

MAIN TITLE: Missing Data in Bioarchaeology II: A Test of Ordinal and Continuous Data Imputation

RUNNING TITLE: Missing Data in Bioarchaeology II

AUTHORS: Amanda Wissler<sup>1</sup>, Kelly E. Blevins<sup>2</sup> and Jane E. Buikstra<sup>3</sup>

AFFILIATIONS:

<sup>1</sup>Department of Anthropology, University of South Carolina, Columbia, SC

<sup>2</sup>Archaeology Department, Durham University, Durham, UK

<sup>3</sup>Center for Bioarchaeological Research, School of Human Evolution and Social Change, Arizona State University, Tempe, AZ

ABSTRACT:

**Objectives:** Previous research has shown that while missing data are common in bioarchaeological studies, they are seldom handled using statistically rigorous methods. The primary objective of this article is to evaluate the ability of imputation to manage missing data and encourage the use of advanced statistical methods in bioarchaeology and paleopathology. An overview of missing data management in biological anthropology is provided, followed by a test of imputation and deletion methods for handling missing data.

**Materials and Methods:** Missing data were simulated on complete datasets of ordinal (n=287) and continuous (n=369) bioarchaeological data. Missing values were imputed using five imputation methods (mean, predictive mean matching, random forest, expectation maximization, stochastic regression) and the success of each at obtaining the parameters of the original dataset compared with pairwise and listwise deletion.

**Results:** In all instances, listwise deletion was least successful at approximating the original parameters. Imputation of continuous data was more effective than ordinal data. Overall, no one method performed best and the amount of missing data proved a stronger predictor of imputation success.

**Discussion:** These findings support the use of imputation methods over deletion for handling missing bioarchaeological and paleopathology data, especially when the data are continuous. Whereas deletion methods reduce sample size, imputation maintains sample size, improving statistical power and preventing bias from being introduced into the dataset.

KEY WORDS: missing data, imputation, bioarchaeology, paleopathology

8344 words, 28 pages text, 12 pages bibliography, 8 figures, 9 tables in supplemental information

Corresponding Author: Amanda Wissler, Department of Anthropology, University of South Carolina, Gambrell Hall Suite 440, 817 Henderson Street, Columbia, S.C. 29208.  
amandacwissler@gmail.com

## INTRODUCTION

Missing data are ubiquitous in the social sciences. Bioarchaeological data may be lost due to myriad factors including differential preservation, selective excavation, post-mortem damage, pathology, transcription errors, and/or computer crashes. When not handled properly, missing values can introduce substantial bias into a dataset, leading to erroneous study results and flawed interpretations (see Stojanowski & Johnson, 2015). Furthermore, most statistical tests require datasets with no missing data. Despite the significance of missing data, their treatment is often unreported in the social sciences, including in anthropology. When missing data are addressed, the least statistically and theoretically rigorous methods are generally used (see companion paper: Missing Data in Bioarchaeology I). The goal of this paper is to explore techniques for handling missing data, focusing on the use of imputation to manage missing bioarchaeological and paleopathological data. Imputation is defined as inserting a plausible value in place of a missing value (Allison, 2001; Schafer, 1999; Schafer & Graham, 2002). Our target audience includes anthropologists who have basic statistical and programming knowledge, but they need not be statistical experts, as methods are explained conceptually rather than mathematically. This paper has two sections: Part I gives a brief overview of classes of missing data and describes commonly used missing data methods in biological anthropology. Part II provides a case study test of seven methods for handling missing ordinal and continuous paleopathology data. This paper is intended as a companion paper to Missing Data in Bioarchaeology I, which reviews current approaches to missing data in bioarchaeology.

## PART I: BACKGROUND

The best way to manage missing data depends on how and why the data are missing. Rubin (1976) described three main categories of missing data: missing completely at random,

missing at random, and missing not at random. Data are described as missing completely at random (MCAR) when the reason the data are missing is unrelated to the pattern of missingness or any other variables of interest in the data set (Graham, Hofer, Donaldson, MacKinnon & Schafer, 1997; Pepinsky, 2018; Quintero & LeBoulluec, 2018). If we have collected two variables (X and Y), data are MCAR “if the probability of missing data on Y is unrelated to the value of Y itself or to the values of any other variables in the data set” (Allison, 2001, p. 3). For example, in a dataset containing information on age (X) and femoral length (Y), data on femoral length would be MCAR if their missingness is unrelated to age or femoral length. MCAR data are likely rare among bioarchaeological datasets but could occur when skeletons are only partially recovered due to an incomplete excavation grid or when taphonomic processes vary stochastically across mortuary deposits, resulting in some poorly preserved skeletal elements or cortical surfaces.

The second category is missing at random (MAR). Data are missing at random if the pattern of missingness depends on some variable in the dataset that is not the variable of interest (Graham, Olchowski & Kilreath, 2007; Pepinsky, 2018; Quintero & LeBoulluec, 2018). Data are MAR if the probability of missing data on Y depends on the variable X but not on the value of Y (Allison, 2001). Using the above example, femoral length (Y) data would be missing at random if the missingness depended on age (X) but not on femoral length (Y) which could occur if older individuals more often had poorly preserved long bones due to osteoporosis.

The third category of missing data is missing not at random (MNAR), also called not missing at random (NMAR). Data are described as MNAR if the probability of missingness is related to the variable of interest i.e., if the probability of missing data on Y depends on Y itself (Pepinsky, 2018; Quintero & LeBoulluec, 2018). For example, data missing under the variable

femoral length (Y) would be MNAR if the data are missing because of femoral length (Y). In practice, this may be because the researcher opted to exclude individuals with unusually short or long femurs, or because only “normal” femurs were accessioned into the collection (see Bhaskaran & Smeeth, 2014 for additional examples of MAR, MCAR, and MNAR variables).

Data that are MCAR or MAR are less problematic than MNAR and are often referred to as “ignorable” (Allison, 2001; Enders, 2010; Graham, 2012; Osborne, 2013). Deleting data that are MCAR or MAR, however, may result in a decline in statistical power due to a decreased sample size. Since MCAR and MAR data are distributed randomly, their absence should not introduce bias into the dataset (Graham, 2009; Howell, 2007; Myers, 2011). Data missing not at random, however, are problematic and referred to as “nonignorable” (Allison, 2001; Graham, 2009; Graham, 2012). The probability of missingness is dependent on the missing data, and it is almost impossible to know the true extent of that relationship. Therefore, it is not possible to control or compensate for data missing not at random (Graham, 2012; Howell, 2007; McKnight, McKnight, Sidani & Figueiredo, 2007). MNAR data can result in a substantially biased dataset, because information vital to answering the research question is absent (De Leeuw, Hox & Hisman, 2003; Finch, 2010; Graham, 2009; Osborne, 2013). In bioarchaeology (and paleopathology in particular), missing data likely fall into a combination of all three categories and it may be impossible to discern which variables belong in which category (Morris, White & Royston, 2014; Myers, 2011).

Overall, the methods bioarchaeologists use to manage missing data can broadly be classified into three categories: deletion, imputation, and maximum likelihood. Here we provide a detailed description of each approach, its advantages, and its disadvantages.

## **Deletion**

*Pairwise Deletion* (aka Available Case Analysis) involves dropping cases or individuals based on variables present for each analysis (Allison, 2001; Graham, 2012; van Buuren, 2018). For example, an individual missing a periodontal disease score will be deleted from any analyses requiring periodontal disease as a variable but included in other tests, such as analyses for femoral length. This approach is easy to perform and has the benefit of making use of all available data, greatly maximizing the sample size. However, each analysis uses a slightly different sample, generating results that may not be comparable or are inconsistent across variables (Myers, 2011; Newman, 2014; van Ginkel, Linting, Rippe & van der Voort, 2020). Published tables may list different sample sizes, which can be misleading if not appropriately explained, and repeatedly running similar analyses on overlapping samples raises concerns of alpha inflation. If the data are not MCAR, pairwise deletion can create bias in the parameter estimates (Allison, 2001; Baraldi & Enders, 2010). Furthermore, when each analysis is based on a slightly different sample, there is no straightforward procedure for calculating the standard error for the entire sample (Graham, 2012).

*Listwise Deletion* (aka Casewise Deletion) involves the removal of an individual and all their data – an entire row in a spreadsheet – if any data for that individual are missing (Allison, 2001; Graham, 2012; van Buuren, 2018). This is the default method employed by statistical software programs SAS, SPSS, and Stata (van Buuren, 2018). Listwise deletion has the advantages of being easy to understand, being simple to execute, and not requiring advanced statistical knowledge or software (Allison, 2001; Meeyai, 2016). It creates a complete dataset that allows one to proceed with statistical analysis (Baraldi & Enders, 2010). If, however, the amount of missing data is even moderate, listwise deletion can result in an enormous decrease in sample size and subsequent loss of statistical power (Baraldi & Enders, 2010; Graham, 2012).

The amount of missing data may be so great that entire variables could be deleted. If the data are not missing completely at random (MCAR), listwise deletion can introduce substantial bias into final p-values and confidence intervals (Allison, 2001; Baraldi & Enders, 2010). Many statisticians consider listwise deletion to be the worst of all possible techniques for handling missing data (Allison, 2001; King, Honaker, Joseph & Scheve, 1998; van Buuren, 2018; Wilkinson, 1999).

## **Imputation**

Imputation, defined as inserting a plausible value in place of a missing value, is an alternative to deletion methods for handling missing data (Allison, 2001; Schafer, 1999; Schafer & Graham, 2002). Imputation is a broad term that encompasses numerous frameworks and mathematical models for producing and selecting the imputed values. Critics of imputation claim these approaches “make up data,” invoking various justifications for continuing to use deletion methods (Osborne, 2013; Schafer, 1999; van Ginkel, Linting, Rippe & van der Voort, 2020). Studies have shown, however, that imputed data are often better able to recover the original parameter estimates and are more easily replicable by other researchers than deletion methods (Fitchman & Cummings, 2003; King, Honaker, Joseph & Scheye, 1998; Osborne, 2013; Pedersen et al., 2017).

*Mean Replacement* is the simplest imputation strategy: the mean of a variable is substituted for each missing data point of that variable (Little & Rubin, 2002). Mean imputation is easy to understand and implement but will decrease the variance of the sample. As such, large amounts of missing data will skew covariate relationships and affect the strength and direction of correlations (Osborne, 2013; Graham, 2012; Musil, Warner, Yobas & Jones, 2002).

*Regression* imputation uses the complete case data for all variables to build a regression model that is then used to predict missing values. Regression imputation is conceptually intuitive and utilizes all variables in the dataset to generate predictions (Graham, 2012; Musil et al., 2002). The imputed values, however, lie on the regression line, so the sample variance is artificially decreased and correlations between variables are spuriously strengthened (Graham, 2012; van Buuren, 2018; Zhang, 2016).

*Stochastic Regression* imputation corrects for the over correlation between variables by adding random “noise” back into the model (Newman, 2003; van Buuren, 2018). One way to add this noise is by randomly selecting from the residuals and adding that value to the predicted missing value (Enders, 2010; Little and Rubin, 2002; van Buuren, 2018). Stochastic regression has the advantage of being able to produce unbiased parameters when the data are MCAR or MAR but will underestimate standard errors (Allison, 2001; Enders, 2010).

*Random Forest* imputation uses a decision tree approach to predict the best values to impute. A bootstrapped random subset of samples is created to build multiple regression trees for each variable (Shah, Bartlett, Carpenter, Nicholas & Hemingway, 2014). The behavior of the data as it is run through the trees predicts the best values for the missing data. Random forest imputation is a commonly used method in epidemiology (Henriksson, Zhao, Dalianis & Boström, 2016; Shah et al., 2014; Weng, Vas, Qureshi & Kai, 2019) and is capable of handling mixed data types and variable interactions (Stekhoven & Bühlmann, 2012; Tang & Ishwaran, 2017; Waljee et al., 2013). Random forest imputation can also be perceived as a black box technique, with little understanding of how the decision trees are being grown (Breiman, 2001).

*Predictive Mean Matching* (PMM) is an expanded form of hot deck imputation. Hot deck imputation is a broad “record matching technique” in which missing values from an individual

(the recipient) are replaced by observed values from a similar case (the donor) (Kaiser, 1983, p. 1). This method requires the selection of an imputation model, such as linear regression. The model is estimated using the complete cases of the predictor variable and the variable to be imputed, and then the model is used to predict *all values of the variable to be imputed*, observed and missing. Subsequently, each predicted missing value is matched to the most similar predicted values from the observed cases; one of these close cases is randomly selected, and the missing value is substituted for the observed value (Bailey, Andridge & Shoben, 2020; Little, 1988; Vink, Frank, Pannekoek & van Buuren, 2014). Because PMM matches values from other donor cases within the dataset, imputed values will always fit with the observed range of values (Kleinke, 2018; Vink et al., 2014). A potential disadvantage to predictive mean matching is that it may not be acceptable for use with small sample sizes, as the pool of available observed outcomes with a similar case prediction will be small (Kleinke, 2018).

### **Maximum Likelihood Estimation**

*Expectation Maximization* (EM) is a common maximum likelihood algorithm that uses a two-step iterative process. In the E-step, a missing value is imputed based on what would be expected given other values in the dataset (Dempster, Laird & Rubin, 1977; Graham, 2012; Newman, 2003). In the M-step, the algorithm checks whether the new value has the highest probability of being a good fit with the rest of data. If not, the process begins again, imputing a more likely value until all missing data have been replaced with the most likely values (Musil et al., 2002). EM procedures generally perform better than mean imputation or deletion methods (Nelwamondo, Mohamed & Marwala, 2007). One potential drawback to EM, however, is that it produces standard errors that may be narrower than those of the true data and thus may artificially increase statistical confidence (Musil et al., 2002).

## Prior Approaches to Imputation in Other Disciplines

While missing data are a problem in nearly all fields of research, some disciplines, such as psychology, ecology, and health sciences, have adopted advanced methods for handling missing data more quickly than others, particularly compared to biological anthropology. We suspect this delay may be due to a lack of awareness that imputation exists or a belief that certain types of data (e.g., age, sex) are not appropriate to impute (McKnight et al., 2007). In other fields, however, scholars regularly impute a wide variety of missing demographic and social variables that are comparable to those used in biological anthropology and which can be used as a model for our field moving forward. Working in the social sciences, Evans and Smokowski (2015) tested how social capital – as measured by proxies such as social support and mental health – predicts the likelihood of intervening in school bullying. They imputed missing survey data on demographic factors such as ethnicity and religion, responses on parental support, school satisfaction, and optimism about the future. Turney (2015) examined how paternal incarceration may be a cause of food insecurity for children, imputing missing survey answers on how often a child has been hungry or how often they skipped meals.

Researchers in the natural and ecological sciences have adopted advanced techniques as standard for dealing with missing data, imputing a diverse array of biological traits – such as leaf area, seed mass, plant height, animal body mass, litter size, diet diversity, sociality, and generation length – that are analogous to data regularly used in many areas of biological anthropology (Bird et al., 2020; Cooke, Eigenbrod & Bates, 2019; Cooke, Eignebrod & Bates, 2020; Grilo, Koroleva, Andrášik, Bíl & González-Suárez, 2020; Ordonez & Svenning, 2017; Pacifici et al., 2013; Taugourdeau, Villerd, Plantureaux, Huguenin-Elie & Amiaud, 2014). Divíšek et al. (2018) for example, searched for patterns of traits that could be indicators of

invasive plant species and imputed missing data on leaf area, plant height, and seed weight. While investigating how certain traits relate to ecological strategies, Cooke et al. (2020) used multiple imputation to manage missing data on body mass, habitat breadth, generation length, diet, and litter/clutch size.

Imputation of missing demographic or health data like those used in sociocultural anthropology and bioarchaeology is commonplace within epidemiological and clinical studies (Barnard & Meng, 1999; Bodnar, Tang, Ness, Harger & Roberts, 2006; Costello et al., 2014; Ferrie, Martikainen, Shipley & Marmot, 2005; Petersen et al., 2014; Zeka, Zanobetti & Schwartz, 2006). Lassale et al. (2018) imputed missing body measurements, testing the association between obesity and coronary heart disease. Dam et al. (2016) imputed missing values for age, education, smoking, and health status to examine whether increased alcohol use in postmenopausal women increases their risk of breast cancer while decreasing their risk of coronary heart disease. In general, disciplines in the social, ecological, and biological sciences employ more sophisticated approaches to missing data, particularly compared to certain areas in biological anthropology.

### **Prior Approaches to Imputation in Biological Anthropology**

Missing data have been identified as a concern in many subfields of biological anthropology including paleoanthropology (Clavel, Merceron & Escarguel, 2014; Gordon, Green & Richmond, 2008; Kramer & Konigsberg, 1999), primatology (Ely, Zavasis & Lammey, 2013; Jardim, Bini, Diniz-Filho, & Villalobos, 2021), paleogenomics (Irving-Pease, Muktupavela, Dannemann & Racimo, 2021; Ishiya, Mizuno, Wang, & Ueda, 2019; Mizuno et al., 2017) and bioarchaeology (Burnett, Irish & Fong, 2013; Kenyhercz & Passalacqua, 2016; Stojanowski & Johnson, 2015). While most scholars in these areas agree that deletion is an unsatisfactory

method for handling missing data, there are few discipline-specific papers providing guidance or suggesting best practices. As discussed in the companion paper, many researchers do not disclose the presence of missing values in their datasets. Imputation seems to have been adopted unequally among various subfields of biological anthropology, likely reflecting the types of questions asked, the statistical analyses employed, and the other disciplines from which each subfield draws.

Paleoanthropology relies on inherently fragmentary data and small sample sizes due to the nature of the fossil record. Paleoanthropologists have therefore been compelled to reconcile with their missing data more so than many other areas of biological anthropology. Expectation maximization is a commonly found missing data method in vertebrate paleontology, having been recommended by Strauss, Atanassov and De Oliveira (2003) and used in numerous other studies in biological anthropology (Athreya & Wu, 2017; Scherer, 2007; Stefan, 2004). The need to reconstruct fragmentary hominin crania has been a major driver of missing data management. Glantz, Athreya and Ritzman, (2009) imputed missing fossil cranial measurements to assess group membership for the Teshik-Tash 1 cranium. Paleoanthropologists employing geometric morphometrics have adopted other approaches for handling missing values. Gunz, Mitteroecker, Neubauer, and Bookstein (2009) for example, proposed using multiple multivariate regression to estimate missing values to reconstruct fragmentary hominin crania.

Due to the fragmentary and low-quantity nature of ancient DNA (aDNA), paleogeneticists must work with incomplete and low coverage genomic data, with most samples below an average 1x depth of coverage. Imputation is used to infer genotype calls (for example, Aa, AA, or aa) across aDNA samples so that there are sufficient genotypes to compare among study samples and analyze concurrently with large comparative aDNA and modern datasets.

Unlike other incomplete biological anthropology data, genomic data are helpfully governed by a well-documented biological mechanism that allows genotypes to be inferred based on their association with other genotypes linkage disequilibrium (LD) (Neale, 2010). LD is the non-random association alleles have with other geographically close alleles. Chromosomes recombine when chunks of the genome are exchanged during meiosis; these chunks are known as haplotypes. Therefore, genotype imputation methods leverage large reference panels of phased (separated by chromosome) haplotypes from hundreds to millions of individuals to infer missing genotypes (Browning & Browning, 2016). While genotype imputation is highly successful for modern genomic data (Browning, Zhou & Browning, 2018; Pasaniuc et al., 2012), there are caveats when working with aDNA. Namely, high-quality comparative reference panels comprise genomes from modern individuals, which may not be representative of ancient genomic diversity and consequently bias results toward the reference alleles (Hui, D'Atanasio, Cassadi, Scheib & Kivisild, 2020). Also, miscoding lesions sequenced from degraded aDNA fragments and sequencing errors in low coverage aDNA data can masquerade as variants, which can penalize real genotype similarities and add additional noise to sparse and stochastically preserved aDNA data. Despite these difficulties, researchers have used imputation workflows to significantly increase aDNA sample sizes for downstream analyses, such as population affinity, genetic relatedness, demographic history, and phenotypic inferences (Gamba et al., 2014; Jensen et al., 2019; Martiniano et al., 2017).

Primateologists have also dealt with missing data in their research, though the qualities of those data and causes of their missingness often differ from those in paleogenomics or paleoanthropology. While investigating the social pairing process among rhesus macaques Capitanio, Blozis, Snarr, Steward, and McCowan (2017) imputed missing values on behavioral

responsiveness and temperament. Grebe, Fitzpatrick, Sharrock, Starling, and Drea (2019) studied ringtail lemurs to understand mechanisms and origins of dominance among females. Using the AMELIA package in R, the authors not only imputed missing endocrine data, but also compared imputed and observed values to ensure they had generated plausible imputed values. Studying the sleeping behaviors of proboscis monkeys, Feilin and Marshall (2014) imputed missing values on tree characteristics and measurements. The authors additionally noted the reasons for missing data that included “malfunction of data collection devices, forest fires, river closure, and storms” (p. 1132) which hindered data collection.

Bioarchaeological data have their own suite of unique characteristics that can make them more challenging to analyze than data from other fields. The data are often a mix of continuous, categorical, and binary variables that are best analyzed together. Many statistical tests do not work well with categorical data or do not accept mixed data types. Unlike continuous data, categorical and binary data have a low range of possible values. For example, according to *Standards for Data Collection from Human Skeletal Remains* (Buikstra & Ubelaker, 1994), porotic hyperostosis should be recorded as 0, 1, 2, 3, or 4 – with 0 as no expression, and 4 as the highest expression. Some of the more statistically complicated methods for imputing missing data do not work well with such a narrow range of allowable values. However, because of this low range, less computationally intensive methods may be successful; for instance, a randomly imputed number selected from 0-4 is more likely to fit than one selected from 0-100. Furthermore, the missing values in a bioarchaeological dataset may fall in different classes of missingness depending on the variable. Cribra orbitalia may be missing completely at random (MCAR), linear enamel hypoplasia may be missing at random (MAR), and periodontal disease

may be missing not at random (MNAR). Each variable may require separate pre-analysis data treatments and procedures for handling missing values (Stojanowski & Johnson, 2015).

Another challenge with bioarchaeological data is that we regularly collect data that are MNAR, yet we fail to account for those biases in our analyses or interpretations. For example, most scoring procedures for periodontal disease code missing teeth as NA or not scorable (e.g., Kerr, 1988). However, in cases of extreme periodontal disease, tooth loss will occur (Lindhe, Haffaiee & Socransky, 1983; Morelli et al., 2018; Ong, 1998; Ramseier et al., 2017). Antemortem tooth loss may therefore be the highest expression of periodontal disease. Scoring teeth missing antemortem as NA introduces MNAR values, creating a biased dataset.

Compared to other subtopics within biological anthropology, bioarchaeologists have made far less use of statistically sophisticated methods for handling missing data and are more likely to rely on deletion methods (see companion article). The areas in which imputation has been used extensively include biodistance analyses and broader investigations of population affinity (Godde & Rangel González, 2022; Paul, Stojanowski & Butler, 2013; Prevedorou & Stojanowski, 2017; Rathmann, Stoyanov & Posamentir, 2022). Noting the limitations of missing data early on, Howells (1973) proposed three options for handling missing biodistance data: mean imputation, regression, and making an educated guess. Working with dental metrics and nonmetrics, Thompson, Hedman, and Slater (2015) imputed missing values to reevaluate evidence surrounding biological relatedness and population movement at Cahokia's mound 72. Similarly, Redfern and Hefner (2019) imputed missing cranial measurements to investigate the presence of individuals with African ancestry in the East Smithfield Black Death cemetery. In recent years, numerous dissertations have emerged that impute missing biodistance data (Bethard, 2013; Bolhofner, 2017; Miller, 2015; Pacheco-Forés, 2020; Paul, 2017).

Despite the regular use of imputation in biodistance, few researchers have assessed the performance of imputed data when analyzed. Kenyhercz and Passalacqua (2016) tested four different imputation methods – hot deck, iterative robust model-based imputation (IRMI), k-nearest neighbor (kNN), and mean – on continuous cranial metric data with 10%, 20%, 50%, and 90% of the values in the dataset missing. Kenyhercz et al. (2019) performed a nearly identical study but assessing imputation of ordinal nonmetric cranial traits. Both papers found all imputation methods perform similarly with low amounts of missing data. At higher percentages of missing data, however, differences emerged; kNN performed well with cranial metric data and IRMI worked well with cranial nonmetric data. Kenyhercz and Passalacqua (2016) however, additionally tested how imputed data affect biodistance analyses by calculating Mahalanobis distances ( $D^2$ ) using both imputed and complete datasets. They found that hot deck, IRMI, and mean imputation artificially decreased the distance between populations, causing them to appear more similar while kNN increased the distance. Fortunately, most methods generated values that were able to classify individuals into correct population groupings with acceptable success. IRMI, however, had the highest levels of group misclassification, with one group only receiving a 7% correct classification.

Beyond biodistance, few researchers have investigated imputation of other bioarchaeological data such as pathology, trauma, age-at-death, or sex. Auerbach and colleagues (Auerbach, 2011; Auerbach, Raxter, & Ruff, 2005; Auerbach & Ruff, 2004) proposed several multiple regression equations to estimate missing skeletal measurements to assess stature. Wissler (2021) imputed missing ordinal paleopathology data to investigate frailty and survival in the 1918 influenza pandemic. A recent paper by Muzzall (2021) proposed a technique for estimating biological sex. Using machine learning techniques and generalized low rank models

to impute missing data, the model achieved a high success rate when cranial interlandmarks and dental metric distances are combined.

## **PART II: A CASE STUDY TEST OF IMPUTATION OF PALEOPATHOLOGY DATA**

The second aim of this paper is to discover which imputation techniques are appropriate for imputing missing ordinal and continuous paleopathology data. Previous researchers have noted that the amount of missing data can have substantial impacts on representativeness of a dataset as well as the success of various imputation approaches (Kleinke, 2018; Leite & Beretvas, 2010; Quintero & Leboulluec, 2018). Furthermore, whether the data are MCAR, MAR, or MNAR will affect imputation performance (King et al., 1998; Musil et al., 2002; Pepinsky, 2018). We therefore examine how different amounts of missing data and how patterns of missingness impact imputation success. To accomplish this, we simulated missing data on two complete bioarchaeological datasets (no missing data) and tested five methods for imputing ordinal paleopathology data and continuous skeletal measurements alongside pairwise and listwise deletion to discover which approach best approximated the parameters of the original dataset.

### **Materials**

Ordinal missing data were simulated on a complete dataset of 287 individuals from the Hamann-Todd Human Skeletal Collection. This sample includes a mix of males, females, African American individuals, and European American individuals ranging in age from 18 to 80 years. Recorded paleopathology data include porotic hyperostosis, cribra orbitalia, periodontal disease, linear enamel hypoplasia, and periosteal lesions of the tibia. The range of ordinal values for each are porotic hyperostosis, 0-2; cribra orbitalia, 0-3; periodontal disease, 0-4; linear enamel hypoplasia, 0-3; periostosis, 0-3.

Continuous missing data were simulated on a complete dataset of 369 individuals from the same collection. Variables include left and right femoral bicondylar lengths – measured in centimeters – and the antero-posterior (AP) and transverse (TR) vertebral neural canal diameters of the first, fifth, and tenth thoracic vertebrae (T1, T5, and T10), and the first and third lumbar vertebrae (L1 and L3) – measured in millimeters.

## Methods

Seven missing data methods were chosen for evaluation: mean imputation, predictive mean matching, stochastic linear regression, random forest, expectation maximization, pairwise deletion, and listwise deletion. These methods were chosen because they are commonly used in the social sciences. Excellent statistical packages are available for each, making these methods easy to implement and more accessible to non-experts. These methods represent a wide range of statistical approaches and range from mathematically simple (e.g., mean imputation), to complex (e.g., expectation maximization).

Several R packages were used to impute the missing data, as no single package worked with all methods and all data types. Mean replacement, predictive mean matching, and stochastic linear regression were achieved using the mice package (v3.11.0; van Buuren & Groothuis-Oudshoorn, 2011). For each,  $m=10$  imputations were performed with 50 iterations. Random forest imputation was executed with the missForest package (v1.4; Stekhoven, 2012) for both ordinal and continuous data; the mice package was unable to form discrete decision trees for the ordinal data given the low range of possible values. Expectation maximization for both ordinal and continuous data was performed using missMethods (v0.2.0; Rockel, 2020). Pairwise deletion was achieved using na.rm=TRUE to remove individuals with missing values by variable.

Listwise deletion used the `na.omit` function, deleting an entire individual from that iteration of the dataset. All analyses were performed in RStudio version 1.1.456 (Rstudio Team, 2016).

To evaluate how the amount of missing data influences the success of the seven approaches five different datasets with 5%, 10%, 20%, 30%, and 40% of the data missing were created using the R package `imputeR` (v2.2; Feng, Moritz, Nowak, Welsh & O'Neill, 2020) resulting in 25 ordinal datasets and 25 continuous datasets with missing data. To assess how patterns in the missing data affect imputation, five additional datasets with percentages of missingness that differ for each variable were created to more accurately reflect patterns of missingness found in a genuine bioarchaeology dataset. Missing data were simulated as MCAR, MAR, and MNAR using the R package `missMethods` (v0.2.0; Rockel, 2020), resulting in a total of 15 additional datasets. For the ordinal data, percentages of missingness were set at porotic hyperostosis = 12.5%, cribra orbitalia = 20%, periodontal disease = 25%, linear enamel hypoplasia = 30%, and periostosis = 10%. For continuous data, the following percentages of missingness were selected: femoral length right = 10%, femoral length left = 10%, T1AP = 15%, T1TR = 10%, T5AP = 12.5%, T5TR = 10%, T10AP = 15%, T10TR = 15%, L1AP = 20%, L1TR = 15%, L3AP = 20%, L3TR = 15%. These percentages mirror the amount of missing values in the first author's dissertation dataset (Wissler, 2021). Note that these data come from a documented osteological collection and therefore these percentages may not represent what would be found in an archaeological assemblage.

## Assessing Success

The success of each imputation method was assessed using the normalized root mean square error (NRMSE), which measures the difference between predicted and observed values; a lower NRMSE indicates a better fit. NRMSE was calculated using the `hydroGOF` package (v0.4-

0; Zambrano-Bigiarini, 2020). As NRMSE is a standard metric for evaluating imputation methods, the results will be broadly comparable to similar studies in other disciplines. Calculating NRMSE requires original and imputed values and thus could not be used to assess success for the two deletion methods. Therefore, percent error of the mean was also used to compare success of imputation and deletion methods. NRMSE and percent error quantify slightly different aspects of imputation success. NRMSE quantifies the difference between paired original and imputed values; percent error evaluates the difference in the overall mean between the original and imputed datasets. The R code for these procedures is available under the first author's GitHub Repository (Wissler, 2022).

## Case Study Results

*Ordinal Data* Summary results are shown in Figures 1-4. Tables with the complete results are available as Supporting Information. Figure 1 shows that when evaluated using NRMSE, all imputation methods performed roughly the same. For the 5%, 10%, 20%, 30%, and 40% missing datasets, pairwise deletion and expectation maximization had slightly better performance compared to the other methods while listwise deletion was the worst when evaluated with percent error (Figure 2). Mean imputation of ordinal data performed poorly compared to all other imputation and deletion methods at 5%, 10%, 20%, and 30% missingness when evaluated with percent error. Predictive mean matching with 30% missingness did worse than 40% missingness when assessed with percent error, which is not a pattern that is found among any other results.

Similarly, for the MCAR, MAR, and MNAR datasets, all methods performed about the same based on the NRMSE (Figure 3). Interestingly, there is no strong difference in imputation success among data that are MCAR, MAR, and MNAR under NRMSE, which contrasts with the

findings of other similar studies (Musil et al., 2002; Pepinsky, 2018). Evaluated with percent error, however, all the missing data methods perform worse on MNAR datasets compared to MCAR or MAR datasets (Figure 4). Overall, no single imputation method was best able to recover the parameters of the original dataset in all categories of missing ordinal data.

*Continuous Data* Summary results are shown in Figures 5-8. Tables with the complete results are available as Supporting Information. The results for continuous data are similar to those of the ordinal data. For 5%, 10%, 20%, 30% and 40% missingness, the percent of missing data was a stronger predictor of imputation success than the imputation method, with the possible exception of mean imputation, which performed worse than all other methods across all levels of missingness (Figure 5). Listwise deletion performed considerably worse compared to all other methods; the percent error for even 5% missing data with listwise deletion exceeded the percent error for 40% missingness with any other method (Figure 6). Overall, data that are MNAR generally have worse imputation success regardless of the method, though the differences are not large when assessed using NRMSE (Figure 7). Figure 8 likewise shows that listwise deletion has the least success at obtaining the parameters of the original dataset, as even MCAR data had a higher percent error when treated with listwise deletion compared to MNAR data with any other missing data method.

*General Findings* Across all the results – ordinal and continuous – there is little variation in performance apart from mean imputation and listwise deletion. When evaluated with NRMSE, mean imputation for continuous data performed worse across all amounts of missingness (5-40%) and all patterns of missingness (MCAR, MAR, and MNAR) (Figures 5 & 7). With ordinal data, however, mean imputation was among the better performing methods (Figures 1 & 3), although the difference is not great. Using percent error, however, mean imputation is among the

best-performing methods, except for data that are MNAR. This discrepancy is due to how NRMSE and percent error are calculated. NRMSE assesses differences between paired original and imputed values while percent error assesses differences in the overall mean between the original and imputed datasets. As most of the imputation methods used here do not calculate imputed values using the mean (e.g. random forest, stochastic regression, expectation maximization), percent error may produce results that are slightly biased against these approaches. Whether NRMSE or percent error is better for evaluating the success of missing data methods will depend on whether one is trying to obtain the exact values of the original dataset or retain overall patterns in the data.

Listwise deletion was undoubtedly the worst method for handling missing data. The percent error for even 5% or 10% missing data exceeded that of 40% missingness for the six other methods. Note that for continuous data, listwise deletion with 40% missingness had a percent error of only 6.14% (exact percent errors are available in Supporting Information). With ordinal data, even the more sophisticated methods had percent errors between 3.9 and 7.4 for 30% missingness. Furthermore, once listwise deletion had been performed with 40% of the values missing there were only a handful of individuals (rows) left in the dataframes, and in two versions the entire dataframe was empty as all rows had at least one NA.

On the whole, all imputation methods were relatively successful at recovering the means of the MCAR and MAR datasets. More sophisticated forms of imputation (predictive mean matching, regression, random forest, and expectation maximization) performed much better than mean imputation or either deletion method.

## Discussion

For both continuous and ordinal data, no imputation or deletion method performed noticeably better than any others across all datasets. Overall, evaluating success of ordinal data proved more difficult than continuous data; the results are more inconsistent and even minor differences in which values were simulated as missing – thus affecting the underlying distribution of the datasets – seemed to have a greater impact on the final results. The success of all seven missing data methods was much worse for ordinal data. Even the lowest percent error for 5% missing ordinal data was greater than the percent error for 40% missing continuous data. This likely reflects problems inherent in ordinal paleopathology data: the low range of possible values and variable percentages of missingness. It is possible that ordinal paleopathology data are not well-suited for imputation. Imputation assumes that there are associations among variables in a dataset. If those associations are poor or absent, imputation will not work. Liao et al. (2014) have devised an “imputability measure” that “provides quantitative evidence of how well each missing value can be imputed by borrowing information from other variables or subjects” (p. 355). The authors warn that researchers should be cautious with variables that have poor imputability measures. While beyond the scope of this paper, it would be worth investigating the performance of ordinal paleopathology data when evaluated with this imputability measure.

An additional concern with the imputation of ordinal data is when scores are collapsed to binary variables such as presence/absence. There is little guidance on whether dichotomization should occur before imputation or after (Demirtas, 2007; Floden & Bell, 2019). Grobler and Lee (2019) found that imputing the original value and dichotomizing at the end can result in biased parameter estimates. They recommend “imput[ing] the binary variable even if intuitively this means throwing away potentially useful data” (p. 476). Furthermore, depending on the method

used, imputed values of ordinal variables may be converted to continuous variables which do not match the values required for analysis. As an example, if the values of an ordinal dataset range from 0-2, imputed values may be 1.2 or 0.7, neither of which are allowable values. In such circumstances, practitioners often round to the nearest whole integer that matches the data (Schafer 1997). Horton, Lipsitz & Parzen (2003) however, found that rounding may lead to inaccurate parameter estimates if the data are not normally distributed. Simple rounding is highly discouraged for nominal datasets because it imposes an order on the data that is not present in the original dataset (Galati, Seaton, Lee, Simpson & Carlin, 2014).

It has been noted that some of the “ordinal” data bioarchaeologists collect may actually be nominal, i.e., scores thought to have inherent ranking may not be so organized. Hens and Godde (2020) found that increasing palatal suture closure scores did not consistently reflect greater age at death and were therefore nominal rather than ordinal scores. Imputation of nominal data has received considerably less attention than ordinal data (Lang & Wu, 2017). As mentioned above, imputing non-ordinal variables may impose an order on the dataset that should not be present. Additional research is needed to clarify how imputation of nominal data impacts the dataset and parameter estimates.

Despite expert caution against pairwise deletion (Allison, 2001; Graham, 2012; Kang, 2013; van Buuren, 2018), this method performed reasonably well for both ordinal and continuous data, though it is difficult to fully test its success as it could not be assessed with NRMSE. Deletion methods cause a high rate of data loss that can be especially problematic if the data are MNAR. As paleopathologists generally have small sample sizes, any method that reduces the data further is suboptimal and may result in a biased dataset and decreased analytical power. The findings of this study agree with prior research on the use of listwise deletion (e.g.,

King, Honaker & Scheve, 1998). Listwise deletion was the worst at recovering the parameters of the original skeletal sample, particularly if the data are MNAR.

How much missing data is too much? There is little clear guidance on the maximum amount of missing data allowed before missing data methods or statistical analyses become too biased (Dong & Peng, 2013; Hardt, Max, Tamara & Wilfired, 2013; Meeyai, 2016; Saunders et al., 2006). The definition of a “small” amount of missingness varies from <5% to <20% missing (Little & Rubin, 2002; Tabachnick, Fidell & Ullman, 2007). Some statisticians caution that bias may occur in samples with more than 10% of the data missing and that samples with over 40% missing should be used for “hypothesis generating” only (Madley-Dowd, Hughes, Tilling & Heron, 2019, p. 64). Others recommend a maximum of 30% missing when imputing missing values and no more than 20% with sample sizes of 50 or lower (Hardt et al., 2013). Under tightly controlled circumstances, however, authors have managed to successfully impute and analyze datasets with much higher percentages of missing data. Madley-Dowd et al. (2019) imputed up to 80% missing MCAR and MAR data employing multiple imputation with auxiliary variables. Meeyai (2016) was able to recover unbiased parameters for MCAR data with 60% of the values missing for samples greater than  $n=1000$ . Unbiased regression coefficients have been obtained with a 90% fraction of missing information<sup>1</sup>, though statistical power dropped considerably after 50% missing even with  $m=20$  imputations (Graham, Olchowski & Gilreath, 2007).

The percent of missing information may not be the most important consideration when faced with missing data. Sample size is an important factor; 20% missing may have a greater impact with a sample size of 50 than with a sample size of 500 (Meeyai, 2016; Saunders et al.,

---

<sup>1</sup> Fraction of missing information (FMI) is “a measure of uncertainty about the values we would impute for missing elements” (Wagner, 2010; p. 224). It quantifies the amount of variance between imputed datasets and is affected by both the quality and quantity of available data. See Wagner (2010) for full discussion of FMI and the mathematical definition.

2006). The class of missingness – MCAR, MAR, or MNAR – will also affect how much missing data is acceptable. Even a small amount of MNAR values may result in a biased dataset no matter what imputation method is used (Dong and Peng, 2013; Tabachnick, Fidell & Ullman., 2007). Whether the missing values are among the independent and/or dependent variables will also impact the success of missing data methods and ultimate statistical analyses (Saunders et al., 2006). Overall, there is no consensus on the maximum amount of missing data and numerous other factors including the sample size, type of data, and patterns of missingness must inform one's approach to dealing with missing data.

Incorporating imputation methods as standard practice for handling missing data in bioarchaeology will contribute to the advancement of the field. Numerous scholars have drawn attention to the dearth of advanced statistical analyses in bioarchaeology (Agarwal, 2016; Konigsberg & Frankenberg, 2013; Zuckerman, Harper & Armelagos, 2016). Zuckerman et al. (2016) argue that paleopathology has been slow to adopt advances from other fields, instead relying on less rigorous methods without critical reflection. These shortcomings have hindered our ability to advance our understanding of human health throughout history in a way that is meaningful to modern populations. While recent years have seen a surge in more advanced analytical methods such as hazards models, survival analysis, and principal components analysis, much research – particularly paleopathology and trauma analysis – still depends on univariate statistical analyses. While summary statistics have their place, reliance on them represents an impediment to the advancement of paleopathology. Analyses such as the t-test or chi-square have strict statistical assumptions about the data such as normal distribution or equal variances; inappropriate use of such tests can result in erroneous results and thus unfounded inferences regarding health in the past. As paleopathology data rarely adhere to these assumptions, often

data must be aggregated or binned in ways that obscure vitally important patterns. More sophisticated analyses allow data to be explored in its complexity rather than by arbitrary bins, contributing to a more nuanced understanding of human health throughout history.

Another serious drawback of less advanced methods is their failure to account for the concerns raised by Wood et al. (1992) in the osteological paradox. Because of selective mortality, straightforward percentages or counts of pathology lesions from a skeletal sample will overestimate disease prevalence in the living population. The authors explain how aggregating data for more simple analyses prevents us from accounting for variation in individual frailty, obscuring not only important variation in disease experience but also the potential presence of subpopulations.

While imputation is a vital new step forward for handling missing data in bioarchaeology, one aspect that should be considered is the use of imputed values to draw conclusions about the lived experiences of past people. A great strength of bioarchaeology is the multiscalar ability to move seamlessly from populations to individuals. Consider, for example, a bioarcheological project that incorporates biological distance and stable isotopes to examine population affinity, migration, and diet. Multidimensional scaling (MDS) plots and carbon-nitrogen graphs show population groupings and diets of the sample. Typically, the researcher would discuss the life histories of aberrant individuals that do not follow the patterns of the rest of the group, such as individuals with unusual diets or who migrated from far away. If, however, missing variables were imputed – which is already standard practice in biodistance – the locations of individual points on the plots would be based on imputed, rather than actual values which may, or may not, represent the true values. Kenyhercz and Passalacqua (2016) recommend that when selecting an imputation technique for this type of individual-level analysis, the objective is to choose the most

accurate method. However, the goal of imputation is not to recover the exact values missing from the dataset, but to “preserve important characteristics of the data set as a whole” (Graham, 2009, p. 559). Imputed data are intended to examine overall patterns in the data, not the life experiences of single individuals. How do we reconcile these two disparate goals? Further research is needed to ensure that as bioarchaeology continues to grow and adopt methods from other fields, we apply those methods appropriately.

This study has several limitations. First, according to Rubin’s Rules (Rubin, 1987), statistical analyses are to be performed on each of the  $m$  datasets, and the final parameters of interest (e.g., p-values, confidence intervals, etc.) pooled at the very end using the equations designed by Rubin. The approach used in this paper, however, pools the multiply imputed datasets and assesses success at the end – a violation of Rubin’s Rules<sup>2</sup>. Failing to adhere to Rubin’s Rules can result in over- or underestimated parameters such as standard errors, confidence intervals and p-values. Second, the success of imputation and deletion methods will depend not only on the percent of missing data, but also on the sample size. The samples used here (ordinal  $n=287$ ; continuous  $n=369$ ) are large compared to most paleopathology datasets. Additional research is needed to compare the results found here with those from smaller samples. Third, this study tests the success of missing data methods on ordinal and continuous data separately, however many bioarchaeologists collect mixed data types and analyze them together. Future research is needed to identify which methods are successful at imputing mixed data including continuous, ordinal, and binary values. Finally, this paper investigates the best methods for imputing missing data, but ultimately bioarchaeologists are interested in how those imputed

---

<sup>2</sup> Recent research has shown that Rubin’s Rules can create biased variance estimates depending on the imputation method used and therefore may not always be the best approach for pooling parameter estimates. See Bartlett (2021) and Seamen et al. (2014) for further discussion.

values can be used to answer questions about human experiences in the past. Further research is needed to examine how imputed values perform in the types of statistical tests we use most often such as survival analysis, principle components analysis, ANOVAs, or even t-tests and chi-square tests to assess how imputed data affect our results and the conclusions we draw about past individuals and populations.

## CONCLUSION

The primary aim of this paper is to provide background on missing data methods, highlighting how imputation can be used to manage missing data in bioarchaeology and paleopathology. There are a number of approaches for handling missing data, including deleting data or imputing missing values. While each technique has advantages and disadvantages, imputation methods are recommended over deletion. Other fields in the natural and social sciences commonly use imputation. However, bioarchaeological research (apart from biodistance analyses) seldom uses imputation to handle missing data. This paper tests the ability of seven methods to yield effective parameter estimates when handling missing ordinal and continuous bioarchaeological data. Results demonstrate that no single method performs best in all circumstances, suggesting there is no “one-size-fits-all” solution to missing data problems. Listwise deletion is not recommended as it performed the worst for both ordinal and continuous data, introducing the most error into the dataset. While pairwise deletion preserved the characteristics of the original dataset, it is not recommended due to the loss of data. Ultimately, the best methods for handling missing data are the more sophisticated methods: stochastic regression, predictive mean matching, random forest, and expectation maximization. Overall, stochastic regression imputation consistently performed well across both ordinal and continuous data when assessed with either percent error or NRMSE.

Future studies should test the success of imputation on mixed continuous and ordinal data as well as how well imputation works with very small sample sizes. We hope these findings encourage the use of more advanced methods to manage missing data in bioarchaeology. With greater understanding of the limitations and structure of our data, bioarchaeologists can explore sources of bias more effectively and implement statistically rigorous analyses.

## **ACKNOWLEDGEMENTS**

We thank Yohannes Haile-Selassie, David Chapman, Gavin Svenson, Amanda McGee, Lyman Jellema, and everyone at the Cleveland Museum of Natural History for access to the Hamann-Todd Collection and support while collecting data during a pandemic. We also thank Matt Peeples for statistical guidance and feedback. Funding support was provided by a Wenner-Gren Dissertation Fieldwork Grant (#1183573640), an American Association of University Women – American Dissertation Fellowship, and a National Science Foundation SBE Postdoctoral Research Fellowship (#2104830) to A. Wissler. We are also grateful to the two anonymous reviewers whose comments significantly improved the quality of this paper. The authors declare no potential conflict of interest.

## **DATA AVAILABILITY STATEMENT**

The R code created for this study is openly available via the first author's GitHub Repository (Wissler, 2022). The raw paleopathology data that support the findings of this study are available from the Cleveland Museum of Natural History with permission from the first author. Data will be available by contacting the Collections Manager of Physical Anthropology.

## REFERENCES

Agarwal, S. C. (2016). Bone morphologies and histories: Life course approaches in bioarchaeology. *American Journal of Physical Anthropology*, 159(S61), 130-149.

Allison, P. D., (2001). *Missing Data*. Sage publications.

Athreya, S., & Wu, X. (2017). A multivariate assessment of the Dali hominin cranium from China: Morphological affinities and implications for Pleistocene evolution in East Asia. *American Journal of Physical Anthropology*, 164(4), 679-701.

Auerbach, B. M. (2011). Methods for estimating missing human skeletal element osteometric dimensions employed in the revised fully technique for estimating stature. *American Journal of Physical Anthropology*, 145(1), 67-80.

Auerbach, B. M., Raxter, M. H., & Ruff, C. (2005). If I only had a...: missing element estimation accuracy using the fully technique for estimating statures. *American Journal of Physical Anthropology*, 126(S40), 70.

Auerbach, B. M., Ruff, C. B. (2004). Human body mass estimation: a comparison of “morphometric” and “mechanical” methods. *American Journal of Physical Anthropology*, 125(4), 331– 342.

Bailey, B. E., Andridge, R., & Shoben, A. B., (2020). Multiple imputation by predictive mean matching in cluster-randomized trials. *BMC Medical Research Methodology*, 20(72),1-16.

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1), 5-37.

Barnard, J., & Meng, X-L. (1999). Applications of multiple imputation in medical studies: from AIDS to NHANES. *Statistical Methods in Medical Research*, 8(1), 17-36.

Bartlett, J. W. (2021). Reference-Based Multiple Imputation—What is the Right Variance and How to Estimate It. *Statistics in Biopharmaceutical Research*, 1-9.

Bethard, J. D. (2013). *The bioarchaeology of Inka resettlement practices: insight from biological distance analysis*. [Doctoral dissertation, University of Tennessee]. [https://trace.tennessee.edu/utk\\_graddiss/2399](https://trace.tennessee.edu/utk_graddiss/2399)

Bhaskaran, K. & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? *International Journal of Epidemiology*, 43(4), 1336-1339.

Bird, J. P., Martin, R., Akçakaya, H. R., Gilroy, J., Burfield, I. J., Garnett, S. T., Symes, A.,

Taylor, J., Şekercioğlu, Ç. H., & Butchart, S. H. (2020). Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology*, 34(5), 1252-1261.

Bodnar, L. M., Tang, G., Ness, R. B., Harger, G., & Roberts, J. M. (2006). Periconceptional multivitamin use reduces the risk of preeclampsia. *American Journal of Epidemiology*, 164(5), 470-477.

Bolhofner, K. L. (2017). *Conquest and conversion in Islamic period Iberia (A.D. 711-1490): A bioarchaeological approach* [Doctoral dissertation, Arizona State University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/conquest-conversion-islamic-period-iberia-d-711/docview/1901472223/se-2?accountid=13965>

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

Browning, B. L., & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *The American Journal of Human Genetics*, 98(1), 116-126.

Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *The American Journal of Human Genetics*, 103(3), 338-348.

Buikstra, J. E., & Ubelaker, D. H. (1994). *Standards for Data Collection from Human Skeletal Remains*. Fayetteville, Arkansas Archaeological Survey.

Burnett, S. E., Irish, J. D., & Fong, M. R. (2013). Wear's the problem? Examining the effect of dental wear on studies of crown morphology. In G.R. Scott & J.D. Irish (Eds.), *Anthropological Perspectives on Tooth Morphology: Genetics, Evolution, Variation*. (pp. 535-554). Cambridge University Press.

Capitanio, J. P., Blozis, S. A., Snarr, J., Steward, A., & McCowan, B. J. (2017). Do "birds of a feather flock together" or do "opposites attract"? Behavioral responses and temperament predict success in pairings of rhesus monkeys in a laboratory setting. *American Journal of Primatology*, 79(1), e22464. <https://doi.org/10.1002/ajp.22464>.

Clavel, J., Merceron, G., & Escarguel, G. (2014). Missing data estimation in morphometrics: how much is too much? *Systematic Biology*, 63(2), 203-218.

Cooke, R. S., Eigenbrod, F., & Bates, A. E. (2019). Projected losses of global mammal and bird ecological strategies. *Nature Communications*, 10(1), 1-8.

Cooke, R. S., Eigenbrod, F., & Bates, A. E. (2020). Ecological distinctiveness of birds and mammals at the global scale. *Global Ecology and Conservation*, 22(e00970), 1-14.

Costello, S., Brown, D. M., Noth, E. M., Cantley, L., Slade, M. D., Tessier-Sherman, B., Hammond, S. K., Eisen, E. A., & Cullen, M. R. (2014). Incident ischemic heart disease and recent occupational exposure to particulate matter in an aluminum cohort. *Journal of Exposure Science & Environmental Epidemiology*, 24(1), 82-88.

Dam, M. K., Hvidtfeldt, U. A., Tjønneland, A., Overvad, K., Grønbæk, M., & Tolstrup, J. S. (2016). Five year change in alcohol intake and risk of breast cancer and coronary heart disease among postmenopausal women: prospective cohort study. *British Medical Journal*, 353(i2314), 1-10.

De Leeuw, E. D., Hox, J. J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, 19:153-176.

Demirtas, H. (2007). Practical advice on how to impute continuous data when the ultimate interest centers on dichotomized outcomes through pre-specified thresholds. *Communications in Statistics – Simulation and Computation*, 36(4), 871-889.

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1), 1-38.

Divíšek, J., Chytrý, M., Beckage, B., Gotelli, N. J., Lososová, Z., Pyšek, P., Richardson, D. M., & Molofsky, J. (2018). Similarity of introduced plant species to native ones facilitates naturalization, but differences enhance invasion success. *Nature Communications*, 9(1), 1-10.

Dong, Y., & Peng, C-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus* 2(1), 1-17.

Ely, J. J., Zavaskis, T., & Lammey, M. L. (2013). Censored data analysis reveals effects of age and hepatitis C infection on C-reactive protein levels in healthy adult chimpanzees (*Pan troglodytes*). *Journal of Biomarkers*, 2013, 1-13.  
<http://dx.doi.org/10.1155/2013/709740>

Enders, C. K. (2010). *Applied Missing Data Analysis*. Guilford press.

Evans, C. B., & Smokowski, P. R. (2015). Prosocial bystander behavior in bullying dynamics: Assessing the impact of social capital. *Journal of Youth and Adolescence*, 44(12), 2289-2307.

Feilen, K. L., & Marshall, A. J. (2014). Sleeping site selection by proboscis monkeys (*Nasalis larvatus*) in West Kalimantan, Indonesia. *American Journal of Primatology*, 76(12), 1127-1139.

Feng, L., Moritz, S., Nowak, G., Welsh, A. H., & O'Neill, T. J. (2020). imputeR: A General Multivariate Imputation Framework. R package version 2.2.

Fichman, M., & Cummings, J. N. (2003). Multiple imputation for missing data: Making the most of what you know. *Organizational Research Methods*, 6(3), 282-308.

Finch, W. H. (2010). Imputation methods for missing categorical questionnaire data: A

comparison of approaches. *Journal of Data Science*, 8(3), 361-378.

Galati, J. C., Seaton, K. A., Lee, K. J., Simpson, J. A., & Carlin, J. B. (2014). Rounding non-binary categorical variables following multivariate normal imputation: evaluation of simple methods and implications for practice. *Journal of Statistical Computation and Simulation*, 84(4), 798-811.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kovári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., & Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications*, 5, 5257.

Glantz, M., Athreya, S., & Ritzman, T. (2009). Is Central Asia the eastern outpost of the Neandertal range? A reassessment of the Teshik-Tash child. *American Journal of Physical Anthropology*, 138(1), 45-61.

Godde, K., & Rangel González, M. (2022). Uncovering the hidden history of transpacific contact between Latin America, the Caribbean, and Asia through skeletal population genetics. *American Journal of Biological Anthropology*, 177(2), 357-364.

Gordon, A. D., Green, D. J., & Richmond, B. G. (2008). Strong postcranial size dimorphism in *Australopithecus afarensis*: results from two new resampling methods for multivariate data sets with missing data. *American Journal of Physical Anthropology*, 135(3), 311-328.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology* 60, 549-576.

Graham, J. W. (2012). *Missing Data: Analysis and Design*. Springer Science & Business Media.

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325-366). American Psychological Association.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation. *Prevention Science*, 8(3), 206-213.

Grebe, N. M., Fitzpatrick, C., Sharrock, K., Starling, A., & Drea, C. M. (2019). Organizational and activational androgens, lemur social play, and the ontogeny of female dominance. *Hormones and Behavior*, 115, 104554. <https://doi.org/10.1016/j.yhbeh.2019.07.002>.

Grilo, C., Koroleva, E., Andrásik, R., Bíl, M., & González-Suárez, M. (2020). Roadkill risk and

population vulnerability in European birds and mammals. *Frontiers in Ecology and the Environment*, 18(6), 323-328.

Grobler, A. C., & Lee, K. (2020). Multiple imputation in the presence of an incomplete binary variable created from an underlying continuous variable. *Biometrical Journal*, 62(2), 467-478.

Gunz, P., Mitteroecker, P., Neubauer, S., Weber, G. W., & Bookstein, F. L. (2009). Principles for the virtual reconstruction of hominin crania. *Journal of Human Evolution*, 57(1), 48-62.

Hardt, J., Max, H., Tamara, B., & Wilfried, L. (2013). Multiple imputation of missing data: A simulation study on a binary response. *Open Journal of Statistics*, 5(3), 370-378.

Henriksson, A., Zhao, J., Dalianis, H., & Boström, H. (2016). Ensembles of randomized trees using diverse distributed representations of clinical events. *BMC Medical Informatics and Decision Making*, 16(2), 85-95.

Hens, S. M., & Godde, K. (2020). New approaches to age estimation using palatal suture fusion. *Journal of Forensic Sciences*, 65(5), 1406-1415.

Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A potential for bias when rounding in multiple imputation. *The American Statistician*, 57(4), 229-232.

Howell, D. C. (2007). The Treatment of Missing Data. In W. Outhwaite & S. P. Turner (Eds.), *The Sage Handbook of Social Science Methodology* (pp. 208-224). Sage Publications.

Howells, W. W. (1973). Cranial variation in man: A study by multivariate analysis of patterns of difference among recent human populations. *Papers of the Peabody Museum of Archaeology and Ethnology*, 67, 1e259.

Hui, R., D'Atanasio, E., Cassidy, L. M., Scheib, C. L., & Kivisild, T. (2020). Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Scientific Reports*, 10(1), 1-8.

Irving-Pease, E. K., Muktupavela, R., Dannemann, M., & Racimo, F. (2021). Quantitative Human Paleogenetics: what can ancient DNA tell us about complex trait evolution? *Frontiers in Genetics*, 12, 703541. <https://doi.org/10.3389/fgene.2021.703541>

Ishiya, K., Mizuno, F., Wang, L., & Ueda, S. (2019). MitoIMP: A computational framework for imputation of missing data in low-coverage human mitochondrial genome. *Bioinformatics and Biology Insights*, 13. <https://doi.org/10.1177/1177932219873884>

Jardim, L., Bini, L. M., Diniz-Filho, J. A. F., & Villalobos, F. (2021). A cautionary note on phylogenetic signal estimation from imputed databases. *Evolutionary Biology*, 48(2), 246-258.

Jensen, T. Z. T., Niemann, J., Iversen, K. H., Fotakis, A. K., Gopalakrishnan, S., Vågene, Å. J., Pedersen, M. W., Sinding, M. H. S., Ellegaard, M. R., Allentoft, M. E., Lanigan, L. T., Taurozzi, A. J., Nielsen, S. H., Dee, M. W., Mortensen, M. N., Christensen, M. C., Sørensen, S. A., Collins, M. J., Gilbert, M. T. P., ... Schroeder, H. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nature Communications*, 10(1).

Kaiser, J. (1983). *The Effectiveness of Hot-deck Procedures in Small Samples*. Proceedings of the Annual Meeting of the American Statistical Association. Ann Arbor, MI.

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), 402.

Kenyhercz, M. W., & Passalacqua, N. V. (2016). Missing data imputation methods and their performance with biodistance analyses. In M. A. Pilloud & J. T. Hefner (Eds.), *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives* (pp. 181-194). Academic Press.

Kenyhercza, M., Passalacquac, N. V., & Hefner, J. T. (2019). Missing data imputation using morphoscopic traits and their performance in the estimation of ancestry. *Journal of Forensic Sciences*, 54(5), 985-995.

Kerr, N. W. (1988). A method of assessing periodontal status in archaeologically derived skeletal material. *Journal of Paleopathology*, 2(2), 67-78.

King, G., Honaker, J., Joseph, A., & Scheve, K. (1998). *List-wise deletion is evil: What to do about missing data in political science*. Annual Meeting of the American Political Science Association, Boston, MA.

Kleinke, K. (2018). Multiple imputation by predictive mean matching when sample size is small. *Methodology*, 14(1), 3-15.

Konigsberg, L. W., & Frankenberg, S. R. (2013). Bayes in biological anthropology. *American Journal of Physical Anthropology*, 152(S57), 153-184.

Kramer, A., & Konigsberg, L. W. (1999). Recognizing species diversity among large-bodied hominoids: A simulation test using missing data finite mixture analysis. *Journal of Human Evolution*, 36(4), 409-421.

Lang, K. M., & Wu, W. (2017). A comparison of methods for creating multiple imputations of nominal variables. *Multivariate Behavioral Research*, 52(3), 290-304.

Lassale, C., Tzoulaki, I., Moons, K. G., Sweeting, M., Boer, J., Johnson, L., Huerta, J. M.,

Agnoli, C., Freisling, H., & Weiderpass, E. (2018). Separate and combined associations of obesity and metabolic health with coronary heart disease: a pan-European case-cohort analysis. *European Heart Journal*, 39(5), 397-406.

Leite, W., & Beretvas, S. N. (2010). The performance of multiple imputation for Likert-type items with missing data. *Journal of Modern Applied Statistical Methods*, 9(1), 64-74.

Liao, S. G., Lin, Y., Kang, D. D., Chandra, D., Bon, J., Kaminski, J., Sciurba, F. C., Tseng, G. C. (2014). Missing value imputation in high-dimensional phenomic data: imputable or not, and how? *BMC Bioinformatics*, 15(1), 1-12.

Lindhe, J., Haffaiee, A., & Socransky, S. (1983). Progression of periodontal disease in adult subjects in the absence of periodontal therapy. *Journal of Clinical Periodontology*, 10(4), 433-442.

Little, R. J. (1988). Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*, 6(3), 287-296.

Little, R. J., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Madley-Dowd, P., Hughes, R., Tilling, K., & Heron, J. (2019). The proportion of missing data should not be used to guide decisions on multiple imputation. *Journal of Clinical Epidemiology*, 110, 63-73.

Martiniano, R., Cassidy, L. M., Ó'Maoldúin, R., McLaughlin, R., Silva, N. M., Manco, L., Fidalgo, D., Pereira, T., Coelho, M. J., Serra, M., Burger, J., Parreira, R., Moran, E., Valera, A. C., Porfirio, E., Boaventura, R., Silva, A. M., & Bradley, D. G. (2017). The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genetics*, 13(7).

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueiredo, A. J. (2007). *Missing data: A Gentle Introduction*. New York: Guilford Press.

Meeyai, S. (2016). Logistic regression with missing data: A comparison of handling methods, and effects of percent missing values. *Journal of Traffic and Logistics Engineering*, 4(2), 128-134.

Miller, K. A. (2015). *Family, "foreigners," and fictive kinship: a bioarchaeological approach to social organization at late classic Copan*. [Doctoral dissertation, Arizona State University]. ProQuest Dissertations & Theses Global.

Mizuno, F., Kumagai, M., Kurosaki, K., Hayashi, M., Sugiyama, S., Ueda, S., & Wang, L. (2017). Imputation approach for deducing a complete mitogenome sequence from low-depth-coverage next-generation sequencing data: application to ancient remains from the Moon Pyramid, Mexico. *Journal of Human Genetics*, 62(6), 631-635.

Morelli, T., Moss, K. L., Preisser, J. S., Beck, J. D., Divaris, K., Wu, D., & Offenbacher, S. (2018). Periodontal profile classes predict periodontal disease progression and tooth loss. *Journal of Periodontology*, 89(2), 148-156.

Morris, T. P., White, I. R., & Royston, P. (2014). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, 14(1), 1-13.

Musil, C. M., Warner, C. B., Yobas, P. K., & Jones, S. L. (2002). A comparison of imputation techniques for handling missing data. *Western Journal of Nursing Research*, 24(7), 815-829.

Muzzall, E. (2021). A novel ensemble machine learning approach for bioarchaeological sex prediction. *Technologies*, 9(2), 23. <https://doi.org/10.3390/technologies9020023>.

Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures*, 5(4), 297-310.

Neale, B. M. (2010). Introduction to linkage disequilibrium, the HapMap, and imputation. *Cold Spring Harbor Protocols*, 5(3).

Nelwamondo, F. V., Mohamed, S., & Marwala, T. (2007). Missing data: A comparison of neural network and expectation maximization techniques. *Current Science*, 91(11), 1514-1521.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, 6(3), 328-362.

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4), 372-411.

Ong, G. (1998). Periodontal disease and tooth loss. *International Dental Journal*, 48(S3), 233-238.

Ordonez, A., & Svenning, J-C. (2017). Consistent role of Quaternary climate change in shaping current plant functional diversity patterns across European plant orders. *Scientific Reports*, 7(1), 1-11.

Osborne, J. W. (2013). *Best Practices in Data Cleaning*. Sage Publication.

Pacifici, M., Santini, L., Di Marco, M., Baisero, D., Francucci, L., Marasini, G. G., Visconti, P., & Rondinini, C. (2013). Generation length for mammals. *Nature Conservation*, 5, 89-94.

Pacheco-Forés, S. I. (2020). *Ritual violence and the perception of social difference: Migration and human sacrifice in the Epiclassic Basin of Mexico*. [Doctoral dissertation, Arizona State University]. ProQuest Dissertations & Theses Global.

Pasaniuc, B., Rohland, N., McLaren, P. J., Garimella, K., Zaitlen, N., Li, H., Namrata, G., Neale, B. M., Daly, M. J., Sklar, P., Sullivan, P. F., Bergan, S., Moran, J.L., Hultman, C.M., Lichtenstein, P., Magnusson, P., Purcell, S. M., Haas, D. W., Liang, L. ... & Price, A. L. (2012). Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genetics*, 44(6), 631-635.

Paul, K. S. (2017). *Developing an infrastructure for biodistance research using deciduous dental phenotypes* [Doctoral dissertation, Arizona State University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/developing-infrastructure-biodistance-research/docview/1973124808/se-2?accountid=13965>

Paul, K. S., Stojanowski, C. M., & Butler, M. M. (2013). Biological and spatial structure of an early classic period cemetery at Charco Redondo, Oaxaca. *American Journal of Physical Anthropology*, 152(2), 217-229.

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology*, 9, 157-166.

Pepinsky, T. B. (2018). A note on listwise deletion versus multiple imputation. *Political Analysis*, 26(4), 480-488.

Petersen, C. B., Bauman, A., Grønbæk, M., Helge, J. W., Thygesen, L. C., & Tolstrup, J. S. (2014). Total sitting time and risk of myocardial infarction, coronary heart disease and all-cause mortality in a prospective cohort of Danish adults. *International Journal of Behavioral Nutrition and Physical Activity*, 11(1), 1-11.

Prevedorou, E., & Stojanowski, C. M. (2017). Biological kinship, postmarital residence and the emergence of cemetery formalisation at prehistoric Marathon. *International Journal of Osteoarchaeology*, 27(4), 580-597.

Quintero, M., & LeBoulluec, A. (2018). Missing data imputation for ordinal data. *International Journal of Computer Applications*, 181(5), 10-16.

Ramseier, C. A., Anerud, A., Dulac, M., Lulic, M., Cullinan, M. P., Seymour, G. J., Faddy, M. J., Bürgin, W., Schätzle, M., & Lang, N. P. (2017). Natural history of periodontitis: Disease progression and tooth loss over 40 years. *Journal of Clinical Periodontology*, 44(12), 1182-1191.

Rathmann, H., Stoyanov, R., & Posamentir, R. (2022). Comparing individuals buried in flexed and extended positions at the Greek colony of Chersonesos (Crimea) using cranial metric, dental metric, and dental nonmetric traits. *International Journal of Osteoarchaeology*, 32(1), 49-63.

Redfern, R., & Hefner, J. T. (2019). "Officially absent but actually present": Bioarchaeological

evidence for population diversity in London during the Black Death, AD 1348–50. In M. Mant & A. Holland (Eds.), *Bioarchaeology of Marginalized People* (pp. 69-114). Academic Press.

Rockel, T. (2020). missMethods: Methods for Missing Data. R package version 0.2.0.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Saunders, J. A., Morrow-Howell, N., Spitznagel, E., Doré, P., Proctor, E. K., & Pescarino, R. (2006). Imputing missing data: A comparison of methods for social work researchers. *Social Work Research*, 30(1), 19-31.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8(1), 3-15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147.

Scherer, A. K. (2007). Population structure of the classic period Maya. *American Journal of Physical Anthropology*, 132(3), 367–380.

Seaman, S. R., White, I. R., & Leacy, F. P. (2014). Comment on “Analysis of longitudinal trials with protocol deviations: A framework for relevant, accessible assumptions, and inference via multiple imputation,” by Carpenter, Roger, and Kenward. *Journal of Biopharmaceutical Statistics*, 24(6), 1358-1362.

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *American Journal of Epidemiology*, 179(6), 764-774.

Stefan, V.H. (2004). Assessing intrasample variation: Analysis of Rapa Nui (Easter Island) museum cranial collections example. *American Journal of Physical Anthropology*, 124(1), 45–58.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest – non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112-118.

Stojanowski, C. M., & Johnson, K. M. (2015). Observer error, dental wear, and the inference of new world sundadonty. *American Journal of Physical Anthropology*, 156(3), 349-362.

Strauss, R. E., Atanassov, M. N., De Oliveira, J. A. (2003). Evaluation of the principal-

component and expectation-maximization methods for estimating missing data in morphometric studies. *Journal of Vertebrate Paleontology*, 23(2), 284–296.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using Multivariate Statistic*. Pearson.

Tang, F., & Ishwaran, H. (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 10(6), 363-377.

Taugourdeau, S., Villerd, J., Plantureux, S., Huguenin-Elie, O., & Amiaud B. (2014). Filling the gap in functional trait databases: Use of ecological hypotheses to replace missing data. *Ecology and Evolution*, 4(7), 944-958.

Thompson, A. R., Hedman, K. M., & Slater, P. A. (2015). New dental and isotope evidence of biological distance and place of origin for mass burial groups at Cahokia's mound 72. *American Journal of Physical Anthropology*, 158(2), 341-357.

Turney, K. (2015). Paternal incarceration and children's food insecurity: A consideration of variation and mechanisms. *Social Service Review*, 89(2), 335-367.

van Buuren, S. (2018). *Flexible Imputation of Missing Data*. CRC press Taylor & Francis Group.

van Buuren, S., and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3):1-67.

van Ginkel, J. R., Linting, M., Rippe, R. C., & van der Voort, A. (2020). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of Personality Assessment*, 102(3), 297-308.

Vink, G., Frank, L. E., Pannekoek, J., & van Buuren, S. (2014). Predictive mean matching imputation of semicontinuous variables. *Statistica Neerlandica*, 68(1), 61-90.

Wagner, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Public Opinion Quarterly*, 74(2), 223-243.

Waljee, A. K., Mukherjee, A., Singal, A. G., Zhang, Y., Warren, J., Balis, U., Marrero, J., Zhu, J., & Higgins, P. D. (2013). Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open*, 3(8).

Weng, S. F., Vaz, L., Qureshi, N., & Kai, J. (2019). Prediction of premature all-cause mortality: A prospective general population cohort study comparing machine-learning and standard epidemiological approaches. *PloS One*, 14(3), e0214365.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.

Wissler, A. (2021). *Engaging the osteological paradox: A study of frailty and survivorship in*

*the 1918 influenza pandemic* [Doctoral dissertation, Arizona State University]. ProQuest Dissertations & Theses Global. <https://www.proquest.com/dissertations-theses/engaging-osteological-paradox-study-frailty/docview/2564838484/se-2?accountid=4485>

Wissler, A. (2022). *Bioarchaeology-missingdata*. GitHub. <https://github.com/acwissler/bioarchaeology-missingdata>

Wood, J. W., Milner, G. R., Harpending, H. C., Weiss, K. M., Cohen, M. N., Eisenberg, L. E., Hutchinson, D. L., Jankauskas, R., Cesnys, G., Česnys, G., Katzenberg, A., Lukacs, J. R., McGrath, J. W., Roth, E. A., Ubelaker, D. H., & Wilkinson, R. G. (1992). The osteological paradox: Problems of inferring prehistoric health from skeletal samples [and comments and reply]. *Current Anthropology*, 33(4), 343-370.

Zambrano-Bigiarini, M. (2020). hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.4-0. <https://github.com/hzambran/hydroGOF>.

Zeka, A., Zanobetti, A., & Schwartz, J. (2006). Individual-level modifiers of the effects of particulate matter on daily mortality. *American Journal of Epidemiology*, 163(9), 849-859.

Zhang, Z. (2016). Missing data imputation: Focusing on single imputation. *Annals of Translational Medicine*, 4(1).

Zuckerman, M. K., Harper, K. N., & Armelagos, G. J. (2016). Adapt or die: Three case studies in which the failure to adopt advances from other fields has compromised paleopathology. *International Journal of Osteoarchaeology*, 26(3), 375-383.

## FIGURE LEGENDS

**Figure 1.** Barplot Showing Imputation Results Using Normalized Root Mean Square Error (NRMSE) for Ordinal Data with Percent Missing Values. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization.

**Figure 2.** Barplot Showing Imputation Results Using Percent Error for Ordinal Data with Percent Missing Values. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization; Pairwise = pairwise deletion; Listwise = listwise deletion.

**Figure 3.** Barplot Showing Imputation Results Using Normalized Root Mean Square Error (NRMSE) for Ordinal Data with MCAR, MAR, and MNAR Missing Data. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization.

**Figure 4.** Barplot Showing Imputation Results Using Percent Error for Ordinal Data with MCAR, MAR, and MNAR Missing Data. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization; Pairwise = pairwise deletion; Listwise = listwise deletion.

**Figure 5.** Barplot Showing Imputation Results Using Normalized Root Mean Square Error (NRMSE) for Continuous Data with Percent Missing Values. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization.

**Figure 6.** Barplot Showing Imputation Results Using Percent Error for Continuous Data with Percent Missing Values. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization; Pairwise = pairwise deletion; Listwise = listwise deletion.

**Figure 7.** Barplot Showing Imputation Results Using Normalized Root Mean Square Error (NRMSE) for Continuous Data with MCAR, MAR, and MNAR Missing Data. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization.

**Figure 8.** Barplot Showing Imputation Results Using Percent Error for Continuous Data with MCAR, MAR, and MNAR Missing Data. Mean = mean; PMM = predictive mean matching; Reg = regression; RF = random forest; EM = expectation maximization; Pairwise = pairwise deletion; Listwise = listwise deletion.