ABSTRACT

SMART, JORDAN NICOLE. Logit Summation as an Alternative Sex Estimation Method. (Under the direction of Dr. Troy Case).

The most common way in which anthropologists estimate sex is by utilizing the skull and/or os coxa. However, there are instances in which these areas are unavailable or are not the best fit for sex estimation due to a variety of reasons including poor preservation, missing bones, pathology, etc. When these instances arise, population-specific alternative sex estimation methods must be developed and tested. Various areas of the skeleton have been tested using a range of methods, including the upper limb bones. Metric traits are most commonly taken from these bones and have shown to be reliable in sex estimation. In particular, one of the statistical methods used when handling metric traits is binary logistic regression. There are two types of binary logistic regression analysis: simple and multiple. The difference between the two being that there is one input for simple binary and multiple inputs for multiple binary. Seeing as simple binary only utilizes one variable, it is not ideal for sex estimation. However, if a group of simple binary outputs are summed together to produce a single sex estimation, that estimation may be reliable. This thesis put this idea to the test using a dataset of 140 Thai individuals and nine upper limb measurements. Both the logit summation method and multiple binary logistic regression method was tested on 24 different variable groups and their accuracy rates were compared. In only 3 of the 24 groups, the logit summation method produced a higher accuracy than did the multiple binary logistic method. However, only 9 of the 48 accuracies fell below 90%, showing that both methods and most groups were reliable for estimating the sex of Thai individuals.

Logit Summation as an Alternative Sex Estimation Method

by
Jordan Nicole Smart

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Arts

Anthropology

Raleigh, North Carolina
2021

APPROVED BY:

Dr. Troy Case
Committee Chair

Dr. Julie Wesp

Dr. John Millhauser

DEDICATION

I dedicate this thesis to my parents who have been so supportive throughout my life in every endeavor I have pursued.

BIOGRAPHY

Jordan Smart was born and raised in Tampa, FL. Throughout high school, her favorite subject was biology and she planned on pursuing a degree in biological sciences. During her time as an undergraduate at the University of Florida, she was drawn to the field of anthropology and decided to pursue it as a minor. However, biological anthropology eventually won her over and she graduated in 2019 with a B.A. in biology and a B.S. in anthropology. In the fall of 2019, Jordan began at North Carolina State University as a master's student pursuing an M.A. in anthropology. As a master's student, her research focused on the viability of an alternative method of sex estimation. Once she graduates, Jordan hopes to take some time off of school to pursue a career outside of academia.

ACKNOWLEDGMENTS

I would like to acknowledge and thank my wonderful adviser Dr. Troy Case for guiding me through this process. I have learned a lot from him and am very grateful to have had someone like him as my advisor.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Sex estimation is a vital tool in the fields of forensic anthropology and bioarchaeology. For the past few decades, research aimed at fine-tuning sex estimation methods has created a strong foundation for developing sex estimation equations from many parts of the skeleton (Case et al. 2015; Colman et al. 2018; Attia and Aboulnor 2020). Among biological anthropologists as a whole, the most widely used approach to sex estimation is non-metric observation of traits from the skull and os coxa. Visual scoring methods have been created for a variety of the non-metric characteristics that have proven to be extremely reliable in sex estimation (Krishan et al. 2016; Gonzalez et al. 2007).

While non-metric sexing of the skull and os coxa is widely accepted and utilized by biological anthropologists, this method is not always a viable option. First, forensic anthropologists must contend with the requirements of evidence presented in criminal court cases, where subjective assessments are not permitted (Nirenberg 2016). Most of the non-metric visual scoring methodologies used to assess the skull and os coxa require experience to attain the highest level of accuracy, and the differences between males and female are often not sufficiently clearly defined to be considered objective methodologies of the sort permitted in court cases (Page and Blenkin 2011). Thus, although forensic anthropologists may begin with visual, non-metric observations of the pelvis or skull, most rely on metric sex estimation techniques for the presentation of evidence in court. Second, in both forensic anthropology and bioarchaeology, there may be cases in which the skull and/or os coxa are missing, or too damaged to be adequately assessed for visual non-metric characteristics indicative of sex. Recovery may be poor, or there are pathologies present, or part of a bone may be missing or damaged while the remainder is intact. In such cases, experts must turn to alternative methods

for sex estimation, using bones from other parts of the skeleton. Non-metric and metric sex estimation methods have been developed and tested for most of the bones of the body in at least some populations, including hand bones (Burrows et al. 2003; Case et al. 2015), upper limb long bones (Duangto and Mahakkanukrauh 2020; Attia and Aboulnor 2020), foot bones (Harris and Case 2012; Case and Ross 2007) and lower limb long bones (Colman et al. 2018; Fliss et al. 2019) among other areas.

Many highly accurate sex estimation equations have been produced by this kind of research, allowing anthropologists to turn to reliable alternative bones for sex estimation when needed. This thesis will focus on the bones of the hand and arm, and in particular, on metric sex estimation methods from this limb. The thesis has two primary goals. The first is to develop sex estimation equations for the upper limbs in order to identify bones and measurements that are likely to produce the most accurate estimates from this region of the skeleton. The second is to test whether a different approach (logit summation) to combining variables for sex estimation will result in similar or better estimates of sex.

The statistical procedure to be utilized in this thesis will be binary logistic regression. It is a specific type of regression that only permits two outcomes – in this case, female or male – hence the term "binary." This approach will be applied to a documented collection of Thai skeletons from the Chiang Mai area in Northwest Thailand, both to produce equations for use in metric sex estimation of modern Thai skeletons, and to compare the two ways of combining variables. Logistic regression is a statistical procedure utilized to predict the relationship between independent variables and a dependent variable (Leech 2011). Along with linear discriminant functions, it is a tool commonly used by biological anthropologists for the development of sex

estimation equations based on measurements, in which the measurements are treated as the independent (predictor) variables, and sex as the dependent variable.

It is possible to predict sex from just a single measurement using logistic regression. This would be an example of simple binary logistic regression. Combining two or three variables together often improves the accuracy of a sex assessment (Case and Ross 2007; Bartholdy et al. 2020). Logistic regression involving two or more variables is called multiple logistic regression. When multiple regression is used, different measurements are combined statistically to create a best fit line that will help to distinguish the two forms of the dependent variable – in this case, females and males. In most cases, the contribution of each variable to the final estimate is different, and this difference is reflected in terms of the size of the coefficient placed in front of the measurements in the final regression equation, relative to the average size of that measurement in the sample. Therefore, if the epicondylar breadth of the humerus is three times larger than the midshaft diameter on average, but the coefficient for midshaft diameter is five times larger, it means that midshaft diameter has an outsized influence on the final sex estimate. In general, if a single variable produces a more accurate sex estimate than other variables when each is tested by means of simple logistic regression, that variable will likely also have an outsized impact in a multiple regression equation if the two variables are combined.

One problem with multiple regression equations for sex estimation from skeletons is the issue of missing data. A researcher may develop a binary logistic regression equation for sexing from the hand, for example, but any multiple regression equations produced are only useful to another researcher if the exact same set of measurements are available in the skeleton of interest. Furthermore, the original researcher who developed the sex estimation equations must have thought to combine the variables that a particular end user wishes to use, or else some of the end

user's data must be left out of multiple regression analysis. As an example, if an original publisher of sex estimation equations for the arm bones analyzes a total of 10 variables but only reports the three most accurate multi-variable equations, with two variables in each, then the end user is limited to only those combinations if they wish to use a (usually) more accurate multiple regression approach. If one of those measurements is not obtainable due to taphonomy or a missing bone, then the equation is useless for that particular case. The end user must be satisfied with what may be a series of single-variable equations, each with somewhat less accuracy. Furthermore, if those single-variable equations do not all agree about the sex estimate, the end user must then decide how to assess the multiple pieces of data and form a conclusion about the sex. In a court case, the forensic anthropologist may have to report, for example, three equations suggestive of female sex and one suggestive of male sex, whereas if the four variables had been combined in a multiple regression, there would be only one outcome without obvious disagreement (although the overall accuracy could, of course, potentially be compromised by the disagreeing variable).

One way to skirt this problem between the original study and the end user's issues with measurement availability is to combine the outputs from single variable equations in a way that provides a summary result. For example, a group of five different measurements will each have their own output from simple binary logistic regression analysis, called a logit, and nearly all logistic regression sex estimation studies report the simple logistic regression results for each variable first before combining variables for multiple logistic regression. The difference between that and multiple binary logistic regression analysis would be that for the multiple analysis, variables are combined in the input, while this new method would combine the outputs to create a single estimation. This can be done simply by adding the logits from each individual equation

together. In binary logistic regression, if the logit is below zero, the indicated sex is female. If it is above zero, the indicated sex is male. The probability of maleness increases with a positive logit, and the probability of female increases with a negative logit. Two negative logits will still be negative, and two positive logits will still be positive. When there is disagreement among individual measurements (one is positive while another is negative), the variable that is most strongly suggestive of a particular sex will determine which sex is suggested. However, when three or more variables are combined, it is possible for two variables suggestive of one sex to counterbalance one variable that is more strongly suggestive of the other sex. Whether or not estimations in those instances become indeterminate or not depends on the magnitudes of the logits involved. The question to be addressed here is whether combining logits in this way will lead to similar or more accurate results than simply using multiple logistic regression equations alone. The goal is to maximize the information gleaned from the data at hand, and to test whether this approach might improve the overall sex estimation accuracy.

The purpose of this thesis, then, is to determine whether utilizing this logit summation method of combining simple binary logistic regression analysis outputs would produce sex estimations that are either equivalent, or significantly more reliable, compared to those from multiple logistic regression approaches. A secondary goal is to assess whether there is a significant issue with sex estimation disagreements among bones from the same general region of the skeleton – the arm and hand, in this case. If there are disagreements, it is possibly due to a difference in activity impacts. The study will make use of several upper limb bones (including hand bones) and their measurements using a dataset of modern Thai individuals from a documented collection.

CHAPTER 2: LITERATURE REVIEW

A lot of important research has already been done on the use of alternative sex estimation methods (Scheuer and Elkington 1993; Falsetti 1995; Case and Ross 2007; Waidhofer and Kirchengast 2015; etc.). Exploration of the usefulness of the appendicular skeleton for sex estimation has been of particular prominence in this research area (Scheuer and Elkington 1993; Smith 1996; Anzellini and Toyne 2019; etc.) The long bones of the upper and lower limbs as well as both hand and foot bones have shown to be at least somewhat reliable for estimation (Case and Ross 2007; Mahakkanukrauh et al. 2016; Case et al. 2015; Colman et al. 2018; etc.). Unlike the skull and os coxa, non-metric features on these bones are minimally dimorphic; therefore, their metric variation has shown to be more informative. Research on these bones using metric data has shown promising results that have allowed scholars in the forensic and archaeology fields to use these areas of the skeleton when the skull and os coxa are not present.

Hand Bones

Research on the viability of the use of hand bones for sex estimation dates back to the 1990s with most of the focus on the use of the metacarpals, but phalanges have been studied as well (Smith 1996). Such studies of hand bones have produced varying results. For example, Scheuer and Elkington (1993) carried out research on sex estimation using metacarpals and the first proximal phalanx. They took six measurements on each of the five metacarpals and the first proximal phalanx from a sample of 60 individuals. With these measurements, they created multiple regression equations – one equation for each bone using all six measurements. The equations they created were subsequently applied to a sample of 20 individuals to test their accuracy. The first metacarpal produced the highest accuracy with a range from 74 to 94 percent.

This study established the validity of using metacarpals and multiple regression equations for sex estimation.

Two years later, Falsetti (1995) conducted research on metacarpals using discriminant functions. In this study, he was only able to test metacarpals 2, 4, and 5, due to metacarpals 1 and 3 showing too much difference between ancestries. Using a sample of 212 individuals from the Terry collection, he created discriminant functions from five metacarpal measurements to subsequently test on two different samples of 33 and 40 individuals. Metacarpals 2, 4, and 5, produced correct sex estimation of 92, 86, and 84 percent respectively. Similar to the Scheuer and Elkington (1993) study, this research validated the use of metric data from metacarpals in sex estimation.

In 1996, Smith conducted a study on the accuracies of metacarpals and phalanges when used for sex estimation and population estimation. Population estimation is similar to sex estimation in that it estimates a particular piece of information about an individual based on their skeleton. Skeletal traits are used to estimate which population an individual originated from, possibly answering where that person lived as well. Smith (1996) took length and width measurements from all 19 bones of the hand and grouped these measurements by side (left/right) as well as by bone type (metacarpals and proximal/intermediate/distal phalanges), resulting in eight different groups, four from each side. She performed a stepwise discriminant function procedure to create models for sex estimation from each of the groups. The left metacarpal group had the greatest power of discrimination (estimation accuracy) at 89 percent, while the right intermediate phalanx group had the lowest power of discrimination at 72 percent. Despite this low power of discrimination for the intermediate phalanx group, the distal phalanx group had the second highest power of discrimination that ranged from 81 to 83 percent. This study reinforced

the notion that metacarpals are useful in the task of estimating sex and suggested that some measurements of the phalanges may be viable options as well.

Sex estimation with metacarpal measurements was put to the test again by Stojanowski (1999) in a forensic-focused study. Stojanowski used six measurements for each metacarpal from a sample of 200 individuals to create 35 different linear discriminant functions based on different taphonomic/pathological preservation scenarios. Of the five metacarpals, measurements from the fourth metacarpal produced the most accurate and consistent functions, with an overall accuracy for the sample ranging from 75 to 90 percent for all functions tested. This result is different from that of Scheuer and Elkington (1993) in which the second metacarpal was most accurate. It is important to note that Scheuer and Elkington (1993) used regression analysis while both Falsetti (1995) and Stojanowski (1999) used discriminant functions, which might explain the difference in results, however, these differences are negligible (Pohar et al. 2004). Stojanowski's (1999) work further demonstrated the usefulness of metacarpal measurements in sex estimation and highlighted the challenges of preservation issues in the development of sex estimation equations that include multiple variables.

Comparative testing of three of the previously mentioned studies (Scheuer and Elkington 1995, Falsetti 1995, and Stojanowski 1999) was carried out by Burrows et al. (2003). She repeated the methods created by each previous study on a sample of 23 adult human cadavers and found that the equations created by Stojanowski (1999) were generally the most accurate with the most accurate equation reaching 95.7 percent correct allocation, thought the lowest rate was 65.2 percent. Scheuer and Elkington (1993) produced the next highest accuracy with a highest rate at 91.1 percent (lowest rate was 63.0 percent). Falsetti's (1995) method produced the lowest accuracy rate (87 percent) but had the smallest accuracy range (lowest rate was 83.3

percent). The only method that produced higher rates than in its original study were some of Stojanowski's (1999) equations. Burrows et al. (2003) concluded from this study that metacarpals can be useful in sex estimation practice, however, they have their limitations if they are used as sole determinants. It is important to note that in repeating all three studies, Metacarpal 2 was consistently the most accurate metacarpal for sex estimation in almost all trials.

While the majority of hand bone studies have focused on the use of metacarpals, hand phalanges have also been considered in some studies. For example, Mahakkanukrauh et al. (2013) tested the use of proximal hand phalanges for sex estimation on a sample of individuals from the Chiang Mai province of Thailand. This study employed six different phalanx measurements and binary logistic regression to test the usefulness of hand phalanges. The authors found that all proximal phalanges exhibited at least an 87 percent accuracy for at least one of their logistic regression equations that included 2-3 measurements. These high accuracy results suggest the possibility that hand phalanges might be more accurate for Thailand populations than are the metacarpals. When this information is combined with the idea that metacarpals might be limited in their accuracy when used as the sole determinants, it seems that using both phalanges and metacarpals simultaneously for sex estimation might produce a more accurate result than using them separately as sole determinants. All of these studies have shown that the use of hand bones (metacarpals and phalanges) does produce adequately accurate results in sex estimation and should therefore be considered and further studied when alternative sex estimations methods are needed.

Upper Limb Long Bones

In addition to hand bones, upper limb long bones (arm bones) have also been researched for their usefulness in sex estimation. Similar to those of hand bones, metric features of arm bones are the useful feature types in sex estimation. Metric dimensions of the arm bones have been studied with regard to the usefulness of the degree of asymmetry in these bones to estimate sex (Waidhofer and Kirchengast 2015). It had been shown in previous studies that right-side bias directional asymmetries existed in the humerus, ulna, and radius. Using 83 individuals from a Khoe-San skeletal sample, Waidhofer and Kirchengast (2015) extracted 40 measurements from those three bones as well as the clavicles to test asymmetry's usefulness in sex estimation. It was found that the measurements that were most directionally asymmetrical for females (humerus length and upper limb length) were not the same measurements that were most directionally asymmetrical for males (breadth and circumference dimensions), showing that there was a greater asymmetry of robustness in males and a greater asymmetry of length in females. It is important to note that this is possibly due to sex-typical labor division in this particular historical population, therefore, those repeating this process for individuals from a different population should be cautious and keep this in mind.

Humeral epiphyses have been used in research aimed at creating population-independent models for sex estimation (Attia and Aboulnor 2020). This study utilized logistic regression to create fitted models that were tested for their viability across seven different populations. The results revealed that the distal bi-epicondylar breadth model showed the most success with an overall accuracy of 90.2 percent and that the vertical humeral head model was the least viable option (especially for European females). This study highlighted the importance of population-specific sex estimation techniques as well as the viability of humeral epiphyses in sex estimation.

While directional asymmetry is not the focus of this thesis, this study provides evidence that arm bone measurements can prove useful in sex estimation practices.

Other research has looked into the reliability of upper limb non-metric traits in sex estimation (Tallman and Blanton 2020). Specifically, Tallman and Blanton (2020) studied the distal humerus and four of its non-metric traits: medial epicondyle angle, olecranon fossa shape, trochlear extension, and cochlear constriction. These traits had previously been shown to be sexually dimorphic among several non-Asian groups, therefore the research was focused on sex estimation of Thai individuals to see if the same was true of that population. Four different methods were employed: population-specific univariate statistics, composite scores, binary logistic regression equations, and binary probit regression. All four of these methods were able to produce accuracies above 90 percent at some point for the Thai population being studied (616 modern individuals). However, it was stated in the conclusion that the four distal humerus traits performed in a "slightly-moderately reliable" way, despite the high success rate of the overall methods. This is because these traits are susceptible to subjectivity when ordinal scores (1-5) are assigned to their expressions. These traits were non-metric and were not tested against metric traits in this particular study.

This thesis uses solely metric traits as they have been shown to be more reliable in sex estimation using the upper limb and this conclusion about the four non-metric distal humerus traits further reinforces that idea. While it did not use metric traits as this thesis does, the Tallman and Blanton (2020) study shows utilizing the humerus (an upper limb bone) has viability in sex estimation. It also brings up multiple relevant points. For example, of the four traits tested, none were reliable when performing alone. An important aspect of this thesis is that when metric variables perform alone in sex estimation, accuracy is lower than when they are

combined. The ways in which they can be combined, and the accuracy of each combination method is what is tested; however, the point is clear that variables must be combined in some way so there is increased reliability in the sex estimation.

Another important aspect of Tallman and Blanton's (2020) study was the emphasis on population-specific methods. The various methods that were created using the four non-metric traits produced a much higher accuracy on the Thai individuals than non-Asian methods did. This conclusion gives further legitimacy that sex estimation methods created from a certain population do not necessarily work on other populations. This creates the need for population-specific methods to be created and is relevant to this thesis as the equations produced are specific to Thai individuals.

Binary Logistic Regression vs. Linear Discriminant Analysis

There are two primary methods used by biological anthropologists to develop metric sex estimation equations, linear discriminant analysis and binary logistic regression analysis. Both have been shown to be useful for sex estimation from the skeleton (Falsetti 1995; Harris and Case 2012; Mahakkanukrauh et al. 2013; Bartholdy et al. 2020). Previous research on sex estimation provides insight into which types of analysis are best based on the constraints of the research project at hand. It is important to look back on this research and their methods when conducting new studies, in order to make the best decision about which method to use for a project that is utilizing metric data for sex estimation.

In his test of metacarpal measurements for sex estimation described above, Falsetti (1995) mentioned something important about the use of discriminant functions. When utilizing linear discriminant analysis, each variable used must satisfy the assumption of normality for the dataset and the assumption of equal variance-covariance matrices (see Baek et al. 2018). Thus, if

any of the variables in a study are non-normal, discriminant function analysis should not be utilized.

In contrast to discriminant functions, binary logistic regression analysis is not restricted to a normal distribution of the data, nor does it assume equality of the variance-covariance matrices (Bartholdy et al. 2020). An example of the use of binary logistic regression is a study performed on a series of tarsal measurements for sex estimation (Harris and Case 2012). Unlike the Falsetti (1995) study, this study chose binary logistic regression analysis over linear discriminant analysis to avoid the more restrictive data requirements of discriminant function analysis. While the variables in the dataset they were using were normally distributed, a total of 10 of the 36 measurements did not meet the equal variance-covariance matrices assumption. In addition to not having these restrictions, binary logistic regression provides various other benefits. It is less sensitive to high correlations among the variables, which is particularly important when measuring bone lengths, or bones of the same type from the hand, for example (Pohar et al. 2004). It is also less sensitive to outliers, which can be particularly important when sample sizes are small, or when measurements are only taken once and the risk of measurement error is high. When compared to using sample sizes over 50, the difference between the two analysis types is negligible (Pohar et al. 2004). In cases where the dataset requires a less restrictive statistical analysis, binary logistic regression is preferred over linear discriminant analysis.

Another advantage of binary logistic regression in comparison to linear discriminant analysis is the ability to calculate classification probabilities straight from the binary logistic equation (Bartholdy et al. 2020). In linear discriminant analysis, a cutoff point is calculated to estimate sex based on which side of the cutoff point the individual's results fall. The outcome is

binary and the equation does not provide much information on the reliability of that result, as a particular measurement could be very near the cutoff point, or extremely far from it, and information on how to interpret distance from the cutoff point is not necessarily provided in discriminant analysis studies. In order to obtain such information, the developer of the equation would either need to calculate posterior probabilities, which is almost never done, or they would need to provide access to the raw data from the reference sample, which is also rare (Bartholdy et al. 2020).

Conversely, logistic regression equations produce a value that can be used directly to calculate a probability of one sex versus the other. In general, results that are further from the cutoff point have a higher probability of being accurate. The equation produces a value, called a logit, that can be plugged into a standard formula to provide a probability of one sex versus the other, so that the reliability for a particular skeleton can be calculated from the equation no matter what other information is provided (Bartholdy et al. 2020).

Other studies have provided evidence that binary logistic regression is a viable option for sex estimation. A study focusing on the use of phalanx measurements utilized binary logistic regression and ROC analysis (Mahakkanukrauh et al. 2013) and found that every phalanx on both sides were able to produce at least one binary logistic equation that had an 87 percent or higher sex estimation accuracy rate. Another study set out to create viable equations that could be applied to an Andean population (Anzellini and Toyne 2019). They successfully created binary logistic regression formulae from osteometric data taken from individuals of known sex recovered at the archaeological site of Kuelap in Chachapoyas, Peru. These equations produced correct classification rates between 82 and 93 percent. This, as the authors pointed out, is around the same classification range that cranial non-metric variables produce. This study is also

important as it focuses on the population-specific aspect of sex estimation methods. The equations created were specifically for individuals from the Andean region. The accuracy of a sex estimation method using metric variables varies greatly between different populations and when using an equation created in the past, it is important to note the origin of the individuals who served as the reference sample for its creation. Binary logistic regression equations are population-specific and should only be used on individuals from the population they were created in.

A study on the creation of metric sex estimation equations for a Dutch population also focused on this population-specific aspect (Colman et al. 2018). The authors of the study created logistic regression formulae specific to the Dutch population using four measurements of the proximal femur. These measurements were taken from Dutch individuals and the resulting equations provided accuracies between 86 and 92 percent. It is important to note that the equations produced in the current thesis works solely with Thai individuals and the resulting equations therefore should only be applied to individuals of the same origin (Thailand).

Metric Sex Estimation

When using traditional areas of the skeleton such as the os coxa and the skull for sex estimation, non-metric traits are most often used. Specific methods have been perfected using these non-metric traits for the skull and os coxa that have proven to be very reliable in sex estimation. However, as previously stated, these bones are not always available when sex estimates are needed. In these cases, anthropologists must turn to other areas of the body such as the upper limb. Unlike the os coxa and the skull, upper limb bones (long bones, metacarpals, phalanges) are used for their metric traits as there are many measurements that can be taken from each bone, allowing the estimator to have as much information as possible. In almost every study

mentioned in this literature review, metric traits from upper limb bones are used. This section will highlight the types of measurements used and how they are imperative for alternative sex estimation methods.

While the use of the upper limb in sex estimation has been established as helpful, there are still some complications with using multiple metric variables from one individual. When multiple variables are present, there will always be the possibility that not every individual measurement will provide the same sex estimate. The difference between length and robusticity measurements heighten that possibility even more. This difference exists because of the influence of lifetime activity. Measurements of robusticity are more prone to be influenced by lifetime activity, such as heavy labor, than measurements of length (Rolian et al. 2011). While this type of influence on bones could follow along gender lines (women being less likely to participate in jobs involving heavy labor), it is not necessarily reliable for sex estimation, especially in this time period where gender norms are becoming more and more a thing of the past. Length, on the other hand, is much less influenced by lifetime activity than is robusticity. Much like sex, it is mostly influenced by genetics and is therefore a useful estimator (Misziewicz and Mahoney 2019) as bone length does not change much throughout one's adult life due to activity.

This difference in reliability between length and robusticity measurements has been discussed in the past (Case and Ross 2007). The sole use of length measurements of hand and foot bones to estimate sex was tested in order to compare their accuracies with results from other studies that considered robusticity and length in combination. The study showed an advantage to using only length measurements when the reference sample used to create equations is temporally or genetically distant from the target sample to which the equation is applied (Case

and Ross 2007). Using data from past studies for comparison, the authors showed that robusticity measurements are more susceptible to lifetime activity changes and could cause confusion in sex estimation, particularly if the behavior of one group is quite different from that of another. Nevertheless, for the current thesis, both length and robusticity measurements are included in the dataset used.

A study by Harris and Case (2012) is also informative in this regard. While they did not use the same bones as this thesis focuses on, their study demonstrated the sexually dimorphic nature of tarsal metric traits. This study used 160 individuals from the Bass collection in order to determine which dimensions of the seven tarsal bones were most reliable for sex estimation. Both types of metric traits (robusticity and length) were represented with eighteen measurements of length, width, and height obtained from the seven tarsals. Logistic regression equations were created for sex estimation, and of the seven tarsals, the talus, cuboid, and cuneiform had the highest accuracies ranging from 88 to 92 percent, thus demonstrating that measurements of bones other than arm bones can provide highly accurate estimations and should be considered when alternative sex estimation methods are required.

Metric traits can also be utilized on bones that are traditionally used for morphological sex estimation, such as the os coxa (Mahakkanukrauh et al. 2016). When six coxal measurements were tested for their efficacy in sex estimation, t-tests revealed that all but one of the measurements were significantly different between males and females. Sex estimation equations from the study produced one equation with an accuracy of 97.5 percent. This provides evidence that metric traits can be extremely accurate and useful when used for sex estimation, even for bones used more commonly in traditional, non-metric sex estimation from the pelvis.

The manner in which metric traits are measured is an important aspect for the reliability of the method used. In 2015, a study was conducted to test the Scheuer and Elkington (1993) measurement technique of using calipers against the use of a mini-osteometric board when taking measurements from hand bones (Case et al. 2015). Both techniques were performed on 20 different skeletal hands in order to test for both intra-observer and inter-observer accuracy. The study found that the technique of using a mini-osteometric board produced lower error rates, with 88 percent of the measurements producing inter-observer rates under 1.5 percent compared to only 60 percent with caliper measurements. Additionally, measurements taken with a mini-osteometric board performed 10-12 percent faster than measurements taken with a caliper (Scheuer and Elkington 1993). These results suggest that a mini-osteometric board is preferable for many hand bone measurements. For this thesis in particular, the measurements used were collected through the use of the mini-osteometric board primarily, with only long bone lengths taken on a traditional osteometric board.

All of the above-mentioned research was vital in building a foundation for my thesis. As has been discussed, sex estimation studies have slowly built up a strong understanding over time of what does and doesn't work when using the skeleton for sex estimation. Alternate parts of the skeleton, such as the upper limb, have been shown to be viable options for sex estimation when needed. Binary logistic regression and linear discriminant analysis have both been used successfully in sex estimation attempts and therefore deserve to be further researched. Specific to the Thai population, some research has been done on metric sex estimation techniques and equations (Mahakkanukrauh et al. 2013 and Mahakkanukrauh et al. 2016), and the hope is that this thesis can add on to that knowledge by creating equations for both simple and multiple binary logistic regression.

CHAPTER 3: MATERIALS AND METHODS

The skeletons used for this study are Thai individuals and come from the Chiang Mai University Skeletal Collection housed at the Forensic Osteology Research Center (FORC), Faculty of Medicine, Chiang Mai University. The sample consists of 192 Thai individuals (96 females, 96 males), and were accessioned between 2005 and 2013. The age of these individuals ranges from 22 to 75, with a mean age of 59. The age breakdown can be viewed in Table 1 below.

Table 1. Age Distribution in Skeletons.

Age	# of Individuals
20-30	3
31-40	7
41-50	36
51-60	54
61-70	61
71-80	31

The bones used in this dataset all derive from the left side of the skeleton and include the humerus, radius, ulna, all five metacarpals, proximal phalanges one, three, and five, intermediate phalanges three and five, and distal phalanges one, three, and five. The full dataset contains 58 different measurements taken from each individual, but most have at least some missing data due to poor recovery after decomposition, damage, or pathology. Measurements taken (in mm) from the phalanges and metacarpals are: length (L), base width (BW), head width (HW), head height (HH), and midshaft diameter (MD); from the long bones: length, proximal breadth (PB), distal breadth (DB), epicondylar breadth (EpiBr) for the humerus, and midshaft diameter (MD).

For all measurements aside from long bone length, a mini-osteometric board (MOB) was used. For long bone measurements, Moore-Jansen's standards were followed (Mahakkanukrauh et al. 2011). In all cases, the bones were measured twice, non-consecutively, in an attempt to avoid measurement error. After the dataset was complete, any repeated measurements that varied

by 1 percent or more were measured for a third time, and the two closest values were retained. Outlier analyses were also conducted using boxplots produced. Any measurements that projected beyond the whiskers of the boxplot were measured again to verify that they were legitimate outliers. All statistical analyses were undertaken on SPSS version 27.

Binary Logistic Regression Analysis

The primary statistical process I used for this study was binary logistic regression. Simply put, binary logistic regression is a technique that is utilized in statistics in order to predict the relationship between independent variables and a dependent variable. The independent variable(s) in this case are the bone measurement(s) taken, while the dependent variable is the estimated sex. What is unique about binary logistic regression is that the dependent variable is always categorical and binary, meaning that there are only two responses possible. This technique is therefore perfect for sex estimation, where there are only two possibilities: male or female. Obviously, the estimation is of an individual's sex rather than their gender. Sex is a biological category that has influence over an individual's skeleton, among other biological entities, while gender is a social construct that has no biological ties. Sex therefore can be defined by only one of two categories: male or female.

Another aspect of binary logistic regression that fits nicely with this study and these data is that it does not assume that the data follows a normal distribution, meaning the data are not limited when it comes to the shape of their distribution. A normal distribution occurs when the majority of the values of a data set are clustered in the middle and taper off at both ends, forming a symmetrical curve (Zhu et al. 2017). In the case of binary logistic regression, the dataset is not assumed to be in that type of distribution, which removes some restrictions from the data (Bartholdy et al. 2020). The main purpose of this study is to assess whether a gain in sex

estimation accuracy can be attained by combining logits from simple binary logistic regression analysis versus the accuracy obtained using multiple binary logistic regression analysis.

Simple binary logistic regression is the use of one independent variable (in this case, one measurement) with a dichotomous output (in this case male or female). Multiple binary logistic regression also has a dichotomous output but differs from simple binary logistic regression due to its use of multiple variables in its input. For the specific process of sex estimation, only one measurement is used each time in simple regression analysis, while at least two measurements are used at a time in multiple regression analysis. The danger with using just one variable to estimate sex is that a single measurement might by chance happen to be unreliable for a specific individual in a forensic context, whereas other parts of the skeleton are not. This study seeks to explore how simple analysis based on single variables could be used in a different way, through combination of the individual results (logits), in order to increase the accuracy of sex estimations. In multiple regression analysis, the variables are combined before the regression is performed, which means that any possible grouping of measurements that might be used for forensic sex estimation must be conceived of ahead of time and published based on tested results from a known sex sample. If single variables are combined instead, then a researcher only needs to have access to forensic sex estimation results for the single variables they happen to have available. This may increase the number of measurements that could be brought to bear on sex estimations in a forensic context.

Simple Binary Logistic Regression and Logits

Simple binary regression will produce an output that can be utilized for the creation of a binary logistic regression equation for a single variable. The equation format is: L = A(variable) – B. “A” and “B” are the two numbers found in the regression output and “L” is the final result

of the equation (Leech 2011). The proper term for "L" is "logit," which is the natural log of odds of the dependent variable (sex, in this case) equal to a certain value. The logit will either be positive or negative. In sex estimation studies, a positive logit is an estimation of male, and a negative logit is an estimation of female. This study examines whether simple binary logistic logits corresponding to different variables can be combined by simple addition to improve sex estimation outcomes, with the idea that the more logits that are combined (and thus, the more variables involved), the stronger the resulting number and subsequent estimation. The main goal of this study is to compare the accuracy of combining these logits from simple binary logistic regression to the accuracy of simply using multiple binary logistic regression.

Determination of Measurement Set

Given the large number of initial measurements in the dataset (58), it proved necessary to reduce the number of variables included in the study in order to make the study manageable, and to focus on the variables with the highest sexual dimorphism. The first step in reducing the number of measurements was to assess the individual sex estimation accuracy using simple binary logistic regression analysis. These data were helpful in determining which measurements showed limited utility for sex estimation and should therefore be excluded from the final dataset. Once the accuracies were obtained, I removed any measurements with predicted correct allocation accuracies below 80 percent, a standard cut off point in forensic sex estimation. These accuracies can be found in Appendix B. Seventeen (17) of the measurements did not meet this standard, leaving 41 measurements. All but five (out of 14) of the length measurements tested for their individual accuracy fell below 80 percent. This could possibly be due to too much influence from the individual's environment; however, we cannot know for

sure. Robusticity measurement accuracies outperformed length measurement accuracies and are therefore heavily represented in this study.

Another reduction in measurements was made due to missing data, and the need to test sex estimation results on a single set of individuals. Not all 58 measurements were able to be taken from each individual due to preservation issues. In the case of some of these measurements, only a fraction of them provided useful data. Thirty-two (32) measurements could not provide data for ten or more individuals (out of an original 192, 96 males, 96 females), with 12 of those measurements not being able to provide data for 15 or more individuals. 15 missing individuals was made the cutoff number, and of the remaining 41 measurements (after the elimination for poor accuracy), eight measurements met this threshold and were taken out of consideration for the final dataset. See Table 2 for sample sizes associated with each variable by sex. The measurements highlighted in the table are the ones that were excluded for having met or exceeded the threshold of 15 missing individuals. The possible number of groupings from the remaining 35 measurements was pretty sizable and therefore the measurements list still needed to be reduced.

Table 2. Measurement Availability.

Measurement	F	M	Measurement	F	M	Measurement	F	M
LhumPB	**92**	**90**	**LradDB**	**89**	**89**	**LPP3BW**	**93**	**94**
LIP3BW	**91**	**95**	**LhumLg**	**91**	**90**	**LMC3HW**	**94**	**94**
LMC3MD	**95**	**95**	**LhumMD**	**90**	**90**	**LPP5BW**	**90**	**90**
LMC1L	90	88	LMC1BW	86	85	LMC1HW	88	92
LMC1HH	88	87	LMC1MD	89	88	LMC2L	93	94
LMC2BW	91	93	**LMC2HW**	**93**	**93**	LMC2HH	91	91
LMC2MD	**93**	**94**	LMC3L	95	95	LMC3BW	96	96
LMC3HH	91	91	LMC4L	93	91	LMC4BW	92	94
LMC4HW	94	94	LMC4HH	93	94	LMC5L	92	92
LMC5BW	92	92	LMC5HW	94	92	LMC5HH	90	92
LMC5MD	92	92	LPP1L	88	86	LPP1BW	87	94
LPP1HW	76	90	LPP3L	93	94	LPP3HW	93	94
LPP5L	90	90	LPP5HW	90	91	LIP3L	87	90
LIP3HW	90	94	LIP5L	79	92	LIP5BW	87	96
LIP5HW	84	93	LDP1L	90	88	LDP1BW	84	88

Table 2 (continued).

LDP1HW	90	87	LDP3L	52	66	LDP3BW	66	81
LDP5L	49	71	LDP5BW	64	80	**LradLg**	**89**	**92**
LulnLg	88	92	LhumEpiBr	87	84	LradPB	83	87
LulnPrBr	91	88	LulnDisBr	89	87	LradMD	89	90
LulnMD	88	90	-	-	-	-	-	-

Using previous research, the number of available data for each measurement, and the accuracy of each measurement, I settled on a total of 12 measurements (in bold in Table 2) that drew from both the upper arm as well as the hand. With these 12, the final step of measurement selection was able to be conducted.

Stepwise Logistic Regression

Once the variables with reasonable accuracies and sample sizes had been selected, binary logistic regression was again used to assess the sexing accuracy of some of these measurements in combination. Backward conditional regression is a type of stepwise regression that serves to select only the variables that contribute most to distinguishing the sexes. The process begins with a set of variables chosen by the researcher, and, through a series of tests, discards the variables that contribute least to classifying the data into the defined groups. Through this process, a group of variables will be narrowed down to only the most significant for whatever kind of classification is being attempted.

In this case, I ran the 12 variables through the stepwise logistic regression process, having categorized them by high, medium, or low accuracy. These accuracy categories were based on the existing variables I was working with. Any variable with an accuracy at or above 87 percent was categorized in the high accuracy group, any variable with an accuracy between 84 percent and 86.9 percent was categorized in the medium accuracy group, and any variables with an accuracy below 84 percent was categorized in the low accuracy group. See Table 3 for a visual breakdown of these categories.

Table 3. Defining of Accuracy Groups.

Accuracy Group	After Backward Conditional Selection	Before Backward Conditional Selection	Accuracy Range
High	LhumPB, LradDB, LPP3BW	LhumPB, LradDB, LPP3BW	87%+
Medium	LIP3BW, LhumLg	LIP3BW, LhumLg, LradLg, LMC2MD	84%-86.9%
Low	LMC3HW, LMC3MD, LhumMD, LPP5BW	LMC2HW, LMC3HW, LMC3MD, LhumMD, LPP5MD	<84%

The high accuracy group of the left humeral proximal breadth, left radial distal breadth, and left PP3 base width remained intact through stepwise regression – all three were deemed to contribute significantly to distinguishing the sexes. This means that each bone contributed some unique information that was helpful in distinguishing the sex. The medium accuracy group that originally comprised four variables was reduced to two – left IP3 base width and left humeral length. The lost variables apparently did not add any unique information that was useful for distinguishing the sexes. The low accuracy group that was originally composed of five variables was reduced to four, leaving out left MC2 head width. At this point, there were nine variables remaining that contributed significantly to classifying the skeletons by sex (Table 4).

Once I had reduced the number of measurements to those best suited for this study, I then created a new dataset of individuals without any missing measurements for the next stage of analysis. If an individual had even one missing measurement among the nine noted above, that skeleton was removed. This ensured that there would be the same exact set of individuals available for each measurement, and that all estimated correct allocation accuracies would be directly comparable. This process reduced the number of individuals available for further analysis from 192 to 140 (70 females and 70 males). It was this reduced dataset that was used to examine the relative ability of each bone to estimate the sex in all remaining analyses.

Groupings Used for Analysis

To create the variable groupings that were to be used in statistical analysis, accuracy groups, skeletal location groups, and Pearson's correlation coefficients were of use. With the sex estimation accuracies of each individual measurement available through SPSS (version 27), accuracy groups were created based on the level of accuracy each variable had. As mentioned above, measurements were sectioned off into high, medium, or low groups depending on their accuracies. This created ready-made groups to combine and compare. With these groups, all possible groupings were created (high/low, high/medium, etc.) for multiple binary logistic regression analysis.

Pearson's correlation coefficient is the measure of correlation between two variables. Through SPSS software, coefficients are able to be computed for each two-measurement combination possible. When there is low correlation between two measurements, it's likely due to the fact that they are in different areas of the skeleton. It seems likely that a group of measurements whose correlations with each other are low would produce more accurate sex estimations. The reason for this is that the variables are from different areas and/or do not have much correlation, so they provide different information about the skeleton because the developmental and behavioral influences upon them are somewhat different. The wider the variety of information there is from each measurement being used, the more likely the estimation made from a certain combination should accurate. For this reason, 15 different variable groupings were created based on Pearson's correlation coefficients. When creating groups, I took one measurement and added any measurements to the group that had a correlation of 0.5 or less. The size of the groups ranged from three to seven and none of the groupings completely overlapped with any others that had been created. The average correlations for each of the

Pearson's groups can be found in Appendix D. The last type of grouping was based on skeletal location – all hand measurements were combined, and all arm measurements were combined. This brought the total number of groupings analyzed to 24.

SPSS Software

Simple binary logistic regression analysis was first performed on each variable using SPSS version 27, which produced the information necessary to create a logistic regression equation specific to that measurement. For each individual in the sample, I then ran all nine of the measurements through their specific equations with the use of Excel to produce individual logits. Subsequent combinations of these logits in the measurement groups I had created based on accuracy, Pearson's correlations, and skeletal location would then produce accuracy information on those groups. Logit combination was by simple addition, and the final sum would indicate whether the estimate was male or female (negative for female, positive for male). Accuracy percentages were then calculated based on the number of correct estimations in each combination. Multiple binary logistic regression analysis was a simpler process than its counterpart. Using SPSS, all variable groupings were run through this analysis type and its output produced the accuracy percentage for each group. Percentage comparisons between the two analysis types as well as among the variable groups can be found in the results chapter.

Logit Summation vs. Multiple Binary Logistic Regression

Once all the variable groupings were created, they went through two separate processes. First, all logits were calculated for each variable and for each individual. Based on the variable groupings that were created using the criteria previously mentioned, these logits were then added together corresponding to the 24 different variable groups, and then the total accuracy of each group was calculated. The second process involved running each variable grouping through

multiple binary regression analyses in which accuracies were produced. There were now two accuracies for each (one from logit summation and one from multiple binary logistic regression analysis) group of variables making it easy to conduct side-by-side comparisons.

CHAPTER 4: RESULTS

The purpose of this study was to produce equations that could be directly compared for sex estimation from the upper limb, and to assess the accuracy of multiple binary logistic regression equations created by combining a variety of measurements. Results of these analyses are reported below. First, descriptive statistics for each variable considered in the study are reported. These statistics for both male and female can be found below in Tables 5 and 6.

Table 4. Male Descriptive Statistics.

Measurement	N	Minimum	Maximum	Mean	Std. Dev.	Variance
LMC3HW	70	12.170	16.370	14.068	0.897	0.805
LMC3MD	70	6.155	8.685	7.697	0.537	0.288
LPP3BW	70	14.190	17.580	15.783	0.728	0.530
LPP5BW	70	11.950	15.140	13.497	0.681	0.464
LIP3BW	70	11.735	15.090	13.659	0.746	0.557
LhumLg	70	280.00	338.5	306.610	13.039	170.019
LhumPB	70	45.480	53.420	49.396	2.0402	4.162
LhumMD	70	16.250	24.765	20.349	1.7453	3.046
LradDB	70	28.285	37.430	33.293	1.8758	3.519

Table 5. Female Descriptive Statistics.

Measurement	N	Minimum	Maximum	Mean	Std. Dev.	Variance
LMC3HW	70	10.930	14.285	12.569	0.6782	0.460
LMC3MD	70	5.990	8.320	6.857	0.4238	0.180
LPP3BW	70	12.730	15.780	14.091	0.6782	0.460
LPP5BW	70	11.180	13.525	12.314	0.5966	0.356
LIP3BW	70	10.685	13.350	12.231	0.5895	0.347
LhumLg	70	257.0	316.50	281.593	13.636	185.933
LhumPB	70	39.410	48.485	43.190	2.109	4.450
LhumMD	70	14.470	22.650	17.144	1.705	2.907
LradDB	70	24.355	32.525	28.543	1.682	2.829

Second, binary logistic regression equations were produced for each variable, and those with at least an 80 percent combined accuracy rate are reported. Third, the variables were divided into three groups based on the range of accuracy obtained from the equations, and these groups were used to guide me in combining variables to determine whether combining of logits or multiple logistic regression produced the greatest sex estimation accuracy. Finally, once direct comparison of the combined logits and multiple regression estimated accuracies were completed,

an analysis of variability within individuals in terms of the accuracy of these equations when applied to different bones and measurements was conducted.

Table 6. Simple Binary Logistic Regression Equations.

Measurement	Equation	Female % accuracy	Male % accuracy	Overall Accuracy	Nagelkerke Score
LMC3HW	L = 2.631(LMC3HW) – 34.850	82.9%	80.9%	81.9%	0.636
LMC3MD	L = 3.439(LMC3MD) – 24.952	82.9%	80.9%	81.9%	0.558
LPP3BW	L = 3.390(LPP3BW) – 50.579	87.1%	86.8%	87.0%	0.76
LPP5BW	L = 2.752(LPP5BW) – 35.469	81.4%	79.4%	80.4%	0.595
LIP3BW	L = 3.357(LIP3BW) – 43.313	84.3%	86.8%	85.5%	0.703
LhumLg	L = 0.138(LhumLg) – 40.520	85.7%	83.8%	84.8%	0.611
LhumPB	L = 1.385(LhumPB) – 64.196	92.9%	92.6%	92.8%	0.857
LradDB	L = 1.485(LradDB) – 45.717	91.4%	88.2%	89.9%	0.813
LhumMD	L = 0.991(LhumMD) – 18.523	80.0%	83.8%	81.9%	0.595

Table 7. Multiple Binary Logistic Regression Equations.

Variable Grouping	Equations
High Accuracy Group: LPP3BW, LhumPB, LradDB	L = 1.6(LPP3BW) + 1.055(LhumPB) + 0.948(LradDB) – 102.14
Medium Accuracy Group: LIP3BW, LhumLg	L = 2.922(LIP3BW) + 0.124(LhumLg) – 74.427
Low Accuracy Group: LhumLg, LMC3HW, LMC3MD, LPP5BW, LhumMD	L = 0.135(LhumLg) + 1.95(LMC3HW) – 0.305(LMC3MD) + 0.519(LPP5BW) + 0.674(LhumMD) – 82.857
LPP3BW, LIP3BW, LMC3HW (High/Medium/Low)	L = 2.617(LPP3BW) + 0.503(LIP3BW) + 0.73(LMC3HW) – 55.269
High/Low Accuracy Group: LhumPB, LradDB, LPP3BW, LMC3HW, LMC3MD, LPP5BW	L = 1.476(LhumPB) + 0.813(LradDB) + 1.957(LPP3BW) + 1.99(LMC3HW) – 0.919(LMC3MD) – 0.666(LPP5BW) – 143.073
High/Medium Accuracy Group: LhumPB, LradDB, LPP3BW, LIP3BW, LhumLg	L = 1.095(LhumPB) + 1.016(LradDB) + 0.32(LPP3BW) + 1.516(LIP3BW) + 0.084(LhumLg) – 131.507

Table 7 (continued).

Medium/Low Accuracy Group: LIP3BW, LhumLg, LMC3HW, LMC3MD, LPP5BW, LhumMD	L = 1.548(LIP3BW) + 0.135(LhumLg) + 1.343(LMC3HW) – 0.394(LMC3MD) – 0.239(LPP5BW) + 0.614(LhumMD) – 83.325
Hand Measurements: LPP3BW, LIP3BW, LMC3HW, LMC3MD, LPP5BW	L = 2.742(LPP3BW) + 0.322(LIP3BW) + 0.584(LMC3HW) + 1.33(LMC3MD) – 0.391(LPP5BW) – 57.343
Upper Arm Measurements: LradDB, LhumLg, LhumMD, LhumPB	L = 1.084(LradDB) + 0.069(LhumLg) – 0.113(LhumMD) + 1.122(LhumPB) – 104.048
LMC3HW, LhumLg, LhumPB, LhumMD	L = 3.516(LMC3HW) + 0.142(LhumLg) + 1.846(LhumPB) – 0.067(LhumMD) – 173.113
LPP3BW, LhumLg, LhumMD, LradDB	L = 1.691(LPP3BW) + 0.084(LhumLg) + 0.401(LhumMD) + 1.0(LradDB) – 88.412
LIP3BW, LhumMD, LhumPB, LhumLg	L = 2.052(LIP3BW) – 0.045(LhumMD) + 1.13(LhumPB) + 0.09(LhumLg) – 104.921
LPP3BW, LhumLg, LhumMD, LMC3MD	L = 2.405(LPP3BW) + 0.111(LhumLg) + 0.576(LhumMD) – 0.178(LMC3MD) – 78.165
LhumPB, LhumLg, LMC3MD, LMC3HW	L = 1.889(LhumPB) + 0.157(LhumLg) – 0.786(LMC3MD) + 3.757(LMC3HW) – 178.583
LPP5BW, LhumLg, LhumMD, LMC3MD	L = 1.294(LPP5BW) + 0.109(LhumLg) + 0.572(LhumMD) + 0.874(LMC3MD) – 65.827
LradDB, LhumMD, LMC3MD, LMC3HW	L = 1.151(LradDB) + 0.318(LhumMD) + 0.554(LMC3MD) + 1.04(LMC3HW) – 59.107
LMC3MD, LhumLg, LPP5BW, LhumMD, LradDB	L = -0.45(LMC3MD) + 0.088(LhumLg) + 1.036(LPP5BW) + 0.396(LhumMD) + 1.129(LradDB) – 81.206
LIP3BW, LMC3MD, LhumLg	L = 2.75(LIP3BW) + 0.404(LMC3MD) + 0.119(LhumLg) – 73.357
LMC3HW, LhumLg, LhumMD, LradDB, LMC3MD, LhumPB	L = 4.209(LMC3HW) + 0.171(LhumLg) – 0.433(LhumMD) + 0.984(LradDB) – 2.383(LMC3MD) + 2.394(LhumPB) – 223.019
LhumPB, LMC3HW, LPP5BW, LhumLg, LIP3BW, LradDB	L = 1.55(LhumPB) + 2.503(LMC3HW) – 0.557(LPP5BW) + 0.116(LhumLg) + 1.348(LIP3BW) + 0.748(LradDB) – 172.512

Table 7 (continued).

LPP5BW, LMC3MD, LradDB, LhumMD, LhumPB, LhumLg	L = 1.132(LPP5BW) – 1.698(LMC3MD) + 1.273(LradDB) – 0.175(LhumMD) + 1.382(LhumPB) + 0.099(LhumLg) – 132.069
LradDB, LhumLg, LhumPB, LPP5BW, LMC3MD, LhumMD	L = 1.273(LradDB) + 0.099(LhumLg) + 1.382(LhumPB) + 1.132(LPP5BW) – 1.698(LMC3MD) – 0.175(LhumMD) – 132.069
LhumMD, LPP5BW, LPP3BW, LradDB, LMC3HW, LMC3MD, LhumLg	L = 0.506(LhumMD) – 0.759(LPP5BW) + 1.857(LPP3BW) + 0.99(LradDB) + 0.851(LMC3HW) – 1.083(LMC3MD) + 0.107(LhumLg) – 93.131
LhumMD, LMC3MD, LMC3HW, LIP3BW, LPP5BW, LradDB, LhumLg	L = 0.425(LhumMD) – 0.739(LMC3MD) + 0.895(LMC3HW) + 1.682(LIP3BW) – 0.508(LPP5BW) + 1.061(LradDB) + 0.115(LhumLg) – 96.334

Comparison with Combined Logits Approach

A quick look at Table 9 below will demonstrate how the two binary logistic regression analysis types compared. Only three of the variable groupings produced a better accuracy by combining logits from simple binary logistic regression than by use of multiple binary logistic regression. The only other time in which multiple binary logistic regression did not produce a higher accuracy was when the accuracies tied at 88.6 percent for the low accuracy group. For the other 20 groupings, multiple binary logistic regression produced a higher accuracy than the combination of single regression logits did. A bolded row indicates the groupings that had a higher accuracy for logit summation than multiple binary logistic regression analysis.

Table 8. Accuracy Rates

Grouping	Logit Summation	Multiple
High Accuracy Group: LPP3BW, LhumPB, LradDB	95%	95.7%
Medium Accuracy Group: LIP3BW, LhumLg	**91.4%**	**90.7%**

Table 8 (continued).

Low Accuracy Group: LhumLg, LMC3HW, LMC3MD, LPP5BW, LhumMD	88.6%	88.6%
LPP3BW, LIP3BW, LMC3HW (High/Medium/Low)	86.4%	88.6%
High/Low Accuracy Group: LhumPB, LradDB, LPP3BW, LMC3HW, LMC3MD, LPP5BW	91.4%	95.7%
High/Medium Accuracy Group: LhumPB, LradDB, LPP3BW, LIP3BW, LhumLg	94.3%	95%
Medium/Low Accuracy Group: LIP3BW, LhumLg, LMC3HW, LMC3MD, LPP5BW, LhumMD	90%	93.6%
Hand Measurements: LPP3BW, LIP3BW, LMC3HW, LMC3MD, LPP5BW	87.1%	88.7%
Arm Measurements: LradDB, LhumLg, LhumMD, LhumPB	93.6%	95%
LMC3HW, LhumLg, LhumPB, LhumMD (Pearson)	92.9%	96.4%
LPP3BW, LhumLg, LhumMD, LradDB (Pearson)	**95%**	**94%**
LIP3BW, LhumMD, LhumPB, LhumLg (Pearson)	**94.3%**	**93.6%**
LPP3BW, LhumLg, LhumMD, LMC3MD (Pearson)	91.4%	92.1%
LhumPB, LhumLg, LMC3MD, LMC3HW (Pearson)	91.4%	95.7%
LPP5BW, LhumLg, LhumMD, LMC3MD (Pearson)	87.1%	89.3%
LradDB, LhumMD, LMC3MD, LMC3HW (Pearson)	90%	92.9%
LMC3MD, LhumLg, LPP5BW, LhumMD, LradDB (Pearson)	92.9%	93.6%
LIP3BW, LMC3MD, LhumLg (Pearson)	87.9%	90.7%
LMC3HW, LhumLg, LhumMD, LradDB, LMC3MD, LhumPB (Pearson)	95.7%	97.1%
LhumPB, LMC3HW, LPP5BW, LhumLg, LIP3BW, LradDB (Pearson)	94.3%	95.7%
LPP5BW, LMC3MD, LradDB, LhumMD, LhumPB, LhumLg (Pearson)	92.9%	95%
LradDB, LhumLg, LhumPB, LPP5BW, LMC3MD, LhumMD (Pearson)	92.9%	95%
LhumMD, LPP5BW, LPP3BW, LradDB, LMC3HW, LMC3MD, LhumLg (Pearson)	91.4%	94.3%
LhumMD, LMC3MD, LMC3HW, LIP3BW, LPP5BW, LradDB, LhumLg (Pearson)	90%	94.3%

Comparisons among all variable groups are a little more detailed. The accuracies of each accuracy group were to be expected – the most accurate percentage for both simple and multiple

binary logistic regression analysis came from the high accuracy group – those individual measurements that had the highest estimated allocation accuracy when used alone – followed by the medium accuracy group, and then the low accuracy group. The difference in accuracy among these three groups of variables are pretty notable, with the highest accuracy being 95.7 percent (based on multiple regression of high accuracy group variables) and the lowest accuracy being 88.6 percent (based on both simple and multiple regression for low accuracy group variables). A total of nine groups were examined, with all but three (low, high/medium/low, hand) producing an accuracy above 90 percent for both simple and multiple regression equations.

When examining each individual skeleton and their logits for each measurement, any errors that occurred for all 9 variables per individual (e.g. the logit addition approach resulted in an incorrect estimation) were logged and analyzed. The estimation of sex is female if the logit is negative and male if the logit is positive. Logits range from negative to positive, and the closer to zero a logit is, the higher the probability of inaccurate sex estimation (see Bartholdy et al. 2020). The farther a logit is from zero, the more confidence one might have in its estimation. I examined each logit result that incorrectly estimated the sex of an individual for each measurement and found this to be certainly true for this dataset. Table 10 will demonstrate the numbers breakdown. For both male and female categories, more than 50 percent of the errors were for logit values between -1 and 1. Above and below these values, the error percentages taper off.

Table 9. Logit Error Breakdown.

	F	M (-)
0-0.5	18 (24.7%)	28 (30.4%)
0.51-1	26 (35.6%)	28 (30.4%)
1.01-1.5	6 (8.2%)	19 (20.7%)
1.51-2	11 (15.1%)	8 (8.7%)
2.01-2.5	4 (5.5%)	4 (4.4%)
2.51-3	5 (6.9%)	3 (3.3%)

Table 9 (continued).

3.01-3.5	2 (2.7%)	0 (0%)
3.51+	1 (1.4%)	2 (2.2%)

When combining all of the single-variable results available for each individual skeleton, the number of sex estimation errors displayed for each individual ranged from zero to seven (out of a possible nine). Thus, while some individuals had all nine logits estimating the same sex, many had one or more logits that indicated an incorrect sex (sometimes resulting in an overall estimation error, sometimes not). These results also mean that all individuals had at least two measurements that were indicative of the correct sex, even in cases where the majority were incorrect. I examined the average age for each number of errors in order to assess whether there was a possible relationship between the number of errors present in an individual and their age. This information is displayed in Table 11 below. The number of individuals who displayed zero sex estimation errors was 68 and the number of individuals who displayed at least one sex estimation error was 72. This means that 51.4 percent of the total sample had at least one measurement that indicated an incorrect sex in its corresponding logit. However, based on this dataset, there doesn't seem to be an obvious relationship between the number of errors an individual's measurements presents and the individual's age, as the average age fluctuates as the number of errors increases and the youngest average is surprisingly for 7 errors.

Table 10. Age vs. Number of Errors.

# of Errors	# of Male Individuals	Average Male Age	# of Female Individuals	Average Female Age	Total # of Individuals	Total Average Age
0	33	54	35	60	68	57
1	17	62	12	55	29	59
2	3	59	9	57	12	58
3	7	51	4	55	11	52
4	3	64	6	64	9	64
5	3	46	2	51	5	47
6	3	63	2	64	5	63
7	1	47	0	N/A	1	47

CHAPTER 5: DISCUSSION

The question being asked in this study was whether or not combining logits produced by simple binary logistic regression analysis was a more or less accurate way to estimate sex than using multiple binary logistic regression analysis. This was tested using measurements from upper limb bones and applying binary logistic regression analysis. The hypothesis was that this new method of combining logits from simple binary logistic regression would produce a more accurate way to estimate sex than multiple binary logistic regression. It would also make application of multiple variables to metric sex estimation much easier in poorly preserved specimens. Based on the results from this study, however, accuracy was not improved by adding logits compared to creating multiple regression equations. For all but 4 of the 24 variable groups tested, the accuracy was higher when multiple binary logistic regression analysis was utilized than when simple binary logistic regression logits were added together. Additionally, in instances in which the accuracy from multiple binary logistic regression analysis was higher, the difference in percentage between the two methods ranged from 0.7 to 4.4. However, when the accuracy from logit summation was higher, the maximum difference was 1 percent. Therefore, it's important to note that when logit summation accuracies were higher, they just barely higher.

Overall Accuracies

20 out of the 24 groups of variables had accuracies above 90 percent for either simple or multiple binary logistic regression (or both). This indicates that, overall, the use of binary logistic regression and upper limb measurements are quite good options for sex estimation when needed. Previous research has already shown this to be true (Scheuer and Elkington 1993, Burrows et al. 2003, Mahakkanukrauh et al. 2013 to reference a few), so this study can now further support the viability of this method (and specifically for Thai individuals). Another important result is that

only one of the 24 groups had a lower accuracy than its variable with the highest single accuracy – this group was LPP3BW, LIP3BW, and LMC3HW (high/medium/low).

Accuracies from Accuracy Groups

As mentioned in the previous section, the accuracies obtained from the three accuracy groups – the high, medium, and low accuracy variables – were as expected for both simple and multiple binary logistic regression. For both logit summation and multiple binary logistic regression analysis, the highest accuracy was found for the high accuracy group, followed by the medium then by the low accuracy group. The same could be said when I combined the accuracy groups in different ways (high/medium, high/low, low/medium) – the accuracies were what one would expect based on the accuracy obtained from the individual measurements. The highest accuracy percentage came from the high/medium group, the second highest accuracy percentage came from the high/low group, and the lowest accuracy percentage came from the low/medium group. These results might have been expected, but the groups were tested in order to determine whether some of the lower accuracy variables might nevertheless provide new information about sex that was not already indicated by the higher accuracy individual variables. It would appear from these results, however, that most of the information about sex is already contained within the measurements from the higher accuracy variables and their combination.

Pearson's Groups

What was less predictable was the level of accuracy that all the other variable categories would produce with both regression types. As mentioned in the materials and methods chapter, the multiple regression variable groups not created based on similar accuracy among the variables were created based on dissimilarity assessed by means of Pearson's correlations. Through SPSS, correlations between each variable for both males and females were produced.

When two measurements are not highly correlated with each other, it might be because those two measurements are influenced by different genes and developmental processes that are influenced by sex, or because they are differently affected by lifetime activity, which may also vary by sex. Due to this difference in information from the measurements, those measurements that are poorly correlated might make for a good combination for the purposes of sex estimation. Rather than combining measurements that provide similar information or are from similar areas, combining measurements that provide different information might increase the chances that the sex estimation will have a high accuracy.

Using the correlation coefficients from SPSS between all variables for both sexes, I created 15 groups of poorly correlated variables. For each sex, each group that was created had a "principle variable" and any variable that had below 50 percent correlation with the principle variable would be included in the group. Group sizes ranged from three to seven and all were checked to make sure none were duplicates of each other. They turned out to be pretty successful in sex estimations, as only two out of the 15 produced an accuracy percentage of less than 90 for at least one of the regression types. Of those two, the lowest accuracy percentage was 87.1 percent, which is still quite high and higher than the low accuracy average. These results suggest that checking the correlations between and among variables may be a good way to determine the most productive variables to combine for multiple logistic regression.

Hand and Upper Arm Groups

The third type of group I tested was based on skeletal location. I grouped together all the hand measurements and all the upper arm measurements I was using into two different variable groups (hand group and upper arm group). There wasn't a specific expectation for this test, however, the hand measurements were all robusticity measurements while there was one length

measurement included in the upper arm group. There were also five measurements in the hand group while there were four in the upper arm group – this could possibly have an effect, as sometimes the more information there is the better (though not always). After testing in SPSS, the upper arm group ended up producing significantly higher accuracy percentages than the hand group did for both simple and multiple binary logistic regression. While I had no expectation one way or the other, it was a little surprising as the upper arm group measurements derived from only two different bones (humerus and radius), while the hand group measurements derived from four different bones (PP3, IP3, MC3, and PP5). However, the arm group included two high accuracy measurements, one medium accuracy measurement and one low accuracy measurement. In contrast, the hand group included one high accuracy measurement, one medium accuracy measurement, and three low accuracy measurements. The arm group included measurements with better accuracies which might explain its overall accuracies being higher than those of the hand group. As mentioned with Pearson's correlations, sometimes bones from different areas of the skeleton or that have a low correlation with each other produce accurate sex estimations because of the differing information they hold (they each bring a different piece of the puzzle). This didn't seem to be the case for this comparison, however, perhaps because of the bones of the hand, and particularly those of the third digit, likely develop and function more as a unit (Honekopp and Watson 2010).

Logit Errors

Each logit produced for each individual and each measurement that indicated an incorrect sex was logged and analyzed. As a reminder, a negative logit is a female estimation while a positive logit is a male estimation. Therefore, when a logit is close to zero, it can be assumed that it might not be as reliable as a logit that is farther away from zero. I tested that assumption and

came up with interesting results. For females, the group that presented the majority of errors was the group of logits that were 0.51 to 1 away from zero, at 36 percent of the errors. The group that presented the second largest number of errors was the group of logits that were 0 to 0.5 away from zero, at 25 percent of the errors. Together, then, 60 percent of all sex estimation errors were found among individuals with logit scores between 1 and negative 1. For males, these two groups tied in their percentages. The rest of the groups, going up by 0.5, followed a similar pattern which was a general decline in error percentage. This is something discussed in Bartholdy (et al. 2020) as well. They decided to test what would happen if they made "determinate" estimates only from those logits with a posterior probability above 80 percent. This meant changing the traditional cutoff point from zero to 0.45 and negative 0.45, a range whose logits that fall within it have a posterior probability under 80 percent. Those in this range produced "indeterminate" estimates. This greatly decreased the misclassifications and demonstrated the relationship between the reliability of a logit and its distance from zero (Bartholdy et al. 2020). While this thesis performed one small study on logit errors, it has further shown that the assumption of logits being more prone to error the closer they are to zero certainly seems to be true for sex estimation. Caution is warranted when estimating sex from a skeleton if all logits one is working with are close to zero.

Age vs. Logit Errors

Another aspect of the results I examined was the relationship between an individual's age and the number of logit errors their measurements produced. The number of errors an individual presented ranged from zero to seven. The assumption was that as an individual gets older, the more likely they are to have misleading measurements for various reasons. However, due to a small sample size, this analysis did not reflect that assumption perfectly. The average age of

individuals with only one error was the third highest average age at 59. From there, a general decline occurs, with only two categories not following the trend – average age of 64 for four errors and average age of 63 for 6 errors. The groups with the lowest average age of 47 were five and seven errors. However, the number of people in each group dropped drastically, as there are 68 in the group with zero errors and less than ten for each group with errors above three. With small samples for these groups, it is hard to conclude that what was reported is true for all Thai samples and datasets. It is also important to note that the average age of the entire sample was 57. While there could be a strong relationship between age and the number of sex estimation errors an individual has, more research with larger sample sizes is needed. It is important to note that sex was also considered when looking at the relationship between age and logit errors. As shown in Table 10, there wasn't much of a difference between male and female for any of the logit error categories. The category with the biggest difference between the two sexes was the 2-logit error category, with the difference being 6. In terms of which sex had more or less individuals for each category, it went back and forth. Similar to the relationship between age and logit errors, there didn't seem to be a strong pattern here, however, a bigger sample size is required to be able look at this relationship more in depth.

Use of Length vs. Robusticity Measurements

It is important to note that the majority (all but one) of the measurements used in this study were robusticity measurements rather than length measurements. Although Case and Ross (2007) argued that length measurements should be preferred over robusticity measurements when the reference and target samples are genetically or temporally more distant, length measurements did not prove to be very accurate in this study. All but five (out of 14) of the length measurements tested for their individual accuracy fell below 80 percent sex estimation

accuracy. This was the threshold I used in order to determine which measurements would be used in this study, leaving only five of the length measurements up for consideration. Only two of those five made it to multiple regression round, during which one of the two (radius length) was taken out of consideration by SPSS because it did not make a significant contribution to the sex estimation. While it is important to keep in mind previous research on this topic, it is also important to pay attention to the information a dataset is providing. I followed the accuracy rates and multiple regression results and ended up with measurements I wasn't expecting. However, these measurements produced promising results using both simple and multiple binary logistic regression.

CHAPTER 6: CONCLUSION

This thesis set out to develop some high accuracy equations for sex estimation, and to determine whether utilizing the method of combining simple binary logistic regression analysis outputs by addition would produce significantly similar or more reliable sex estimations than does standard multiple binary logistic regression analysis. A comparison was made between the accuracies of these two approaches to combining variables using nine different upper limb metric variables and 140 modern Thai individuals of known sex.

While several high-accuracy simple and multiple logistic regression equations were indeed developed, the overall conclusion from this comparison is that the method of combining the outputs (logits) of simple binary logistic regression analysis does not produce a higher accuracy than does multiple binary logistic regression analysis a majority of the time. Out of the 24 groups of variables created for statistical analysis, only three produced a higher accuracy from logit summation versus multiple binary logistic regression analysis. One additional group of variables produced the same accuracy for both simple and multiple logistic regression. The other 20 groups produced a higher accuracy for multiple regression than for logit summation.

It is now evident that for at least the Thai population, it might not be necessary to pursue this alternative method of combining outputs of simple binary logistic regression analysis over simply using multiple binary logistic regression analysis. Aside from this comparison, which was the main purpose of my thesis, light was shed on other important pieces of information regarding binary logistic regression analysis as a sex estimation method. Twenty-four (24) different groups were made from upper limb bone measurements, resulting in 48 binary logistic equations (24 simple, 24 multiple). Of those 48, sex estimation equations, only nine produced accuracies below 90.0 percent (accuracies for simple regression were from after logits were combined). This is an

extremely high success rate for both simple binary logistic regression analysis logit addition, and multiple binary logistic regression analysis in estimating sex. The lowest accuracy out of all 48 was 86.4 percent from logit summation for the high/medium/low group. While this accuracy percentage is not ideal when other options are available, it is still quite reliable for sex estimation in general.

This study also considered the relationship between age and the number of logit errors and the relationship between the distance a logit is from zero and how likely it is to result in an error. The latter analysis produced a fairly predictable result – the farther a logit was from zero, the lower the chance was that it was an error.

For the relationship between age and logit error, the results were less clear-cut. The prediction for this analysis was that as age increased, so did the number of logit errors. However, the average age for each category based on number of errors fluctuated as the number of errors increased. This analysis did not paint a clear picture of the relationship between age and number of errors present. This result might be due to the relatively small sample size (140) and/or the small number of young people present in the sample (the youngest individual was 22, the oldest individual was 75, but average age was 57).

This study highlighted an important comparison between two different methods of binary logistic regression analysis in sex estimation and showed conclusive evidence about which method produced a higher accuracy a majority of the time. In the process, 24 simple and 24 multiple binary logistic regression equations were created. The multiple binary logistic regression equations produced reliable accuracies while the various combinations of the logits produced by the 24 simple binary logistic regression equations produced less reliable but still

respectable accuracies. This information can be utilized in future studies as well as by anyone in need of sex estimation methods/equations that utilize upper limb bones.

It is important to note that this study and these equations/variable groups were conducted using an entirely Thai dataset and should therefore be used with caution if being considered for any other population. While this thesis has produced many sex estimation equations and shed light on smaller aspects of binary logistic regression in sex estimation, there is much more research needed in this area (particularly for various populations that have been poorly studied). I hope that this research will add to the growing number of studies on this topic and will help further the discussion and knowledge on alternative sex estimation methods.

REFERENCES

Anzellini, A., & Toyne, J. M. (2019). Estimating sex using isolated appendicular skeletal elements from chachapoyas, peru. *International Journal of Osteoarchaeology, 29*(6), 961-973. doi:10.1002/oa.2810

Attia, M.H., & Aboulnor, B.A.E. (2020). Tailored logistic regression models for sex estimation of unknown individuals using the published population data of the humeral epiphyses. Legal Medicine, 45, 101708. doi:https://doi-org.prox.lib.ncsu.edu/10.1016/j.legalmed.2020.101708

Baek, S., Komori, O., and Ma, Y. (2018) An Optimal Semiparametric Method for Two-group Classification. *Scand J Statist*, 45: 806– 846. https://doi-org.prox.lib.ncsu.edu/10.1111/sjos.12323.

Bartholdy, B.P., Sandoval, E., Hoogland, M.L.P. and Schrader, S.A. (2020), Getting Rid of Dichotomous Sex Estimations: Why Logistic Regression Should be Preferred Over Discriminant Function Analysis. J Forensic Sci, 65: 1685-1691. https://doi.org/10.1111/1556-4029.14482

Bongiovanni, R., & LeGarde, C. B. (2018). A univariate approach to sex estimation for the fragmentary upper limb. Journal of Forensic Sciences, 63(2), 356-360. doi:10.1111/1556-4029.13530

Burrows, A., Zanella, V., & Brown, T. (2003). Testing the validity of metacarpal use in sex assessment of human skeletal remains. Journal of forensic sciences. 48:17.

Case, D. T., & Ross, A. H. (2007). Sex determination from hand and foot bone lengths. Journal of Forensic Sciences, 52(2), 264-270. doi:10.1111/j.1556-4029.2006.00365.x

Case, D. T., Rawlins, C. M., & Mick, C. B. (2015). Measurement standards for human metacarpals: Measurement standards for metacarpals. *American Journal of Physical Anthropology, 157*(2), 322-329. doi:10.1002/ajpa.22700

Colman, K. L., Janssen, M., Stull, K. E., RR, v. R., Oostra, R. J., de Boer, H., & AE van, d. M. (2018). Dutch population specific sex estimation formulae using the proximal femur. Forensic Science International (Online), 286, 268.e1-268.e8. doi:http://dx.doi.org.prox.lib.ncsu.edu/10.1016/j.forsciint.2017.12.029

Duangto, P., & Mahakkanukrauh, P. (2020). Sex estimation from upper limb bones in a Thai population. Anatomy & cell biology, 53(1), 36–43. https://doi.org/10.5115/acb.19.179

Falsetti, A. (1995). Sex assessment from metacarpals of the human hand. Journal of Forensic Sciences, 40(5), 774-776. doi:10.1520/JFS15382J

Fliss, Barbara, Marcel Luethi, Philipp Fuernstahl, Angi M. Christensen, Ken Sibold, Michael Thali, and Lars C. Ebert. (2019). "CT-Based Sex Estimation on Human Femora using Statistical Shape Modeling." *American Journal of Physical Anthropology* 169 (2): 279-286. doi:10.1002/ajpa.23828. https://doi.org/10.1002/ajpa.23828.

Gonzalez, P.N., Bernal, V., Ivan Perez, S., & Barrientos, G. (2007). Analysis of dimorphic structures of the human pelvis: its implications for sex estimation in samples without reference collections. *Journal of Archaeological Science, 34*(10), 1720-1730.

Harris, S. M., & Case, D. T. (2012). Sexual dimorphism in the tarsal bones: Implications for sex determination. Journal of Forensic Sciences, 57(2), 295-305. doi:10.1111/j.1556-4029.2011.02004.x

Honekopp, J. and Watson, S. (2010), Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand. Am. J. Hum. Biol., 22: 619-630.

Krishan, K., Chatterjee, P.M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International, 261,* 165.e1-165.e8.

Leech, N. L., Barrett, K.C., & Morgan, G. A. (2011). Logistic regression and discriminant analysis. (pp. 140-160). Routlege.

Mahakkanukrauh, P., Khanpetch P., Prasitwattanseree, S., Vichairat, K., & Troy Case, D. (2011). Stature estimation from long bone lengths in a thai population. *Forensic Science International, 210*(1), 279.e1-279.e7.

Mahakkanukrauh, P., Khanpetch, P., Prasitwattanseree, S., & Case, D. T. (2013). Determination of sex from the proximal hand phalanges in a thai population. Forensic Science International, 226(1), 208-215. doi:10.1016/j.forsciint.2013.01.024

Mahakkanukrauh, P., Ruengdit, S., Tun, S. M., Case, D. T., & Sinthubua, A. (2016;2017;). Osteometric sex estimation from the os coxa in a thai population. Forensic Science International, 271, 127.e1-127.e7. doi:10.1016/j.forsciint.2016.11.043

Misziewicz, J.J., Miszkiewicz, J.J., Mahoney, P., & Mahoney, P. (2019). Histomorphometry and cortical robusticity of the adult human femur. *Journal of Bone and Mineral Metabolism, 37*(1), 90-104.

Nirenberg, M. (2016). Meeting a forensic podiatry admissibility challenge: A daubert case study. *Journal of Forensic Sciences, 61*(3), 833.841.

Page, M., Taylor, J., & Blenkin, M. (2011). Forensic identification science evidence since Daubert: Part 1 – A quantitative analysis of the exclusion of forensic identification science evidence. *Journal of Forensic Sciences, 56*(5), 1180-1184.

Pohar M, Blas M, Turk S. (2004) Comparison of logistic regression and linear discriminant analysis: a simulation study. MetodoloÐki Zvezki;1: 143–61.

Rolian, C., Lieberman, D.E., & Zermeno, J.P. (2011). Hand biomechanics during simulated stone tool use. *Journal of Human Evolution, 61*(1), 26-41.

Scheuer, J., & Elkington, N. (1993). Sex determination from metacarpals and the first proximal phalanx. Journal of Forensic Sciences, 38(4), 769-778. doi:10.1520/JFS13472J

Smith, S. (1996). Attribution of hand bones to sex and population groups. Journal of Forensic Sciences, 41(3), 469-477. doi:10.1520/JFS13937J

Stojanowski, C. M. (1999). Sexing potential of fragmentary and pathological metacarpals. American Journal of Physical Anthropology, 109(2), 245-252. doi:10.1002/(SICI)1096-8644(199906)109:2<245::AID-AJPA8>3.0.CO;2-T

Stull, Kyra E., Ericka N. L'Abbé, and Stephen D. Ousley. 2017. "Subadult Sex Estimation from Diaphyseal Dimensions." *American Journal of Physical Anthropology* 163 (1): 64-74. doi:10.1002/ajpa.23185. https://doi.org/10.1002/ajpa.23185.

Tallman, S. D., & Blanton, A. I. (2020). Distal humerus morphological variation and sex estimation in modern thai individuals, Journal of Forensic Sciences, 65(2), 361-371. doi:https://doi-org.prox.lib.ncsu.edu/10.1111/1556-4029.14218

Waidhofer, M. and S. Kirchengast. 2015. *Sexual Dimorphism in Directional Asymmetry of the Upper Limb Bones among Khoe-San Skeletons*. Vol. 66. doi:https://doi.org/10.1016/j.jchb.2015.08.001. http://www.sciencedirect.com/science/article/pii/S0018442X15000736.

Zhu, X., Ma, Z., Wang, T., & Teetranont, T. (2017). Plausibility regions on the skewness parameter of skew normal distributions based on inferential models. (pp. 267-286). Springer International Publishing.

APPENDICES

Appendix A

Abbreviation	Full Name	Abbreviation	Full Name
LMC1L	Left Metacarpal 1 Length	LMC1BW	Left Metacarpal 1 Base Width
LMC1HW	Left Metacarpal 1 Head Width	LMC1HH	Left Metacarpal 1 Head Height
LMC1MD	Left Metacarpal 1 Midshaft Diameter	LMC2L	Left Metacarpal 2 Length
LMC2BW	Left Metacarpal 2 Base Width	LMC2HW	Left Metacarpal 2 Head Width
LMC2HH	Left Metacarpal 2 Head Height	LMC2MD	Left Metacarpal 2 Midshaft Diameter
LMC3L	Left Metacarpal 3 Length	LMC3BW	Left Metacarpal 3 Base Width
LMC3HW	Left Metacarpal 3 Head Width	LMC3HH	Left Metacarpal 3 Head Height
LMC3MD	Left Metacarpal 3 Midshaft Diameter	LMC4L	Left Metacarpal 4 Length
LMC4BW	Left Metacarpal 4 Base Width	LMC4HW	Left Metacarpal 4 Head Width
LMC4HH	Left Metacarpal 4 Head Height	LMC5L	Left Metacarpal 5 Length
LMC5BW	Left Metacarpal 5 Base Width	LMC5HW	Left Metacarpal 5 Head Width
LMC5HH	Left Metacarpal 5 Head Height	LMC5MD	Left Metacarpal 5 Midshaft Diameter
LPP1L	Left Proximal Phalanx 1 Length	LPP1BW	Left Proximal Phalanx 1 Base Width
LPP1HW	Left Proximal Phalanx 1 Head Width	LPP3L	Left Proximal Phalanx 3 Length
LPP3BW	Left Proximal Phalanx 3 Base Width	LPP3HW	Left Proximal Phalanx 3 Head Width
LPP5L	Left Proximal Phalanx 5 Length	LPP5BW	Left Proximal Phalanx 5 Base Width
LPP5HW	Left Proximal Phalanx 5 Head Width	LIP3L	Left Intermediate Phalanx 3 Length
LIP3BW	Left Intermediate Phalanx 3 Base Width	LIP3HW	Left Intermediate Phalanx 3 Head Width
LIP5L	Left Intermediate Phalanx 5 Length	LIP5BW	Left Intermediate Phalanx 5 Base Width
LIP5HW	Left Intermediate Phalanx 5 Head Width	LDP1L	Left Distal Phalanx 1 Length
LDP1BW	Left Distal Phalanx 1 Base Width	LDP1HW	Left Distal Phalanx 1 Head Width
LDP3L	Left Distal Phalanx 3 Length	LDP3BW	Left Distal Phalanx 3 Base Width
LDP5L	Left Distal Phalanx 5 Length	LDP5BW	Left Distal Phalanx 5 Base Width
LHumLg	Left Humerus Length	LUlnLg	Left Ulna Length
LRadLg	Left Radius Length	LHumPB	Left Humerus Proximal Breadth
LEpiBr	Left Humerus Epicondylar Breadth	LRadPB	Left Radius Proximal Breadth
LRadDB	Left Radius Distal Breadth	LUlnPB	Left Ulna Proximal Breadth
LUlnDB	Left Ulna Distal Breadth	LHumMD	Left Humerus Midshaft Diameter
LRadMD	Left Radius Midshaft Diameter	LUlnMD	Left Ulna Midshaft Diameter

Measurement Abbreviations

Appendix B

Measurement	Accuracy	Measurement	Accuracy	Measurement	Accuracy
LMC1L	78.7%	LMC1BW	86.0%	LMC1HW	83.8%
LMC1HH	80.0%	LMC1MD	78.0%	LMC2L	80.7%
LMC2BW	86.4%	LMC2HW	83.3%	LMC2HH	83.5%
LMC2MD	86.6%	LMC3L	77.9%	LMC3BW	81.3%
LMC3HW	80.9%	LMC3HH	88.5%	LMC3MD	83.7%
LMC4L	76.1%	LMC4BW	81.2%	LMC4HW	81.4%
LMC4HH	88.8%	LMC5L	77.2%	LMC5BW	81.5%
LMC5HW	79.0%	LMC5HH	80.2%	LMC5MD	75.5%
LPP1L	77.6%	LPP1BW	89.5%	LPP1HW	84.3%
LPP3L	74.3%	LPP3BW	87.2%	LPP3HW	84.5%
LPP5L	71.7%	LPP5BW	79.0%	LPP5HW	85.1%
LIP3L	73.6%	LIP3BW	86.0%	LIP3HW	86.4%
LIP5L	70.8%	LIP5BW	80.9%	LIP5HW	86.4%
LDP1L	80.9%	LDP1BW	82.6%	LDP1HW	75.1%
LDP3L	75.4%	LDP3BW	85.0%	LDP5L	75.0%
LDP5BW	82.6%	LHumLg	84.4%	LUlnLg	84.4%
LRadLg	86.2%	LHumPB	93.4%	LEpiBr	91.2%
LRadPB	93.5%	LRadDB	91.6%	LUlnPB	85.5%
LUlnDB	88.1%	LHumMD	82.1%	LRadMD	76.0%
LUlnMD	78.7%	-	-	-	-

Measurement Accuracies for Dataset of 192 Individuals

Appendix C

Measurement	Accuracy (192 dataset)	Measurement	Accuracy (140 sample)
LMC3HW	80.9%	LMC3HW	81.9% (+)
LMC3MD	83.7%	LMC3MD	81.9% (-)
LPP3BW	87.2%	LPP3BW	87.0% (-)
LPP5BW	79.0%	LPP5BW	80.4% (+)
LIP3BW	86.0%	LIP3BW	85.5% (-)
LHumLg	84.4%	LHumLg	84.8% (+)
LHumPB	93.4%	LHumPB	92.8% (-)
LHumMD	82.1%	LHumMD	81.9% (-)
LRadDB	91.6%	LRadDB	89.9% (-)

Measurement Accuracies Comparison Before and After Dataset Reduction

APPENDIX D

Pearson's Grouping	Correlation Average
LMC3HW, LHumLg, LHumPB, LHumMD	0.35
LPP3BW, LHumLg, LHumMD, LRadDB	0.36
LIP3BW, LHumMD, LHumPB, LHumLg	0.39
LPP3BW, LHumLg, LHumMD, LMC3MD	0.36
LHumPB, LHumLg, LMC3MD, LMC3HW	0.38
LPP5BW, LHumLg, LHumMD, LMC3MD	0.34
LRadDB, LHumMD, LMC3MD, LMC3HW	0.45
LMC3MD, LHumLg, LPP5BW, LHumMD, LRadDB	0.38
LIP3BW, LHumLg, LMC3MD	0.32
LMC3HW, LHumLg, LHumMD, LRadDB, LMC3MD, LHumPB	0.39
LHumPB, LMC3HW, LPP5BW, LHumLg, LIP3BW, LRadDB	0.46
LPP5BW, LMC3MD, LRadDB, LHumMD, LHumPB, LHumLg	0.42
LRadDB, LHumLg, LHumPB, LPP5BW, LMC3MD, LHumMD	0.41
LHumMD, LPP5BW, LRadDB, LPP3BW, LMC3HW, LMC3MD, LHumLg	0.43
LHumMD, LMC3MD, LMC3HW, LIP3BW, LPP5BW, LRadDB, LHumLg	0.42

Correlation Averages for Each Pearson's Grouping