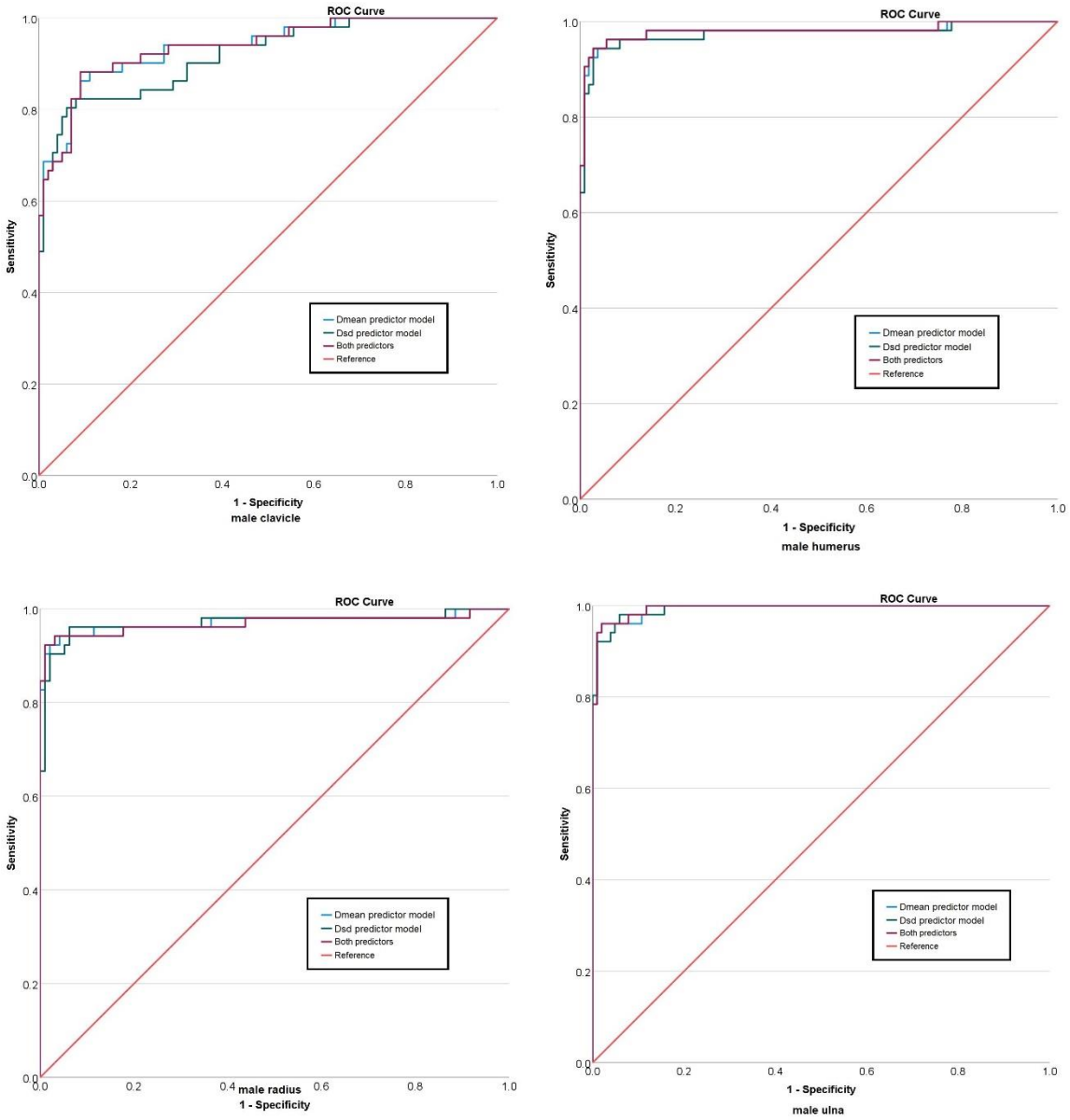

Figure 7

ROC curves for male models for upper body.

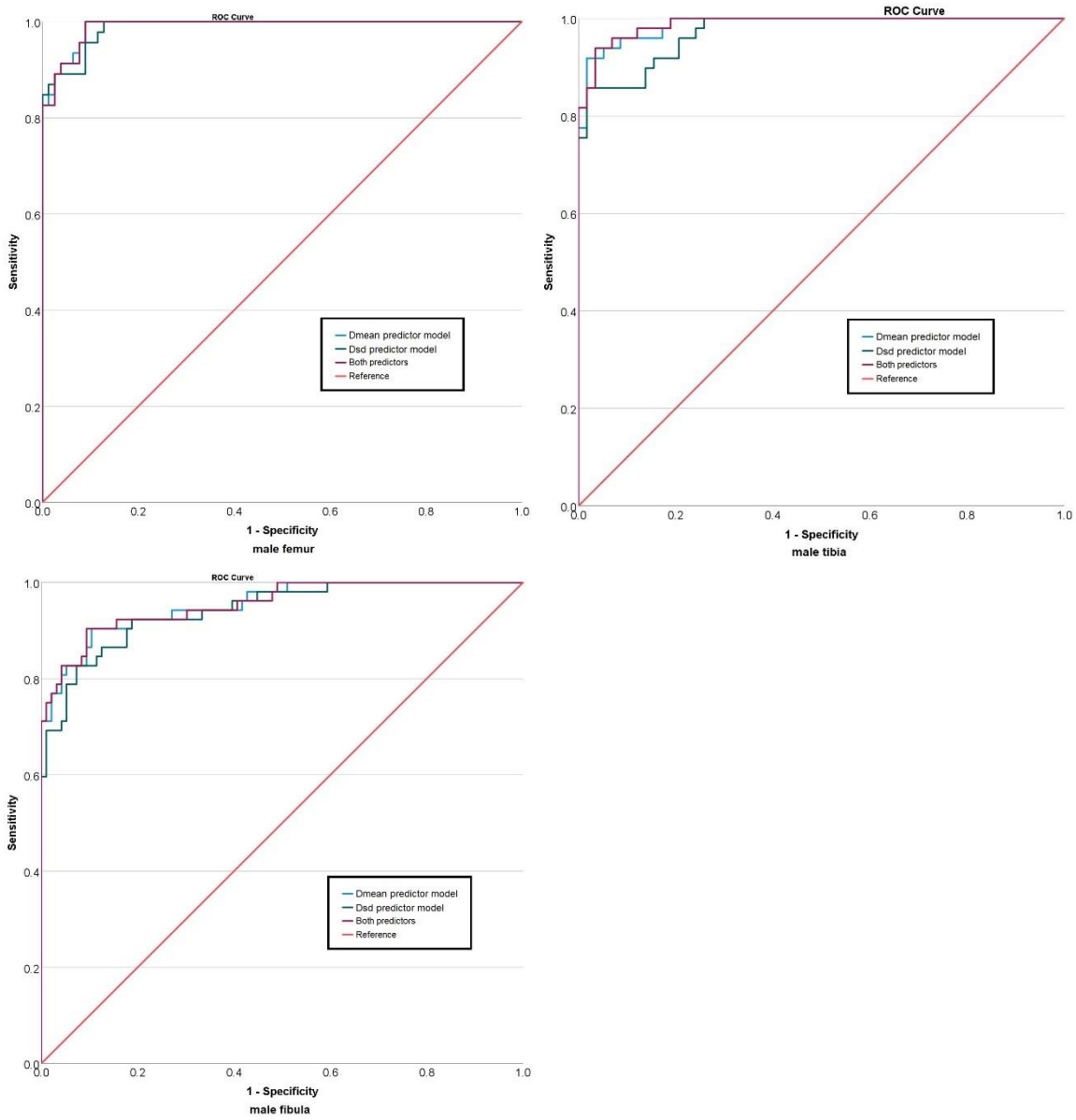

Figure 8

ROC curves for male models for lower body.

CHAPTER FOUR: DISCUSSION AND CONCLUSIONS

4.1 Summary of research design and results

A total of 1,656 digital comparisons were made for this research, comprising 7 different long bones for each sex. Each comparison consisted of a right bone compared to a mirrored left that belonged to either the same individual or a different individual. Forty-two logistic regression models were used to sort commingled skeletal remains by pair matching. The results showed high cross-validation accuracies (85 – 98%).

4.2 Sample size

In previous research, the ten events per variable (EPV) rule has been applied to determine an adequate sample size in binary logistic regression models, meaning that for each regression coefficient that is multiplied by a predictor in the model, there needs to be at least ten events of the lowest occurring outcome in the sample (van Smeden et al., 2019). Even considering the smallest number of events, which would be the female clavicle and humerus matches (27 events each), there are an adequate number of samples following the ten EPV rule for the model with the highest number of predictors (= 2). It has been argued that the claims that the ten EPV is adequate are baseless and that a much larger sample size is needed (van Smeden et al., 2016). When applied in medical settings, Bujang et al. (2018) concluded that a minimum sample size of 500 ensured an accurate reflection of the population. Future research and validation tests using these methods could continue to increase the number of comparisons until there are 500 for each bone by sex and ancestry.

4.3 Bilateral Asymmetry in Human Skeletons

One of the reasons for false negatives in pair matching is because bilateral asymmetry is common throughout nature, including in humans. Bones may be excluded as a possible match when they are a match because they are more dissimilar than expected due to naturally occurring asymmetries. The degree of bilateral asymmetry in humans varies by population and sex (Auerbach & Ruff, 2006). There are three types of asymmetries when considering population trends: directional, fluctuating, and antisymmetry (Klingenberg, 2015). Directional symmetry refers to expected variations caused by natural development, such as the heart being more to the left side of the body, while fluctuating asymmetry refers to variations in an individual that are outside of what is expected based off population norms (Klingenberg, 2015). Antisymmetry includes traits that are asymmetrical but can occur on either side, such as handedness. All of these asymmetries can occur at the same time in one trait (Klingenberg, 2015).

Previous research on human bilateral asymmetry has shown that, in general, right upper limbs and left lower limbs tend to have larger dimensions and that both males and females are more similar in asymmetry in lower than in upper limbs (Auerbach & Ruff, 2006). These discrepancies could be indicative of division of labor between sexes, though sexual dimorphism in limb asymmetries has reduced throughout human history (Ruff, 1987). The results presented in this study are consistent with previous research, indicating that lower limbs were similar in D_{mean} and D_{sd} between sexes while there were dissimilarities between sexes for D_{mean}s and D_{sd}s for matched humeri, D_{mean}s for matched ulnae, and D_{mean}s and D_{sd}s for nonmatched humeri, radii, and ulna. Asymmetries can be complex and the methodology employed here using 3D image analysis of bones ventures

outside of simple 2D characteristics like length while utilizing models that consider potential differences due to sexual dimorphism.

4.4 Advantages of Using Logistic Regression

Previous methods of pair matching, particularly osteometric pair-matching, rely on hypothesis testing, with the null hypothesis stating that the bones belong to the same individual. Rejection of the null hypothesis can be valuable for excluding two bones from belonging to the same individual. However, since only two outcomes are possible (rejecting the null hypothesis or failing to reject), this can lead to accidental elimination of a true match that may have had higher than expected bilateral asymmetry. With logistic regression models, the probability of bones belonging to the same individual is estimated, allowing for comparison of probabilities between different pairings in a commingled assemblage.

With logistic regression, the lower the cutoff threshold (the probability at which you accept a comparison as a match), the higher the sensitivity, or the ability of the model to detect a true match. However, this also means that the FPR (stating something is a match when it is not) is also high. Increasing the cutoff will lead to a lower FPR, minimizing the risk of pairing the wrong bones together. Moving the cutoff value can also allow for the consideration of antemortem lifestyle behaviors when known that may have influenced bilateral asymmetry. Further research can combine that of bilateral asymmetry as it relates to human behavior and factor that into cutoff value decisions. Similarly, analyses of the ROC curves for each model allows for the adjustment of sensitivity and specificity according to specific cases and desired levels of risk.

4.5 Digital Bone Loss in Females

The same threshold was used to isolate bone tissue from the foam material it was sitting on in 3D slicer for both males and females. This could result in digital bone loss depending on the density of the hard tissues. This was most commonly seen in female individuals. The methodologies could be refined to minimize digital bone loss by establishing different thresholds for males and females. Also, a threshold range for isolating hard tissue from soft tissue or foam could also be established so that the value that results in the least amount of bone loss could be selected on an individual basis. More research would need to be established to create a method to select the best range objectively instead of based upon visual cues only.

4.6 Comparing Bone Fragments

This research used fully intact bones. However, human skeletal remains can be fragmented or have various types of peri or post mortem trauma. Further research could apply these models to fragments to test the classification accuracy and validity of the models in these samples. Preparing the segmented .stl files would require establishing cutoff locations for the larger fragment such that the parts of the smaller fragment are completely contained within the larger fragment. Then, the larger fragment can be used as the reference bone when comparisons are made in CloudCompare.

4.7 Further Research in Different Sample Sets

Because of the sample set available for this research, only models for one ancestry, white American, were generated. First, validation tests are needed on the same ancestral groups and bones to determine the accuracy of the models on other sample sets. Next, more models are needed from bones of the skeletal remains of individuals with

other ancestries. Additionally, comparisons need to be made between sexes and ancestries for generalized models for assemblages when these variables are unknown. Until generalized models are developed, ancestry-specific models can be applied to unknown bones separately, with their respective results reported.

4.8 Bones of Disparate Sizes

When using these proposed pair matching methods, it is recommended to first eliminate bones as matches that are highly disparate in size. Not only does it take considerably longer to align the bones and compute D_{mean} and D_{sd} in CloudCompare, but there are larger than expected discrepancies between the two output sets when the "compared" and "reference" are swapped. This could indicate that the methods are not as reliable when bones are very different in size.

4.9 Conclusion

Pair matching via mesh-to-mesh comparisons through CT generated 3D renderings in conjunction with logistic regression analysis provides an inexpensive, accurate, objective, and efficient method of pair matching. The research presented here overcame limitations of dimension reduction, subjectivity, and the limits of hypothesis testing associated with previous osteometric pair matching methods. Additionally, the techniques presented require little training to reach high proficiency, eliminating error due to the experience level of the practitioner. These methods can contribute to the successful segregation of commingled human remains and subsequent identification and return to next of kin.

REFERENCES

Adams, B., & Byrd, J. E. (2014). *Commingled Human Remains: Methods in Recovery, Analysis, and Identification*. Elsevier Inc. https://doi-org.ezproxy.mtsu.edu/10.1016/C2012-0-02768-8

Adams, B. J., & Byrd, J. E. (2008). Models and Methods for Osteometric Sorting. In: Adams, B. J., Byrd, J. E. (eds) *Recovery, Analysis & Identification of Commingled Human Remains* (pp. 199–220). https://doi-org.ezproxy.mtsu.edu/10.1007/978-1-59745-316-5_10

Adams, B. J., & Konigsberg, L. W. (2008). How Many People? Determining the Number of Individuals Represented by Commingled Human Remains. In: Adams, B. J., Byrd, J. E. (eds) *Recovery, Analysis, and Identification of Commingled Human Remains* (pp. 241–253). Humana Press. https://doi-org.ezproxy.mtsu.edu/10.1007/978-1-59745-316-5_12

Adams, B. J., & Konigsberg, L. W. (2004). Estimation of the most likely number of individuals from commingled human skeletal remains. *American Journal of Physical Anthropology, 125*(2), 138.

Andersen, M. M., & Balding, D. J. (2018). How many individuals share a mitochondrial genome? *PLoS genetics*, *14*(11), e1007774. https://doi.org/10.1371/journal.pgen.1007774

Auerbach, B. M., & Ruff, C. B. (2006). Limb bone bilateral asymmetry: variability and commonality among modern humans. *Journal of Human Evolution, 50*(2), 203–218. https://doi-org.ezproxy.mtsu.edu/10.1016/j.jhevol.2005.09.004

Autodesk, Inc. (2020). *Autodesk Meshmixer*. meshmixer.com. https://meshmixer.com/

Bertsatos A., & Chovalopoulou, M. E. (2020). Advances in Osteometric Sorting: Utilizing Diaphyseal CSG Properties for Lower Limb Skeletal Pair-Matching. *Journal of Forensic Sciences. 65*(5), 1400–1405. doi:10.1111/1556-4029.14480

Brigham and Women's Hospital and 3D Slicer contributors. (2022, November 22). 3D Slicer image computing platform. https://www.slicer.org/

Bujang, M. A., Sa'at, N., Sidik, T. M. I. T. A. B., & Joo, L. C. (2018). Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. *The Malaysian Journal of Medical Sciences*, *25*(4), 122–130. https://doi.org/10.21315/mjms2018.25.4.12

Buikstra, J. E., & Gordon, C. C. (1980). Individuation in forensic science study: decapitation. *Journal of forensic sciences*, *25*(1), 246-259.

Byrd, J. E., & Adams, B. J. (2003). Osteometric sorting of commingled human remains. *Journal of Forensic Sciences*, *48*(4), 717–724.

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. (2019). Chapter 7 - Processing, resolving commingling, and preserving remains. In: Christensen, A. M., Passalacqua, N. V., Bartelink, E. (eds) *Forensic Anthropology (Second Edition)* (pp. 217–242). Academic Press. https://doi.org/10.1016/B978-0-12-815734-3.00007-5

CloudCompare. (2019, February 20). *Entities*. Entities - CloudCompareWiki. https://www.cloudcompare.org/doc/wiki/index.php/Entities

Court, S. D. (2021). Mitochondrial DNA in forensic use. *Emerging topics in life sciences*, *5*(3), 415–426. https://doi.org/10.1042/ETLS20210204

Damann, F. E., & Edson, S. M. (2008). Sorting and Identifying Commingled Remains of U.S. War Dead: The Collaborative Roles of JPAC and AFDIL. In: Adams, B. J., Byrd, J. E. (eds) *Recovery, Analysis, and Identification of Commingled Human Remains* (pp. 301–314). Humana Press. https://doi-org.ezproxy.mtsu.edu/10.1007/978-1-59745-316-5_16

Defense POW/MIA Accounting Agency. (2022). *USS oklahoma project*. DPAA USS Oklahoma. https://dpaa-mil.sites.crmforce.mil/Projects/WWII/USSOklahoma/InfoSheet

Fancourt, H., Lynch, J. J., Byrd, J. E., & Stephan, C. N. (2021). Next-generation osteometric sorting: Using 3D shape, elliptical Fourier analysis, and Hausdorff distance to optimize osteological pair-matching. *Journal of Forensic Sciences, 66*(3), 821–836. https://doi.org/10.1111/1556-4029.14681

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J-C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F. M., Sonka, M., Buatti, J., Aylward, S. R., Miller, J. V., Pieper, S., & Kikinis, R. (2012). 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging, 30*(9), 1323–1341. PMID: 22770690. PMCID: PMC3466397

Finlayson, J. E., Bartelink, E. J., Perrone, A. and Dalton, K. (2017). Multimethod Resolution of a Small-Scale Case of Commingling. *Journal of Forensic Sciences, 62*, 493-497. https://doi.org/10.1111/1556-4029.13265

Gaudio, D., Fernandes, D. M., Schmidt, R., Cheronet, O., Mazzarelli, D., Mattia, M., O'Keeffe, T., Feeney, R. N. M., Cattaneo, C., Pinhasi, R. (2019). Genome–Wide DNA from Degraded Petrous Bones and the Assessment of Sex and Probable Geographic Origins of Forensic Cases. *Scientific Reports, 9*(8226). https://doi.org/10.1038/s41598-019-44638-w

Girardeau-Montaut, D. (2023, May 29). *CloudCompare – Open source project.* CloudCompare. http://www.cloudcompare.org/

Jeong, Y., Lee, S., Choi, I., Min, S., Ali, O., & Woo, E. J. (2021). A three–dimensional (3D) approach to estimating sex based on the subpubic angle of the contemporary Korean population. *Australian Journal of Forensic Sciences,* 1–11. https://doi-org.ezproxy.mtsu.edu/10.1080/00450618.2021.1998627

Jessee, E., & Skinner, M. (2005). A typology of mass grave and mass grave–related sites. *Forensic Science International, 152*(1), 55–59. https://doi-org.ezproxy.mtsu.edu/10.1016/j.forsciint.2005.02.031

Karell, M. A., Langstaff, H. K, Kranioti, E. F., Halazonetis, D. J., Minghetti, C., & Frelat, M. (2016). A novel method for pair-matching using three–dimensional digital models of bone: mesh-to-mesh value comparison. *International Journal of Legal Medicine, 130*(5), 1315–1322. https://doi-org.ezproxy.mtsu.edu/10.1007/s00414-016-1334-3

Keerti, A., & Ninave, S. (2022). DNA Fingerprinting: Use of Autosomal Short Tandem Repeats in Forensic DNA Typing. *Cureus, 14*(10), e30210. https://doi.org/10.7759/cureus.30210

Klingenberg, C. P. (2015). Analyzing Fluctuating Asymmetry with Geometric Morphometrics: Concepts, Methods, and Applications. *Symmetry*, *7*(2), 843–934. https://doi-org.ezproxy.mtsu.edu/10.3390/sym7020843

Kowalczyk, M., Domaradzki, P., Staniszewski, A., Kamiñska, K., & Horecka, B. (2021). Advantages, Possibilities, and Limitations of Mitochondrial DNA Analysis in Molecular Identification. *Folia Biologica (Poland)*, *69*(3), 101–111. https://doi-org.ezproxy.mtsu.edu/10.3409/fb_69-3.12

Kuhn, M. (2008). "Building Predictive Models in R Using the caret Package." *Journal of Statistical Software, 28(5)*, 1–26. doi:10.18637/jss.v028.i05, https://www.jstatsoft.org/index.php/jss/article/view/v028i05.

Latham, K. E., Miller, J. J. (2018). DNA recovery and analysis from skeletal material in modern forensic contexts. *Forensic Sciences Research*, *4*(1), 51–59. https://doi: 10.1080/20961790.2018.1515594

LeGarde, C. B. (2019). Preliminary Findings from a Visual Pair-Matching Study in a Large Commingled Assemblage. *Forensic Anthropology, 2*(2), 65. https://doi-org.ezproxy.mtsu.edu/10.5744/fa.2019.1001

Linacre, A., & Ottens, R. (2016). DNA: hair analysis. *Encyclopedia of Forensic and Legal Medicine (Second Edition)* (pp. 337–342). Elsevier. https://doi.org/10.1016/B978-0-12-800034-2.00157-9

Lynch, J. J. (2018). An automated two-dimensional form registration method for osteological pair-matching. *Journal of Forensic Sciences 63*, 1236–1242.

Mundorff, A. Z., Shaler, R., Bieschke, E. T., & Mar-Cash, E. (2014). Chapter 12 - Marrying Anthropology and DNA: Essential for Solving Complex Commingling Problems in Cases of Extreme Fragmentation. In: Adams, B. J., Byrd, J. E. (eds) *Recovery, Analysis, and Identification of Commingled Human Remains* (pp. 257–273). Humana Press. https://doi-org.ezproxy.mtsu.edu/10.1016/B978-0-12-405889-7.00012-5

R Core Team. (2022). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., & Müller, M. (2011). "pROC: an open-source package for R and S+ to analyze and compare ROC curves." *BMC Bioinformatics*, *12*, 77.

Ruengdit, S., Troy Case, D., & Mahakkanukrauh, P. (2020). Cranial suture closure as an age indicator: A review. *Forensic Science International*, *307*, 110111. https://doi.org/10.1016/j.forsciint.2019.110111

Ruff, C. (1987). Sexual dimorphism in human lower limb bone structure: relationship to subsistence strategy and sexual division of labor. *Journal of Human Evolution*, *16*(5), 391–416. https://doi-org.ezproxy.mtsu.edu/10.1016/0047-2484(87)90069-8

Schaefer, M. (2008). Patterns of Epiphyseal Union and Their Use in the Detection and Sorting of Commingled Remains. In: Adams, B.J., Byrd, J.E. (eds) *Recovery, Analysis, and Identification of Commingled Human Remains* (pp. 341–240). Humana Press. https://doi-org.ezproxy.mtsu.edu/10.1007/978-1-59745-316-5_11

Shukla, S., Mahato, P. K., Agarwal, P., & Shukla, S. (2022). A Morphometric study of different parameters of greater sciatic notch relation to sexual dimorphism in north Indian population. *European Journal of Molecular and Clinical Medicine, 9*(1), 1435. https://link.gale.com/apps/doc/A698246018/HRCA?u=anon~c6e15caa&sid=googleScholar&xid=f67e55e0

Tuller, H., & Hofmeister, U. (2014). Chapter 2 - Spatial Analysis of Mass Grave Mapping Data to Assist in the Reassociation of Disarticulated and Commingled Human Remains. *Commingled Human Remains*, 7–32. https://doi-org.ezproxy.mtsu.edu/10.1016/B978-0-12-405889-7.00002-2

Ubelaker, D. H., & Khosrowshahi, H. (2019). Estimation of age in forensic anthropology: historical perspective and recent methodological advances. *Forensic Sciences Research, 4*(1), 1–9. https://doi.org/10.1080/20961790.2018.1549711

University of Tennessee (2022, August 11). *Fast Facts*. Forensic Anthropology Center. Retrieved May 29, 2023, from http://fac.utk.edu/fast-facts/

van Smeden, M., de Groot, J. A., Moons, K. G., Collins, G. S., Altman, D., Eijkemans, M. J. C., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology, 16*, 163. https://doi.org/10.1186/s12874-016-0267-3

van Smeden, M., Moons, K. G. M., de Groot, J. A. H., Eijkemans, M. J. C., Reitsma, J. B., Collins, G. S., & Altman, D. G. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research, 28*(8), 2455-2474. https://doi-org.ezproxy.mtsu.edu/10.1177/0962280218784726

Vickers, S., Lubinski, P. M., Henebry Deleon, L., & Bowen, J. T. (2015). Proposed Method for Predicting Pair Matching of Skeletal Elements Allows Too Many False Rejections. *Journal of Forensic Sciences, 60*(1), 102–106. https://doi-org.ezproxy.mtsu.edu/10.1111/1556-4029.12545

Yazedjian, L., & Kešetović, R. (2008). The Application of Traditional Anthropological Methods in a DNA-Led Identification Process. In: Adams, B. J., Byrd, J. E. (eds) *Recovery, Analysis & Identification of Commingled Human Remains* (pp. 271–284). https://doi-org.ezproxy.mtsu.edu/10.1007/978-1-59745-316-5_14

APPENDIX

APPENDIX 1: ADDITIONAL TABLES

Appendix 1-1

Mean comparison between match and non-match groups using untransformed data

	D_{mean}				D_{sd}			
	female							
	SW	**Bart**	**t-test**	**MWU**	**SW**	**Bart**	**t-test**	**MWU**
cla	1.6e–4, 5.3e–5	0.34†	1.8e–6*	2.8e–7*	3.7e–5, 3.8e–5	0.28†	2.7e–5*	5.5e–7*
hum	2.9e–3, 1.1e–5	3.5e–11	< 2e–16*	1.0e–13*	8.7e–6, 2.9e–6	7.9e–12	< 2e–16*	4.8e–13*
rad	2.1e–4, 1.1e–4	5.2e–3	9.4e–15*	2.6e–12*	3.1e–4, 1.2e–4	4.1e–5	1.3e–13*	1.9e–11*
uln	6.7e–6, 7.2e–4	9.0e–8	< 2e–16*	2.2e–15*	6.3e–3, 1.4e–6	1.8e–14	< 2e–16*	1.4e–15*
fem	1.3e–3, 4.4e–6	3.2e–14	< 2e–16*	1.9e–14*	0.09†, 6.5e–6	< 2e–16*	2.5e–16*	2.0e–14*
tib	4.4e–5, 1.7e–4	8.5e–6	< 2e–16*	1.4e–12*	6.4e–5, 7.5e–5	1.1e–5	1.6e–13*	2.7e–11*
fib	2.4e–4, 8.9e–3	5.7e–4	< 2e–16*	1.2e–12*	2.9e–4, 8.4e–3	2.4e–6	< 2e–16*	1.5e–12*
	male							
cla	0.011, 4.6e–4	1.6e–4	< 2e–16*	< 2e–16*	7.1e–4, 9.7e–6	2.6e–6	< 2e–16*	< 2e–16*
hum	7.9e–9, 4.7e–4	1.1e–8	< 2e–16*	< 2e–16*	1.7e–10, 3.6e–4	8.6e–11	< 2e–16*	< 2e–16*
rad	2.8e–10, 9.7e–9	2.5e–6	< 2e–16*	< 2e–16*	2.6e–10, 2.0e–10	5.6e–12	< 2e–16*	< 2e–16*
uln	0.10†, 7.0e–10	< 2e–16*	< 2e–16*	< 2e–16*	0.052†, 1.2e–10	< 2e–16*	< 2e–16*	< 2e–16*
fem	5.8e–5, 3.1e–6	< 2e–16*	< 2e–16*	< 2e–16*	1.4e–5, 4.1e–6	< 2e–16*	< 2e–16*	< 2e–16*
tib	1.1e–5, 1.6e–4	3.5e–13	< 2e–16*	< 2e–16*	1.3e–6, 7.8e–4	3.5e–10	< 2e–16*	< 2e–16*
fib	3.5e–4, 7.6e–4	1.7e–8	< 2e–16*	< 2e–16*	5.1e–4, 2.9e–5	9.0e–15	< 2e–16*	< 2e–16*

Note. Shapiro-Wilks (SW) results list match then non-match. For SW and Bartlett (Bart) tests, p values greater than 0.05 indicate satisfaction of the normality and homogeneity of variance assumptions respectively and are annotated with †. For the t-tests and Mann-Whitey U tests (MWU), p-values less than 0.05 indicate a significant difference between the compared groups and are annotated with *.

Appendix 1-2

Mean comparison between male and female data sets of either match or non-match groups using untransformed data sets

	D_{mean}				D_{sd}			
	match							
	SW	**Bart.**	**t-test**	**MWU**	**SW**	**Bart**	**t-test**	**MWU**
cla	1.6e–4	1.1e–6	0.66	0.16	3.7e–5	5.2e–8	0.39	0.34
	0.011				7.1e–4			
hum	2.9e–3	1.1e–4	3.0e–4*	2.6e–4*	8.7e–6	5.6e–3	0.021*	3.2e–3*
	7.9e–9				1.7e–10			
rad	2.1e–4	0.70†	0.84	0.44	3.1e–4	0.66†	0.84	0.52
	2.8e–10				2.6e–10			
uln	6.7e–6	0.65†	9.3e–4*	1.2e–4*	6.3e–3	0.20†	6.9e–3*	4.9e–3*
	0.10†				0.052†			
fem	1.3e–3	0.44†	0.59	0.97	0.090†	0.17†	0.86	0.24
	5.8e–5				1.4e–5			
tib	4.4e–5	0.21†	0.26	0.029*	6.4e–5	0.53†	0.71	0.48
	1.1e–5				1.3e–6			
fib	2.5e–4	0.47†	0.62	0.41	2.9e–4	0.35†	0.95	0.62
	3.5e–4				5.1e–4			
	non-match							
cla	5.3e–5	2.3e–5	0.21	0.83	3.8e–5	4.1e–5	0.11	0.57
	4.6e–4				9.7e–6			
hum	1.1e–5	0.31†	1.8e–3*	1.11e–3*	2.9e–6	0.61†	0.051	0.017*
	4.7e–4				3.6e–4			
rad	1.1e–4	0.051†	2.1e–5*	1.6e–6*	1.2e–4	1.6e–3	6.0e–4*	4.0e–4*
	9.7e–9				2.0e–10			
uln	7.2e–4	6.6e–4	3.7e–6*	1.6e–6*	1.4e–6	0.056†	4.3e–3*	3.6e–4*
	7.0e–10				1.2e–10			
fem	4.4e–6	0.30†	0.65	0.87	6.5e–6	0.49†	0.61	0.66
	3.1e–6				4.1e–6			
tib	1.7e–4	0.51†	0.089	0.035*	7.5e–5	0.85†	0.041*	0.013*
	1.6e–4				7.8e–4			
fib	8.9e–3	0.57†	0.94	0.97	8.4e–3	0.24†	0.72	0.997
	7.6e–4				2.9e–5			

Note. Shapiro-Wilks (SW) results list match then non-match. For SW and Bartlett (Bart) tests, p values greater than 0.05 indicate satisfaction of the normality and homogeneity of variance assumptions respectively and are annotated with †. For the t-tests and Mann-Whitey U tests (MWU), p-values less than 0.05 indicate a significant difference between the compared groups and are annotated with *.

Appendix 1-3

Mean comparison between match and non-match groups using log10 transformed data

	D_{mean}				D_{sd}			
	female							
	SW	**Bart**	**t-test**	**MWU**	**SW**	**Bart**	**t-test**	**MWU**
cla	0.015, 0.021	0.13†	3.2e–9*	2.8e–7*	4.8e–3, 0.11†	0.22†	2.1e–7*	1.1e–6*
hum	0.079†, 0.012	2.1e–3	< 2e–16*	1.0e–13*	2.6e–3, 0.10†	6.4e–4	< 2e–16*	4.8e–13*
rad	0.088†, 0.10†	0.53†	< 2e–16*	2.6e–12*	0.08†, 0.037	0.39†	5.1e–14*	1.9e–11*
uln	1.7e–3, 0.42†	0.087†	< 2e–16*	2.2e–15*	0.34†, 0.044	2.0e–4	< 2e–16*	1.4e–15*
fem	0.068†, 0.026	3.6e–3	< 2e–16*	1.9e–14*	0.30†, 0.031	1.5e–3	< 2e–16*	2.0e–14*
tib	1.4e–3, 0.26†	0.51†	< 2e–16*	1.4e–12*	5.5e–3, 0.38†	0.39†	2.0e–15*	2.7e–11*
fib	0.018, 0.08†	0.995†	< 2e–16*	1.2e–12*	0.016, 0.49†	0.28†	< 2e–16*	1.5e–12*
	male							
cla	0.50†, 0.33†	0.79†	< 2e–16*	< 2e–16*	0.10†, 0.027	0.30†	< 2e–16*	< 2e–16*
hum	2.4e–4, 0.33†	0.14†	< 2e–16*	< 2e–16*	1.5e–5, 0.21†	0.012	< 2e–16*	< 2e–16*
rad	3.9e–6, 2.0e–3	0.99†	< 2e–16*	< 2e–16*	1.0e–5, 7.2e–4	0.026	< 2e–16*	< 2e–16*
uln	0.85†, 2.4e–4	0.032	< 2e–16*	< 2e–16*	0.67†, 5.8e–5	7.4e–4	< 2e–16*	< 2e–16*
fem	3.4e–3, 0.13†	1.4e–4	< 2e–16*	< 2e–16*	2.1e–3, 0.12†	2.6e–4	< 2e–16*	< 2e–16*
tib	1.4e–3, 0.70†	0.011	< 2e–16*	< 2e–16*	1.4e–4, 0.65†	0.16†	< 2e–16*	< 2e–16*
fib	0.059†, 0.24†	0.26†	< 2e–16*	< 2e–16*	0.10†, 0.32†	3.4e–3	< 2e–16*	< 2e–16*

Note. Shapiro-Wilks (SW) results list match then non-match. For SW and Bartlett (Bart) tests, p values greater than 0.05 indicate satisfaction of the normality and homogeneity of variance assumptions respectively and are annotated with †. For the t-tests and Mann-Whitey U tests (MWU), p-values less than 0.05 indicate a significant difference between the compared groups and are annotated with *.

Appendix 1-4

Mean comparison between male and female data sets of either match or non-match groups using log10 transformed data

	D_{mean}				D_{sd}			
	match							
	SW	**Bart**	**t-test**	**MWU**	**SW**	**Bart**	**t-test**	**MWU**
cla	0.015 0.50†	7.5e–5	0.81	0.16	4.8e–3 0.10†	4.0e–5	0.89	0.34
hum	0.079† 2.5e–4	0.032	4.0e–4*	2.6e–4*	2.6e–3 1.5e–5	0.21†	7.4e–3*	3.2e–3*
rad	0.088† 3.9e–6	0.33†	0.70	0.44	0.08† 1.0e–5	0.33†	0.70	0.52
uln	1.7e–3 0.85†	0.79†	3.0e–4*	1.2e–4*	0.34† 0.67†	0.45†	5.6e–3*	4.9e–3*
fem	0.068† 3.4e–3	0.58†	0.62	0.97	0.30† 2.1e–3	0.45†	0.70	0.24
tib	1.4e–3 1.4e–3	0.15†	0.13	0.03*	5.5e–3 1.4e–4	0.50†	0.57	0.48
fib	0.018 0.059†	0.51†	0.50	0.41	0.02 0.10†	0.49†	0.83	0.62
	non-match							
cla	0.021 0.33†	6.7e–4	0.46	0.83	0.11† 0.027	1.4e–3	0.43	0.69
hum	0.012 0.33†	0.91†	1.0e–3*	1.1e–3*	0.10† 0.21†	0.41†	0.020*	0.017*
rad	0.10† 2.0e–3	0.58†	1.0e–6*	1.6e–6*	0.037 7.2e–4	0.99†	1.5e–4*	4.0e–4*
uln	0.42† 2.4e–4	0.64†	9.3e–7*	1.6e–6*	0.044 5.8e–5	0.69†	5.5e–4*	3.6e–4*
fem	0.026 0.13†	0.30†	0.81	0.87	0.031 0.12	0.47†	0.71	0.66
tib	0.26† 0.70†	0.93†	0.064	0.035*	0.38† 0.65†	0.64†	0.020*	0.013*
fib	0.079† 0.24†	0.78†	0.89	0.95	0.49† 0.32†	0.45†	0.87	0.997

Note. Shapiro-Wilks (SW) results list match then non-match. For SW and Bartlett (Bart) tests, p values greater than 0.05 indicate satisfaction of the normality and homogeneity of variance assumptions respectively and are annotated with †. For the t-tests and Mann-Whitey U tests (MWU), p-values less than 0.05 indicate a significant difference between the compared groups and are annotated with *.

APPENDIX 2: R Code

```
### Descriptive Statistics


### import data and change data file names


#### PASTE VECTORS


#### Sample statistics and z scores for Dmean to include on table


Dmean_match <- c()

Dmean_nonmatch <- c()

Dsd_match <- c()

Dsd_nonmatch <- c()



mean(Dmean_match)


zscore <- (Dmean_match - mean(Dmean_match)) / sd(Dmean_match)

zscore


mean(Dsd_match)
```

```
zscore <- (Dsd_match - mean(Dsd_match)) / sd(Dsd_match)
zscore

mean(Dmean_nonmatch)
zscore <- (Dmean_nonmatch - mean(Dmean_nonmatch)) / sd(Dmean_nonmatch)
zscore

mean(Dsd_nonmatch)

zscore <- (Dsd_nonmatch - mean(Dsd_nonmatch)) / sd(Dsd_nonmatch)
zscore




#### Other Sample Statistics for reference

sd(Dmean)
median(Dmean)
min(Dmean)
max(Dmean)

mean(Dsd)
```

```
sd(Dsd)

median(Dsd)

min(Dsd)

max(Dsd)

### turn on the scientific notation of values

options(scipen=0)


#### Import data and change data name for bartlett tests

Dmean_match_f <-

        c(0.546275,0.612877,0.673546,0.646071,0.683779,0.65922,0.637481,0.64906,0.556
        608,0.552828,1.1013,0.672942,0.612774,1.31289,0.643273,0.658032,0.520874,0.658
        79,0.930515,0.510116,0.688061,0.629134,0.522593,0.796553,0.882964,0.647266,1.1
        9878,0.531099,0.859754)

Dmean_nonmatch_f <-

        c(1.13744,1.74585,1.37437,1.2303,1.49932,1.47569,0.955332,1.60789,1.09433,1.45
        033,1.2454,1.52862,0.886379,1.70984,2.19604,0.998772,1.08283,1.36758,1.27166,1.
        82436,1.6643,1.27169,1.31345,1.32537,1.18271,1.71422,1.94101,1.80609,3.10021,1.
        58145,1.27406,1.02842,1.17473,1.28603,1.91092,2.24228,1.3645,1.45047,1.17552,1.
        50116,2.449,1.34958,1.17125,1.92206,1.65134,1.03536,2.99355,1.99276,1.61571,1.1
        9152,1.76544,1.24722,2.25418,1.8446)

Dsd_match_f <-

        c(0.486163,0.529261,0.63466,0.634369,0.65963,0.701146,0.527944,0.675509,0.473
        254,0.69399,1.14927,0.676215,0.552704,1.2652,0.636229,0.58255,0.450906,0.61785
```

,0.759275,0.484056,0.681691,0.506043,0.486127,0.689731,0.888487,0.632823,1.346
42,0.490662,0.919024)

Dsd_nonmatch_f <-

c(1.00563,1.49196,1.37469,1.13422,1.29243,1.4486,0.749418,1.52697,0.915121,1.3
5003,1.09149,1.6384,0.741931,1.43438,2.14238,0.808443,0.907085,1.29826,1.11245
,1.81912,1.32482,0.892751,1.36121,1.19987,0.909272,1.66922,1.57446,1.54736,3.33
312,1.64201,1.28489,0.833444,1.05208,1.14892,1.74705,2.19725,1.06976,1.34698,1.
17423,1.65734,2.40627,1.05875,1.14794,1.52925,1.52557,0.825974,2.84604,1.86501
,1.32573,0.982136,1.66218,1.07953,2.52663,1.70197)

Dmean_match_m <-

c(0.681415,0.61047,0.693564,0.924804,1.26216,0.836007,0.638858,0.813115,0.666
401,1.03249,0.724109,0.818505,0.57003,0.608321,0.595935,0.759476,0.590799,1.13
878,0.738614,0.662897,0.84802,0.641689,0.920519,0.728902,0.900896,0.641276,0.7
3822,0.801745,0.604807,0.66913,0.717077,0.680833,0.731051,0.747368,0.634808,0.
796724,0.583867,0.690884,0.70534,0.867624,0.604117,0.712795,0.671584,0.599316
,0.675003,0.937255,1.29874,0.714649,0.887937)

Dmean_nonmatch_m <-

c(1.62822,2.57967,1.55795,1.90935,1.79207,1.33526,1.66595,1.57018,2.0154,1.975
41,1.61228,1.979,1.67459,1.21086,1.54641,1.69301,1.95426,1.61689,1.70966,1.2439
8,2.31133,1.33988,1.52924,1.46361,2.32163,1.81992,1.80167,1.17923,2.10163,1.483
53,1.45482,2.13372,2.24269,0.995093,1.54279,2.08106,1.40767,1.64125,1.37884,1.5

```
        2382,1.79775,1.90056,1.96865,2.39471,1.92659,0.952594,1.41704,1.13108,1.0461,3.
        75768,0.837076,1.20231,1.65754,1.37528,1.25863,1.27959,1.71038,3.22259)
Dsd_match_m <-
        c(0.582268,0.505074,0.693295,0.752464,1.22304,0.814594,0.538505,0.709866,0.56
        3755,1.20868,0.630379,0.732887,0.538255,0.638008,0.516707,0.668897,0.478914,1.
        26898,0.580819,0.54221,0.743032,0.633008,1.03017,0.779752,0.788822,0.576085,0.
        612712,0.774681,0.518483,0.570236,0.717374,0.569245,0.678618,0.665854,0.55061
        2,0.71118,0.596971,0.54289,0.608751,1.08415,0.513433,0.587086,0.620301,0.54157
        6,0.56594,0.855844,1.20608,0.602445,1.00287)
Dsd_nonmatch_m <-
        c(1.71013,2.65067,1.78528,1.98498,1.77731,1.34435,1.53589,1.42268,1.55741,2.11
        267,1.47828,1.87481,1.53291,1.1346,1.54849,1.5232,2.12968,1.55442,1.56689,0.992
        444,2.32504,1.44506,1.66824,1.2184,2.31531,1.69906,1.86042,1.04308,2.01323,1.71
        655,1.57165,2.26072,1.78113,0.974546,1.46271,1.96839,1.18249,1.37149,1.2102,1.6
        2067,1.63066,1.90916,2.0507,2.3552,1.59759,0.991762,0.981574,1.11358,0.975682,
        3.59559,0.773011,0.870415,1.48009,1.34874,0.99218,1.22931,1.42746,3.18108)


### Paste in Vectors, include:
##### Dmean_match_f <- c(
##### Dmean_match_m <- c(
##### Dmean_nonmatch_f <- c(
##### Dmean_nonmatch_m <- c(
##### Dsd_match_f <- c(
```

```
##### Dsd_match_m <- c(

##### Dsd_nonmatch_f <- c(

##### Dsd_nonmatch_m <- c(



########## DMEAN T TEST FEMALES MATCH VS NONMATCH############

# H0: Dmean of match vs non match are the same

# H1: Dmean of match vs non match are different.

# Test normality Dmean

par(mfrow=c(1,2))

plot(density(Dmean_match_f))

plot(density(Dmean_nonmatch_f))

shapiro.test(Dmean_match_f)

shapiro.test(Dmean_nonmatch_f)

# test for the homogeneity of the variances

### ENSURE YOU HAVE LOADED CORRECT DATA SET

bartlett.test(Dmean~match, tib_f)
```

```
# Dmean t test, change var.equal according to Bartlett test result

t.test(Dmean_match_f,Dmean_nonmatch_f,var.equal=FALSE,paired=FALSE)

### Mann Whitney U Test ###

wilcox.test(Dmean_match_f,Dmean_nonmatch_f,paired=FALSE)



#################################################################



########## Dsd T TEST FEMALES MATCH VS NONMATCH############

# H0: Dsd of match vs non match are the same
# H1: Dsd of match vs non match are different.

#Test normality Dsd
par(mfrow=c(1,2))
plot(density(Dsd_match_f))
```

```
plot(density(Dsd_nonmatch_f))

shapiro.test(Dsd_match_f)

shapiro.test(Dsd_nonmatch_f)



# test for the homogeneity of the variances

### ENSURE YOU HAVE LOADED CORRECT DATA SET


bartlett.test(Dsd~match, tib_f)


# Dsd t test, change var.equal according to Bartlett test result

t.test(Dsd_match_f,Dsd_nonmatch_f,var.equal=FALSE,paired=FALSE)



### Mann Whitney U Test ###


wilcox.test(Dsd_match_f,Dsd_nonmatch_f,paired=FALSE)



############################################################
```

```
########## DMEAN T TEST MALES MATCH VS NONMATCH############

# H0: Dmean of match vs non match are the same

# H1: Dmean of match vs non match are different.


# Test normality Dmean

par(mfrow=c(1,2))

plot(density(Dmean_match_m))

plot(density(Dmean_nonmatch_m))

shapiro.test(Dmean_match_m)

shapiro.test(Dmean_nonmatch_m)


# test for the homogeneity of the variances

### ENSURE YOU HAVE LOADED CORRECT DATA SET


bartlett.test(Dmean~match, tib_m)


# Dmean t test, change var.equal according to Bartlett test result
```

```
t.test(Dmean_match_m,Dmean_nonmatch_m,var.equal=FALSE,paired=FALSE)

### Mann Whitney U Test ###

wilcox.test(Dmean_match_m,Dmean_nonmatch_m,paired=FALSE)



########## Dsd T TEST MALES MATCH VS NONMATCH############

# H0: Dsd of match vs non match are the same

# H1: Dsd of match vs non match are different.

#Test normality Dsd

par(mfrow=c(1,2))

plot(density(Dsd_match_m))

plot(density(Dsd_nonmatch_m))

shapiro.test(Dsd_match_m)

shapiro.test(Dsd_nonmatch_m)



# test for the homogeneity of the variances
```

```
### ENSURE YOU HAVE LOADED CORRECT DATA SET

bartlett.test(Dsd~match, tib_m)

# Dsd t test, change var.equal according to Bartlett test result

t.test(Dsd_match_m,Dsd_nonmatch_m,var.equal=FALSE,paired=FALSE)



### Mann Whitney U Test ###

wilcox.test(Dsd_match_m,Dsd_nonmatch_m,paired=FALSE)
```

######### T TESTS BETWEEN SEXES ##############

#########DMEAN MATCH F with MATCH M #######

```
# H0: Dmean of male match vs female match are the same
# H1: Dmean of male match vs female match are different

#Test normality Dmean
shapiro.test(Dmean_match_f)
shapiro.test(Dmean_match_m)


# test for the homogeneity of the variances
### ENSURE YOU HAVE LOADED CORRECT DATA SET

bartlett.test(Dmean~sex, tib_both_match)


# Dmean t test, change var.equal according to Bartlett test result
```

```
t.test(Dmean_match_f,Dmean_match_m,var.equal=TRUE,paired=FALSE)



### Mann Whitney U Test ###

wilcox.test(Dmean_match_f,Dmean_match_m,paired=FALSE)



####### Dsd MATCH F VS Dsd MATCH M ##########

# H0: Dsd of male match vs female match are the same

# H1: Dsd of male match vs female match are different

#Test normality

shapiro.test(Dsd_match_f)

shapiro.test(Dsd_match_m)



# test for the homogeneity of the variances

### ENSURE YOU HAVE LOADED CORRECT DATA SET

bartlett.test(Dsd~sex, tib_both_match)
```

```
# Dsd t test, change var.equal according to Bartlett test result
t.test(Dsd_match_f,Dsd_match_m,var.equal=TRUE,paired=FALSE)

### Mann Whitney U Test ###

wilcox.test(Dsd_match_f,Dsd_match_m,paired=FALSE)



########## Dmean OF F NONMATCH VS MALE NONMATCH #####

# H0: Dmean of male nonmatch vs female nonmatch are the same
# H1: Dmean of male nonmatch vs female nonmatch are different

#Test normality Dmean
shapiro.test(Dmean_nonmatch_f)
shapiro.test(Dmean_nonmatch_m)


# test for the homogeneity of the variances
### ENSURE YOU HAVE LOADED CORRECT DATA SET
```

```
bartlett.test(Dmean~sex, tib_both_nonmatch)

# Dmean t test, change var.equal according to Bartlett test result
t.test(Dmean_nonmatch_f,Dmean_nonmatch_m,var.equal=TRUE,paired=FALSE)


### Mann Whitney U Test ###

wilcox.test(Dmean_nonmatch_f,Dmean_nonmatch_m,paired=FALSE)


########## Dsd OF F NONMATCH VS MALE NONMATCH #####


# H0: Dsd of male nonmatch vs female nonmatch are the same
# H1: Dsd of male nonmatch vs female nonmatch are different

#Test normality Dsd
shapiro.test(Dsd_nonmatch_f)
shapiro.test(Dsd_nonmatch_m)


# test for the homogeneity of the variances
```

```
### ENSURE YOU HAVE LOADED CORRECT DATA SET

bartlett.test(Dsd~sex, tib_both_nonmatch)

# Dsd t test, change var.equal according to Bartlett test result

t.test(Dsd_nonmatch_f,Dsd_nonmatch_m,var.equal=TRUE,paired=FALSE)

# Mann Whitney U Test

wilcox.test(Dsd_nonmatch_f,Dsd_nonmatch_m,paired=FALSE)



###First load data set and replace all data set names###

##should be 20 occurrences replaced###

### turn off the scientific notation of values

options(scipen=999)

##### Check Pearson's Correllation between Dmean and Dsd ####

Dmean.vector <- as.vector(tib_f$Dmean)
```

```
Dsd.vector <- as.vector(tib_f$Dsd)

cor(Dmean.vector,Dsd.vector)




######### JUST DMEAN AS PREDICTOR #################

### fit the logistic regresion model with Dmean as predictor
fit.Dmean=glm(match~Dmean, data=tib_f,family=binomial)
summary(fit.Dmean)

##ROC Curve for Dmean model
library(pROC)
predict.prob=predict(fit.Dmean,tib_f,type="response")
Dmean_roc = roc(tib_f$match ~ predict.prob, plot = TRUE, print.auc = TRUE)
Dmean_cutoff=as.numeric(Dmean_roc$auc)
as.numeric(Dmean_roc$auc)

### obtain the classification table using cutoff value from as.numeric test
```

```
predicted=ifelse(predict.prob>=Dmean_cutoff,"Yes","No")
table(predicted,tib_f$match)

### obtain the classification table using 0.5 as cutoff
predicted=ifelse(predict.prob>=0.5,"Yes","No")
table(predicted,tib_f$match)



######## JUST DSD AS PREDICTOR ###################

### fit the logistic regresion model with Dsd as the predictor
fit.Dsd=glm(match~Dsd, data=tib_f,family=binomial)
summary(fit.Dsd)

##ROC Curve for Dsd model
library(pROC)
predict.prob=predict(fit.Dsd,tib_f,type="response")
Dsd_roc = roc(tib_f$match ~ predict.prob, plot = TRUE, print.auc = TRUE)
Dsd_cutoff=as.numeric(Dsd_roc$auc)
as.numeric(Dsd_roc$auc)

### obtain the classification table using cutoff value from as.numeric test
predicted=ifelse(predict.prob>=Dsd_cutoff,"Yes","No")
```

```
table(predicted,tib_f$match)

### obtain the classification table using cutoff value of 0.5
predicted=ifelse(predict.prob>=0.5,"Yes","No")
table(predicted,tib_f$match)



######## MODEL WITH BOTH PREDICTORS ###################

### logistic regresion with both Dmean and Dsd as predictors
fit.2=glm(match~Dmean+Dsd, data=tib_f,family=binomial)
summary(fit.2)

##ROC Curve for 2 predictor model
library(pROC)
predict.prob=predict(fit.2,tib_f,type="response")
fit.2_roc = roc(tib_f$match ~ predict.prob, plot = TRUE, print.auc = TRUE)
fit.2_cutoff=as.numeric(fit.2_roc$auc)
as.numeric(fit.2_roc$auc)

### obtain the classification table using cutoff value from as.numeric test
```

```
predicted=ifelse(predict.prob>=fit.2_cutoff,"Yes","No")
table(predicted,tib_f$match)

### obtain the classification table using cutoff value of 0.5
predicted=ifelse(predict.prob>=0.5,"Yes","No")
table(predicted,tib_f$match)



######## cross-validation ####################

library(caret)
set.seed(123)

# cross validation for model with Dmean as predictor
train.control=trainControl(method="cv", number=5) ###
model.Dmean=train(as.factor(match)~Dmean, data=tib_f, method="glm", trControl =
        train.control)

# Summarize the results for Dmean predictor model
print(model.Dmean)
```

```
# cross validation for model with Dsd as predictor
train.control=trainControl(method="cv", number=5) ### for leave one out cross validation,
        put number to sample size
model.Dsd=train(as.factor(match)~Dsd, data=tib_f, method="glm", trControl = train.control)

# Summarize the results for Dsd predictor model
print(model.Dsd)

# Train the model with 2 predictor
model.2=train(as.factor(match)~Dmean+Dsd, data=tib_f, method="glm", trControl =
        train.control)
# Summarize the results
print(model.2)
```