TWO FACTORS AFFECTING MULTI-ELEMENT QUANTIFICATION OF COMMINGLED HUMAN SKELETAL ASSEMBLAGES

A Thesis

Presented

to the Faculty of

California State University, Chico

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Anthropology

---

by

Valerie Sgheiza

Spring 2018

TWO FACTORS AFFECTING MULTI-ELEMENT QUANTIFICATION OF COMMINGLED HUMAN SKELETAL ASSEMBLAGES

A Thesis

by

Valerie Sgheiza

Spring 2018

APPROVED BY THE INTERIM DEAN OF GRADUATE STUDIES:

_____

Sharon Barrios, Ph.D.

APPROVED BY THE GRADUATE ADVISORY COMMITTEE:

_____

Colleen Milligan, Ph.D., Chair

_____

Eric Bartelink, Ph.D.

_____

Katharine Gray, Ph.D.

ACKNOWLEDGEMENTS

I am sincerely grateful to my committee, Dr. Colleen Milligan, Dr. Eric Bartelink, and Dr. Katharine Gray. Without your support and guidance, my thesis would likely have been unintelligible.

This project would not have been possible without access to datasets and skeletal collections. I am therefore deeply grateful to Alex Perrone of the CSU, Chico Human Identification Lab, Dr. Jason King of the Center for American Archeology, and Dr. Angela Soler and Dr. Brad Adams of the Forensic Anthropology Unit at the New York City Office of the Chief Medical Examiner for their assistance in that regard. I would also like to thank the CSU, Chico Honors Program, which funded my travel to New York through a Research and Creativity Grant, and the CSU, Chico Graduate Equity Fellowship Program for their support throughout this process.

Although I came into the program a year later, the Dream Team welcomed me with open arms. Matt, Lynsey, Maria, Martha, and Sarah, you truly have been my graduate school family.

I would like to thank my family for their support throughout my educational journey and their unwavering patience in reading my thesis.

Maria, aside from the endless drafts of this thesis you have read, in research you remind me to think big and not bury the lead; in life you have taught me to look up and live a little. My world is brighter with you in it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

TWO FACTORS AFFECTING MULTI-ELEMENT QUANTIFICATION

OF COMMINGLED HUMAN SKELETAL ASSEMBLAGES

by

Valerie Sgheiza

Master of Arts in Anthropology

California State University, Chico

Spring 2018

Skeletal quantification is critical to the analysis of skeletal assemblages. The use of multiple skeletal elements to estimate numbers of individuals can increase viability of estimation methods. Quantification, however, is both necessitated and constrained by skeletal attrition. Two factors involved in skeletal attrition that are identified here are independence of recovery of elements within the skeleton and the interaction between specific physical properties of skeletal elements and their recovery probabilities. The hypotheses tested were that (1) dependence of recovery of skeletal elements would be correlated to anatomical proximity and (2) element-specific recovery rates would be correlated with bone mass, length, and mineral density. Skeletal attrition is viewed here as a probabilistic process.

Skeletal inventories of 15 appendicular skeletal elements were collected from the Forensic Data Bank (FDB), CSU, Chico Human Identification Lab (HIL), Gibson Md. 2 from the Lower Illinois River Valley (LIV), Santa Clara Valley Medical Center historic cemetery (VMC), and unidentified forensic anthropology cases from the New York City

Office of the Chief Medical Examiner (NYC). Mass and length data were collected from non-overlapping cases from the HIL and NYC as well as the CSU, Chico anthropological teaching collection. Bone mineral density values were drawn from Kendell and Willey (2014). Independence of all possible 2-element combinations (n = 600) was tested for FDB, NYC, HIL, VMC, and LIV using Holm's adjusted Fisher's exact tests. Anatomical patterns were uncovered using Wilcoxon rank sum tests of phi values from all but VMC. Spearman's correlation was used to test relationships between element-specific recovery probabilities and mass, length, and mineral density for NYC, HIL, VMC, and LIV.

The results indicate that there is a relationship between anatomical proximity and independence of recovery. When selecting elements for inclusion in assemblage size estimation it is therefore advisable to choose elements on both limbs and avoid those that are attached at a joint. This will maximize agreement between actual and expected degrees of freedom, generating a reliable estimate of confidence in assemblage size approximation. Testing of hypothesis 2 showed a distinct pattern in significant skeletal properties for each dataset. This means that there is no one most important property for determining patterns of attrition, but rather that the most important property is the case-by-case result of taphonomic variables. This study indicates the need for further testing using collections affected by different taphonomic conditions and with a wide range of assemblage recovery rates.

CHAPTER I

INTRODUCTION

Background

One of the first questions that must be asked about a commingled skeletal assemblage is "how many individuals are represented?" We have to ask this question because a commingled assemblage is often the result of loss of anatomical context of skeletal elements between individuals and the process of skeletal attrition. This thesis addresses anatomical relationships, element dimensions, and taphonomy as interrelated contributors to skeletal attrition. Defining the roles of each of these factors allows for improving the application of existing estimators of number of individuals, as well as informing sample selection and field recovery.

This chapter introduces the topics of commingled remains and skeletal quantification, beginning with definitions of the specific terms used throughout the thesis. The second and third sections will cover existing methods, their limitations, and the ways in which limitations might be addressed. The fourth section presents the hypotheses tested, and how they will be evaluated and potentially falsified. The final section provides an outline of the remaining chapters of the thesis.

Definitions

Commingled remains are defined here as the combined skeletal elements of two or more individuals. This thesis will focus on skeletal quantification, the undertaking of estimating

the number of individuals contained in a commingled assemblage. Individuation, the separation and assignment of skeletal parts to specific individuals, will not be addressed.

When estimating the number of people represented in a commingled skeletal assemblage, it is important to distinguish between assemblage size and population size. Assemblage size is the number of individuals required to account for the elements observed in the physical assemblage. Population size is the true number of complete individuals whose remains have contributed to the assemblage. As a result of skeletal attrition, assemblage size is often smaller than population size. If at least one of every element type has been lost through attrition, the assemblage size will be at least one individual smaller than the population size. The definition of population used here is distinct from a living population, a group of living organisms.

"Element" will be used to refer to the individual bones of the skeleton. Any references to chemical elements will be specified as such. "Element type" will refer to the specific name of an element: humerus, femur, etc. Antimeres are lefts and rights of the same element type. Skeletal attrition will be used broadly here to mean the random or systematic loss of elements from a skeletal assemblage.

Elements can be physically lost or destroyed, or lost due to destructive processes that render elements unidentifiable as a specific element type even when much of the bone is still physically present. These processes fall into the larger category of taphonomy. "Taphonomy" is used here in the recent sense to refer to processes of human, environmental, or animal origin that alter remains after the death of an organism, not in the paleontological sense to refer to the process of fossilization.

Methodology

Skeletal quantification is a critical step in any skeletal analysis that may possibly involve more than a single individual. In faunal studies, quantification of taxa is necessary when studying patterns in resource distribution, site seasonality, and subsistence practices. Quantification of human remains is also necessary in bioarchaeology as a first step in reconstructing the demography of commingled assemblages. In modern contexts such as mass disasters and human rights investigations, an estimate of the number of deceased individuals can inform decisions related to resource allocation and investigative strategy.

The questions asked in each subfield may be quite different but there is considerable overlap in the methods employed. Most methods that can be used to quantify a single taxon of faunal remains can be applied to human remains with equal success and vice versa. These methods can be broadly categorized into two types: those that estimate the size of the assemblage and those that estimate the size of the population that created the assemblage.

Generally, methods that estimate population size are more statistically viable, while methods that estimate assemblage size are more forgiving under conditions of poor preservation. This distinction follows from the fact that in order to estimate population size it is usually necessary to use methods that employ capture-recapture principles, that is, pair-matching of right and left elements. As preservation declines, pair-matching accuracy will fall, but methods that depend on element counts alone may still be viable.

Research Utility

A further consideration is that most methods of both types rely on a single element type to produce their final result. The test statistic is computed for all element types present but the final number used is from the element that produced the highest value or the highest value for each taxon. As a result, most of the data available do not factor into the final calculation. The

Most Likely Number of Individuals (MLNI) is one exception. MLNI is a population-size estimator that can be performed on multiple element types to create a probability region (Adams and Konigsberg 2004).

Unfortunately, without knowing whether or not the element types selected are independent of one another, such an extension is not defensible. That is, does recovering a tibia from someone make it more likely that the corresponding fibula from the same person will also be recovered. Lack of independence will result in underestimation of interval size or overestimation of confidence. Any estimator that relies on multiple element types will encounter this same problem.

This decrease in probability interval size occurs because the degrees of freedom used to calculate the interval size is larger than the actual degrees of freedom. MLNI follows a hypergeometric distribution, a discrete distribution where sampling occurs without replacement. Despite many differences between hypergeometric and *t* distributions, the effect of degrees of freedom on the central tendency of probability density can be modeled using the *t* distribution. Figure 1A shows the base of a Student's *t* distribution with one degree of freedom. The unshaded region represents 95% of the area under the curve. Due to the size of the interval, most of the shaded area extends beyond the x-axis limits of the graph.

Figure 1B shows a Student's *t* distribution with 3 degrees of freedom, again with 95% of the area unshaded. This distribution has greater central tendency, corresponding to thinner tails and a much smaller confidence interval ([-12.71,12.71] vs. [-3.81,3.18]). This is analogous to the advantage conferred by using multiple elements to estimate MLNI.

Figure 1C shows a Student's *t* distribution with two degrees of freedom in which critical values have been calculated for three degrees of freedom. This is analogous to the

situation in which elements are not independent, so degrees of freedom are overestimated. Comparing figure 1B and figure 1C shows that the tails of figure 1C are thicker. As a result, the unshaded area contains less than 95% of the total area under the curve.

A reduction in area under the curve is apparent in figure 1D, a Student's *t* distribution with two degrees of freedom where critical values have been calculated correctly. The resulting confidence interval becomes -4.30,4.30. Figure 1C demonstrates the importance of knowing whether samples are independent. In this case, what is reported as a 95% confidence interval has less than a 95% chance of containing the true value.

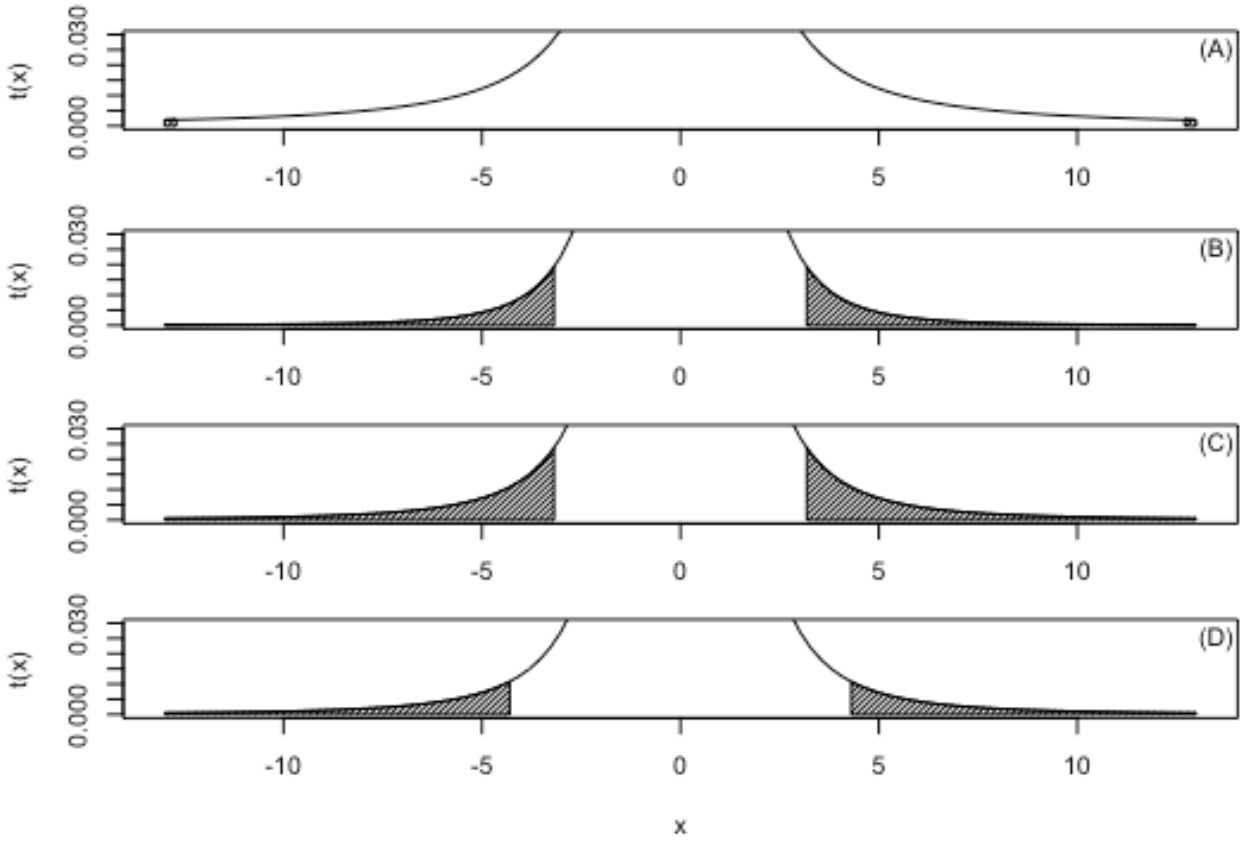

Figure 1. Student's *t* distributions with 1 (A), 3 (B), 2 (C), and 2 (D) degree of freedom. 95% confidence interval is unshaded for 1 (A), 3 (B), 3 (C), and 2 (D) degrees of freedom.

An ideal estimator would be one that uses multiple element types but does not rely on pair-matching. Such an estimator would be capable of estimating population size rather than assemblage size but would also be viable even when preservation is poor. There are two major obstacles that must be surmounted before such an estimator can be constructed. As mentioned above, independence must be tested because data from element types that are not independent should not be used to construct probability intervals. The second consideration is that many of the parameters of the statistical distribution of commingled skeletal remains have not been systematically characterized. Coming to grips with these issues would both improve existing estimators such as MLNI and open doors to deriving new estimators that are both statistically defensible and robust in the face of poor preservation.

Hypotheses

The following work was undertaken toward these two goals: first, to test element types for independence of recovery to inform the use of existing estimators, and second, to gain a better understanding of how and why assemblage recovery rates vary. Toward these ends, the following two pairs of hypotheses will be tested:

> **H_{o1}: within the appendicular skeleton, recovery of every element is independent of every other element.**
> **H_{a1}: dependence is correlated to anatomical proximity.**
>
> **H_{o2}: element attrition is random with respect to element characteristics.**
> **H_{a2}: element attrition will co-vary with secondary variables, specifically, intrinsic properties of bone.**

H_{o1} will be rejected if Fisher's exact tests demonstrate significant relationships between recovery incidences of element types. **H_{a1}** will be accepted if the probability of a significant relationship between element types differs according to proximity within the skeleton. Skeletal inventories were collected from several collections and used to test for independence

between combinations of element types. For example, would recovering a femur from an individual increase the probability of recovering the corresponding tibia from the same individual?

Inventories were additionally used to generate element-specific recovery probabilities. These recovery probabilities were tested for correlation with bone mass, length, and mineral density to look for interactions between intrinsic properties of bone and taphonomic processes of skeletal attrition. $\mathbf{H_{o2}}$ will be rejected and $\mathbf{H_{a2}}$ accepted if correlation tests demonstrate a significant correlation between element recovery rates and any of the three element properties tested.

Element-specific recovery probabilities are considered within the larger context of the hypergeometric distribution, the statistical distribution that best describes the probability of recovering elements in a skeletal assemblage. The hypergeometric distribution is similar to the binomial distribution except that sampling occurs without replacement. This distribution will be covered in more detail in chapter 3.

Outline of Thesis

This thesis is comprised of seven chapters. Chapter two will discuss the development of quantification methods in zooarchaeology, faunal studies, and human osteology. Chapter three will cover the statistical theory of skeletal quantification, including the development of estimators based on the hypegeometric distribution and the relationship between recovery rate and estimation bias. Chapter four will describe the materials and methods used in the study for independence testing and recovery rate characterization. Chapter five will present the results of both segments. Chapter six will discuss each of these segments as well as the limitations and

potential biases of this work. Chapter seven will cover the implications of this work for pair-match estimators and skeletal recovery, as well as identify possible directions for future research.

CHAPTER II

BACKGROUND

Introduction

This chapter will begin by discussing the development of skeletal quantification from its beginnings in archaeology to its use in osteology in the present day. The second section will examine the taphonomic processes that lead to skeletal attrition. It is crucial to consider skeletal quantification within the context of taphonomy. These are the forces that cause skeletal attrition, creating a need for quantification methods.

Skeletal Quantification

Skeletal quantification is a problem that is not restricted to a single area of anthropology. The historical development of the methods used by human osteologists today traces through several disciplines and many decades of research. Documented use of the Minimum Number of Individuals first appeared in Russian archaeology at the end of the 19th century (Casteel 1977a; Иностранцев 1882). It became a contentious topic in zooarchaeology in the 1970s and 1980s as different strategies for calculating MNI abounded and researchers debated the use of some strategies over others (Grayson 1984). Controversy over best practices in the adoption of methods developed by ecologists to quantify populations of living animals continues in human osteology today (Nikita 2014; Konigsberg and Adams 2014).

Zooarchaeology

Systematic skeletal quantification has its roots in zooarchaeology. The Number of Identified Specimens (NISP) has been employed for decades as a tool for comparing taxonomic frequencies of faunal remains at archaeological sites. NISP is defined as the number of specimens that can be identified to a specific taxonomic group (Grayson 1984). Beyond this basic definition there is wide variation in how this estimator is computed. Some researchers attempt to refit fragments before computing the final NISP (Clason 1972), while others do not (Chaplin 1971:65). NISP can be applied to taxonomic groups as narrow as species or as broad as class (Bobrowsky 1982; Reitz and Wing 1999:167). There is also dispute over the size of fragment that should be counted. This has caused problematic ripple effects in the archaeological record, such as the ascendance of hunting debate among California archaeologists (Hildebrandt and McGuire 2002). The apparent boom in artiodactyl abundance in faunal assemblages during the Middle Archaic in California was revealed to be entirely due to a single researcher including thousands of minute tooth enamel fragments in their NISP calculations (Fisher 2015).

NISP is also criticized in faunal studies because different taxa have different numbers of skeletal elements and these elements may have different levels of identifiability. Different taxa may also be processed in ways that impact their preservation, such as pulverizing long bones of mammals for marrow extraction. Furthermore, NISP does not account for interdependence. Dozens of specimens from the same taxon may represent anywhere from one organism to as many organisms as there are specimens (Grayson 1984).

The Minimum Number of Individuals (MNI) was proposed as a solution to many of these problems (White 1952). The MNI is defined as the greatest number of a single element from the same side of the body (White 1953). When applied conservatively, this method avoids the NISP pitfall of counting an unknown number of individuals multiple times. Unfortunately,

the methods for computing MNI are even more variable than for NISP. Since MNI counts individuals, the idea of finding ways to further differentiate counts from different groups of individuals can be very enticing.

One variation on MNI is to compute several MNI values from different units or levels within an archaeological site and add these values together. Aggregation of MNI values can result in very different numbers from the same skeletal data depending on the specific strategy used. If MNI values are computed for each unit or stratum and then added together, the result will likely be larger than a single MNI computed from the entire site. If multiple MNI values are computed within the same site, it is possible that some individuals may be counted twice, since their remains could be spread between multiple counted areas. As a result, MNI only escapes the problem of independence if it is computed on a site-by-site basis without further subdivision by unit or level (Grayson 1984; Casteel 1977a).

Grayson was initially a proponent of MNI over NISP, but in later work he reversed this position. In addition to interdependence, there is a mathematical relationship between MNI and NISP. For assemblages including multiple taxa this takes the form MNI = a(NISP)^b where $a > 0$ and $0 < b \leq 1$. For assemblages containing a single taxon, the relationship is often linear (b = 1) (Grayson 1984). This relationship raises the further issue of sample size. When MNI/NISP is plotted against MNI an exponential decay curve is produced. This means that with increasing NISP, the relative contribution to MNI from each newly identified element decreases. With every additional element counted, it becomes less likely that the new element will belong to an individual not accounted for by the elements already present in the assemblage.

MNI is therefore also a function of sample size. The problem with this is that the relative abundance of low frequency taxa will tend to be overinflated compared to taxa

represented by larger numbers of elements (Grayson 1984, 1978). This occurs as a direct result of the ratio of MNI to NISP decreasing as NISP increases. As a result, the larger of two assemblages might contain twice as many elements as the smaller one, but have an MNI that is only 50% larger. To control for this effect, Grayson suggests either not using MNI until MNI/NISP is less than 0.15 or adding 0.15 to the MNI/NISP value for the largest value of NISP and then computing MNI from the resulting adjusted ratio. These are not perfect solutions and in some cases it may be best to only compare relative abundance between taxa of similar NISP values (Grayson 1978).

Human osteology

Some skeletal quantification problems are less concerning in studies limited to a single taxon. Osteologists don't generally need to worry about inflation of relative taxonomic abundance because they are only looking at one species. It is a serious issue in zooarchaeology that snakes, for example, have very large numbers of similar looking vertebrae and ribs that will lead to high NISP values without increasing MNI. When comparing relative abundance of snakes and mice, individual mouse bones will tend to add more to MNI than individual snake bones because mouse skeletons have a higher proportion of uniquely identifiable skeletal elements than snake skeletons. Comparing humans to humans avoids this problem since all humans have similar skeletal anatomy.

Osteologists are more concerned with population size, the number of individuals that contributed to the observed skeletal assemblage, than they are with assemblage size, the number of elements physically present. Most work in human osteology has a direct connection to living people, either to assess the number of people missing, or to estimate population parameters for

past groups. Assemblage size is not directly comparable to numbers of living people because it is modified through skeletal attrition.

Accurately estimating population size is a point where many estimators fall short. MNI is a function of recovery rate, meaning that the lower the recovery rate is, the lower the MNI will be for the same population size. As a result, MNI is designed to estimate assemblage size, making it a biased estimator of population size (Adams and Konigsberg 2008). This bias is not a flaw in MNI, but rather a flaw in applications and expectations of the estimator.

The observation of bias in estimates of population size using MNI has spurred several studies aimed at reducing bias. Parmentier (2010) calculates MNI by using osteometric sorting to exclude element associations to the same individual. The modified MNI is the number of groups of elements that cannot be excluded from other members of the group (Parmentier 2010). Age has also been used as a means of accounting for more individuals. This approach has been recognized in zooarchaeology (Bokonyi 1970), but similar explicit applications in human osteology have been limited (Cruse 2002).

By taking advantage of potential diversity of biological profiles present in an assemblage, it may be possible to account for more individuals than can be measured through element counts alone. These methodological modifications are conceptually similar to the site aggregate MNI scores vilified by Grayson. The important difference is that aggregation strategies based on biological parameters are less likely to double count individuals than those that are based on spatial parameters. An individual can have skeletal remains in disparate physical locations after death, but barring antemortem dismemberment, simultaneous postmortem occupation of different biological states in the same skeleton (such as age, sex, or body size) should be minimal. If the researcher is adequately conservative in how they separate

or don't separate elements with similar attributes, such as adjacent age categories, double counting individuals should largely be avoided.

Oftentimes in human osteology, particularly forensic cases, the goal is to individuate rather than simply quantify. Reliable individuation without molecular methods quickly becomes impossible as population size increases. However, several methods have been developed for detecting and resolving small-scale commingling. Snow and Folk (1970) address the simplest case of the potential commingling of two individuals. If two individuals with similar biological profiles have potentially been commingled but the recovered remains do not contain any duplicate elements it is possible to compute the probability of recovering the number of skeletal elements present from two individuals without encountering duplicates. Under this method, as the number of elements present increases, the probability of their representing more than one individual decreases. At the extreme, it is highly unlikely that an assemblage would naturally contain 206 bones from 206 people without any duplication of elements. Similar to Snow and Folk, methods of osteometric sorting are exclusive rather than inclusive. This includes Schaefer's (Schaefer and Black 2007; Schaefer 2008, 2014) method for detecting commingling of juveniles using epiphyseal fusion, Buikstra et al. (1984) and Byrd (Byrd and Adams 2003; Byrd 2008) on joint articulation, and Thomas et al. (2013) on antimeres.

Outside of closed populations it is only with atomic and molecular methods that the probability of another individual giving the same result becomes small enough to justify definitively assigning skeletal parts to the same individual. Since the World Trade Center terrorist attack and the subsequent developments in DNA technology, DNA testing has become sufficiently practical for large-scale application, particularly in mass disasters (Mundorff 2009;

Mundorff et al. 2014). It is a destructive method, however, and its cost remains prohibitive to institutions with limited means.

The continued financial and destructive barriers to blanket DNA testing of remains has led to exploration of other chemical methods, one of which is X-ray fluorescence (XRF). XRF is non-destructive and has been shown to be capable of differentiating individuals based on atomic elemental signatures (Gonzalez-Rodriguez and Fowler 2013). XRF is particularly helpful when individuals are only partially commingled because elements in question can be grouped with elements of known assignment based on grouping of chemical element concentration values (Perrone et al. 2014).

Applications from ecology

In addition to crossover between the two fields, both zooarchaeology and human osteology have borrowed methods from faunal ecology to address the problem of skeletal quantification. These are the capture-recapture methods that were originally developed to estimate population sizes of living animals through tagging. First, a group of animals is captured, tagged, and released. After a period of time sufficient for complete mixture of tagged animals back into the population but not so long that tagged animals will lose their tags or die, a second group is captured. The proportion of tagged animals in the second group should be equal to the proportion of tagged animals in the total population since the first captured group should have completely mixed back into the population.

Since the number of tagged animals is known, this can be used to estimate total population size (Seber 1982). This can be easily transferred to skeletal studies by taking the number of left elements as the first capture group, the number of rights as the second capture group, and the number of pairs as the number of tagged animals in the second group (Fieller and

Turner 1982; Adams 1996). Individuals are recaptured by pairing right elements to previously counted left elements.

Krantz (1968) was the first to apply capture-recapture methods to faunal assemblages with the estimator $\left\{N = {(R^2 + L^2)}/{2P}\right\}$, where L and R are the number of lefts and rights for a single element type and P is the number of matched pairs. His discussion of the development estimator is not extensive, advising only that one should be cautious of small assemblage sizes as this could cause large errors (Krantz 1968:288). Casteel (1977b) harshly criticizes Krantz for the lack of empirical evidence and statistical support for his estimator. Since the behavior of Krantz's estimator is unknown, Casteel suggests using pair-match data to compute Chaplin's (1971) Grand Minimum Total (GMT) $\{N = L + R - P\}$. This estimator is similar in behavior to MNI, but less prone to underestimating assemblage size because it accounts for unpaired elements (Casteel 1977b).

Fieller and Turner (1982) disagree with Casteel's suggestion to continue using minimum numbers such as GMT. They acknowledge that Krantz's estimator is flawed, but agree that it is better to estimate population size rather than assemblage size. The authors recommend against zooarchaeologists creating their own estimators that are poorly understood. Instead, it would be better to borrow established estimators from faunal studies, particularly the Lincoln-Petersen Index (LI) $\left\{N = {LR}/{P}\right\}$, which is well-characterized (Fieller and Turner 1982). Krantz's goal of estimating population size had merit, but his strategy of creating a new estimator was problematic.

Horton (1984) takes a more moderate stance, presenting all three methods—MNI/GMT, Krantz's estimator, and the LI—as valid as long as they are applied under appropriate circumstances. If only part of a site is excavated, it would be best to use minimum

numbers because element pairs might be missed in unexcavated areas of the site. Casteel's wholesale dismissal of mark-recapture methods is unfounded, however, because the LI is a statistically valid means of estimating population size. The problem occurs when researchers expect estimators that are intended to calculate different things to produce the same answer. Assemblage size and population size are two very different things, and so an estimator should be chosen that is specific to the parameter desired (Horton 1984). This position is further disputed by Turner and Fieller (1985) who maintain that minimum number estimates are trivial and the LI will provide the best estimate when intelligently applied.

While human osteologists may be interested in quantifying assemblage size, they are often more concerned with population size, making capture-recapture methods particularly suited for human osteology. In addition, human osteologists are only looking at one taxon, and so don't have to be concerned about population estimators skewing relative taxonomic abundance, a problem for zooarchaeologists, as discussed above (Horton 1984).

The LI has been successfully applied to human skeletal assemblages (Adams 1996), but more recently the LI has been replaced by Chapman's Estimator $\{N = [(L + 1)(R + 1)/(P + 1)] - 1\}$, a modification of LI (Chapman 1951). Chapman's estimator has since been shown to be the maximum likelihood estimator of LI and renamed the Most Likely Number of Individuals (MLNI). This property makes MLNI preferable to LI because it has reduced bias (Adams and Konigsberg 2004). Both MLNI and LI are biased at low recovery rates, but MLNI is biased below 0.3 while LI is biased below 0.6 (Konigsberg and Adams 2014; Adams and Konigsberg 2008).

Current limitations

The greater challenge when using these estimators is that they require pair-matching. Under conditions of high fragmentation or otherwise poor preservation, pair-matching may be impossible. Additionally, in large assemblages the number of potential pairs becomes prohibitive (Adams and Konigsberg 2004). Research has instead focused on more efficient means of reducing the number of possible pairs to be compared when analyzing large assemblages (Nikita and Lahr 2011; Nikita 2014; McCormick 2016). Osteometric methods can be used to exclude element comparisons where dimensions differ widely, generating a more manageable subset of elements to be compared visually.

Such methods rely on good preservation and can only reduce the analytical barrier of large sample size, rather than eliminate it. When preservation is poor, the ability to take accurate measurements is reduced and visual pairmatching becomes difficult. Generally, if it is not possible to compute MLNI in an accurate and efficient manner, MNI is used instead because while it is not ideal, its problems are reasonably well understood and the investigator can at least be certain of not overestimating absolute number of individuals (Adams and Konigsberg 2004).

A further concern is element independence. This includes both antimeres as well as different element types. Independence of different element types is the more pressing issue at this time. Independence of antimeres, while a problem (Kendell and Willey 2014), has largely been addressed (Konigsberg and Adams 2014). If antimeres that are not independent are used to compute MLNI, for example, this will result in an estimate that has downward bias proportional to the level of dependence of the antimeres. Under the worst-case scenario that the dependence is 100%, the MLNI will be equal to the MNI. While analysts should be aware that antimere dependence will introduce downward bias into MLNI estimates, the bias of MLNI will always be

less than or equal to that of MNI, making it an overall better estimator (Konigsberg and Adams 2014).

Taphonomy and Skeletal Attrition

Outside of paleontology, taphonomy refers to all processes that affect the body after death including depositional activity, environmental effects, recovery, and curation. This section will focus primarily on taphonomic factors that lead to skeletal attrition. While effects such as bone staining and root etching do fall within taphonomy, they are less focal to the study. Taphonomic factors can be broadly divided by cause into the categories of animal, human, and environmental.

Animals

Both large and small scavengers can cause considerable damage to bone as well as disrupt context. Large carnivores may puncture, crush, or break bone in the process of removing consumable flesh from the skeleton. They may also remove and transport entire body portions, such as limbs, a considerable distance away from the original location of deposition. These actions can result in commingling of elements from multiple bodies, loss of elements, and loss of identifiability of remaining elements (Bright 2011; Haglund 1997).

Smaller scavengers, such as rodents and birds, can cause skeletal attrition as well. Rodents will chew on dense, upraised areas of bone to consume the available minerals. This can lead to loss of important features. Burrowing rodents such as gophers can greatly disrupt buried remains, leading to commingling and loss of smaller elements (Perino et al. 2006). Scavenging birds, such as vultures, can disturb remains on the surface and remove elements (Spradley et al. 2012).

Humans

Humans intentionally and unintentionally contribute considerably to skeletal attrition. Intentional activities include dismemberment, repeated reburial, and burning. Dismemberment may be performed either to obscure identity of the deceased or to facilitate disposal of the body, resulting in destruction and loss of skeletal parts. Repeated reburial due to concern over discovery of the remains or as part of mortuary ritual may be performed on a scale as small as a single body or as large as hundreds. Repeated reburial with heavy machinery of mass graves from the genocide in the former Yugoslavia resulted in extensive commingling of bodies and the redistribution of parts of individuals within multiple graves. This is not uncommon as elements are often missed in the process of reburial (Adams and Byrd 2008:7-29).

Well-intentioned humans may still disturb or destroy human remains. Accidental discovery of human remains can be quite destructive, particularly when heavy machinery is involved. Remains may be damaged during intentional recovery as well, via backhoes, shovels, hand tools, and even brushes. Processing of remains to remove flesh or grease from bone can damage bone directly or make it more vulnerable to future destructive forces. Some degreasing processes, for example, may weaken bone if applied overzealously (Lussi and Geistlich 1995). Curation of skeletal remains is an additional avenue for destruction; remains may be misplaced, stolen, or broken through poor handling practices.

Environment

Natural environmental processes have a variety of destructive effects on bone. Weathering from the sun causes bleaching, longitudinal cracking, and surface flaking (Behrensmeyer 1978). Fluvial transport of remains can lead to disarticulation of the body and abrasion of elements, causing loss of identifying features (Haglund and Sorg 2002). Repeated wetting and drying or freezing and thawing can cause cracking due to expansion and contraction,

and the very presence of water makes bone vulnerable to the effects of soil pH because damaging ions must be dissolved in order to be active.

Soil or substrate pH is an important factor in skeletal preservation as substrates of low pH can be very destructive to bone (Gordon and Buikstra 1984). The mineral matrix of bone is composed of hydroxyapatite ($Ca_5(PO_4)_3(OH)$). The exact mechanism of acid dissolution of hydroxyapatite is debated but the reactants and products are known. Addition of hydrogen ions removes the ^{-}OH leaving group, leading to the final reaction products of calcium ions, dihydrogen phosphate ions ($H_2PO_4^{-}$), and water, all of which are soluble in water (Dorozhkin 2012). Lastly, burning of bone removes its organic components, causing color changes, warping, cracking, and shrinkage, all of which can destroy bone or make it less identifiable (Schmidt and Symes 2015).

Summary

This chapter described the history and development of skeletal quantification in zooarchaeology and human osteology, including substantial contributions from ecology. This included a discussion of limitations to current methods that will be of interest in this study. The second section was an overview of taphonomic processes that lead to skeletal attrition, many of which likely affected the remains examined in this study. The next chapter will examine the theoretical models underlying existing methods in more detail and related these models to the hypotheses of the thesis.

CHAPTER III

THEORY

Introduction

The subject of skeletal quantification is not a new one, but its approaches have undergone extensive changes in the past century. This chapter will place human skeletal quantification within its greater theoretical framework: the intersection of skeletal attrition and statistics. The first part of this chapter will discuss skeletal attrition and how this affects estimation. The second section will address the use of the hypergeometric distribution as a tool for viewing the attritional process and introduce the statistical function of testing hypothesis 1. The third section will consider the challenges of pair-matching and its relation to recovery rates, as well as return to non-random attrition to develop hypothesis 2. The chapter will conclude by tying the theoretical framework developed to the specifics of the two hypotheses.

Skeletal Attrition and Estimation

The purpose of this thesis is to advance skeletal quantification, but such work would not be necessary without skeletal attrition. It is only the partial loss of elements from an assemblage that provides both a question to answer (how many individuals?) and the material with which to answer it (the elements remaining).

For the purposes of this section, skeletal attrition is the complete physical loss of an element or the loss of its uniquely identifying characteristics. There is a vast array of taphonomic processes that contribute to skeletal attrition. Some of these processes are specifically addressed in this thesis, but most are not. Depending on the processes at work, their relative effects, and the duration of time in which they are able to operate, the degree of attrition may be negligible or total, and its effect random or non-random. Both of these points will be addressed, but setting aside for the moment the question of random vs. non-random attrition allows for a closer examination of degree of attrition.

MNI, recovery rate and bias

The MNI is the largest count of a single element from the same side of the body **(1)**. All identifiable elements are counted and the largest count is used.

(1) $$N = max(L, R)$$

MNI is straightforward to use, but it does not accurately estimate the initial number of individuals when the degree of attrition is high and the recovery rate is low. Recovery rate is defined here as the ratio of the number of elements in the recovered assemblage to the total number of elements in the population that created the assemblage. When the recovery rate is low, MNI will show a distinct downward bias in estimation. This occurs because unless every element of at least one type is recovered, some individuals will not be counted.

These uncounted individuals may have other elements present in the assemblage, but MNI is a conservative estimator. MNI is designed to avoid double counting of the same individual above all else. There is therefore no mechanism within a strict MNI for deciding which elements might belong to additional individuals not represented in the highest count. Modifications to the MNI, such as divisions based on age or size, seek to account for these

individuals while minimizing the risk of double counting individuals. Since even a modified MNI should not overestimate the number of idividuals, modified MNI methods will still show downward bias. Bias in MNI occurs even at high recovery rates, showing a nearly linear relationship with recovery rate (Figure 2, 'MNI').

The slight degree of downward concavity seen in the MNI bias curve is the result of using the maximum of several elements. If an average of all counts were used instead, the relationship would be perfectly linear (Figure 2, 'Mean Count'). By using the maximum instead of the average, MNI slightly reduces bias relative to the mean.

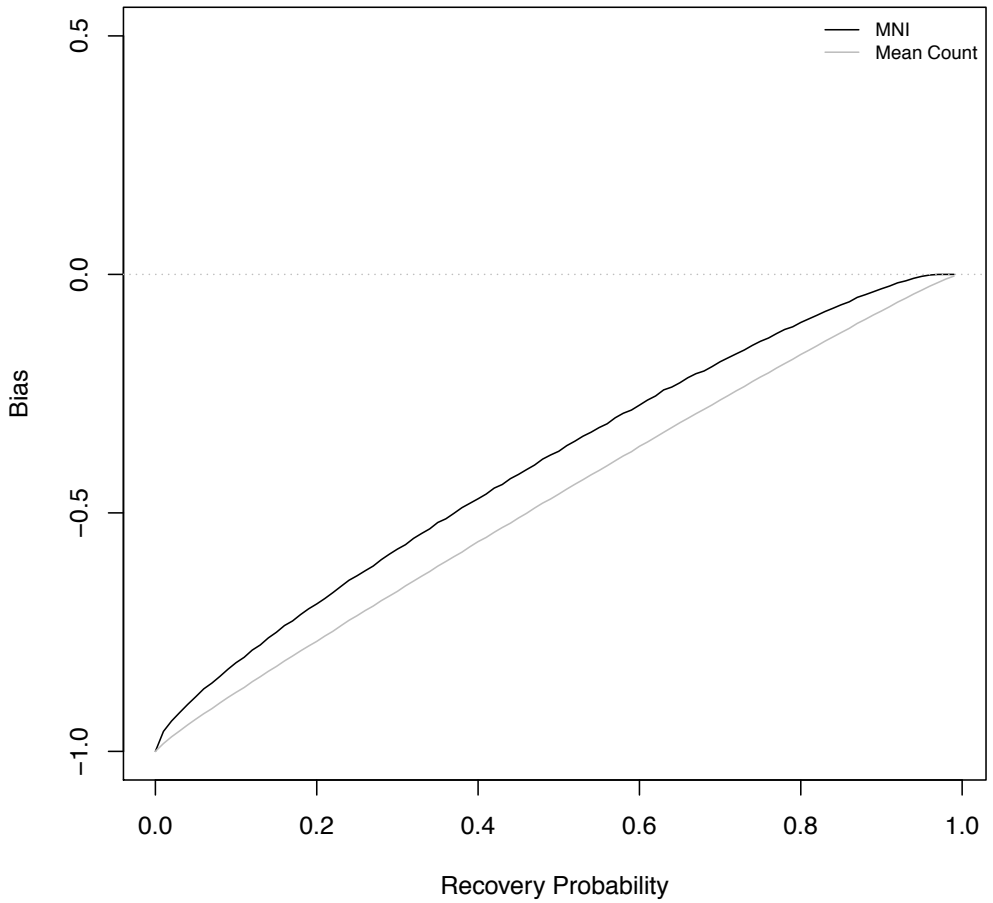

Figure 2. Percent bias of MNI and average counts (N=50 individuals per assemblage, 10 element types per individual, 1000-replicate simulation per 1% step in recovery probability).

Assemblage and population estimators

The problem of downward bias of MNI occurs because MNI is an assemblage estimator and not a population estimator. The true number of individuals (*N*) that osteologists are concerned with is a parameter of the initial population. The function of MNI is to count the size of the recovered assemblage. The problem is not that MNI is a poor estimator, but that it is used to estimate the wrong parameter. MNI and population size are related to one another by recovery rate, as shown above. A population estimator must account for the effect that recovery rate has on assemblage size.

The challenge of estimating the size of a complete population from an incomplete sample is one often encountered in studies of living animals. It is not possible to capture and count every member of a species or even every member of a local population. Instead, a wildlife biologist tags and releases a certain number, captures a second group some time later, and counts the number of tagged individuals in the second group. Capture-recapture methods are able to get at population size because the larger the invisible total population is, the more it will dilute the initial tagged group and therefore the lower proportion they will have in the second sample.

Population size is assessed by quantifying the effect of that invisible component. The proportion of tagged animals in the second sample should be approximately equal to the proportion of tagged animals in the total population. This means that *N:K* should be approximately equal to *n:x* (Figure 3). Capture-recapture methods lend themselves well to skeletal quantification because elements that can be pair-matched can be viewed as representing individuals who are both “captured” on the left and “recaptured” on the right.

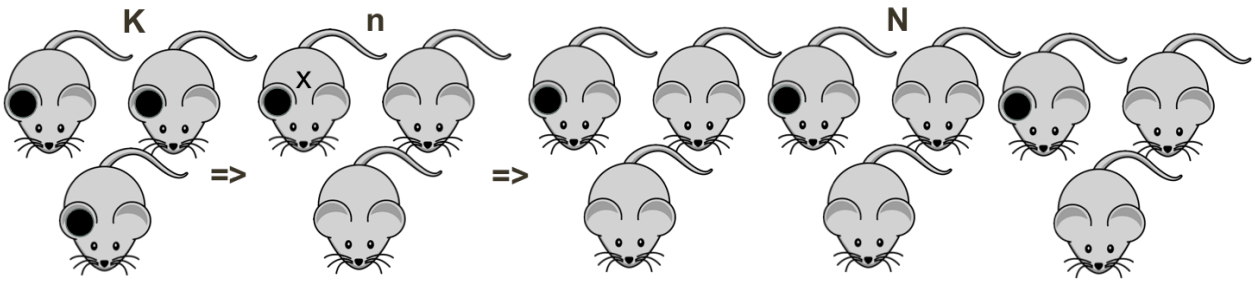

Figure 3. Diagram of population size estimation using capture-recapture methods.

The Hypergeometric Distribution

The initial population and the recovered assemblage can be viewed together using the hypergeometric distribution. This distribution models the behavior of a single element type in a skeletal assemblage. This discrete distribution describes the probability of obtaining *x* successes in *n* draws from a finite population *N* that contains *K* successes **(2)**. The hypergeometric distribution is similar to the binomial distribution except that sampling is done without replacement.

(2)
$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

Osteologists are primarily concerned with estimating *N*. In the approach taken for deriving pair-match estimators, *N* is the original number of individuals who produced the assemblage. *K*, the number of successes present in the population, is the number of lefts of a single element type present in the assemblage. The number of draws *n* is the number of rights of the same element in the assemblage. *x*, the number of successes in *n* draws, is the number of pairs recovered. Since the ratio of *N* to *K* should be similar to the ratio of *n* to *x* and only *N* is unknown, it is possible to estimate *N*. This estimate is performed using the equations in the next section.

LI and MLNI

The relationship between *N:K* and *n:x* is the basis of the Lincoln Index (LI). Setting N/K equal to n/x and solving for N produces equation **3a**, which is the LI. In osteology this is rewritten as equation **3b** to show which element counts are used.

(3a) $$N = \frac{nK}{x}$$

(3b) $$N = \frac{LR}{P}$$

For the purposes of this thesis, the osteological letter system *L, R,* and *P* will be used instead of the statistical letters *K, n,* and *x.*

Chapman (1951) modified the LI to produce equation **4**. Adams and Konigsberg (2004) showed that this modification is the Maximum Likelihood estimator of the LI and named it the Most Likely Number of Individuals (MLNI). MLNI is less biased than LI over the full range of recovery rates, however both MLNI and LI perform better than MNI in estimating *N*. As shown in Figure 3, MLNI becomes unbiased at around r = 0.3, while LI shows downward bias below 0.2, upward bias between 0.2 and 0.6, and is unbiased above 0.6 (Figure 4). This reduced bias makes MLNI the best estimator of the three when pair-matching can be performed.

(4) $$N = \frac{(L+1)(R+1)}{P+1} - 1$$

MLNI can be easily expanded to multiple elements. In this scenario, there are *M* independent hypergeometric distributions, where *M* is the number of element types used. *N* is the same for all distributions but *L, R,* and *P* may be different. In this case, it does not matter that the true number of successes (*L*) is not constant across all distributions because the number of draws (*R*) and the number of successes obtained (*P*) should all undergo similar attritional effects. Using several element types allows for an estimate of probability density around the mean point estimate, a useful tool in representing accuracy.

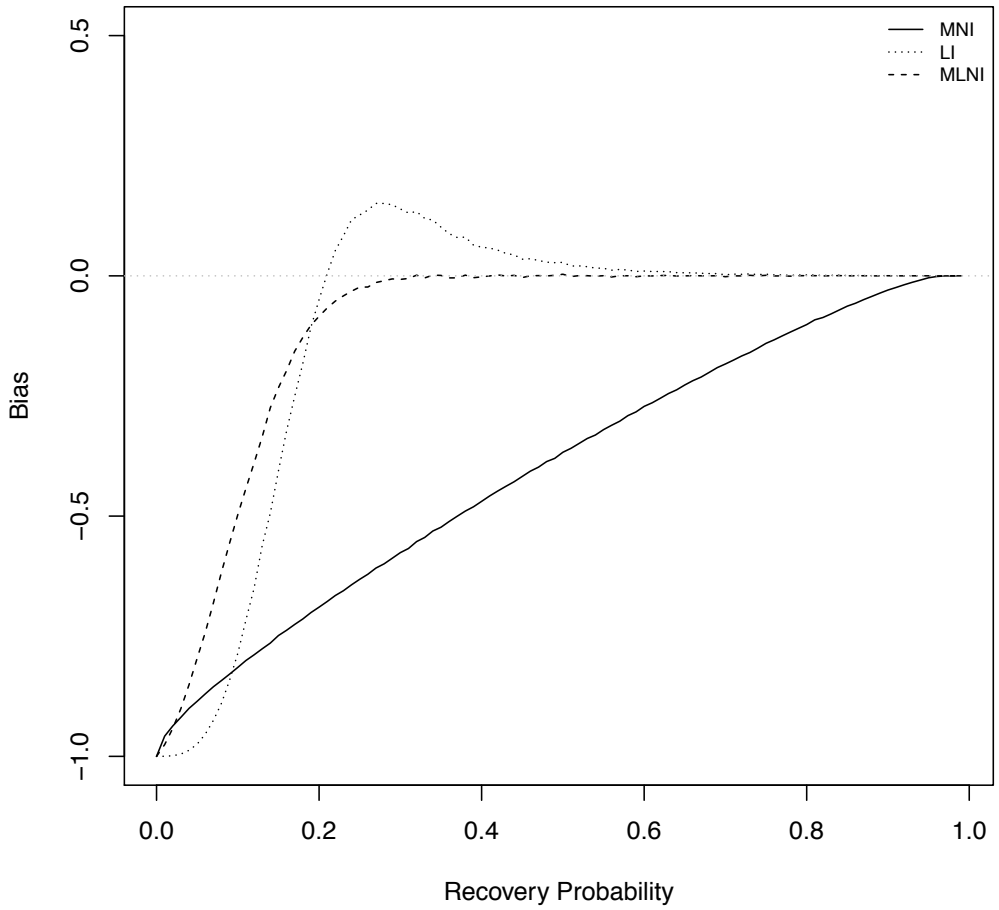

Figure 4. Comparison of percent bias of MLNI, LI, and MNI (N=50 individuals per assemblage, 10 element types per individual, 1000-replicate simulation per 1% step in recovery probability).

Interval estimation and independence

This probability region is not technically a confidence interval because the hypergeometric distribution is not symmetrical (Adams and Konigsberg 2004). Confidence intervals are an inefficient means of representing probability densities for asymmetrical distributions because the bound on the thinner tail will be too long and the bound on the thicker tail too short. This results in an interval that fails to capture the region of highest probability density. Asymmetrical distributions require asymmetrical upper and lower bounds. Adams and Konigsberg (2004) propose using Highest Density Regions (HDR) as an alternative means of

probability region estimation. An HDR captures the desired proportion of area under the curve (90%, 95%, etc.) in the smallest x-axis distance possible.

The key piece of information for expanding MLNI to multiple elements is that all element types must be independent from one another. This is critical to appropriately estimating the size of the HDR. In estimating HDRs, just as in estimating confidence intervals, the more independent values that fall within a certain area, the greater the probability that the true value for the system is captured within that area. The total size of the area is therefore directly proportional to the spread of the data points (variance) and inversely proportional to the number of independent points (degrees of freedom). If the points are not independent, the degrees of freedom is overrepresented and the resulting region will be smaller than what is actually justified by the available data. For this reason, one half of this thesis is devoted to the study of independence of recovery of skeletal elements.

Quantification Beyond Pair-matching

Non-random attrition with respect to element type is not necessarily problematic for the application of the ecological estimators. If each *L, R,* and *P* for a single element type is experiencing attrition in the same way (e.g. lefts and rights have equal probability of recovery, as do elements within individuals vs. elements between individuals), MLNI should be unaffected by attritional differences between element types. Neither one of these conditions is a given, but testing them is outside of the scope of this work.

Non-random attrition is of interest, however, when looking beyond pair-matching for avenues of skeletal quantification. Pair-matching can be precluded by either poor preservation or a very large number of potential pairs. Either scenario can make pair-matching extremely time consuming or impossible to perform accurately. Without pair-matching, the only analytical tool

remaining for estimating *N* is MNI. This is not ideal for the reasons stated above. Even without the option of pair-matching, the highest count element is only a small fraction of the total data available.

Estimation using variance in attrition

LI and MLNI operate by using the recovery rate to estimate the initial value. Information about the assemblage recovery rate should be available by studying the other element counts aside from the MNI. The variability of element counts should generally be low when overall recovery rate is high, high when recovery rate is intermediate, and low when the recovery rate is extremely low. This trend is expected because at extremely low recovery rates there are not enough elements in the assemblage to create much variability, whereas at high recovery rates most element counts will be close to the true size of the population. At intermediate recovery rates element counts should be least restricted by recovery rate.

The mean element count divided by the variance of the counts is related to the recovery rate *r* by $r = 1 - \frac{\sigma^2}{\bar{x}}$ (Figure 5). Since $r = \frac{\bar{x}}{N}$, it follows by setting the two equations for recovery rate equal and solving for *N* that $N = \bar{x}/\left(1 - \frac{\sigma^2}{\bar{x}}\right)$ **(5)**. Testing on simulated assemblages shows that equation **5** is not a viable estimator. The variance in element counts is itself much too variable to make a good basis for an estimator (Figure 5, 'N.var'). The estimator is wildly variable below recovery rates of 0.6 and is unbiased only above recovery rates of 0.8 (Figure 6). Compared to MLNI and even MNI (Figure 4), equation **5** performs very poorly.

(5) $$r = 1 - \frac{\sigma^2}{\bar{x}} = \frac{\bar{x}}{N} \Rightarrow N = \frac{\bar{x}}{1 - \frac{\sigma^2}{\bar{x}}}$$

Returning to Figure 5, it is apparent why this is the case. While there is a clear inverse relationship between variance and recovery rate, the point scatter in Figure 5 also becomes

narrower as recovery rate increases. It therefore stands to reason that putting a highly variable relationship in the denominator of the estimator for N will produce a result that is even more unpredictable. Bias fluctuates between downward and upward, and rather than decreasing linearly, the variability of the estimator drops off rapidly around a recovery rate of 0.6 (Figure 6).

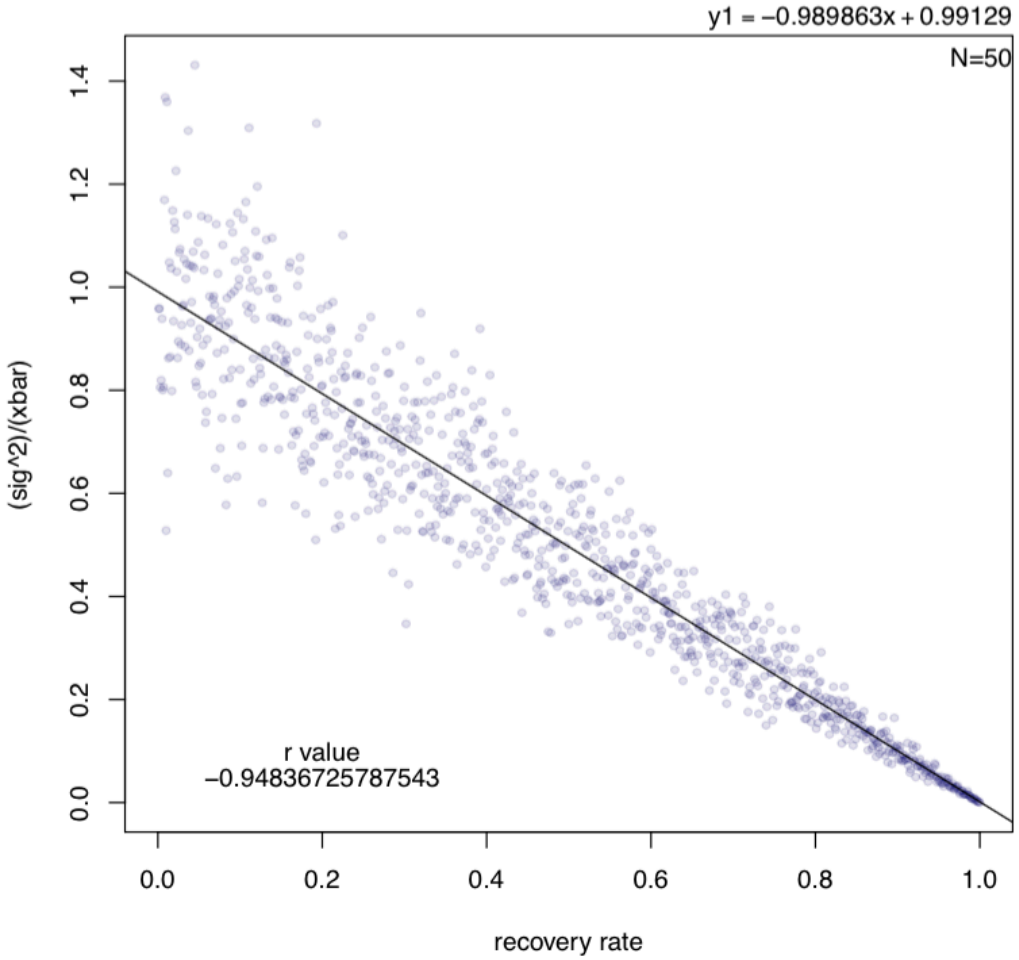

Figure 5. Variance/mean element count vs. recovery probability (N=50 individuals, recovery probability in steps of 0.1% from 0-100%).

One source of this variability is differences in attrition according to element type. Such differences may be the result of intrinsic properties of skeletal elements, such as size or shape variables, as well as the interaction of these variables with different extrinsic taphonomic

processes. For this reason, several intrinsic variables will be tested in assemblages from widely differing environments.

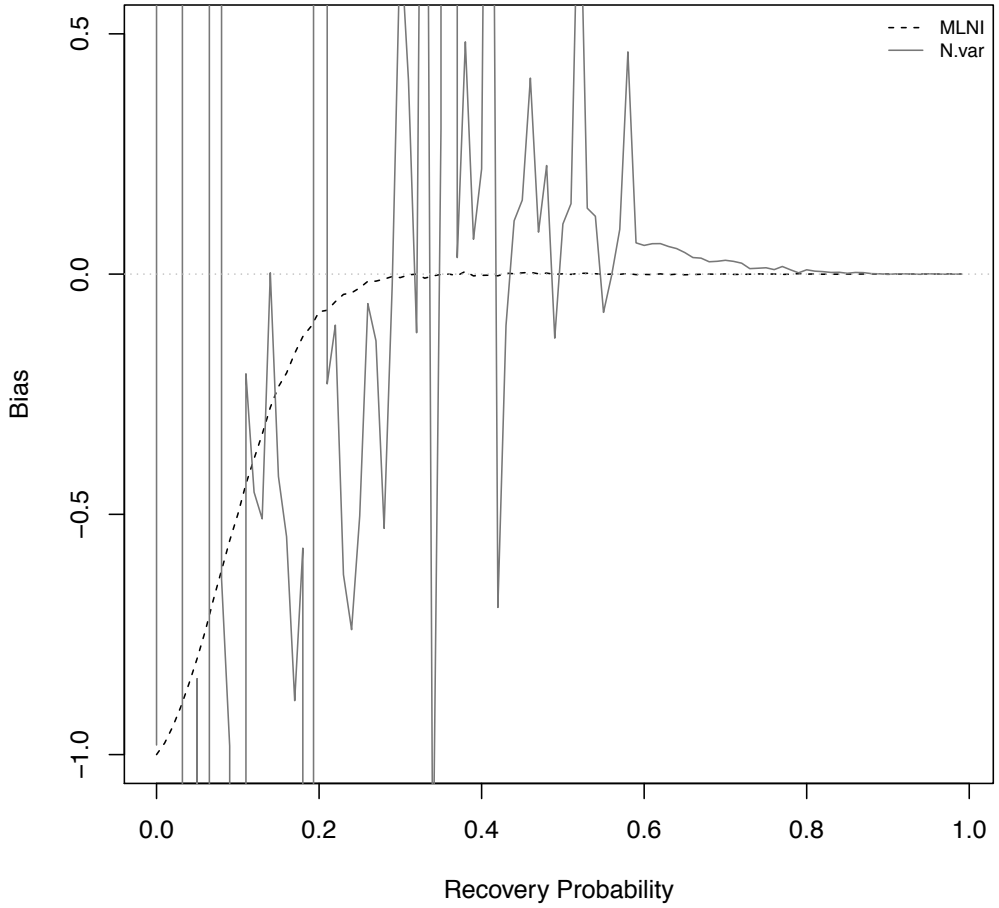

Figure 6. Percentias of MLNI and variance based estimator (N=50 individuals per assemblage, 10 element types per individual, 1000-replicate simulation per 1% step in recovery probability).

Estimating P

There is a third avenue of estimation that involves using a proxy for *P* in the original MLNI equation. If the two highest count elements are used instead of *L* and *R*, an appropriate value for *P* should fall somewhere between the third highest count element and the lowest. In fact, when the first, and second ranked counts are used for *L* and *R* in the MLNI equation and the 3rd-20th ranked values for *P* **(6)**, the average of these 18 estimates is an estimator that performs

similarly to Grand Minimum Total **(7)**, although it does show upward bias at recovery rates close to 1 (Figure 7, 'N.rank').

(6) $$N = \sum_{i}^{n-2} \left[\frac{(y_n+1)(y_{n-1}+1)}{y_{n-1-i}} - 1 \right] \cdot \frac{1}{n-1-i}$$

where $y_1 \ldots y_n$ are element counts ranked least to greatest

(7) $$N = L + R - P$$

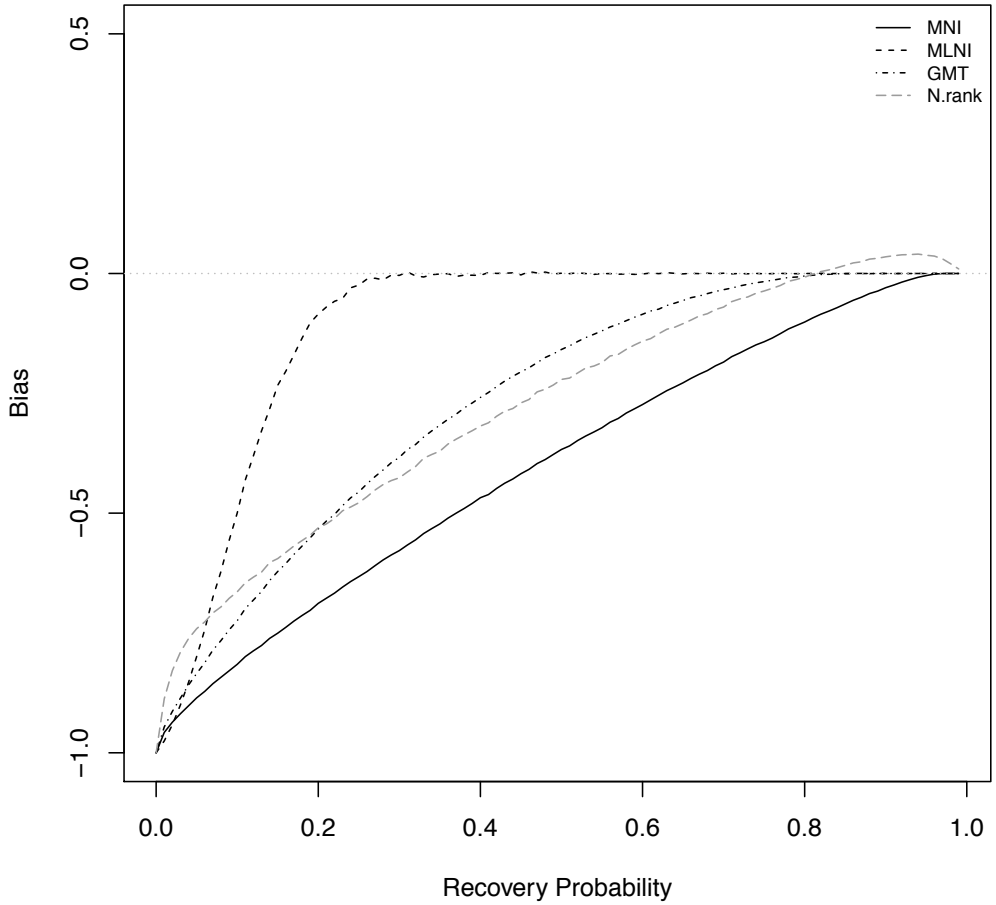

Figure 7. Percent bias of rank-based estimator compared to established methods (N=50 individuals per assemblage, 10 element types per individual, 1000-replicate simulation per 1% step in recovery probability).

Equation **6** is not a perfect estimator as it is biased at almost all recovery rates. It is encouraging, however, that an estimator that is developed strictly from counts **(6)** is able to perform similarly to one that incorporates pair-match data **(7)**. A better understanding of

variation in recovery by element type could inform the selection of appropriate counts to use in place of *P*.

Hypothesis 1

As discussed previously, the first half of this thesis will test for independence of recovery of appendicular skeletal elements. $\mathbf{H_{o1}}$ states: within the appendicular skeleton, recovery of every element is independent of every other element. The alternative hypothesis is $\mathbf{H_{a1}}$**:** dependence is correlated to anatomical proximity. Systematically testing for pairwise relationships in element recovery will allow for construction of MLNI interval estimates of a range that is supported by the data at hand.

Hypothesis 2

Estimators that use a larger proportion of assemblage data are a potential alternative to pair-match estimators in situations where pair-matching is not possible. Developing these estimators requires a better understanding of the process of element attrition, therefore the second set of hypotheses are as follows:

> **$\mathbf{H_{o2}}$: element attrition is random with respect to element characteristics.**
> **$\mathbf{H_{a2}}$: element attrition will co-vary with secondary variables, specifically, intrinsic properties of bone.**

Relating attrition to skeletal properties and taphonomic variables will provide a more robust picture of skeletal attrition. In addition to assisting with skeletal quantification, characterizing these relationships is of interest for fieldwork and sample selection for research projects.

Summary

Skeletal attrition as a process can be characterized by the hypergeometric distribution. Skeletal attrition is a complex process in which the recovery or loss of a specific skeletal element may be affected by the recovery of other elements in the skeleton, physical properties of the specific element, and the taphonomic environment. The hypotheses presented here are designed to develop an understanding of how these processes operate and interact to produce observed skeletal assemblages. The following chapter will outline the specific procedures used to test these hypotheses.

CHAPTER IV

MATERIALS AND METHODS

Introduction

This chapter presents the datasets included in this project and the methods used for testing the research hypotheses. The first section will discuss the criteria used to select skeletal regions for inclusion in the study. The second section will cover the datasets included and how they were chosen. Sections three and four will present the methods used to test the independence and recovery probability hypotheses, respectively.

Selection of Skeletal Regions

Twelve major elements of the appendicular skeleton were chosen for this study: clavicle, scapula, humerus, radius, ulna, innominate, femur, patella, tibia, fibula, talus, and calcaneus. These elements were selected based on the criteria of uniqueness within the skeleton, utility for pair-matching, and type of attachment to adjacent elements. A unique element is an element that can be identified in isolation as a single specific bone. C1 and C2, for example, are unique, however, C3 is not because its number can't be definitively assigned without comparison and articulation with additional elements.

The criterion of uniqueness excludes most of the elements of the axial skeleton as well as many of the elements of the hands and feet. While many of these elements can be

distinguished when in good condition or when found with other elements, such ideal circumstances are often not met, making it impossible to establish the exact positional number of a vertebra, rib, or even a metapodial. Knowing whether or not an element is unique is critical to quantification. If a unique element type is identified, then all further identifications of that type constitute additional individuals. Elements that cannot be uniquely identified may contribute probabilistically, but not concretely, to quantification.

Utility for pair-matching was considered because part of the purpose of this work is to establish independence relationships to improve application of existing pairmatch estimators. As such, elements of interest must be bilaterally occurring and must retain discriminating features even when preservation is imperfect. This second condition is somewhat subjective, but broadly speaking, smaller elements will have more subtle features to begin with and will tend to lose them more rapidly. For this reason, hand and foot elements, with the exception of the larger talus and calcaneus, were excluded.

Method of attachment was the final factor considered. In order to be included in the study, elements generally had to be attached to their anatomical neighbors by soft tissue, not bone, barring pathological exceptions. This excluded the cranium and mandible; despite the uniqueness and pairmatching utility of many of these elements, in adulthood they are attached to their antimeres by bone. It also necessitated the consideration of the innominate as a single bone rather than the combination of the ilium, ischium, and pubis. Although these elements are separate during development and uniquely identifiable, they are fused in adults. The separate elements of the innominate were included in Fisher's tests only, to test for the effect of attachment via bone.

Selection of Datasets

Inventory data were collected from five separate collections and datasets. The most important consideration was that individual cases not be commingled. In order to test for independence of recovery of elements within the skeleton, it had to be known that each case represented only a single skeleton. Using datasets that were not commingled also meant that the true number of individuals in each dataset was known. This allowed for calculation of the true recovery rate (relative to the actual number of individuals), rather than the relative recovery rate (relative to the MNI). Relative recovery rates are greater than or equal to true rates because the MNI is less than or equal to the true number of individuals.

Inventories from a variety of sources, including both archaeological and modern remains, were used to capture the effects of diverse taphonomic processes on skeletal assemblages. Bone mass and length data were collected from remains that had undergone minimal taphonomic change. The inventories of the Forensic Data Bank were chosen first because of the large sample size, which would not be readily available in most collections. Recovery context and measurements are not tied to inventory for this dataset, however, so four additional datasets with known recovery context and measureable characteristics were chosen for skeletal inventory collection. A sixth dataset was used for skeletal measurements only.

The inventory datasets were chosen for their wide variety of time periods and recovery contexts in the hopes of capturing as much variation of taphonomic processes as possible. The measurement dataset was chosen for its skeletal completeness and uniform taphonomy.

Modern remains

The modern collections and data sets used in this study were the Forensic Data Bank Skeletal inventories (FDB), California State University, Chico Human Identification Lab donated

forensic cases (HIL), California State University, Chico anthropology teaching collection (ATC), and New York City Office of the Chief Medical Examiner forensic anthropology cases (NYC). The FDB is a national database of forensic anthropology cases that was created in 1986 and is curated through the University of Tennessee, Knoxville. Additional cases are added each year, however the public domain version of the FDB used in this study only contains cases dated between 1962 and 1991. Each data type (inventory, measurements, etc.) is curated separately, so FDB inventories were entirely without further case information (Jantz and Moore-Jansen 2006). This necessitated the examination of data from other sources.

The HIL donated collection is a subset of the forensic cases that have come through the CSU, Chico Human Identification Lab since the 1970s. Mass and length data were collected from six of these cases. Skeletal inventories were drawn from the 22 cases used by Bright (2011) to study animal scavenging patterns. The surrounding area is primarily rural with hot, dry summers and cool, wet winters. Cases were included from the Sacramento Valley as well as the surrounding mountainous areas. Both large and small carnivores are present and active in the region (Bright 2011). The CSU, Chico ATC consists of individual skeletons that have been commercially processed for grease removal by an anatomical supply company. Elements are generally in good condition with minor changes caused by student handling over many years of use as a teaching collection.

NYC remains are curated skeletal cases from the five boroughs of New York City. The surrounding area is highly urbanized so taphonomic effects are highly variable. These effects may include city or clandestine burial followed by disinterment, fluvial transport, dismemberment, and laboratory processing of fleshed remains.

Archaeological remains

Two archaeological assemblages were used in this study: the Santa Clara Valley Medical Center historic cemetery (VMC) and the Gibson mound group from the Lower Illinois River Valley (LIV). VMC remains were discovered during expansion of the San Jose hospital campus in 2012 and excavated from 2012 to 2014. The cemetery was used prior to 1937 primarily as a pauper cemetery for patients (URS 2015). Individuals were interred in redwood coffins, which led to poor bone quality due to acid erosion from the redwood.

The Gibson mounds (11C-120) are located in Calhoun County, IL on the bluff crests overlooking the west bank of the Illinois River about 30 miles north of the confluence between the Illinois and Mississippi Rivers. This mound group was excavated in 1969 by Northwestern University and the Gilcrease Institute and has been dated to the Middle Woodland Period (50 B.C.-A.D. 400). Individuals were typically buried fully articulated or as disarticulated "bundle burials" after processing to remove flesh. Preservation at the site was generally good with some rodent disturbance (Perino et al. 2006).

Independence Testing

H_{o1}: within the appendicular skeleton, recovery of every element is independent of every other element.
H_{a1}: dependence is correlated to anatomical proximity.

The objective of the first phase was to assess independence of skeletal elements. The datasets used for this section of the project were the FDB, HIL donated forensic cases, NYC forensic anthropology cases, and LIV prehistoric burials. VMC inventories were included in initial testing for independence but not in additional testing. There were no cases with missing femora in the VMC collection, resulting in undefined correlation values for these comparisons. Anatomical relatedness of element groupings within VMC was therefore not directly comparable to other datasets.

The two forensic data sets were selected due to the difference in taphonomic environment. NYC Office of the Chief Medical Examiner forensic anthropology cases were recovered from a large metropolitan area, while the carnivore scavenged HIL cases were primarily from more rural environments. The archaeological datasets were similarly chosen for contrasts. LIV remains generally underwent minimal chemical diagenesis but were often disturbed by burrowing rodents.

All supplementary datasets were coded to match the coding of FDB cases, with 1 and 2 as present and 3 as absent. This made it possible to process all datasets using the same Perl/R script. Inventory data were processed through a Perl script to create a 2x2 contingency table for each element-to-element comparison. Each element was assessed against its antimere and to the remaining eleven elements from the same side of the body. The left humerus, for example was assessed against the right humerus as well as the left clavicle, scapula, etc. Contingency tables for the same from opposite sides (left humerus and left radius, right humerus and right radius) were added together, for a total of 120 comparisons per dataset.

Lack of depositional information for FDB cases presented a challenge. Many cases contained only 1-2 elements, did not contain postcranial elements at all, or had high proportions of missing data. With all cases included, all results were significant but it was apparent that this was more an artifact of case inclusion and data entry rather than a true reflection of dependency relationships within the skeleton. To address this issue, FDB cases with >70% missing elements (indicated as ‘3’) or any missing data (indicated as ‘9’) from the 24 bilateral elements of interest were excluded from analysis.

Cases with missing data could be easily excluded, however, excluding all cases with missing elements would negate the study. The 70% exclusion cutoff was determined graphically

by looking for points of greatest curvature of adjusted Fisher's exact p-values when plotted across ranges of missing elements. Many cases in the FDB database contained only one or two elements present with the rest coded as missing elements or missing data. A graphical test was chosen due to the large number of comparisons (n=120). The point of greatest curvature was used to indicate the point at which nearly all p-values transitioned from not significant to significant to an unreasonable degree. The resulting contingency tables from all four datasets were tested for independence using adjusted Fisher's exact tests (Holm) in R (R Core Team 2015). Holm's adjustment was used to correct for the fact that since 600 Fisher's tests were performed, 30 significant values could be expected due to random chance alone at an alpha of 0.05.

Due to the large number of comparisons, correlation values rather than p-values were tested for patterns based on anatomical relationships of element groupings. Phi was used to measure correlation from a 2x2 contingency table. Table 1 shows a hypothetical contingency table with sums of rows and columns. The equation for phi **(8)** is derived from this table according to Nyholt (2004).

Table 1. 2x2 contingency table for element comparisons.

	Bone A		
Bone B	Present	Absent	Total
Present	a	c	g
Absent	b	d	h
Total	e	f	

(8) $$\varphi = \frac{ad-bc}{(efgh)^{1/2}}$$

Phi values from each dataset were tested for normality using Shapiro-Wilk tests and quantile-quantile plots. Datasets were analyzed separately due to their very different sample sizes

(N=21-105). This separation also facilitated comparison of results between datasets. At this point, comparisons using the separate parts of the innominate were removed, for a total of 78 comparisons per dataset. Element comparisons were categorized as related (pairs, elements attached at a joint, elements on the same limb), and not related (elements that did not meet the preceding criteria). Group mean phi values were tested for significant difference using Wilcoxon rank sum tests. Patterns across datasets were assessed graphically. Unless otherwise stated, all computer scripts for both segments of this project were written by the author.

Recovery Probability

H_{o2}: element attrition is random with respect to element characteristics.
H_{a2}: element attrition will co-vary with secondary variables, specifically, intrinsic properties of bone.

The second phase of the project was concerned with element-specific recovery probabilities. This required the use of all of the inventory data included in the previous section as well as measurements of mass, length, and mineral density from other sources. Mass and length data were collected from the ATC, HIL cases, and NYC cases. Cases from HIL and NYC were selected based on completeness; the ATC was sampled in its entirety. Bone mineral density data were drawn from Kendell and Willey (2014).

Remains in the ATC were commercially degreased whereas HIL cases were processed with dermestid beetles or via boiling. NYC cases were processed using a variety of methods over a number of years including burial and hot water maceration. In order to account for such differences in processing, mass and length measurements were converted to ratios of total mass and total length. Treating each element as a ratio of the total for a single individual means that only variables that are not constant within the individual should affect the ratio.

Variables such as age, sex, and processing primarily vary between individuals and therefore should have less effect on ratio values than they do on raw data.

A pilot study was conducted on two males and three females from both the HIL and ATC collections to test for significant differences in mass ratios due to processing. To expose the highest degree of processing contrast possible, HIL cases were selected for the pilot study on the basis of minimal post-depositional change prior to defleshing. An equal number of both males and females was then sampled from ATC in letter order. The two samples were tested for significant difference using a Wilcoxon signed rank test in R.

For individuals in all three datasets, left and right element values were averaged before converting to ratios. This was done to make the study more robust to missing data and because this study is not testing differences between the right and left side of the body. If one side was unobservable, only the measurable side was used. If both sides were unobservable, the ratio could not be calculated. Only 27 individuals had adequate data for mass ratios and 29 individuals had adequate data for length ratios. Imputation, the use of existing data to estimate missing data, was performed on mass data for five individuals and length data for three individuals.

Only individuals with no more than one missing data point were included in imputation. Individuals with more than one missing element type could not be used because the remaining 11 elements had to be tested for correlation to the absent one. The three measurements that required imputation were patella mass (n=3), patella length (n=3), and femur mass (n=2). Patellae were commonly absent, particularly from ATC, and two sets of femora exceeded the weight limit of the measuring scale.

Imputation of the mean value from the same collection (ATC, HIL, or NYC) a random value from the same collection, or regression on the most highly correlated element from the same case were tested for best fit using simulation in R. The three datasets were treated separately because imputation had to be performed on raw data, not ratio data. All individuals with missing data were removed from the dataset, mean and variance for the element of interest were taken from the remaining values, and the most closely related element to the element of interest was determined using Pearson's correlation. One measurement of the element of interest was randomly removed from the dataset and imputation values were generated via the three methods.

The resulting mean and variance were collected for 10000 replicates of random removal. The mean square errors and sample variances of all three estimates of the mean were compared to test for bias. The sample mean of the means and variances were compared to the true mean and variance using a 2-sample t-test. The imputation method with the least difference in MSE and sample variance of the mean and the highest p-value for sample variance was chosen as the optimal method for that measurement.

Choosing the correct imputation method is a challenge because any method has the potential to alter the mean of the sample, the variance, or both, from the true value. Imputation of the mean has the highest potential to reduce variance through increasing the central tendency of the distribution, but will not bias the value of the sample mean from the true mean unless data loss is systematic rather than random (Figure 7, "mean"). Imputation of a random value may similarly reduce variance because the added value is not independent from the rest of the sample (Figure 7, "random"). If the relationship between values of interest and related values is weak or expressed incorrectly, imputation of values using regression may bias the mean but should not

reduce variance since the added value is independent from the rest of the sample (Figure 7, "regression").

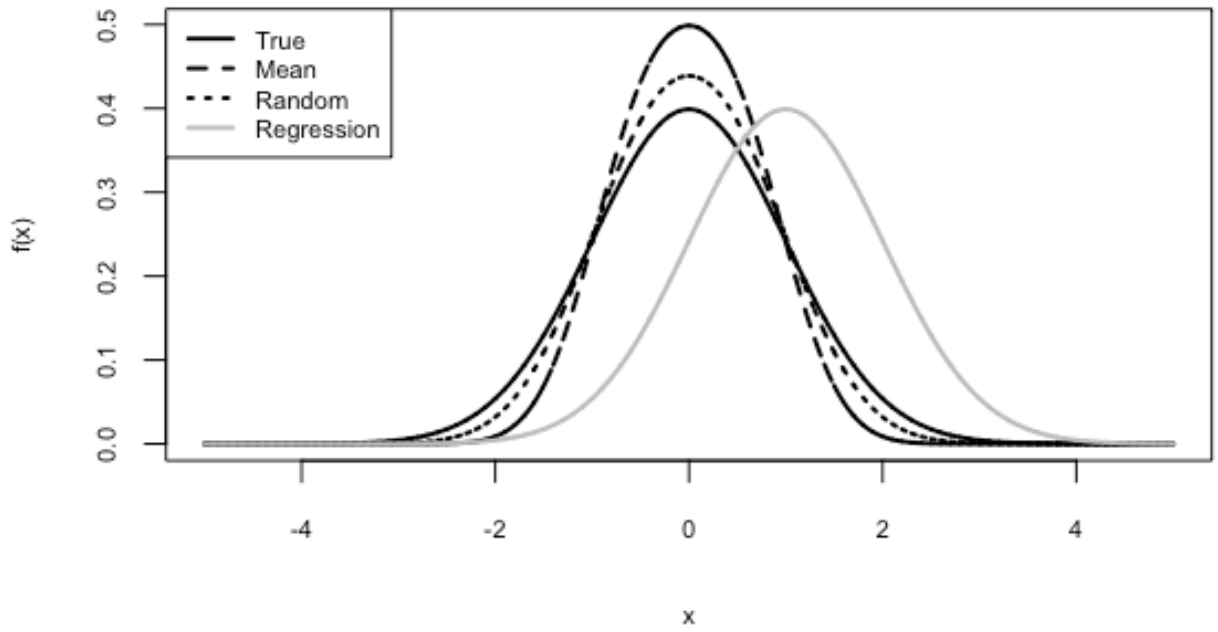

Figure 8. Hypothetical distributions for mean, random, and regression imputation methods.

Simulation was chosen over traditional leave-one-out cross validation (LOOCV) because of the need to treat missing values from each dataset differently. LOOCV systematically removes each value from the dataset and generates an estimate with each model, repeating until each value has been removed once and only once. This method would not partition the data by origin but instead treat the dataset as a whole. E.g. a random value used to estimate missing data from NYC might be drawn from either ATC or HIL values. LOOCV could be performed on each original dataset individually, but this would greatly reduce sample size and sample size between datasets would be unequal, making it difficult to compare model performance. For these reasons, a simulation procedure similar to LOOCV but allowing for customized partitioning was used instead.

To test for relationships between element recovery probabilities and properties of bone, recovery probabilities were generated for all 12 elements in each of the inventory datasets used in independence testing. Again, inventories were selected to represent a diverse range of post-depositional processes. FDB was excluded from this portion of the analysis since depositional information for this dataset is unknown. Degree of relationship between recovery probability and skeletal property was tested using Spearman's correlation in R for each inventory dataset. Bone mineral density data was only available for six elements: humerus, radius, ulna, femur, tibia, and fibula. Correlations for mass, length, and mineral density were compared to each other using data from only this subset of elements, while mass and length correlations were compared to each other using all 12 elements.

Summary

This chapter provides details of sample selection and the methods used to test each hypothesis. Sample selection included both the selection of skeletal regions as well as datasets. The second half of the chapter covered methods for testing each hypothesis, with the objective of using methods capable of falsifying either the null or alternative hypothesis. The following chapter will present the results obtained via these methods.

CHAPTER V

RESULTS

Introduction

This chapter describes the results of independence testing and recovery probability correlation tests. First the results of exclusion testing of the Forensic Databank inventories are presented, as this is relevant to both independence testing and recovery probability correlations. The next section provides the results of independence testing, including a closer look at possible patterns within and across datasets. The final section covers the recovery probability portion of this thesis, including the results of imputation optimization as well as correlation tests for recovery probabilities and intrinsic properties of bone.

Forensic Databank Exclusion Testing

Initial testing of the full set of FDB inventories (n = 1523) produced significant results in 92/120 comparisons after adjusting for number of comparisons. The most likely cause of this inflation of significant values was a large number of FDB cases with only one or two elements present. Without knowing the circumstances surrounding each case, it was not possible to deliberately exclude cases whose element composition was an artifact of reporting methods rather than taphonomy. Instead, exclusions were performed by number of elements present.

A suitable cutoff point for absent elements in the FDB inventories was estimated graphically. Contingency tables were collected from the FDB inventories for 0-28 elements missing in steps of one element. The resulting plot of *p*-values from each cycle showed a clear point of curvature in *p*-values for Fisher's exact tests with 16 of 28 elements missing (Figure 8). FDB cases with more than 16 of 28 elements of interest missing were therefore excluded from both sections of analysis. After exclusions, the final sample size for the FDB was 105 cases. Exclusion testing was not necessary for the other four datasets because all inventory data were of known origin.

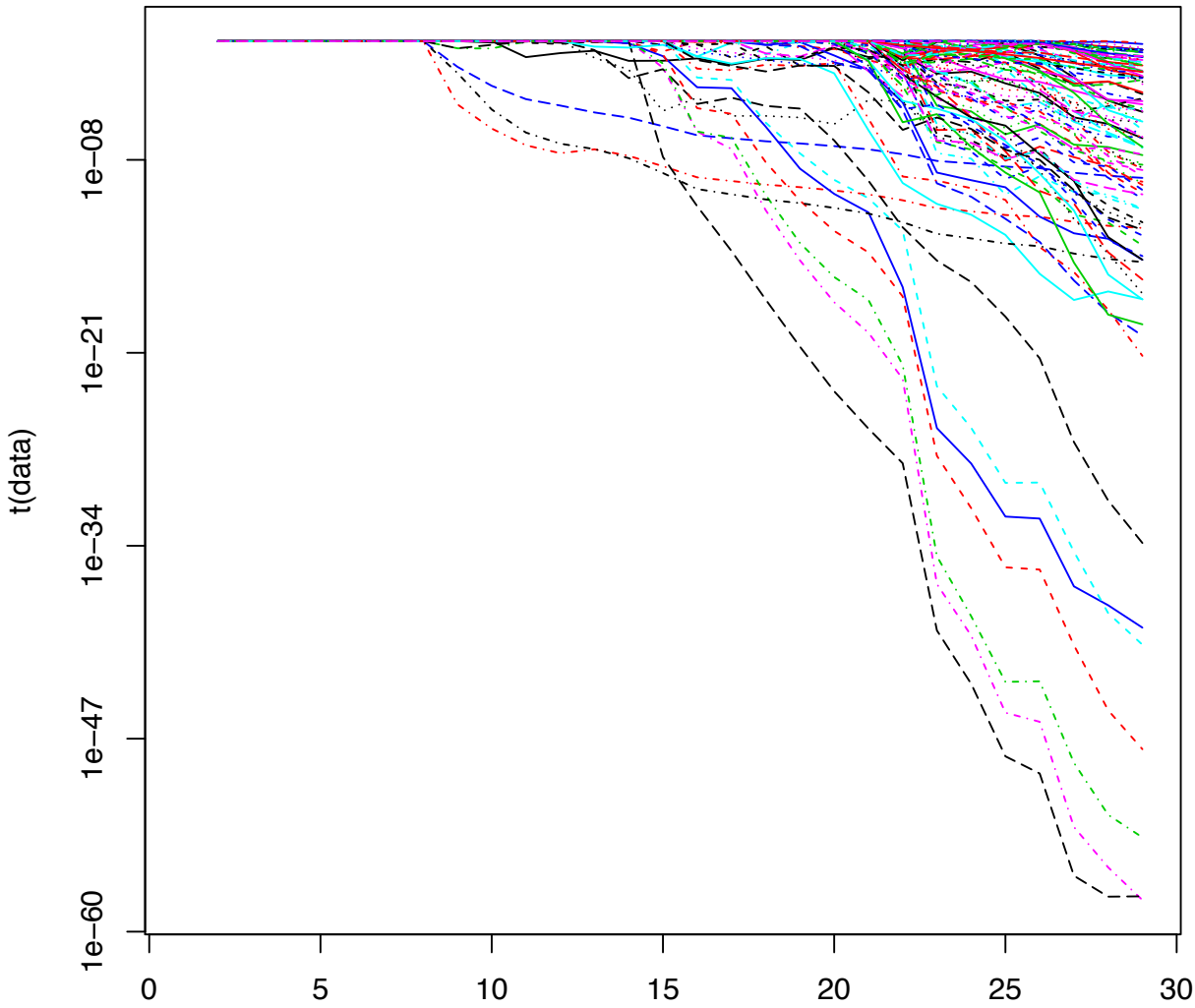

Figure 9. Decline of *p*-values after n =16 elements missing in FDB dataset, used for exclusions.

It is important to note that as a result of this exclusionary strategy, exact *p*-values become somewhat arbitrary for the FDB. The graphical test served as a means of optimizing maximum capture of relationships that were taphonomically significant against minimal capture of spurious relationships. For this reason, the second of the three drop-off points visible was chosen (Figure 9).

Unfortunately, as with any nonspecific optimization strategy, some spurious relationships were likely captured, and some significant relationships were likely lost. The subsequent use of phi values to compare strength of correlations between and within datasets should serve to ameliorate this somewhat. Unlike *p*-values, which use an arbitrary cut-off point (0.05) to determine significance or non-significance, mean phi values demonstrate continuous differences.

Independence Testing

Fisher's exact tests were used to test the first null hypothesis **H_{o1}:** within the appendicular skeleton, recovery of every element is independent of every other element. The results of these individual comparisons and their corresponding phi values are provided in Appendix A. Fisher's exact tests revealed significant relationships (n=124) in element groupings in all five datasets. This result falsifies the null hypothesis of independence (Appendix A).

The corresponding alternative hypothesis **H_{a1}:** dependence is correlated to anatomical proximity, was tested graphically and via Wilcoxon rank-sum tests. Shapiro-Wilk tests showed that phi values from HIL and LIV potentially followed a normal distribution ($p = 0.5671$ and $p = 0.0907$) while FDB and NYC did not ($p < 0.001$ and $p = 0.0115$, Table 2). These results were supported by histograms and quantile-quantile plots of phi values (Figure 10).

Table 2. Summary statistics and Shapiro-Wilk results for phi values of each dataset.

Dataset	Min.	1st Qu.	Median	Mean	3rd Qu.	Max	Shapiro-Wilk	
							W	p
FDB	-0.0784	0.0944	0.1416	0.1791	0.2229	0.8733	0.7861	<0.001
HIL	0.0000	0.2680	0.3845	0.3944	0.5173	0.8575	0.9862	0.5671
NYC	0.0555	0.2380	0.3850	0.4008	0.5123	0.9376	0.9580	0.0115
LIV	0.0676	0.2687	0.3328	0.3621	0.4782	0.7604	0.9726	0.0907

The histogram for FDB is right-skewed and the Q-Q plot shows a large deviation from linear, supporting that the phi values for this dataset are not normally distributed (Figure 10). The histogram for NYC is slightly bimodal and the Q-Q plot deviates somewhat from linear, also supporting that this dataset does not follow a normal distribution (Figure 10).

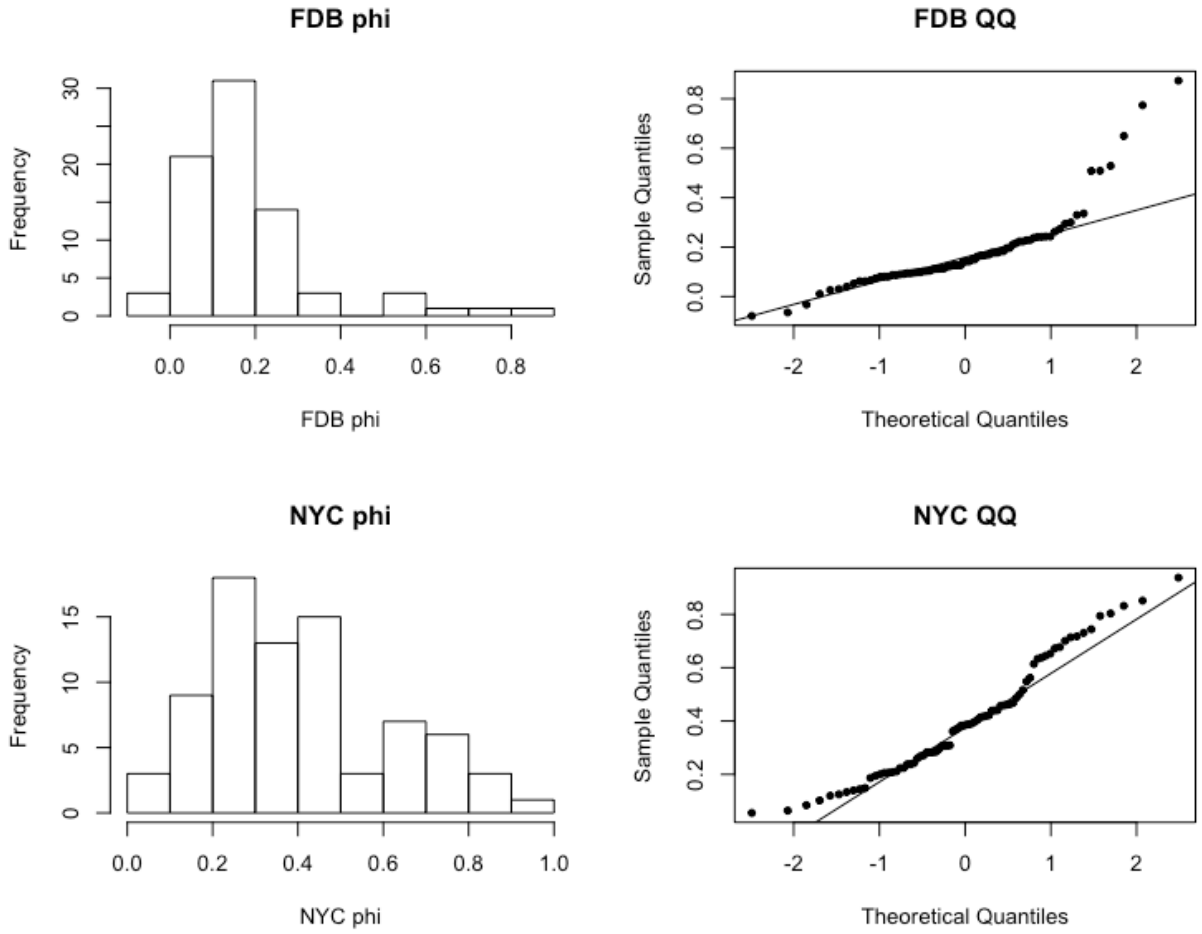

Figure 10. Histograms and Q-Q plots for FDB and NYC.

Histograms for both HIL and LIV are largely symmetrical and their Q-Q plots are generally linear, supporting that these two datasets have normally distributed phi values (Figure 11). Since two out of four datasets had non-normally distributed phi values, Wilcoxon signed-rank tests rather than *t*-tests were used to test for significant differences between levels.

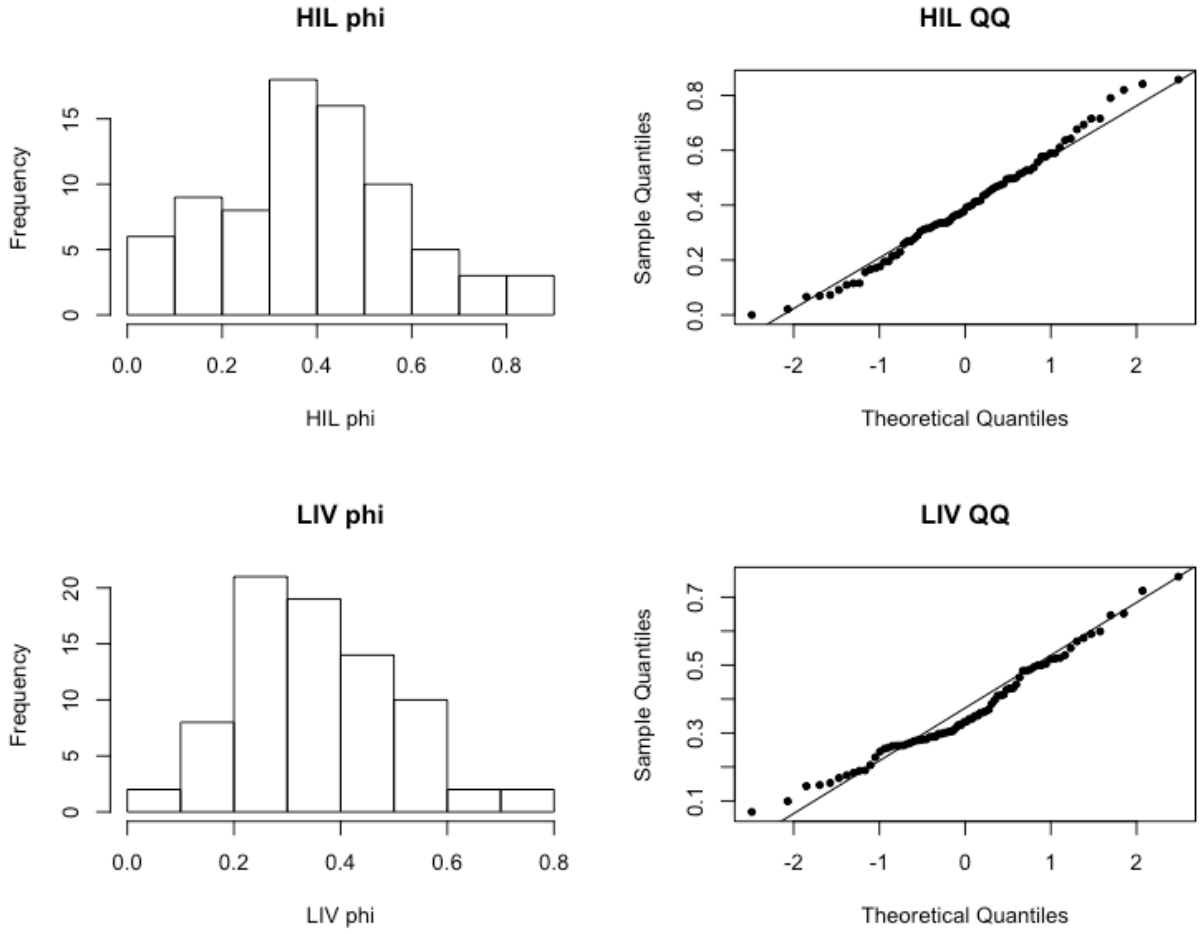

Figure 11. Histograms and Q-Q plots for HIL and LIV.

Wilcoxon rank sum tests of group mean phi values showed significant differences ($p < 0.05$) between related (pair, joint, or same limb) and unrelated (opposite limb) element groupings for all datasets (Table 3). Displaying these results graphically reveals trends in directionality of differences (Figure 12). Related element groupings had higher mean phi values for all datasets (Figure 12), even though the difference in means was not significant in all cases.

Table 3. Wilcoxon rank sum tests for related (R) vs. unrelated (U) element groupings.

Dataset	W	p	Mean phi		phi SE	
			R	U	R	U
FDB	963	0.0349	0.2205	0.1282	0.0304	0.0120
HIL	1023.5	0.0066	0.4552	0.3196	0.0324	0.0264
NYC	1158	<0.001	0.4914	0.2894	0.0337	0.0230
LIV	978	0.0238	0.3988	0.3170	0.0238	0.0198

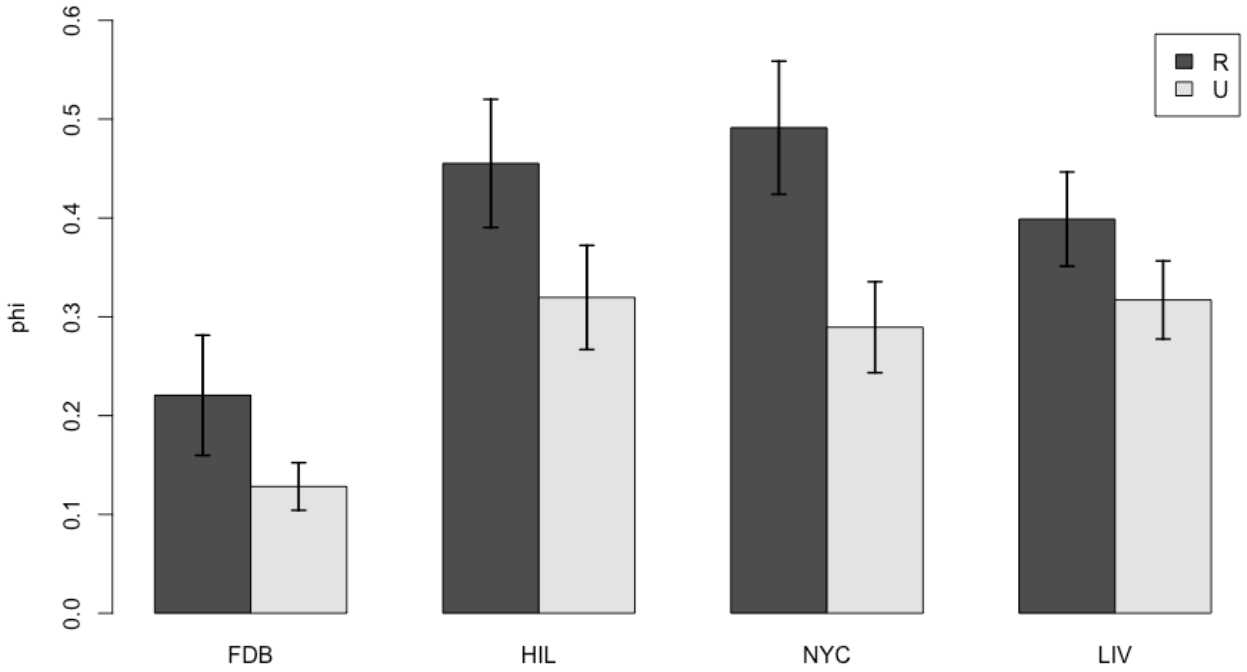

Figure 12. Comparison of group mean phi values for Related (paired elements, elements attached at a joint, or elements on the same limb) vs. Unrelated (upper vs. lower limb) categories of element groupings, with standard error bars.

When related elements were subdivided into "Pair," "Joint," and "Limb" categories to examine the effect of anatomical proximity more closely, results were mixed (Table 4, Figure 13). No datasets showed significant difference between Pair and Joint (Table 4, column "PJ"), all but FDB showed significant differences between Pair and Limb (Table 4, column "PL"), and

NYC and LIV showed significant differences between Joint and Limb while FDB and HIL did not (Table 4, column "JL"). Graphically, paired elements consistently show the highest mean phi values, followed by elements attached at a joint (Figure 13). Elements on the same limb not sharing a joint had the lowest means for all datasets (Table 5). Overall, in order from most correlated to least correlated, groupings were paired, joint, limb, and not related (Figure 13).

Table 4. Wilcoxon rank sum tests for P, J, and L

Dataset	PJ W	PJ p	PL W	PL p	JL W	JL p
FDB	87	0.8977	120	0.4445	141	0.3934
HIL	103	0.3412	153	0.0253	156	0.1493
NYC	88	0.8596	169	0.0022	182	0.0116
LIV	88	0.8596	148	0.0426	174.5	0.0290

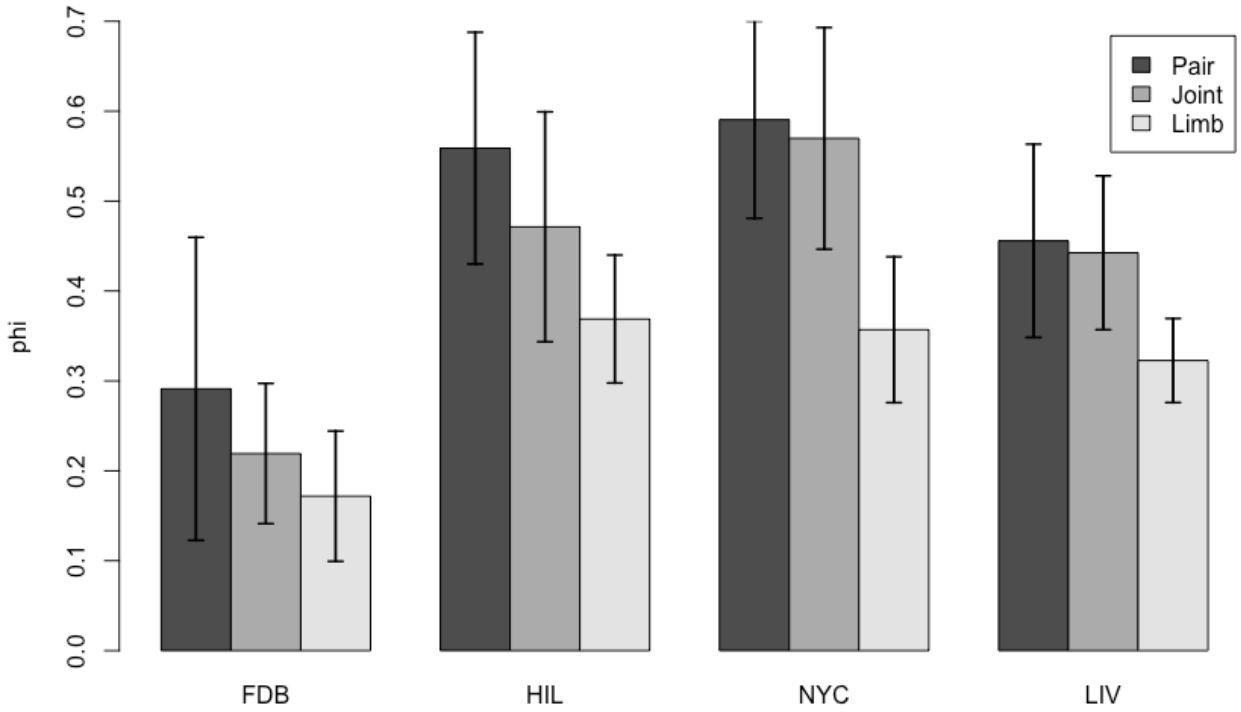

Figure 13. Comparison of group means for pair, joint, and limb categories of element groupings, with standard error bars.

Table 5. Group means and standard errors (SE) for pair, joint, and limb groupings.

Dataset	P mean phi	P phi SE	J mean phi	J phi SE	L mean phi	L phi SE
FDB	0.2912	0.0843	0.2192	0.0390	0.1718	0.0362
HIL	0.5589	0.0644	0.4714	0.0639	0.3688	0.0356
NYC	0.5905	0.0548	0.5698	0.0616	0.3569	0.0406
LIV	0.4558	0.0537	0.4425	0.0427	0.3227	0.0234

Recovery Probability

This section details the results of procedures used to test the second hypothesis **H_{o2}:** element attrition is random with respect to element characteristics, and its corresponding alternative **H_{a2}:** element attrition will co-vary with secondary variables, specifically, intrinsic properties of bone. This includes the imputation procedures tested and employed to achieve adequate sample sizes for mass and length measurements as well as the results of correlation testing between recovery rates and bone mass, length, and mineral density.

Imputation

The most highly correlated element measurements were tibia mass to both femur mass (Figure 14) and patella mass (Figure 15) and calcaneus length to patella length (Figure 16).

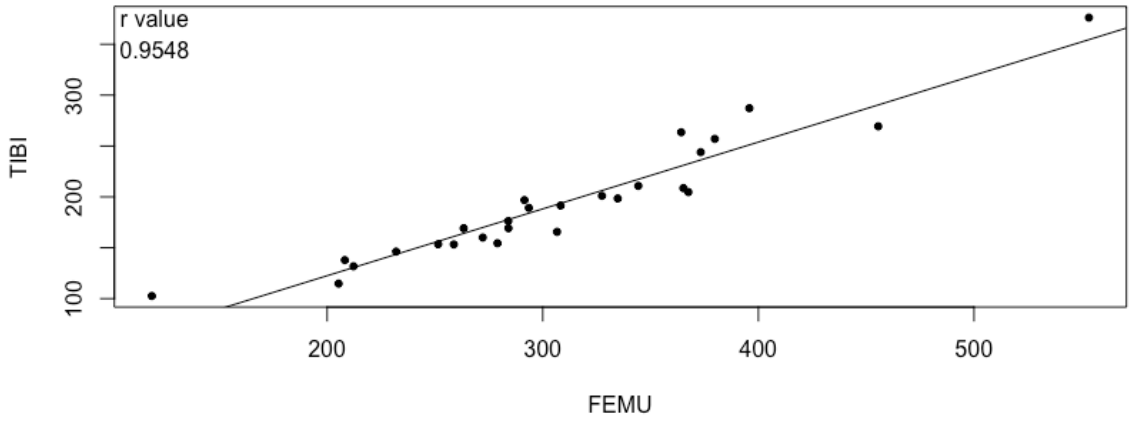

Figure 14. Correlation between femur and tibia mass.

For imputation of patella mass, the mean, random, and regression imputation methods produced a sample mean and sample variance that were not significantly different from the target

values. This showed that these imputation strategies did not significantly alter the parameters of interest from the true values.

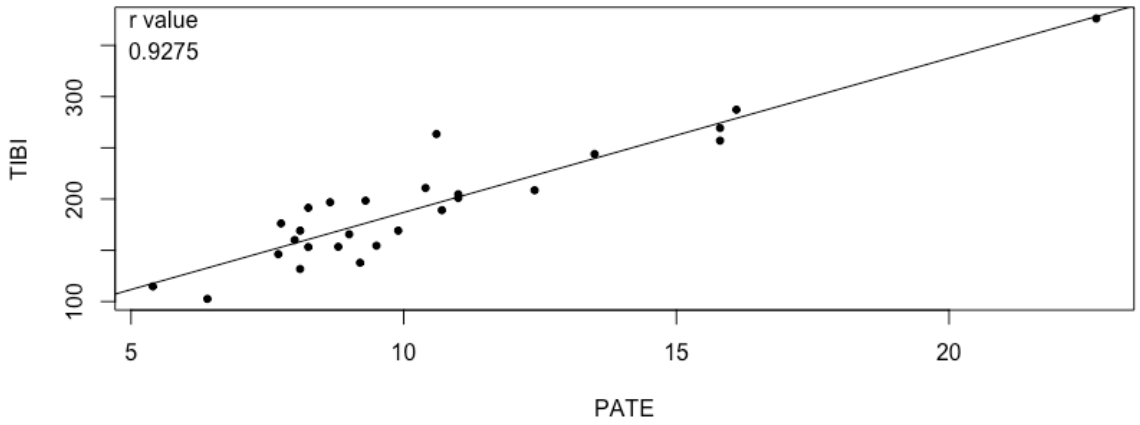

Figure 15. Correlation between patella and tibia mass.

Regression had the smallest difference between mean-square error (MSE) and variance of the sample mean and was therefore chosen as the optimal imputation strategy (Table 6). MSE is composed of the sum of the variance and the squared bias of the variable, so for an unbiased estimator, MSE and variance will be the same. Therefore, the difference between MSE and variance were used to indicate bias of the sample mean.

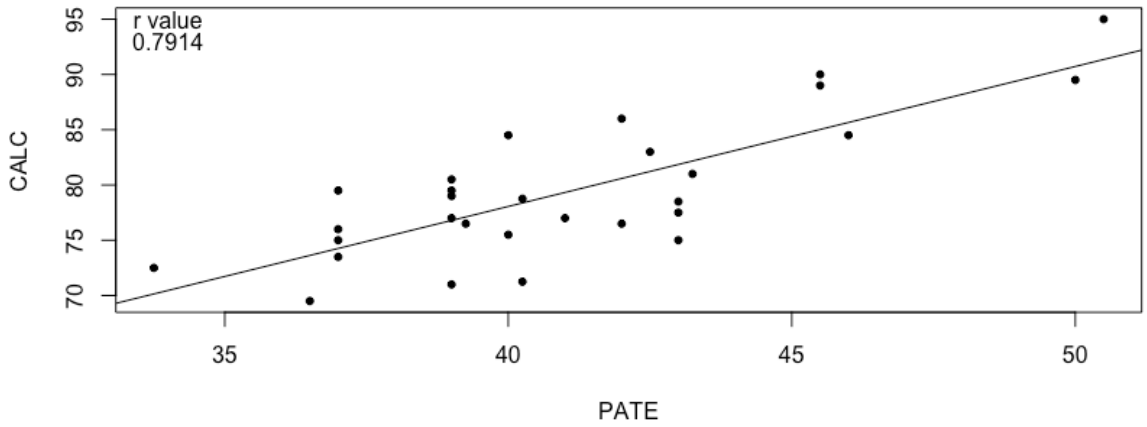

Figure 16. Correlation between patella and calcaneus length.

For imputation of patella length, again, all three methods produced a sample mean not significantly different from the target value. Sample variance was significantly different for imputation using the mean, showing a downward bias. As discussed in chapter 4, imputation of the mean artificially increased the central tendency of the distribution. Regression showed a smaller difference than random imputation between MSE and variance of the sample mean, indicating reduced bias. This made regression the optimal strategy for patella length as well (Table 7).

Table 6. Patella mass imputation results for imputation method selection.

Parameter	True value	Mean	Random	Regression
Mean	10.456	10.456	10.440	10.456
Variance	13.410	12.999	13.345	13.338
MSE		0.019246	0.029878	0.0024762
Sample variance		0.019987	0.030762	0.0025714
p value: mean		1	0.63985	1
p value: variance		0.082057	0.84346	0.34061

Table 7. Patella length imputation results for imputation method selection.

Parameter	True value	Mean	Random	Regression
Mean	41.043	41.043	41.073	41.043
Variance	15.036	14.518	15.205	14.836
MSE		0.022303	0.044672	0.0064502
Sample variance		0.023099	0.045315	0.0066805
p value: mean		1	0.44947	1
p value: variance		0.0023342	0.50864	0.058647

For imputation of femur mass, both mean and random imputation produced sample means and variances that were significantly different from the target values. Sample mean had downward bias and sample variance had upward bias. Regression was the only method that

produced a sample mean and variance not significantly different from the target values and was chosen as the optimal imputation strategy (Table 8).

Table 8. Femur mass imputation results for imputations method selection

Parameter	True value	Mean	Random	Regression
Mean	308.53	297.49	297.49	308.53
Variance	7465.3	10469	10470	7440.0
MSE		131.60	131.56	0.87151
Sample variance		10.104	10.027	0.90503
p value: mean		3.16E-16	2.88E-16	1
p value: variance		9.03E-21	1.12E-20	0.51763

Recovery probability and intrinsic property correlations

LIV had the highest mean recovery rate at r = 0.7986, followed by NYC (r = 0.7586), VMC (r = 0.7482), and HIL (r = 0.6381, Table 9). Violation of normality was graphically apparent in histograms of recovery rates for each dataset. VMC showed a clearly separated bimodal distribution. LIV and NYC were also bimodal but with less separation. HIL appeared more uniform (Figure 17). Based on a Shapiro-Wilks test, recovery rates were consistent with a normal distribution for HIL (W = 0.9337, p = 0.06146). Recovery rates were not normally distributed for VMC (W = 0.7010, p < 0.001), NYC (W = 8586, p < 0.001), or LIV (W = 0.8975, p = 0.007303, Table 9).

Table 9. Shapiro-Wilk tests for normality of recovery rate distributions.

Dataset	Mean r	Variance	Shapiro-Wilk	
			W	p
VMC	0.7482	0.0205	0.7010	1.61E-06
NYC	0.7586	0.0290	0.8586	0.0009448
HIL	0.6381	0.0318	0.9337	0.06146
LIV	0.7986	0.0077	0.8975	0.007303

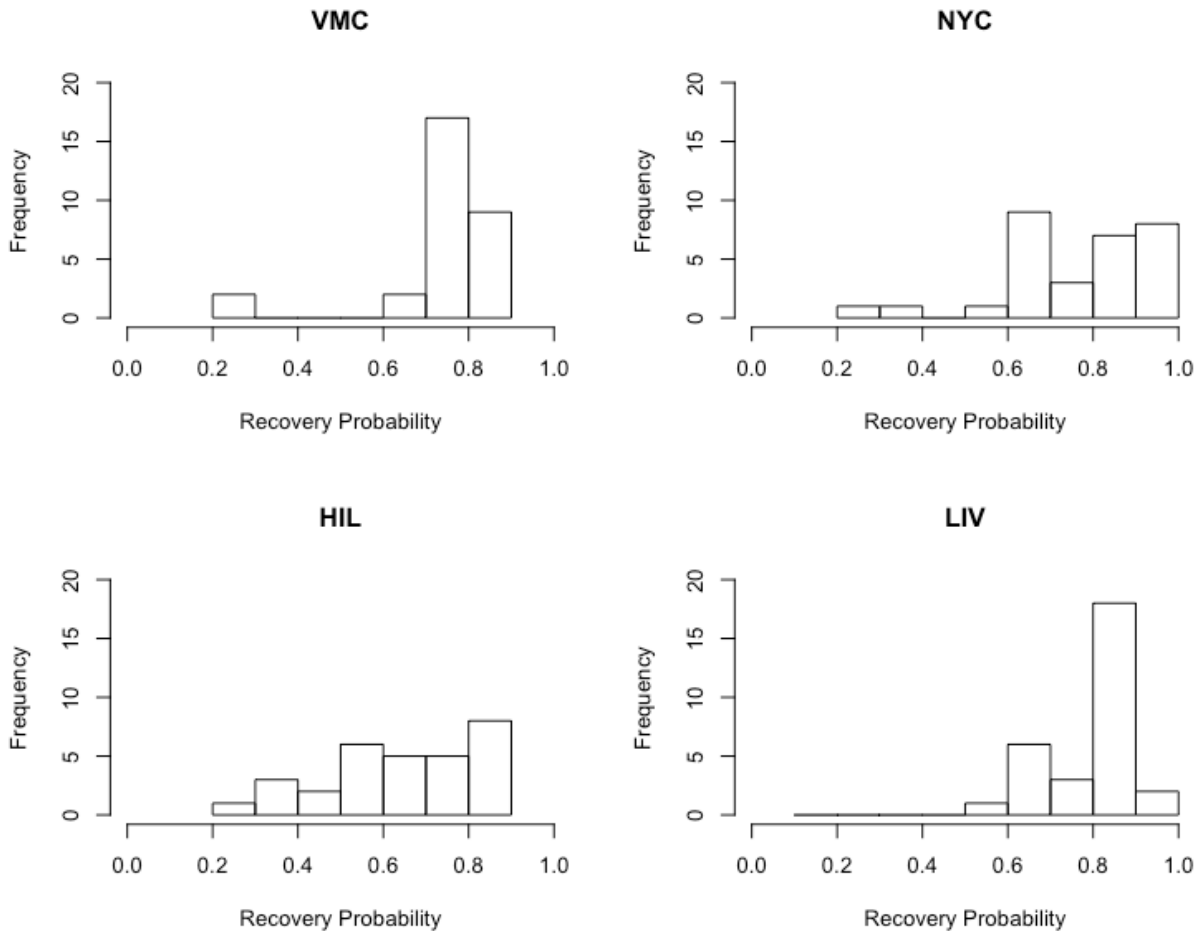

Figure 17. Histograms of recovery probabilities for all datasets.

Results from the Shapiro-Wilks tests were supported by normal quantile-quantile (Q-Q) plots (Figure 18). The Q-Q plot for HIL showed the linear pattern expected for normally distributed data, while the other datasets had Q-Q plots with large deviations from linearity (Figure 18). Since normality was violated in most cases, Spearman's rho was used in place of Pearson's r for correlation tests.

Spearman's correlation tests were performed between element recovery probabilities for each datasets and the specific intrinsic properties of those elements (mass, length, and mineral density). Intrinsic properties are expressed as a ratio of the measurement for the element and the sum of all measurements for the same property. This controls for interpersonal variation.

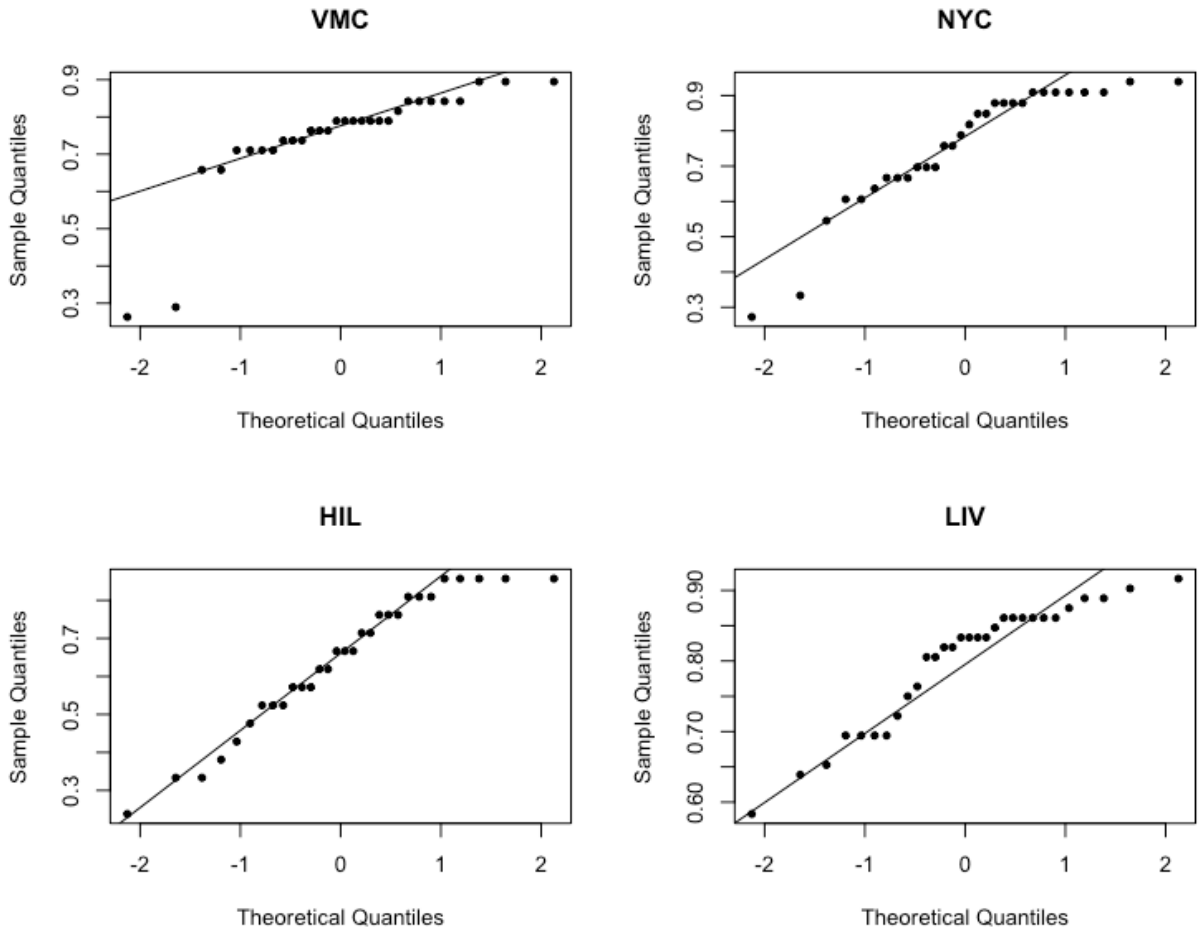

Figure 15. Normal Q-Q plots of recovery rates for VMC, NYC, HIL, and LIV datasets.

Testing of all 12 elements (Table 10) showed that mass was significantly correlated to recovery for NYC (rho = 0.6993, p = 0.0145, Figure 19), HIL (rho = 0.7075, p = 0.0101, Figure 20), and LIV (rho = 0.779, p = 0.0028, Figure 21). Length showed significant correlation for VMC (rho = 0.655, p = 0.0208, Figure 22) and LIV (rho = 0.7404, p = 0.0059, Figure 21).

Table 10. Spearman's results for full inventories.

Dataset	Mass		Length	
	rho	p	rho	p
VMC	0.5604	0.0581	0.655	0.0208
NYC	0.6993	0.0145	0.5035	0.0988
HIL	0.7075	0.0101	0.5009	0.0972
LIV	0.779	0.0028	0.7404	0.0059

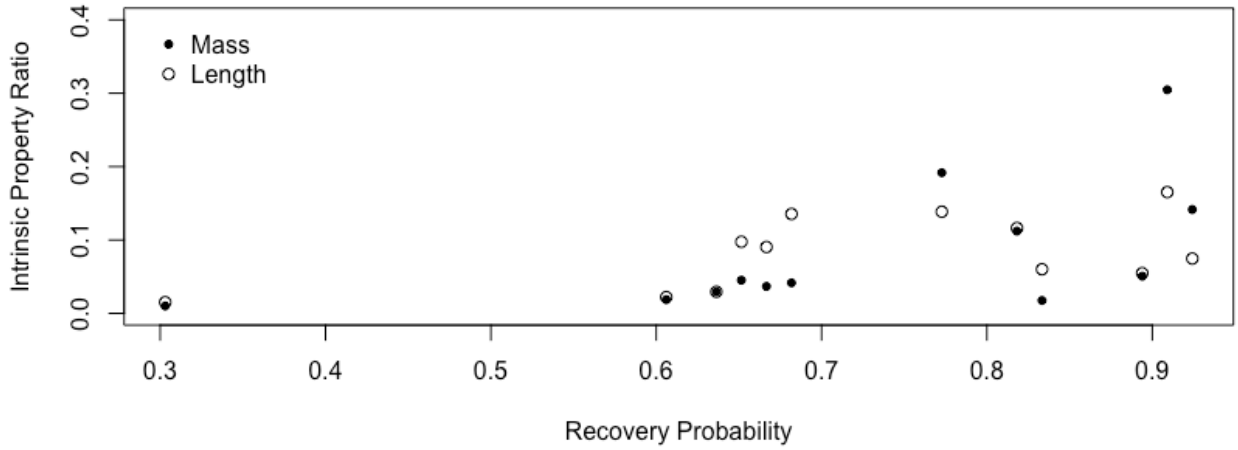

Figure 19. Intrinsic property values vs. recovery probabilities for full NYC inventory (n = 12).

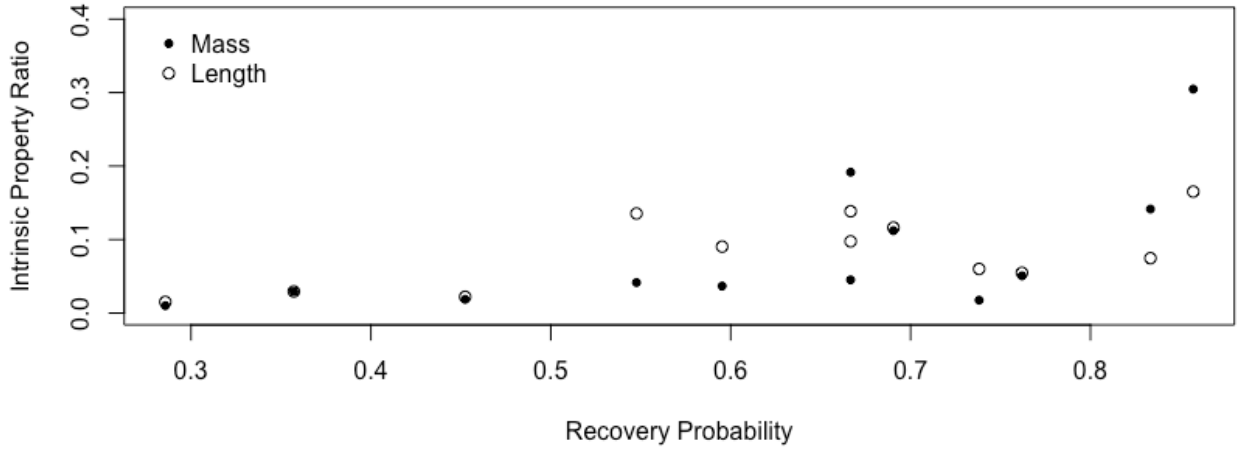

Figure 20. Intrinsic property values vs. recovery probabilities for full HIL inventory (n = 12).

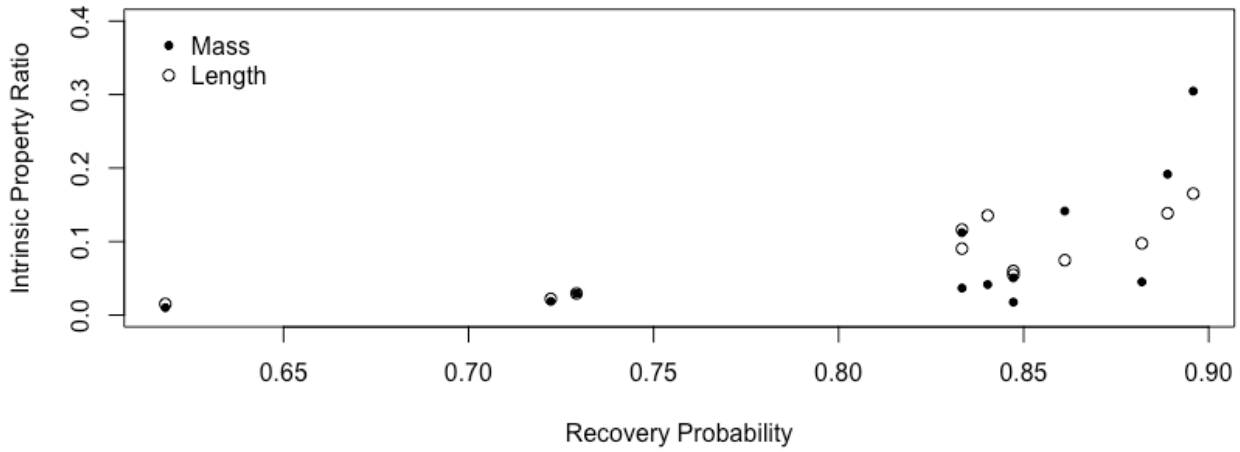

Figure 21. Intrinsic property values vs. recovery probabilities for full LIV inventory (n = 12).

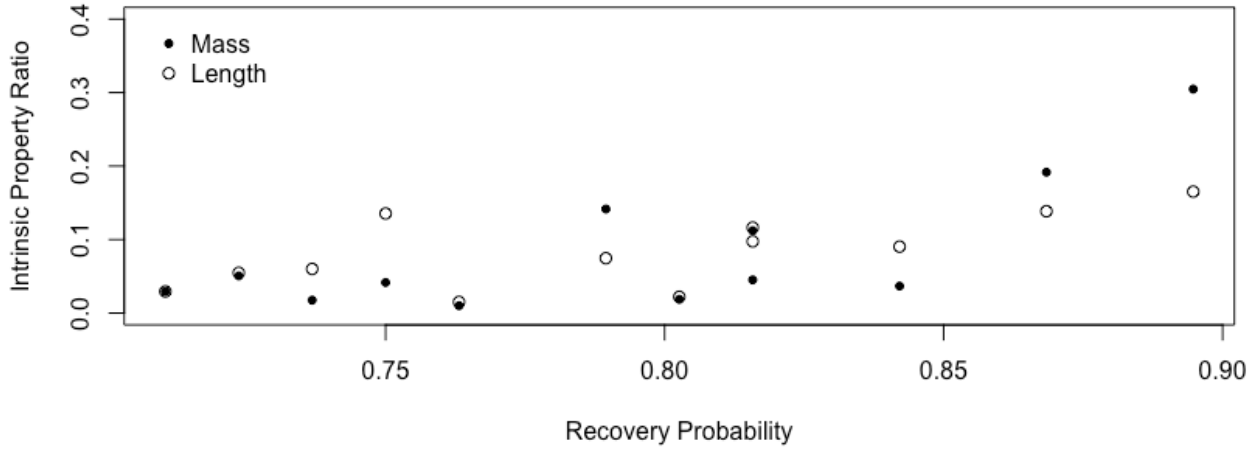

Figure 22. Intrinsic property values vs. recovery probabilities for full VMC inventory (n = 12).

For testing of the six major long bones only (Table 11), mass was significantly correlated for HIL (rho = 0.8407, p = 0.0361, Figure 23). Mineral density was significantly correlated for VMC (rho = 0.8117, p = 0.0499, Figure 24) and HIL (rho = 0.8986, p = 0.0149, Figure 23). Length did not show a significant correlation to recovery rate for any of the long bone subsamples. LIV and NYC subsamples did not show significant correlations with mineral

density (rho = 0.6667, 0.7143, p = 0.1481, 0.1361), mass (rho = 0.7537, 0.7714, p = 0.0835, 0.1028), or length (rho = 0.7827, 0.7714, p = 0.0657, 0.1028, Table 11, Figures 25 and 26).

Table 11. Spearman's results for long bones (n = 6).

Dataset	BMD		Mass		Length	
	rho	p	rho	p	rho	p
VMC	0.8117	0.0499	0.6377	0.1731	0.4638	0.3542
NYC	0.7143	0.1361	0.7714	0.1028	0.7714	0.1028
HIL	0.8986	0.0149	0.8407	0.0361	0.4638	0.3542
LIV	0.6667	0.1481	0.7537	0.0835	0.7827	0.0657

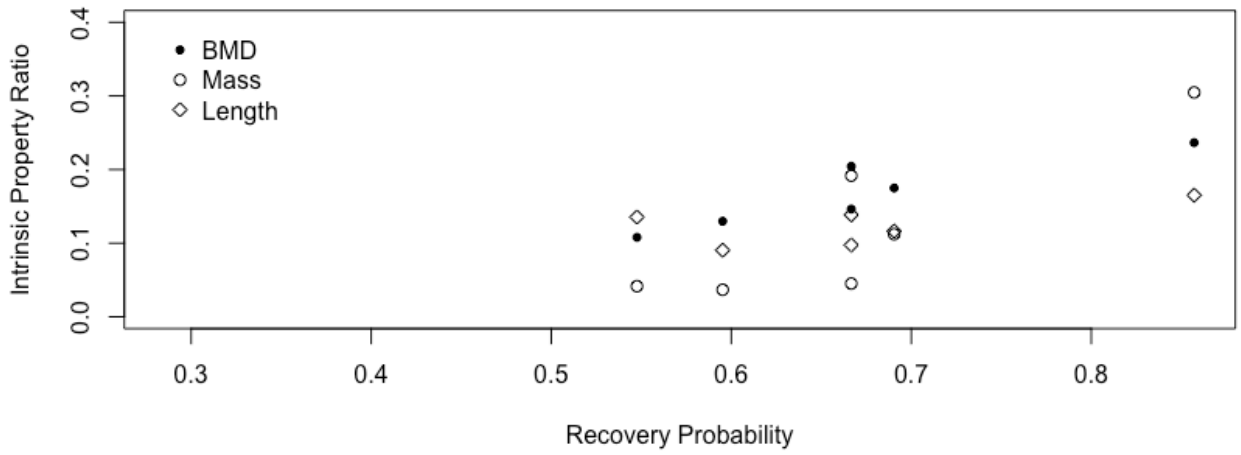

Figure 23. Intrinsic property values vs. recovery probabilities for HIL long bones (n = 6).

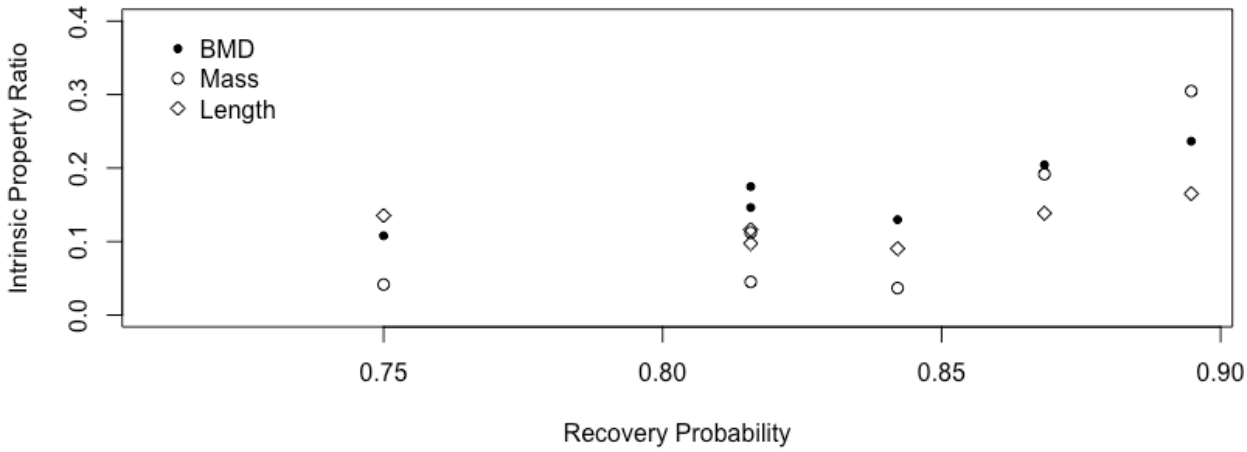

Figure 24. Intrinsic property values vs. recovery probabilities for VMC long bones (n = 6).

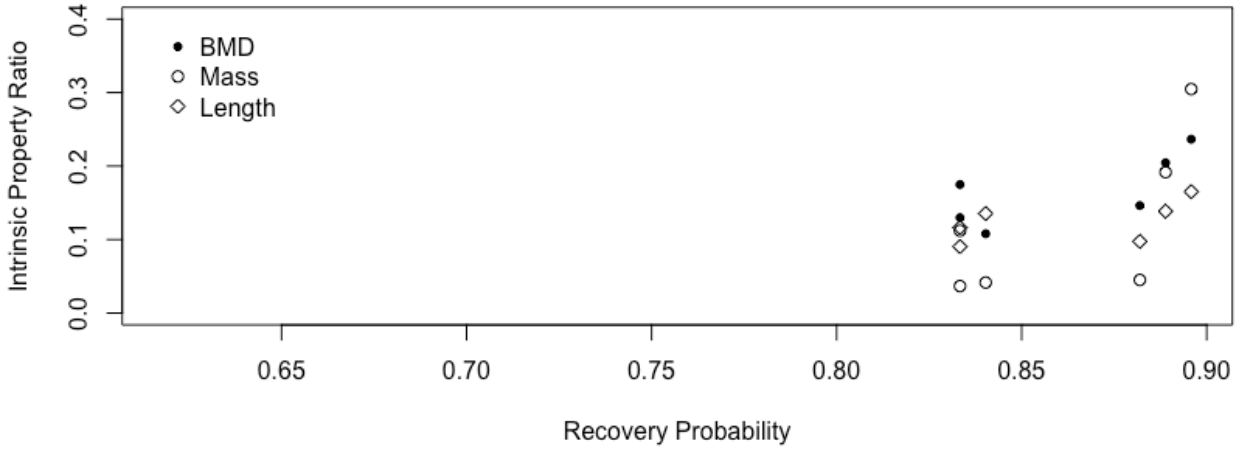

Figure 25. Intrinsic property values vs. recovery probabilities for LIV long bones (n = 6).

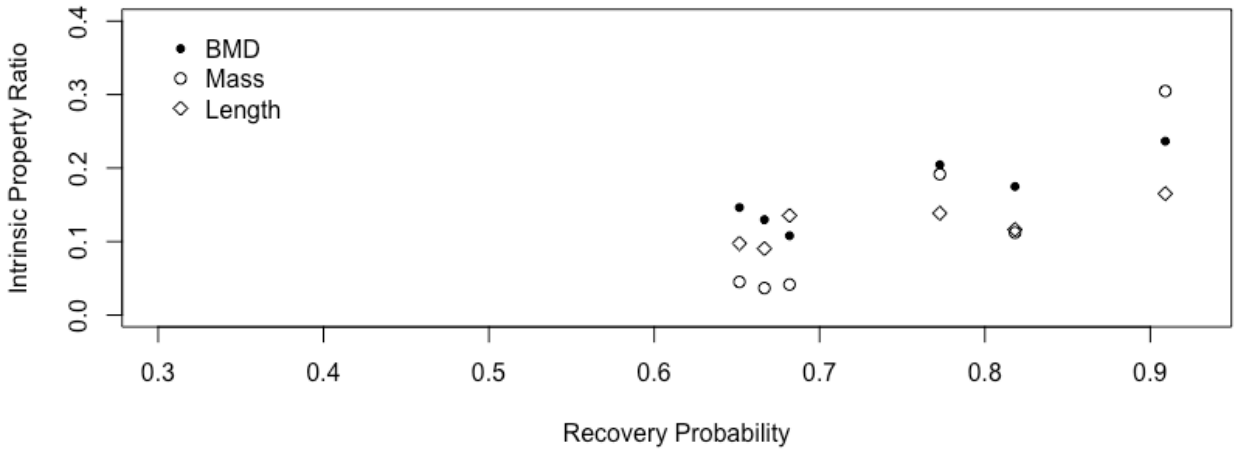

Figure 26. Intrinsic property values vs. recovery probabilities for NYC long bones (n = 6).

Summary

Out of a total of 600 Fisher's exact tests performed to test hypothesis 1, 124 showed significant values, indicating a lack of independence of these element groupings. Phi values were not normally distributed for FDB and NYC, necessitating non-parametric testing for differences in group means. Group mean phi values were consistently significantly higher for related elements than for unrelated elements. Within the related category, paired elements had the highest mean, followed by elements attached at a joint, with elements on the same limb but not attached at a joint having the lowest mean. Differences in paired and joint group means were not significant. Paired and limb group means were significantly different for all but FDB, and joint and limb means were significantly different for NYC and LIV.

For testing hypothesis 2, element-specific recovery rates deviated from the normal distribution for all but HIL, necessitating the use of Spearman's correlation. When all 12 elements were included, mass significantly affected recovery probability for NYC, HIL, and LIV, and length had a significant effect on VMC and LIV. When only the six long bones were tested, mass had a significant effect on HIL and mineral density had a significant effect on VMC and HIL Interpretation of these results will be covered in the next section.

CHAPTER VI

DISCUSSION

Introduction

This chapter discusses the results of the independence testing and recovery-property correlations. Results from each section are considered against their corresponding null and alternative hypotheses. Independence test results are compared both anatomically and by dataset. Reasons for imputation results are examined in light of the collections used and the mathematical properties of the different strategies tested. Correlations between element properties and recovery are discussed, with consideration of the taphonomic environment of each collection. Results from both sections are examined together for ways in which element recovery rates and properties might affect independence. Lastly, potential limitations and biases are identified and discussed.

Independence Testing

H_{o1}: within the appendicular skeleton, recovery of every element is independent of every other element.
H_{a1}: dependence is correlated to anatomical proximity.

Element grouping variability

Fisher's exact tests showed that for 124 out of 600 element comparisons, recovery was not independent. Within each of these two-element groupings, recovery of one element affected the likelihood of recovering the other. The null hypothesis is therefore rejected

(Appendix A). Evidence for the alternative hypothesis was ambiguous. While there was a significant difference in all datasets between related and unrelated groups, when the related group was broken down into subcategories, results did not conclusively correlate directly with anatomical proximity.

Anatomically, antimere pairs are more distant than elements attached at a joint or elements on the same limb; elements attached at a joint are the closest. The fact that paired element groupings had the highest mean phi values contradicts hypothesis 1. In contrast, the trend of joint groupings having higher phi values than limb groupings supports the hypothesis, as does the trend of larger phi values in the overall category of related element groups than in the overall category of unrelated element groups.

Despite this interesting contradiction, the results of independence testing are informative for selecting elements for inclusion in pair-match estimation of number of individuals. The data show that an optimal collection of elements is one that maximizes anatomical distance. If two elements are used, it would be best to use one from the lower limb and one from the upper limb. If more than two elements are used, ideally, elements from the same limb should not share a joint.

Based on the individual values from Fisher's tests, elements that should not be included together even if inclusion of elements attached at a joint cannot be avoided are the ulna and radius, and calcaneus and talus. These groupings had significant *p*-values with phi values above 0.5 for at least four out of five datasets. Results for the parts of the innominate indicate that elements attached by bone should also not be included together. Ilium-ischium, ilium-pubis, and pubis-ischium showed significant results for at least three of five datasets.

Recovery Probability

Imputation

The results of imputation strategy testing were not surprising, particularly for patella length and femur mass. Patella mass was not overly sensitive to the method used. While regression appeared to be the best method, all three methods produced accurate representations of the target values, making all of them valid strategies.

Mean imputation was not a valid strategy for patella length due to downward bias of the sample variance. Imputation of the mean will increase the size of the center of the sampling distribution relative to the tails. Imputation of a random value is less vulnerable to this problem because while it does not increase the spread of the data, the value may occur anywhere along the data range. Regression does not face this issue because regression generates values that are not dependent on the rest of the dataset; degrees of freedom are accurately represented and variance is not reduced.

Regression was the only imputation method that accurately represented the target values for femur mass. Femur mass varied widely within the sample and seemed to be greatly affected by overall individual size and post-depositional factors such as processing. Only imputation of regression values was able to control for the high level of interpersonal variation in this measurement. Overall, regression showed the best performance in all three cases.

Recovery probability and intrinsic property correlations

H_{o2}: element attrition is random with respect to element characteristics.
H_{a2}: element attrition will co-vary with secondary variables, specifically, intrinsic properties of bone.

Results of correlation testing showed that element recovery was random with respect to certain element characteristics in certain cases but not others. Mineral density was significantly correlated with recovery probability for two of the four datasets (Figure 27),

indicating that it is an important factor in recovery of remains from diverse taphonomic contexts. In the long bone subset, length showed no significant correlation with recovery for any dataset, while in the full 12-element group, length was significantly correlated with recovery probability for two of the four datasets (Figure 27).

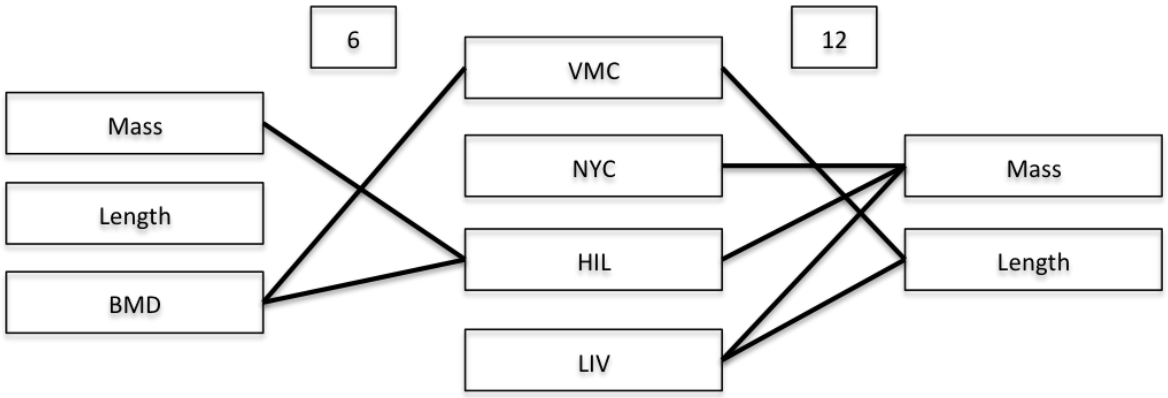

Figure 27. Significant relationships between recovery probability and intrinsic property by dataset. Results for the six long bones are shown on the left and all 12 elements on the right.

These results make sense considering that length is much more variable in the full group than in the long bone subset. The longest and shortest long bones are the femur and radius respectively, compared to the longest and shortest elements overall, the femur and patella. The standard deviations in mean length were 131.2 mm for all 12 elements and 74.9 mm for long bones only.

Each dataset showed a unique pattern of which properties had a significant effect on skeletal attrition (Figure 27). Mass showed significant correlations with HIL long bones and NYC, HIL, and LIV all elements. The inconsistency of the mass results for NYC and LIV indicates that long bones may experience attrition differently than elements of irregular shape. The two buried collections (LIV and VMC) were the only instances where length had a significant effect. LIV was also affected by length while VMC was affected by mineral density, possibly relating to the differing impacts of acid erosion (VMC) and rodent disturbance (LIV).

Altogether, the alternative hypothesis that element attrition correlates with intrinsic properties is supported, however the specific properties that affect element attrition vary based on the taphonomic environment and elements considered.

It is also important to consider that mass, length, and bone mineral density may have overlapping effects. Pearson's correlation tests show a significant correlation between long bone mass and mineral density ($r = 0.9543$, $p = 0.0031$, Figure 28), but no significant correlation between long bone length and mineral density ($r = 0.6622$, $p = 0.1519$, Figure 29). The lack of significant correlation between length and mineral density is caused by the fibula as an outlier. The fibula is the longest non-weight bearing bone in the human body. It has the lowest mineral density of the six long bones, but is similar in length to the tibia and femur. With the fibula removed, the correlation between length and mineral density for the subset of long bones is significant ($r = 0.9962$, $p = 0.0003$, Figure 30). For all 12 elements, there was a significant correlation between mass and length ($r = 0.7405$, $p = 0.0059$, Figure 31).

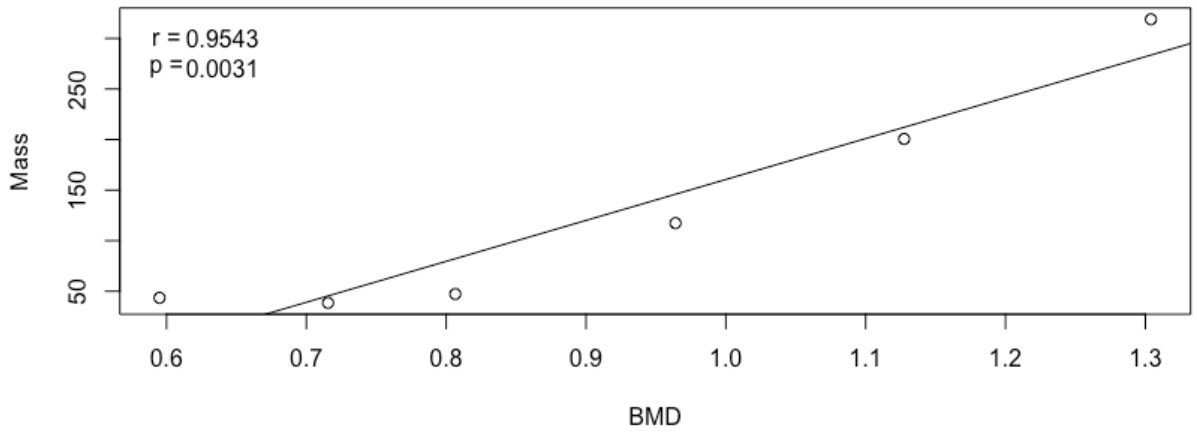

Figure 28. Correlation between mass and bone mineral density for long bones.

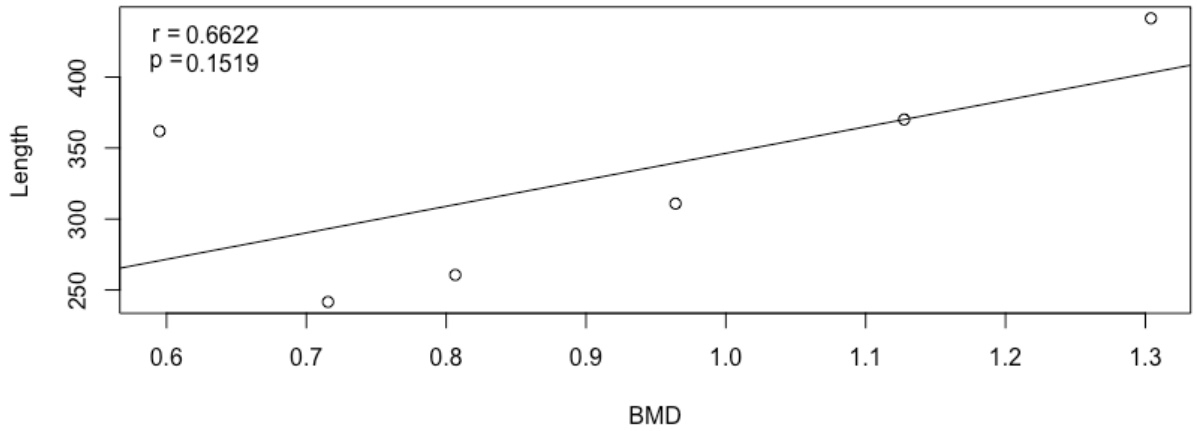

Figure 29. Correlation between length and bone mineral density for long bones.

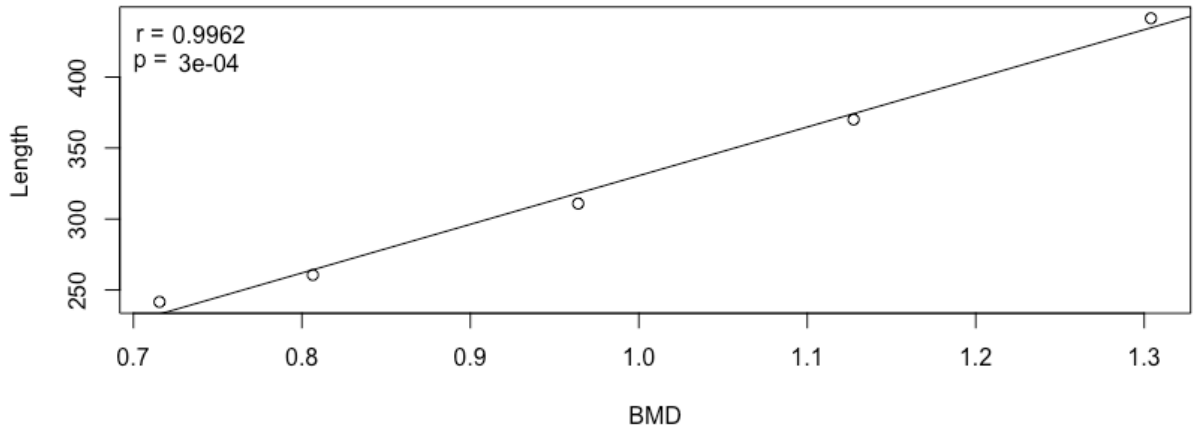

Figure 30. Correlation between long bone length and mineral density, fibula removed.

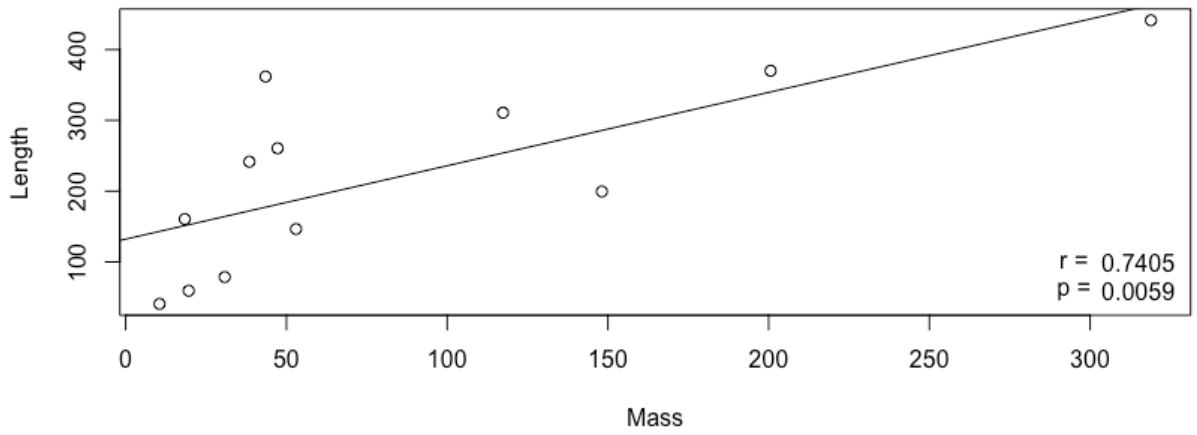

Figure 31. Correlation between mass and length for all 12 elements.

Mass and mineral density show a positive correlation because elements that have thicker cortical bone will be heavier and more highly mineralized. The correlation between length and mineral density is likely related to limb functionality. Elements of the upper limb are shorter and not used for weight bearing, so they have thinner cortical bone and are less mineralized compared to the femur and tibia. The fibula is the exception because it is similar in length to the other lower limb elements but does not bear weight and so does not have the corresponding mineral density adaptation.

Correlation between length and mass is similar to that between length and mineral density in that lower limb elements are both longer and heavier than upper limb elements. In addition, adding length will necessarily add some mass as more bone is needed to make an element longer if shape is held constant. This relationship has a lower correlation than the long bone comparisons because the full set of 12 elements includes irregularly shaped elements such as the innominate and scapula.

The considerable interaction between the intrinsic properties studied indicates that measuring all of these properties is somewhat redundant in terms of information gained. This is not overly problematic as the results indicate that the level of redundancy depends on the

taphonomic environment. Mass and length appear to have similar interactions with acid erosion (VMC) and carnivore scavenging (HIL), but not rodent disturbance and reburial (LIV). There is still important information to be gained by considering all three of these properties.

Independence and Recovery Probability: Cross-Implications

One possible explanation for the high correlations for recovery of pairs is that paired elements should experience taphonomic processes in relatively similar ways. Rather than being tightly correlated with one another, paired elements may co-vary due to their similar properties. Part of the apparent strength of relationship may instead be caused by covariance with these other variables. The VMC collection would be an ideal dataset on which to test this since elements would have likely stayed in place from the time of soft tissue decomposition until excavation.

Limitations and Biases

While there is some overlap between quantification methods and individuation methods, the ideas presented here will fall strictly on the side of quantification. Furthermore, this project is taxonomically limited. The results of both independence testing and recovery probability assessment should only be applied to humans, the taxon from which they were derived. Anatomical differences in other taxa could significantly alter both interdependence and relative recovery probabilities. For example, the radius and ulna are fused in many quadrupedal mammals, so these element types might be expected to show dependence in artiodactyls even if they are independent in humans.

This project is also limited as much as possible to adult elements of the appendicular skeleton. While it would be worthwhile to examine both interdependence and recovery

probability within juvenile skeletons, it is beyond the scope of this study to do justice to the enormous amount of variability introduced by skeletal development. Not only does the number of skeletal parts vary throughout development as the epiphyses appear and fuse, but their relative masses and lengths would likely vary as well.

The uniquely identifiable elements of the appendicular skeleton are chosen over the axial skeleton. This study is targeted toward improving pair-match estimators, which rely mainly on bilateral elements of the appendicular skeleton. Furthermore, assessing interdependence of elements that are not uniquely identifiable would be difficult as interdependence relationships could vary between elements that could not be differentiated, particularly under conditions of poor preservation.

The FDB inventory presents a potentially significant source of unknown variation. The depositional context, case context, ancestry, age, and sex of these individuals are all unknown variables. Other data sets in which these variables are known or can be estimated were used to estimate the effect that this variation may have on independence and recovery probability. In data sets where age is known, this study will be limited to adults.

There are some concerns with using a modern collection to estimate skeletal properties that will then be tested against archaeological data sets. Recent trends have shown that modern populations are generally taller and heavier than their prehistoric counterparts. Normalizing the raw data is one approach that will be used to circumvent this effect. Relative mass or length of elements within the individual skeleton is more important than absolute values, so it should be less problematic that modern populations are taller and heavier than prehistoric ones. In addition, higher body fat may affect not only the amount of lipid in the skeleton but also

its relative distribution. If this is true, normalizing values will not erase the problem because the relative differences between elements will remain.

Summary

It is important to consider the interconnected effects of recovery rates, recovery independence, properties of bone, and properties of the depositional environment when making conclusions about either independence or probability of recovery. Independence testing has shown here that skeletal quantification procedures must account for anatomy. Anatomical relationships play a vital role in the statistical viability of quantification approaches. It is not possible to explain skeletal attrition using only a single factor. Extrinsic and intrinsic factors interact extensively to determine which elements are lost from the assemblage.

CHAPTER VII

CONCLUSIONS

Introduction

This chapter reviews the results obtained for each hypothesis and discusses implications for current practices and future directions. The first two sections summarize results for each hypothesis. The third section offers some potential avenues for future research to further develop the work started here. The final section discusses the insights gained from this work.

Hypothesis 1

$\mathbf{H_{o1}}$ was rejected, as recovery was not independent for multiple element groupings in each collection. $\mathbf{H_{a1}}$ was inconsistently supported, as element groups on the same limb had lower mean phi values than elements connected at a joint, but paired elements had the highest group phi values. Altogether, these results suggest that if multiple elements are being combined in an estimator such as MLNI, selecting elements from both the upper and lower limbs is preferred, and they should not be adjacent if possible.

Particularly troublesome groupings would be those that include multiple elements of the pelvis, the radius and ulna, or the calcaneus and talus. Each of these groupings was significant for at least four out of five datasets with large effect sizes (0.5-0.938). No other

consistent patterns were observed, suggesting that accurate probability density intervals can likely be obtained when non-adjacent elements are used.

Relatedness of recovery is not unexpected, particularly for elements attached at joints. Tuller and Hofmeister (2014) found that for elements sharing an anatomical joint but disarticulated within a highly disturbed mass grave, elements were matched via DNA to the spatially closest appropriate element 88% of the time. Matches were made to the fourth rank distance or less 100% of the time for n = 32 instances tested (Tuller and Hofmeister 2014). This supports that even when disturbance is high, anatomical relationships may still have some bearing on element deposition.

Hypothesis 2

H_{o2} was rejected and **H_{a2}** accepted, as all datasets showed significant correlations between recovery rate and one or more intrinsic properties. Differences in taphonomic context likely explain differences in which correlations were strongest for each dataset. These results show that knowledge of taphonomy can inform expectations for recovery. In particular, burial may have a relationship with length, while unburied remains are more affected by mass and mineral density.

` Having an idea of what element attrition should look like in a particular taphonomic environment can also be helpful in cases where these expectations are not met. If attrition follows an unexpected pattern, this might be suggestive of possibilities such as reburial. For sample selection from existing collections, knowledge of how the remains were recovered and from what environment can guide in choosing the sample even if prior knowledge of inventory data is incomplete. For example, remains buried in neutral or alkaline soils might be suitable for

a stature study even if there was rodent disturbance, as good preservation of long bones would still be possible.

Future Research

This study would be more robust through the inclusion of more datasets, including those where taphonomy could be expected to be similar to those included here, as well as different. Particular areas that are lacking are remains affected by water and by fire. Collection of bone mineral density data for all 12 elements would also be useful, as this would include a wider range of element lengths than the subset of long bones.

Another area of research identified here is investigation of rank-based methods of estimating population size. The data collected here are insufficient to adequately test the rank-based estimator demonstrated through simulation in chapter 3. Recovery rates of all datasets used in this thesis are very similar, and recovery rate has a large impact on the final estimate. Further testing is needed outside the recovery rate range of 0.6-0.8. As shown in the simulation graphs in chapter 3, estimators can display erratic behavior at low recovery rates that is not predictable through examining behavior at higher rates.

Summary

This study is an incremental addition to the study of human skeletal quantification. The independence testing performed has generated guidelines to complement existing pair-match methods by making it possible to choose statistically defensible element combinations to compute these estimators. As a result, it will be possible to include a larger proportion of available data in skeletal quantification. The information gained here will permit the construction

of defensible estimations of uncertainty of estimators. This is a step in the vital transition away from point estimation and toward the estimation of ranges as general best practice.

By understanding the behavior of skeletal element recovery probabilities it is possible to assess the variance in assemblage recovery rate. This is of interest from the standpoints of both skeletal quantification and taphonomy. Estimation of assemblage recovery rate is potentially a viable avenue for estimating population size. Studying the impact that element properties have on element attrition under different depositional contexts can inform both field recovery and curation of skeletal material. The results of this study suggest that there is no single skeletal property that is most important in skeletal attrition, but rather that the most important property will vary according to taphonomic conditions.

REFERENCES

REFERENCES

Adams, Bradley J.
1996 The Use of the Lincoln/Petersen Index for Quantification and Interpretation of Commingled Human Remains. PhD dissertation, University of Tennessee, Knoxville.

Adams, Bradley J., and John E. Byrd
2008 Recovery, Analysis, and Identification of Commingled Human Remains. Totowa, NJ: Humana Press.

Adams, Bradley J., and Lyle W. Konigsberg
2004 Estimation of the Most Likely Number of Individuals from Commingled Human Skeletal Remains. American Journal of Physical Anthropology 125(2): 138–151.
2008 How Many People? Determining the Number of Individuals Represented by Commingled Human Remains. *In* Recovery, Analysis, and Identification of Commingled Human Remains Pp. 241–255. Totowa, NJ: Humana Press.

Behrensmeyer, Anna K.
1978 Taphonomic and Ecologic Information from Bone Weathering. Paleobiology 4(2): 150–162.

Bobrowsky, Peter T.
1982 An Examination of Casteel's MNI Behavior Analysis: a Reductionist Approach. Midcontinental Journal of Archaeology 7(2): 171–184.

Bokonyi, Sandor
1970 A New Method for the Determination of the Number of In-dividuals in Animal Bone Material. American Journal of Archaeology 74(3): 291–292.

Bright, Lisa
2011 Taphonomic Signatures of Animal Scavenging in Northern California: a Forensic Anthropological Analysis. California State University, Chico.

Buikstra, Jane E., Claire C. Gordon, and L. St Hoyme
1984 The Case of the Severed Skull: Individuation in Forensic Anthropology. *In* Human Identification: Case Studies in Forensic Anthropology Pp. 121–135. Springfield, IL: Charles C. Thomas Ltd.

Byrd, John E.
2008 Models and Methods for Osteometric Sorting. *In* Recovery, Analysis, and Identification of Commingled Human Remains Pp. 199–220. Totowa, NJ: Humana Press.

Byrd, John E., and Bradley J. Adams
2003 Osteometric Sorting of Commingled Human Remains. Journal of Forensic Sciences 48(4): 717–724.

Casteel, Richard W.
1977a A Consideration of the Behaviour of the Minimum Number of Individuals Index: a Problem in Faunal Characterization. Ossa 3(4): 141–151.
1977b Characterization of Faunal Assemblages and the Minimum Number of Individuals Determined from Paired Elements: Continuing Problem in Archaeology. Journal of Archaeological Science 4(2): 125–134.

Chaplin, Raymond E.
1971 The Study of Animal Bones from Archaeological Sites. London: Seminar Press.

Chapman, Douglas George
1951 Some Properties of the Hypergeometric Distribution with Applications to Zoological Sample Censuses, vol.1. University of California Publications in Statistics, 7. Berkeley, CA: University of California Press.

Clason, Antje Trientje
1972 Some Remarks on the Use and Presentation of Archaeozoological Data. Helinium 12(2).

Cruse, Karen Lynn
2002 Commingled Prehistoric Skeletal Remains: An Undying Story. Bios 73(4): 120–126.

Dorozhkin, Sergey V.
2012 Dissolution Mechanism of Calcium Apatites in Acids: A Review of Literature. World Journal of Methodology 2(1): 1.

Fieller, N. R. J., and A. Turner
1982 Number Estimation in Vertebrate Samples. Journal of Archaeological Science 9(1): 49–62.

Fisher, Jacob L.
2015 Faunal Quantification and the Ascendance of Hunting Debate: Reevaluation of the Data from Southeastern California. American Antiquity 80(4): 767–775.

Gonzalez-Rodriguez, Jose, and Gillian Fowler
2013 A Study on the Discrimination of Human Skeletons Using X-ray Fluorescence and Chemometric Tools in Chemical Anthropology. Forensic Science International 231(1): 407. e1–407. e6.

Grayson, Donald K.
1978 Minimum Numbers and Sample Size in Vertebrate Faunal Analysis. American Antiquity 43(1): 53–65.

1984 Quantitative Zooarchaeology: Topics in the Analysis of Archaelogical Faunas. Studies in Archaeological Science. Orlando, FL: Academic Press.

Haglund, William D.
1997 Dogs and Coyotes: Postmortem Involvement with Human Remains. Forensic Taphonomy: The Postmortem Fate of Human Remains: 367–381.

Haglund, William D., and Marcella H. Sorg
2002 Human Remains in Water Environments. Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives: 201–218.

Hildebrandt, William R., and Kelly R. McGuire
2002 The Ascendance of Hunting During the California Middle Archaic: An Evolutionary Perspective. American Antiquity: 231–256.

Horton, D. R.
1984 Minimum Numbers: a Consideration. Journal of Archaeological Science 11(3): 255–271.

Jantz, Richard J., and Peer H. Moore-Jansen
2006 Database for Forensic Anthropology in the United States, 1962-1991. Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. http://doi.org/10.3886/ICPSR02581.v1.

Kendell, Ashley, and P. Willey
2014 Crow Creek Bone Bed Commingling: Relationship Between Bone Mineral Density and Minimum Number of Individuals and Its Effect on Paleodemographic Analyses. *In* Commingled and Disarticulated Human Remains Pp. 85–104. Springer.

Konigsberg, Lyle W., and Bradley J. Adams
2014 Estimating the Number of Individuals Represented by Commingled Human Remains: A Critical Evaluation of Methods. *In* Commingled Human Remains: Methods in Recovery, Analysis, and Identification Pp. 193–220. Oxford, UK: Elsevier.

Krantz, Grover S.
1968 A New Method of Counting Mammal Bones. American Journal of Archaeology 72(3): 286–288.

McCormick, Kyle
2016 A Biologically Informed Structure to Accuracy in Osteometric Reassociation. PhD dissertation, University of Tennessee, Knoxville.

Mundorff, Amy
2009 Human Identification Following the World Trade Center Disaster: Assessing Management Practices for Highly Fragmented and Commingled Human Remains. PhD dissertation, Simon Fraser University.

Mundorff, Amy, Robert Shaler, Erik T. Bieschke, and Elaine Mar-Cash
2014 Marrying Anthropology and DNA: Essential for Solving Complex Commingling Problems in Cases of Extreme Fragmentation. *In* Commingled Human Remains: Methods in Recovery, Analysis, and Identification Pp. 257–273.

Nikita, Efthymia
2014 Estimation of the Original Number of Individuals Using Multiple Skeletal Elements. International Journal of Osteoarchaeology 24(5): 660–664.

Nikita, Efthymia, and Marta M. Lahr
2011 Simple Algorithms for the Estimation of the Initial Number of Individuals in Commingled Skeletal Remains. American Journal of Physical Anthropology 146(4): 629–636.

Nyholt, Dale R.
2004 A Simple Correction for Multiple Testing for Single-nucleotide Polymorphisms in Linkage Disequilibrium with Each Other. The American Journal of Human Genetics 74(4): 765–769.

Parmentier, Sandy
2010 Une Nouvelle Méthode D'estimation Du Nombre Minimum d'individus(NMI) Par Une Approche Allométrique : Le NMI Par Exclusions. : Applications Aux Séries Ostéologiques De La Région Provence-Alpes-Côte d'Azur. PhD dissertation, Universite de la Mediterranee. http://oatd.org/oatd/record?record=star-france\:2010AIX20698&q=commingled%20anthropology.

Perino, Gregory, Kenneth B. Farnsworth, and Michael D. Wiant
2006 Illinois Hopewell and Late Woodland Mounds: The Excavations of Gregory Perino, 1950-1975. Illinois Transportation Archaeological Research Program.

Perrone, Alexandra, Janet E. Finlayson, Eric J. Bartelink, and Kevin Dennis Dalton
2014 Application of Portable X-ray Fluorescence (XRF) for Sorting Commingled Human Remains. *In* Commingled Human Remains: Methods in Recovery, Analysis, and Identification Pp. 145–165. Oxford, UK: Elsevier.

R Core Team
2015 R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Reitz, Elizabeth J., and Elizabeth S. Wing
1999 Zooarchaeology. Cambridge University Press.

Schaefer, Maureen
2008 Patterns of Epiphyseal Union and Their Use in the Detection and Sorting of Commingled Remains. *In* Recovery, Analysis, and Identification of Commingled Human Remains Pp. 221–240. Totowa, NJ: Humana Press.

2014 A Practical Method for Detecting Commingled Remains Using Epiphyseal Union. *In* Commingled Human Remains: Methods in Recovery, Analysis, and Identification Pp. 123–144. Oxford, UK: Elsevier.

Schaefer, Maureen C., and Sue M. Black
2007 Epiphyseal Union Sequencing: Aiding in the Recognition and Sorting of Commingled Remains. Journal of Forensic Sciences 52(2): 277–285.

Schmidt, Christopher W., and Steven A. Symes
2015 The Analysis of Burned Human Remains. Academic Press.

Seber, George AF
1982 The Estimation of Animal Abundance and Related Parameters. 2nd edition. New York, NY: MacMillan Publishing Company.

Snow, Clyde C., and Earl D. Folk
1970 Statistical Assessment of Commingled Skeletal Remains. American Journal of Physical Anthropology 32(3): 423–427.

Spradley, M. Katherine, Michelle D. Hamilton, and Alberto Giordano
2012 Spatial Patterning of Vulture Scavenged Human Remains. Forensic Science International 219(1): 57–63.

Thomas, Richard M., Douglas H. Ubelaker, and John E. Byrd
2013 Tables for the Metric Evaluation of Pair-Matching of Human Skeletal Elements. Journal of Forensic Sciences 58(4): 952–956.

Tuller, Hugh, and Ute Hofmeister
2014 Spatial Analysis of Mass Grave Mapping Data to Assist in the Reassociation of Disarticulated and Commingled Human Remains. *In* Commingled Human Remains Pp. 7–32. Elsevier.

Turner, A., and N. R. J. Fieller
1985 Considerations of Minimum Numbers: a Response to Horton. Journal of Archaeological Science 12(6): 477–483.

White, Theodore E.
1952 Observations on the Butchering Technique of Some Aboriginal Peoples: I. American Antiquity 17(4): 337–338.
1953 A Method of Calculating the Dietary Percentage of Various Food Animals Utilized by Aboriginal Peoples. American Antiquity 18(4): 396–398.

Иностранцев, А.
1882 Доисторическій Человѣкъ Каменнаго Века Побережья Ладожскаго Озера. St. Petersberg: Стасюлевич.

APPENDIX A

Fisher's exact test results, cells with significant values are shaded.

	VMC	FDB	NYC	HIL	LIV
CLAVL_CLAVR	0.3194	0.4979	0.1472	0.8173	0.2688
SCAPL_SCAPR	0.0444	0.3549	0.0681	1.0000	1.0000
HUMEL_HUMER	0.8663	1.0000	1.0000	1.0000	0.0220
RADIL_RADIR	1.0000	1.0000	0.9164	1.0000	0.0207
ULNAL_ULNAR	0.8663	1.0000	0.3517	1.0000	0.2003
ILIML_ILIMR	0.5280	1.0000	0.9173	0.8173	0.0001
PUBSL_PUBSR	0.0011	0.0488	0.0128	1.0000	0.0002
ISCHL_ISCHR	0.0225	0.4352	0.6818	0.3543	0.0000
FEMUL_FEMUR	1.0000	1.0000	0.9173	1.0000	0.9614
PATEL_PATER	0.9547	0.0000	1.0000	0.1063	0.5505
TIBIL_TIBIR	1.0000	1.0000	0.0171	0.9195	0.0004
FIBUL_FIBUR	1.0000	1.0000	0.0416	0.1886	0.0004
CALCL_CALCR	0.1377	0.0000	0.0681	0.2975	0.0089
TALUL_TALUR	1.0000	0.0000	0.5147	0.1886	0.0002
INOML_INOMR	0.2426	1.0000	0.4032	0.2857	0.0008
CLAV_SCAP	0.0000	0.0518	0.0036	0.1083	0.0002
CLAV_HUME	0.0434	1.0000	0.2372	1.0000	0.0006
CLAV_RADI	0.0626	1.0000	0.3350	1.0000	0.0001
CLAV_ULNA	0.6050	1.0000	0.0681	1.0000	0.0008
CLAV_ILIM	0.0028	1.0000	0.4032	1.0000	0.1033
CLAV_PUBS	1.0000	1.0000	1.0000	1.0000	0.5597
CLAV_ISCH	0.0152	1.0000	0.0285	0.8096	0.9412
CLAV_FEMU	1.0000	1.0000	1.0000	1.0000	0.0207
CLAV_PATE	1.0000	1.0000	1.0000	1.0000	0.0371
CLAV_TIBI	1.0000	1.0000	1.0000	1.0000	0.1332
CLAV_FIBU	0.0216	1.0000	1.0000	1.0000	0.0223
CLAV_CALC	1.0000	0.2375	1.0000	1.0000	0.9080
CLAV_TALU	1.0000	1.0000	1.0000	1.0000	0.1477
CLAV_INOM	0.0216	0.1537	0.1729	1.0000	0.0358
SCAP_HUME	0.0676	1.0000	0.0001	1.0000	0.1174
SCAP_RADI	0.0869	1.0000	0.3350	1.0000	0.0006
SCAP_ULNA	0.0676	1.0000	0.4063	1.0000	0.0538
SCAP_ILIM	0.0002	1.0000	0.0639	0.8173	0.0001
SCAP_PUBS	1.0000	1.0000	0.0267	1.0000	0.3382
SCAP_ISCH	0.0001	1.0000	0.0639	0.4553	0.0005
SCAP_FEMU	1.0000	1.0000	0.7785	1.0000	0.1033
SCAP_PATE	1.0000	1.0000	1.0000	1.0000	0.6105
SCAP_TIBI	1.0000	1.0000	1.0000	0.6423	0.0345
SCAP_FIBU	0.3783	1.0000	1.0000	1.0000	0.0974
SCAP_CALC	1.0000	0.0196	1.0000	0.5963	0.1332
SCAP_TALU	1.0000	1.0000	1.0000	0.8729	0.9080

SCAP_INOM	0.0014	1.0000	0.0238	0.4154	0.0000
HUME_RADI	0.1495	0.5857	0.0136	1.0000	0.0000
HUME_ULNA	0.0278	0.0094	0.0013	0.9499	0.0000
HUME_ILIM	0.1936	0.5879	0.0466	1.0000	0.0628
HUME_PUBS	1.0000	1.0000	0.4970	1.0000	0.9412
HUME_ISCH	0.3675	1.0000	0.5297	1.0000	0.5438
HUME_FEMU	1.0000	1.0000	0.5297	1.0000	0.0050
HUME_PATE	1.0000	1.0000	1.0000	1.0000	0.0332
HUME_TIBI	1.0000	1.0000	1.0000	1.0000	0.2011
HUME_FIBU	0.4399	1.0000	1.0000	1.0000	0.0486
HUME_CALC	1.0000	0.2817	1.0000	1.0000	0.7040
HUME_TALU	1.0000	1.0000	1.0000	1.0000	0.1174
HUME_INOM	1.0000	1.0000	0.2433	1.0000	0.0656
RADI_ULNA	0.1495	0.0000	0.0000	0.0000	0.0000
RADI_ILIM	0.5242	1.0000	0.7873	0.3806	0.0145
RADI_PUBS	1.0000	1.0000	0.3517	0.8135	0.9412
RADI_ISCH	0.3419	1.0000	0.7873	0.5993	0.2003
RADI_FEMU	1.0000	1.0000	1.0000	0.2265	0.0000
RADI_PATE	1.0000	1.0000	1.0000	0.8729	0.1119
RADI_TIBI	1.0000	1.0000	0.2975	1.0000	0.0001
RADI_FIBU	0.9324	0.5938	0.1376	1.0000	0.0000
RADI_CALC	1.0000	0.2375	0.0287	0.0991	0.9412
RADI_TALU	1.0000	1.0000	0.1027	0.3902	0.0470
RADI_INOM	1.0000	1.0000	1.0000	0.8411	0.0016
ULNA_ILIM	0.1936	1.0000	0.0978	0.0890	0.0587
ULNA_PUBS	1.0000	1.0000	0.4469	1.0000	1.0000
ULNA_ISCH	0.3675	1.0000	0.9173	0.0867	0.7040
ULNA_FEMU	1.0000	1.0000	1.0000	0.0630	0.0000
ULNA_PATE	1.0000	1.0000	1.0000	1.0000	0.0571
ULNA_TIBI	1.0000	0.3344	0.3856	0.4154	0.0370
ULNA_FIBU	0.4399	1.0000	0.1988	0.2975	0.0001
ULNA_CALC	1.0000	1.0000	0.0862	0.0513	1.0000
ULNA_TALU	1.0000	1.0000	0.2485	0.5993	0.4718
ULNA_INOM	1.0000	1.0000	1.0000	0.2975	0.0272
ILIM_PUBS	1.0000	0.0000	0.0087	0.9499	0.0000
ILIM_ISCH	0.0001	0.0000	0.0004	0.0000	0.0000
ILIM_FEMU	1.0000	1.0000	1.0000	0.6048	0.0983
ILIM_PATE	1.0000	1.0000	1.0000	1.0000	0.5646
ILIM_TIBI	1.0000	1.0000	1.0000	0.0890	0.1332
ILIM_FIBU	1.0000	1.0000	0.6619	0.9499	0.0084
ILIM_CALC	1.0000	0.5106	1.0000	1.0000	0.0811
ILIM_TALU	1.0000	1.0000	1.0000	1.0000	0.0006
ILIM_INOM	0.0000	0.0000	0.0108	0.0000	0.0000

PUBS_ISCH	1.0000	0.0000	0.0087	0.2216	0.0000
PUBS_FEMU	1.0000	1.0000	1.0000	1.0000	1.0000
PUBS_PATE	1.0000	1.0000	1.0000	1.0000	0.1001
PUBS_TIBI	1.0000	1.0000	1.0000	1.0000	0.7553
PUBS_FIBU	1.0000	0.2030	1.0000	1.0000	0.3767
PUBS_CALC	1.0000	0.1172	1.0000	1.0000	0.2688
PUBS_TALU	1.0000	1.0000	1.0000	1.0000	0.1332
PUBS_INOM	1.0000	0.0000	0.0736	0.1868	0.0000
ISCH_FEMU	1.0000	1.0000	1.0000	1.0000	0.9587
ISCH_PATE	1.0000	1.0000	1.0000	1.0000	0.3542
ISCH_TIBI	1.0000	1.0000	1.0000	0.6423	0.4368
ISCH_FIBU	0.4573	1.0000	0.6619	1.0000	0.3382
ISCH_CALC	1.0000	0.1176	1.0000	0.5963	0.1332
ISCH_TALU	1.0000	1.0000	1.0000	1.0000	0.0269
ISCH_INOM	0.0007	0.0000	0.0108	0.0005	0.0000
FEMU_PATE	1.0000	1.0000	1.0000	1.0000	0.9412
FEMU_TIBI	1.0000	0.0001	0.1472	0.0630	0.0161
FEMU_FIBU	1.0000	1.0000	0.0573	0.4293	0.0034
FEMU_CALC	1.0000	0.9471	1.0000	1.0000	1.0000
FEMU_TALU	1.0000	1.0000	1.0000	1.0000	0.2003
FEMU_INOM	1.0000	1.0000	1.0000	1.0000	0.0650
PATE_TIBI	1.0000	1.0000	1.0000	1.0000	0.1001
PATE_FIBU	1.0000	1.0000	1.0000	0.4060	0.0929
PATE_CALC	1.0000	0.0478	1.0000	0.1437	0.2279
PATE_TALU	1.0000	0.0000	1.0000	0.0321	0.0650
PATE_INOM	1.0000	1.0000	1.0000	1.0000	0.1119
TIBI_FIBU	1.0000	0.3576	0.0000	0.0014	0.0000
TIBI_CALC	1.0000	0.1176	0.0000	0.0513	0.0929
TIBI_TALU	1.0000	1.0000	0.0000	0.0031	0.0000
TIBI_INOM	1.0000	1.0000	0.5282	0.2975	0.0170
FIBU_CALC	0.0758	0.0031	0.0000	0.2934	0.0779
FIBU_TALU	0.8663	1.0000	0.0000	0.0705	0.0038
FIBU_INOM	1.0000	1.0000	0.1867	1.0000	0.0010
CALC_TALU	0.0004	0.0158	0.0000	0.0000	0.0000
CALC_INOM	1.0000	1.0000	1.0000	1.0000	0.0538
TALU_INOM	1.0000	1.0000	1.0000	1.0000	0.0154