

Article

A Categorical Approach to Age Estimation Using the Auricular Surface of the Ilium

Elena Sierra-Serrano ¹  and John Albanese ^{2,*} ¹ Department of Archaeology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada; elena_sierra@sfu.ca² Department of Integrative Biology, University of Windsor, 401 Sunset Avenue, Windsor, ON N9B 3P4, Canada

* Correspondence: albanese@uwindsor.ca

Abstract

Background/Objectives: Age estimation from skeletal remains is an important step in reconstructing the identity of deceased individuals. The auricular surface of the ilium, known to be a strong predictor of age at death, could benefit from methodological refinements that have previously shown potential using only the retroauricular area. The aim of this paper is to showcase the strengths and weaknesses of an approach that requires greater complexity than using only the retroauricular area. **Methods:** A series of logistic regressions were performed to classify 105 individuals from the Terry Collection into three age groups: younger, middle-aged, and older adult. Stratified sample selection provided a more all-encompassing reflection of the diverse experiences of those individuals who carried the morphologies observed. The features scored included microporosity, macroporosity, presence of dense bone, morphology of the auricular rim, linear organization, and retroauricular area. **Results:** Overall, the substantial increase in time and effort does increase accuracy of the classifications, but not by the same magnitude. Results reflect an improved classification of younger individuals compared to using the retroauricular area alone. With the present method, individuals were classified into their correct group 66% of the time, and results that narrowed down classification to two of the three age groups were obtained in a further 26%. **Conclusions:** The utility of this method is based on clarity and reproducibility, addressing the known issue of “attraction to the middle”, and its potential to correctly identify adults as beyond 65 years old.

Keywords: forensic anthropology; skeletal age at death; skeletal age indicators; retroauricular area; age mimicry; sacroiliac joint; logistic regression; Terry Collection; intraobserver error



Academic Editor: Sara C. Zapico

Received: 13 May 2026

Revised: 25 July 2026

Accepted: 27 July 2026

Published: 29 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The advantages of the auricular surface of the ilium as a crucial source of data for age estimation have long been recognized [1–3]. The analysis of age-at-death estimation from skeletal remains constitutes a description of personal identity that is fundamental for preliminary identification of human remains. The first systematic method for estimating age at death from the auricular surface was proposed by Lovejoy et al. [2]. They identified five age skeletal features that together informed eight possible phases of the aging process. Nonetheless, their method did not calculate prediction intervals of age, being most applicable to paleodemography. Next, Osborne et al. [3,4] suggested to collapse the eight phases from Lovejoy et al. into six phases and created a method for calculating 95% confidence intervals for each of the phases. This modification revised the Lovejoy et al. method making it more applicable to forensic contexts. Buckberry & Chamberlain [1] noticed how

chronological change occurs at a different age for each of the individual skeletal features. Their data-collection was feature-specific in that a quantitative system assigned a score to each feature independently, eliminating age phases. They created age ranges specific to certain composite scores using their own Bayesian approach to correct age mirroring.

Various areas and morphologies of the auricular surface have been observed as part of age estimation assessments. Lovejoy et al. [2] proposed the analysis of grain and density, macroporosity, billowing, striations, apex, transverse organization, and the retroauricular area to investigate age at death from skeletal remains. On the other hand, Buckberry & Chamberlain [1] described the observation of microporosity, macroporosity, dense bone, apical changes, and transverse organization. The retroauricular area was discarded by these authors due to concerns of being a “poor” indicator of age. More recently [5], it was demonstrated that variation in the retroauricular area could be recorded consistently, and its score is one of the single best sources of information for estimating age.

Several modifications of the methodology contributed to the utility of the retroauricular area as an indicator of age [5]. Firstly, an all-sexes, age-balanced group of individuals was selected to represent a wide range of age-related variation. Secondly, logistic regression was used to classify individuals into one of three age groups (younger, middle-aged, and older adults). Thirdly, the data collection protocol for this feature was developed to prioritize clarity and reproducibility of this process. The results obtained demonstrated the ability of this feature to provide quick and accurate estimates of age-at-death for older adults. However, younger and middle-aged adults could not be distinguished from each other with high certainty. Hence, the possibility of missing necessary information from adjacent skeletal features by using one single age indicator cannot be eliminated. Nonetheless, these three refinements proved key, and contributed to addressing several historical challenges in methodology and practice.

For decades, anthropologists have been grappling with the selection of diverse reference samples for age estimation methods [6,7]. In recent iterations of these discussions, it has been exposed how common sources of reference material, skeletal collections, are inherently biased [8–11]. Among the issues raised, age mimicry has been subject to much debate [6,7,12,13]. In such instances, age estimation outside the initial reference sample mimics the distribution of ages in the reference group. According to Masset [7], it is assumed that in any reference population, over- and under-estimation errors should cancel each other out. However, this is not the case since younger individuals can only be overestimated and older individuals can only be underestimated. This trend has been referred to as “attraction to the middle” [7,14]. Moreover, a typical flaw in sample selection is an overrepresentation of individuals who died between 50–70 years old. This asymmetry further prevents random errors from truly cancelling each other out. The sampling strategy and statistical approach in this paper aim to mitigate the issue of “attraction to the middle” by maximizing the age-related variation included in the reference sample.

Rather than stemming from biologically informative factors, “attraction to the middle” has been associated with the use of regression-based techniques [15–18]. Systematic overestimation of younger individuals and underestimation of older ones creates a pattern of non-normally distributed errors, which does not satisfy the requirements of linear regression [15,16]. Despite its name, logistic regression is primarily a classification technique rather than a regression. This calculation proposes an individualized model that explores how the morphological information collected can infer which age group an individual belongs to and with what certainty. Instead of attempting to fit a linear trend to a disperse set of data, logistic regression classifies each of the observations as belonging or not to a set of mutually exclusive nominal categories (yes/no, above/below, present/absent, etc.) [19]. The certainty of the classification is also recorded through a probability value.

Logistic regression is a consistent and easy-to-interpret technique that has proven to fit age estimation using three different age groups [5].

Another solution to the potential pitfalls of regression-based methods are Bayesian approaches. Through Bayes' theorem, the age at death that is tied to the observed skeletal morphology can be estimated if two other parameters are available: (1) a prior distribution of how skeletal morphologies are tied to age in reference data, and (2) the mortality distribution for said reference [20]. Even though some priors may be more appropriate to use as a reference than others, this approach has still been considered as more suitable than a method created from a reference sample "lacking any corrections" [12]. In addition, in a survey by Garvin & Passalacqua [21] only 8.9% of the respondents reported using the application of Bayes' theorem (transition analysis) to determine the final age estimate based on input from multiple skeletal regions, when given a complete human skeleton. Although a discussion of the applicability of transition analysis for age-at-death estimation is beyond the scope of this paper, it is worth noting that the retroauricular area approach [5] contains a reference sample *with* corrections.

Lastly, the direct impacts of reproducibility errors have not been assessed in detail. Reproducibility (sometimes alluded to as "reliability") refers to the ability of researchers to obtain consistent results in the subsequent use of a method. Notably, lack of consistency in the age estimation would be undesirable. In addition, the concepts "reproducibility" or "reliability" are sometimes not explicitly mentioned. A variety of other terms have also been used, such as "blind test" [2], "blindfold test" [22], "double determination" [23], or "concordance between intraexaminer [. . .]" [24]. Moreover, age-estimation methods do not necessarily report key aspects of reproducibility testing, including who conducted the assessment, which data were used, and what results were obtained. Greater standardization and transparency in these steps could be useful for understanding reproducibility of a method. In this context, the retroauricular area approach [5] created clear scoring criteria and included a simple reproducibility assessment consistent with the simplicity of the procedure itself. The present paper expands this evaluation through a more detailed analysis.

Building on the method and results obtained when using only the retroauricular area [5], the main objective of this paper is to apply the approach to the entire auricular surface (including the retroauricular area). This approach includes a balanced sampling strategy to mitigate known biases in skeletal collections, as well as the use of logistic regression and an in-depth reproducibility analysis. As the auricular surface method includes the observation of six skeletal features, it requires significantly more effort than using only the retroauricular area. Hence, a secondary aim is to present the strengths and weaknesses of this more time-consuming approach. This re-assessment has the potential to increase reproducibility of scoring variation in the auricular surface and contribute to the age-informative value of this area.

2. Materials and Methods

2.1. A Stratified Approach for Sample Selection

The individuals of this study were selected from the Terry Collection, curated at the Smithsonian Institution [11]. Although the majority of individuals in this collection died during the first half of the 20th century, this skeletal repository has been carefully and successfully sampled to develop methods for sex and stature estimation that are applicable in forensic contexts [25–28]. This same subgroup of individuals has contributed to reviewing the use of the retroauricular area of the ilium as an indicator of age at death [5]. The group comprises 105 individuals (Figure 1), whose age at death ranges from 22 to 101 years old, and with 58 individuals documented as female and 47 as male.

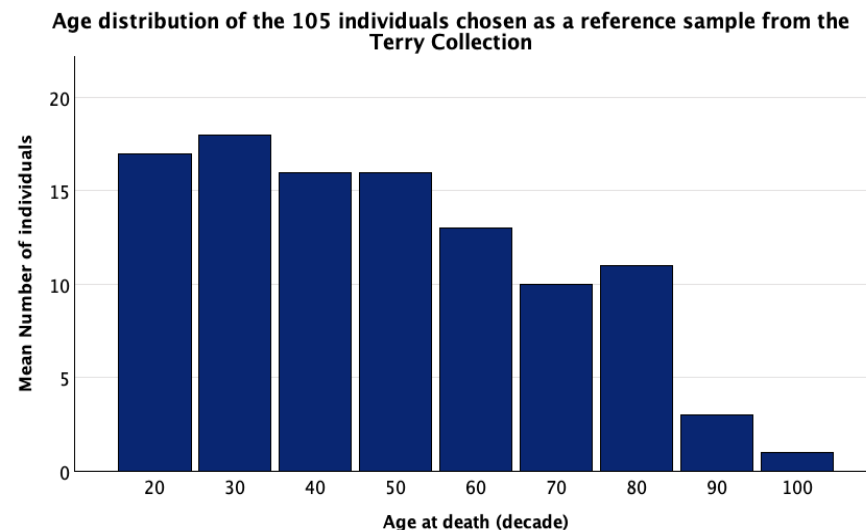


Figure 1. Age distribution of the individuals who comprise the reference sample in this study. Each bar represents a 10-year interval of age, 20–29 years old, 30–39 years old, etc. Individuals selected are curated at the Terry Collection housed at the Smithsonian Institution.

This group of 105 individuals was obtained through a stratified sampling approach, targeting specific subgroups to ensure their presence in the final sample [19]. These subgroups were created by incorporating all available information about the individuals in the sample, such as documented sex, verified ages, and preservation status. The goal of this sampling approach was to create a reference group of individuals that reflects the wide diversity associated with the aging process. This sampling approach was also informed by the known recommendations for identified reference collections [8,9,27]. Individuals with ages not verified were avoided, as well as those with verified ages that were multiples of five, due to a risk of “age heaping” [9]. Individuals were also excluded when there were too many available for a specific age range, or if the auricular surface was damaged. Similar number of individuals documented as female or male were included in each decade of age. Documented sex was used in this case as a proxy to include a broad range of variation, and not with the aim of creating a distinction of aging patterns based on biological sex. If there were multiple possible individuals that met all criteria, a random selection was performed from that pool.

2.2. Data Collection Protocol

The protocol for this research was developed as a two-stage process. An initial pilot study was performed using the Grant Collection housed at the University of Toronto [29]. The purpose of this step was to generate a baseline of the range of variation for each trait. To this end, 25 to 40 individuals distributed between 18 and 85 years old, mostly documented as male and with undamaged sacroiliac joints, were seriated in several rounds following different morphologies. Upon a systematic revision of the traits observed by Lovejoy et al. [2] and Buckberry & Chamberlain [1], the present study created a new data collection protocol (Supplementary Materials). This procedure was then further revised based on seriation and analysis of morphological features on individuals from the Terry Collection curated at the Smithsonian Institution. The two-stage process ensured that a broad array of variation representing individuals in a variety of contexts was captured in the scoring criteria.

The final protocol was also designed to prioritize ease of use and reproducibility of the results. Existing scoring instructions for known indicators of age were refined to fit the variation observed during the seriation studies. The written protocol systematically provides

both a description of the feature being observed (including its anatomical location [30], and general pattern of aging), as well as specific scoring instructions. If the observer appears to be undecided between two scores, specific guidance is provided on which score to choose depending on the feature. In many instances, it is instructed to choose the lowest of the two scores the observer is hesitant about. This attempts to introduce a known directionality and predictability to potential discrepancies between observers. Photographs of reference morphologies are also provided with the protocol, together with the associated score that the first author (ESS) assigned to those individuals. This strategy also attempted to maximize observer reproducibility and overall usability of the method.

Six age-related features were observed on the auricular surface of the ilium: microporosity, macroporosity, dense bone, formation of the auricular rim, linear organization, and retroauricular area. Each of the first five features was scored in three different sections of the auricular surface: upper, middle, and lower. On the other hand, the retroauricular area was scored in two sections: upper and lower. Features were divided in different sections to facilitate observer reproducibility by focusing on very specific areas during data collection. Hence, the total number of variables recorded was 17. The Supplementary Materials can be consulted for reference photographs, scoring instructions, and a data collection sheet.

Several parameters were kept constant during the analysis. Scores were taken from the left side if both sides of the os coxa were available and not damaged. In the case of those eventualities, data were obtained from the right os coxa. Data were collected from all areas in the data collection protocol that were not damaged. If an area was too damaged to obtain a score, this was also indicated on the collection sheet. Light conditions remained constant throughout the study, and no magnification lenses were used.

Intraobserver error was examined on a subgroup of 29 individuals (15 documented as female and 14 as male), aged 23 to 97. Scores from these individuals were collected two times and in both instances, the researcher performing data collection was not aware of the documented age of each individual. To obtain this second round of data over a wide range of dates, and because of the timing of the collection process, between 2 and 14 days had passed between the first and second round of scores depending on the individual. This strategy ensured that intraobserver error is applicable to any time span ranging from two days to two weeks. Intraobserver error was assessed by calculating percentage of agreement and Cohen's kappa statistic.

2.3. Data Transformation

The protocol requires the user to score 17 variables separately for ease and consistency of the data collection. Subsequently, these variables are summed to lessen the impact of variation, and to allow for a more robust statistical approach containing fewer predictor variables in logistic regression. Hence, composite variables were created to contain the information of the different features in all sections ('IIAurMiSUM', 'IIAurMaSUM', 'IIAurDenSUM', 'IIAurRimSUM', 'IIAurRetSUM', 'IIAurLinSUM'). Table 1 below comprises a list of all variables scored in the protocol. All variables, except for linear organization were known to be positively correlated with age, meaning that higher scores typically correspond to older individuals. On the other hand, linear organization is known to be negatively correlated with age, as higher scores in this variable typically correspond to younger individuals. Subsequently, the scores obtained on linear organization were reversed and a new variable was created ('IIAurLinSUM_r'). After all data were collected and before performing further analyses, individuals were removed if data in any variable were missing.

Table 1. The different skeletal features and variables scored, each with a brief description and an indication of which composite variable they are summed into. Full scoring details can be consulted in the Supplementary Materials.

Skeletal Feature	Variable Scored	Description	Variable in the Equation
Microporosity	IIAurMiUp	This scale scores presence of microporosity in the upper section of the auricular surface.	IIAurMiSUM
	IIAurMiMid	This scale scores presence of microporosity in the middle section of the auricular surface.	
	IIAurMiLow	This scale scores presence of microporosity in the lower section of the auricular surface.	
Macroporosity	IIAurMaUp	This scale scores presence of macroporosity in the upper section of the auricular surface.	IIAurMaSUM
	IIAurMaMid	This scale scores presence of macroporosity in the middle section of the auricular surface.	
	IIAurMaLow	This scale scores presence of macroporosity in the lower section of the auricular surface.	
Dense bone	IIAurDenUp	This scale scores presence of dense bone in the upper section of the auricular surface.	IIAurDenSUM
	IIAurDenMid	This scale scores presence of dense bone in the middle section of the auricular surface.	
	IIAurDenLow	This scale scores presence of dense bone in the lower section of the auricular surface.	
Rim formation	IIAurRimAp	This scale scores the morphology of the auricular rim at the point where the arcuate line meets the auricular surface (the apex).	IIAurRimSUM
	IIAurRimSup	This scale scores the morphology of the superior portion of the auricular rim.	
	IIAurRimInf	This scale scores the morphology of the inferior portion of the auricular rim.	
Linear organization	IIAurLinUp	This scale scores linearity in the distribution of certain patterns in the upper section of the auricular surface.	IIAurLinSUM_r *
	IIAurLinMid	This scale scores linearity in the distribution of certain patterns in the middle section of the auricular surface.	
	IIAurLinLow	This scale scores linearity in the distribution of certain patterns in the lower section of the auricular surface.	
Retroauricular area	IIAurRetUp	This scale scores morphology of a section on the upper retroauricular area.	IIAurRetSUM
	IIAurRetLow	This scale scores morphology of a section on the lower retroauricular area.	
Total: 6 features	Total: 17 variables scored		Total: 6 variables in the analysis

* Linear organization variables are summed (IIAurLinSUM), after which scores are reversed to create IIAurLinSUM_r.

2.4. Logistic Regression

While the retroauricular area approach [5] employed simple, binary logistic regression, the present method uses multiple, binary logistic regression. Individuals were classified into age groups based on the scores collected with the protocol. Logistic regression involves a transformation of age at death from a continuous numerical variable into a nominal variable that can take two mutually exclusive values. Individuals were assigned to one

of three age groups according to the age at death that was recorded for them: younger, middle-aged, and older adults (Table 2). These age groups were selected, in part, because they were the best fit for the patterns of variation observed in the seriation analyses. It was also intended to select cohorts that are representative of aging in the social groups of younger, middle-aged, and older adults. As a result, the intervals in each group differ in their range, containing a 20-year range, 26-year range, and over 35-year range, respectively.

Table 2. The boundaries of the three age groups to which individuals were assigned based on their verified age at death.

Age Group (Nominal Variable)	Age at Death (Continuous Variable)
Younger adults	20–39 years old
Middle-aged adults	40–65 years old
Older adults	66 years old and above

The results from three different logistic regression equations were combined to calculate age group membership for each individual. Initially, 70 models were explored with the goal of maximizing correct age group allocations and optimizing the number and combinations of variables included. One model was selected as best balancing these criteria and included six variables: microporosity ('IAurMiSUM'), macroporosity ('IAurMaSUM'), dense bone ('IAurDenSUM'), rim formation ('IAurRimSUM'), linear organization ('IAurLinSUM_r'), and retroauricular area ('IAurRetSUM'). Next, three different logistic regression equations (each containing the six variables named above) were obtained. Each of the three equations calculated the probability of one individual belonging to one of the two age groups being compared at a time (younger and middle-aged group, middle and older group, or younger and older group).

Although the equations were initially generated with individuals known to belong to either of the age groups compared at a time, they were later applied to all individuals in the reference sample. After completing the equations with the scores obtained for each individual, a probability value (*p*-value) was obtained. If the *p*-value was 0.5 or higher, the individual was allocated to the second of the two groups being compared. If the *p*-value was lower than 0.5 the individual was allocated to the first of the two age groups. Lastly, the mode among the three allocations was taken as the final age group assigned. For instance, an individual with the following *p*-values: probability (younger VS middle-aged) = 0.65; probability (middle-aged VS older) = 0.44; and probability (younger VS older) = 0.35 would be classified into age groups middle-aged, middle-aged, and younger, with the mode and final determination being into the middle-aged group.

The fit of this classification was assessed in two ways. Firstly, allocation accuracy showed how many times an individual was correctly assigned to an age group. Correct classification was noted in cases that had the same observed age group (documented age at death) and predicted age group (mode of the three *p*-values calculated by logistic regression). Additionally, results that eliminated one of the three age groups with certainty and accuracy even though they could not provide a correct classification between the remaining two age groups, were labeled as partly correct. Incorrect results clearly placed an individual in the wrong age group, and indeterminate results were selected if a mode did not exist. Secondly, a set of three *p*-values was calculated for each individual, regardless of what group the individual belonged to. The *p*-values were calculated by one of the authors (ESS) and later used to classify individuals into age groups (by JA) without any *a priori* assumptions about group membership.

2.5. Ethical Management of the Human Remains and Data

The following ethical statement outlines certain steps taken to minimize potential damage and respect the individual human remains [31]. Firstly, bones were always transported inside paper boxes, sometimes on a cart for longer distances. Every surface was covered with foam pads. Individuals were observed outside of their box one at a time to avoid commingling or potential contact between bones. When seriation needed to be performed, bones were placed on wide tables and kept several inches apart. The only part of the skeleton examined in this research was the os coxa, with other skeletal elements remaining in their respective boxes. No destructive analyses were performed. Furthermore, individuals' identities were maintained anonymous to the researcher who obtained the scores (ESS). Various photographs of the remains are incorporated in the Supplementary Materials, an inclusion deemed justifiable as they describe precise scoring instructions and anatomical landmarks for a new osteological protocol. All images shared are of individuals from the Grant Collection.

The existence of individuals in identified reference collections is by no means coincidental, and researchers owe them the highest level of acknowledgement and respect. Anatomical laws that led to the creation of anatomical collections, such as the Anatomy Act of Ontario (for the Grant Collection [29]) and the Anatomy Act in Missouri (for the Terry Collection [11,32]) specifically targeted those considered “deviant” by society [33]. Active ways of disengaging with this practice have been suggested through acknowledging the very individuals that make up skeletal collections, as well as the many forms of structural violence that led to their curation [33]. Acknowledging these structural biases directly contributed to the development of a stratified sampling strategy in this research. Furthermore, conscious wording efforts are made in this paper aiming to humanize the individuals discussed, instead of referring solely to “the remains scored” or “the sample used”. Apart from being an ethical commitment, considering the historical and sociocultural context that influenced the creation of skeletal collections results in more fruitful research and furthers academic exchange [34–39].

3. Results

Reproducibility in the scores collected was assessed in two blind scoring sessions by one observer (ESS). Percentage of agreement between both tests was 78% on average. While some features like the retroauricular area ('IIAurRetSUM') showed very high reproducibility, others such as dense bone ('IIAurDenSUM') caused more disagreement. The specific values for each feature are shown in Table 3. Cohen's kappa, a measure of how much of the agreement between both scores is due to chance, was reported as 0.52 on average, indicating moderate agreement [40]. Again, retroauricular area showed the highest results, with other areas such as linear organization ('IIAurLinSUM') having lower values.

Table 3. Intraobserver error was calculated for the six variables in the equations. Percentage of agreement and Cohen's kappa are given. Results on average indicate a 78% percentage of agreement and moderate agreement as calculated by Cohen's kappa.

Skeletal Variable	Percentage of Agreement (%)	Cohen's Kappa
IIAurRetSUM	91	0.82
IIAurMaSUM	82	0.55
IIAurLinSUM	77	0.31
IIAurRimSUM	75	0.50
IIAurMiSUM	74	0.48
IIAurDenSUM	70	0.46
Average:	78	0.52

A series of three equations were used for calculating to which age groups an individual belonged. Probability of an individual belonging to either the younger or middle-aged group was calculated with Equation (1):

$$\text{Probability (younger VS middle)} = \frac{1}{1 + e^{-(a)}} \quad (1)$$

where:

$$a = (-3.742 + 0.018 \times \text{IIAurMiSUM} + 0.228 \times \text{IIAurMaSUM} + 0.091 \times \text{IIAurDenSUM} + 0.162 \times \text{IIAurRimSUM} + 0.458 \times \text{IIAurLinSUM}_r + 0.42 \times \text{IIAurRetSUM})$$

To classify whether an individual belonged to either the middle or older age groups, Equation (2) was used:

$$\text{Probability (middle VS older)} = \frac{1}{1 + e^{-(b)}} \quad (2)$$

where:

$$b = (-6.353 - 0.414 \times \text{IIAurMiSUM} - 0.136 \times \text{IIAurMaSUM} + 0.211 \times \text{IIAurDenSUM} + 0.163 \times \text{IIAurRimSUM} + 1.103 \times \text{IIAurLinSUM}_r + 1.116 \times \text{IIAurRetSUM})$$

Lastly, to test if an individual belonged to either the younger or older age groups, Equation (3) was used:

$$\text{Probability (younger VS older)} = \frac{1}{1 + e^{-(c)}} \quad (3)$$

where:

$$c = (-10.729 - 0.094 \times \text{IIAurMiSUM} - 0.043 \times \text{IIAurMaSUM} + 0.522 \times \text{IIAurDenSUM} + 0.585 \times \text{IIAurRimSUM} + 1.556 \times \text{IIAurLinSUM}_r + 0.789 \times \text{IIAurRetSUM})$$

Equation (1) calculated whether an individual belonged to either the younger or the middle-aged group. This equation was developed on those known to be either younger or middle-aged adults and could classify them with 73% accuracy (Table 4). Similarly, Equation (2) generated the probability of an individual belonging to either the middle or the older age group and could do so with 77.1% accuracy (Table 5). Lastly, Equation (3) calculated whether an individual belonged to either the younger or the older age groups. These could be distinguished with 87.9% accuracy (Table 6).

Table 4. Classification table showing the accuracy with which individuals in the younger and middle-aged groups can be distinguished *.

Observed		Predicted		Percentage Correct
		Age Group		
		Younger	Middle-Aged	
Age group	Younger	26	9	74.3
	Middle-aged	11	28	71.8
Overall percentage				73.0

* The classification is made based on the *p*-value obtained from (1). If the *p*-value is 0.5 or higher, the individual is classified as middle-aged. If it is lower than 0.5 the individual is classified into the younger age group.

Table 5. Classification table showing the accuracy with which individuals in the middle-aged and older groups can be distinguished *.

Observed		Predicted		Percentage Correct
		Age Group		
		Middle-Aged	Older	
Age group	Middle-aged	31	8	79.5
	Older	8	23	74.2
Overall percentage				77.1

* The classification is made based on the *p*-value obtained from (2). If the *p*-value is 0.5 or higher, the individual is classified into the older age group. If it is lower than 0.5 the individual is classified as middle-aged.

Table 6. Classification table showing the accuracy with which individuals in the younger and older age groups can be distinguished *.

Observed		Predicted		Percentage Correct
		Age Group		
		Younger	Older	
Age group	Younger	31	4	88.6
	Older	4	27	87.1
Overall percentage				87.9

* The classification is made based on the *p*-value obtained from (3). If the *p*-value is 0.5 or higher, the individual is classified into the older age group. If it is lower than 0.5 the individual is classified into the younger age group.

This accuracy is further illustrated in Figures 2–4. The three scatterplots represent the calculated *p*-value for each individual compared to their documented age at death. The threshold value on the y-axis $p = 0.5$ indicates which of the two age groups in the graph the individual is being allocated to. Other threshold values on the x-axis indicate the boundaries that were selected for each of the age groups (with 40 and 66 being the lowest ends of their respective age groups). Therefore, individual dots on the upper left and lower right quadrants of each scatterplot are classified into the wrong age group. It should also be noted how the three age groups differ not only in age range, but in sample size. The younger age group contains 35 individuals, the middle-aged group 39, and the older age group, 31. Yet, the visual trend with more dots in the correct quadrants aligns with the classification tables (Tables 4–6), in that accuracy increases when individuals on the older age group are included.

When these equations were tested blind in the entire sample, final correct allocation occurred in 66% of the 105 cases (Table 7). In addition, partly correct results could be obtained in a further 26% of the cases (Table 8). This information still narrows down the estimation, as it can eliminate one of the age groups with certainty, even if it cannot provide an accurate classification into one of the remaining two age groups. As observed in Tables 7 and 8, the classification of middle-aged individuals benefits the most from considering partly correct information. Lack of a mode was observed in one individual who was middle-aged and was counted as incorrect in the calculations.

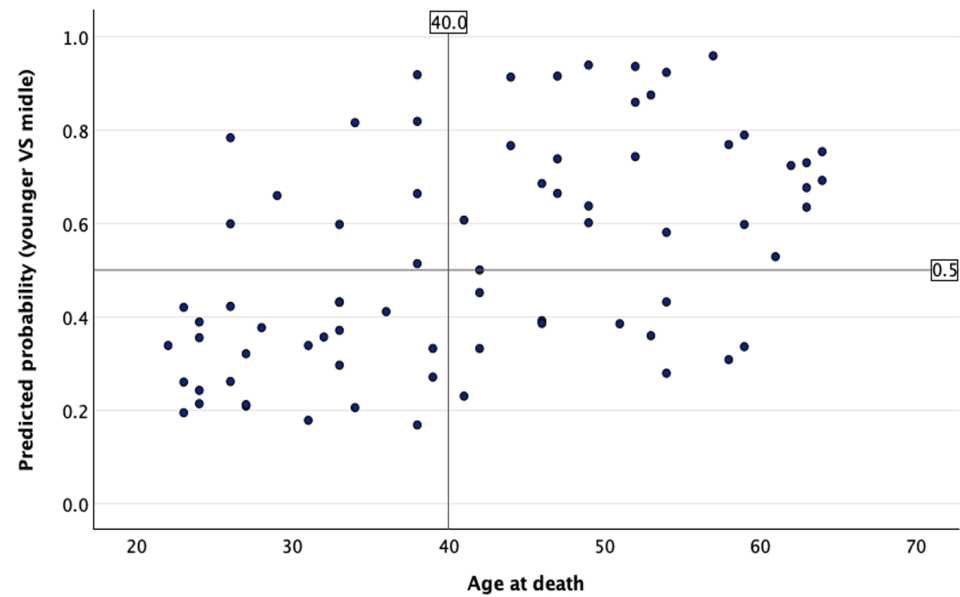


Figure 2. Scatterplot showing the fit of the classification in individuals used to create (1). The threshold value $p = 0.5$ on the y-axis indicates to which of the two age groups is each individual allocated through logistic regression. Values equal or higher than 0.5 classify an individual to the middle-aged group, and lower than 0.5 to the younger group. The threshold value age = 40 on the x-axis indicates the age at which an individual would be documented as middle-aged. Ages lower than 40 correspond to the younger age group. Hence, individual dots on the upper left and lower right quadrants of the scatterplot show those who are classified into the wrong age group.

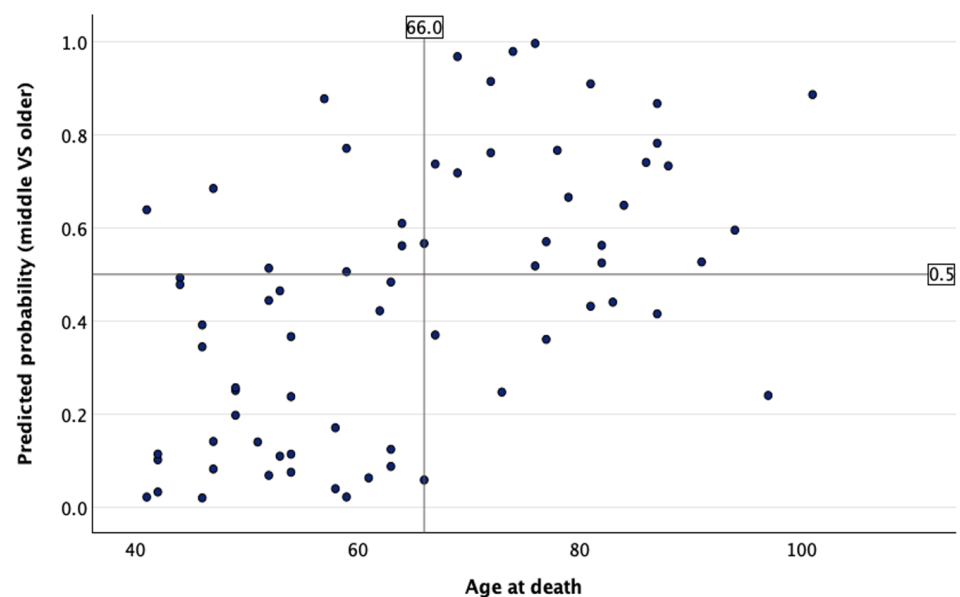


Figure 3. Scatterplot showing the fit of the classification in individuals used to create (2). The threshold value $p = 0.5$ on the y-axis indicates to which of the two age groups is each individual allocated through logistic regression. Values equal or higher than 0.5 classify an individual to the older age group, and lower than 0.5 to the middle-aged group. The threshold value age = 66 on the x-axis indicates the age at which an individual would be documented as belonging to the older-age group. Ages lower than 66 correspond to the middle-aged group. Hence, individual dots on the upper left and lower right quadrants of the scatterplot show those who are classified into the wrong age group.

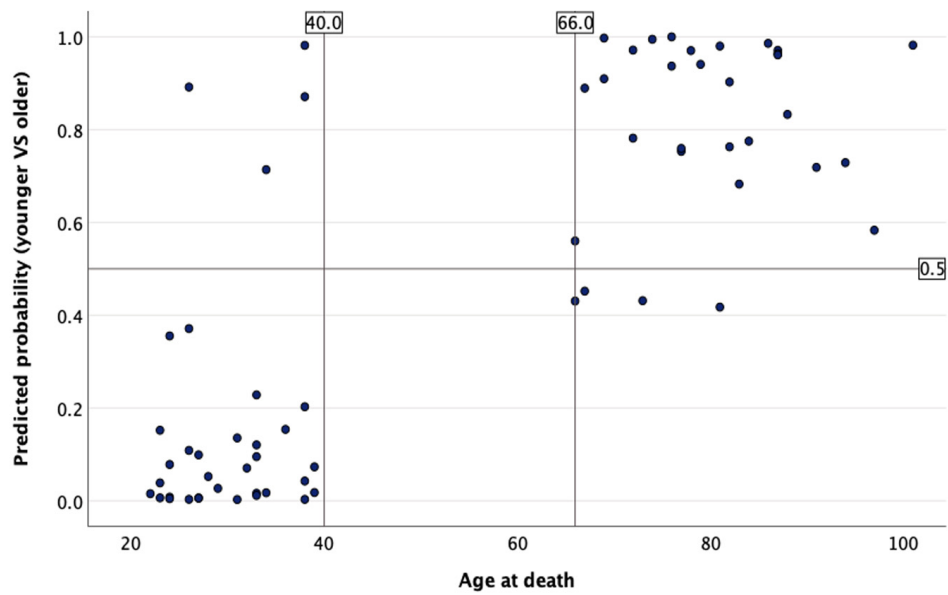


Figure 4. Scatterplot showing the fit of the classification in individuals used to create (3). The threshold value $p = 0.5$ on the y-axis indicates to which of the two age groups is each individual allocated through logistic regression. Values equal or higher than 0.5 classify an individual to the older age group, and lower than 0.5 to the younger group. The threshold values age = 40 and age = 66 on the x-axis indicate ages that mark the beginning of a different age group. Individuals with age documented between 40 and 65 would be middle-aged individuals and were not observed to create (3). Hence, individual dots on the upper left and lower right quadrants of the scatterplot show those who are classified into the wrong age group.

Table 7. Count and percentage of individuals (out of 105) that were assigned to the correct age group with the presented auricular surface method *.

Observed	Predicted		
	Younger Individuals	Middle-Aged Individuals	Older Individuals
Younger individuals	26 (25%)	6 (6%)	3 (3%)
Middle-aged individuals	11(10%)	20 (19%)	7 (7%)
Older individuals	1 (1%)	7 (7%)	23 (22%)

* Correct results do not contain individuals for whom partly correct information was obtained. There is one individual observed to be middle-aged for whom the calculations did not yield a mode. This individual was counted as incorrectly classified but is not included in this table.

Table 8. Count and percentage of individuals (out of 105) that provided correct or partly correct information with the presented auricular surface method.

Observed	Predicted		
	Younger Individuals	Middle-Aged Individuals *	Older Individuals
Younger individuals	31 (30%)	1 (1%)	3 (3%)
Middle-aged individuals	0 (0%)	38 (36%)	0 (0%)
Older individuals	1 (1%)	3 (3%)	27 (26%)

* In all middle-aged individuals, the estimation can provide a classification that excludes one of the age groups with high certainty. There is one individual observed to be middle-aged for whom the calculations did not yield a mode. This individual was counted as incorrectly classified but is not included in this table.

4. Discussion

Traditionally, the issue of “attraction to the middle” has made the estimation of age of younger and older adults more challenging than that of middle-aged adults [7,14]. Among these well documented issues, a tendency for error to increase with advancing age has also been noted [41,42]. This increase sometimes leads to skeletal indicators ceasing to be correlated with age in older adults. Even though age-related skeletal changes in older adults exist, individual variation seems to make statistical trend detection unfeasible [42]. Conversely, the results shown here become more robust when older-age individuals are incorporated (as seen in Tables 4–6 and Figures 2–4). Individuals who belong to the older age group and are not classified correctly represent 8% of all cases. On the other hand, younger individuals have also been known to experience inaccurate age estimation due to “attraction to the middle” [3,42]. Yet, the classification for younger individuals here is comparable to that of older individuals, being those who are younger adults and classified into the wrong age group in 9% of the total cases.

Even though previous research using only the retroauricular area [5] showed how younger and middle-aged adults cannot be distinguished from each other with high accuracy, an improvement is noted here. With the current method, 18 younger and middle-aged individuals that were not classified correctly using the retroauricular area, can now be correctly classified when using all six variables. A total of 28 individuals out of 105 were correct using either method. In addition, 16 younger and middle-aged individuals who were not classified correctly do not experience any change to their results irrespective of the method. Lastly, 11 individuals previously correct with the retroauricular area method, now provide partly correct information. One individual was correct with the previous method but is indeterminate with the current one.

Middle-aged individuals seem to present more challenges to accurate classification than those from other age groups. In fact, middle-aged individuals who are not classified into the correct age group comprise 17% of the total cases. However, the *p*-values generated for these individuals can still exclude all of them from being either younger or older aged. Only one exception arises with a middle-aged individual for whom the results were indeterminate due to the lack of a mode. In a real-case scenario, age cannot be reasonably estimated if the result is indeterminate and this result reveals that the method did not work for that specific case. Previously [5], the notion of partly correct information was introduced to articulate a more practical and nuanced view of the results. While correct information eliminates two of the three age groups accurately, partly correct information eliminates one of the three. The classification of younger and older individuals can also produce information that is partly correct, as noted in Table 8. However, some individuals from these age groups were classified incorrectly with both methods.

It is likely that idiosyncratic reasons other than age are at play for the seven cases that are always incorrectly classified (using either only retroauricular area or the entire auricular surface). These individuals belong either to the younger or older age groups, but none are middle-aged. Incorrect cases may respond to a pattern in which the four younger individuals present morphologies that overestimate their age, and the three older adults exhibit morphologies that underestimate their age.

While experience and training are important, it is also possible that increased standardization of scoring procedures may aid with interpretation and translate into less observer error. However, it is not uncommon that discussions of reproducibility describe the researcher’s experience as the only possible source of observer differences. Firstly, this creates a need for establishing, unambiguously, when a practitioner is experienced enough. Secondly, perception may also change throughout the researcher’s life. Reproducibility should be a measure of observer differences at a concrete point in time. If only a selected number of

researchers can obtain a very accurate and precise answer, there is room to create techniques that broaden applicability. Deliberate attempts to clarify and simplify the methods used would be beneficial for less experienced observers, while the more knowledgeable ones would be able to apply the method consistently anyway.

Using multiple binary logistic regression responds to a commitment to make the methodology accessible. Previous results used the most uncomplicated approach to logistic regression, which is simple and binary [5]. Binary logistic regression can be either simple (obtaining input from a single variable) or multiple (based on input from more than one variable, six in the case of the current method). Therefore, using multiple binary logistic regression presents an increase in time and complexity of the procedure. However, binary logistic regression is still chosen over other more complicated approaches such as multinomial logistic regression. A description of the methodological complications that this technique would create has been provided elsewhere [5]. This decision prioritizes ease of use and accessibility of the method, which would make it more widely applicable.

While other approaches such as random sampling will necessarily reflect the initial bias in the collection sampled, stratified sampling broadens applicability of this method. All skeletal collections are an inherently biased representation of human variation, and the Terry and Grant collections are not an exception [10,11,29,43]. In fact, this possibility for bias permeates most aspects of bioanthropological research. In principle, the ideal reference sample is one that perfectly represents the population under study. In practice, this perfect representation does not exist. Furthermore, individuals' identities and specifically how old they were at the time of death has been affected by a historical lack of academic attention, among other factors. For instance, this is most frequently the case for older adults or female individuals. In this context, stratified sampling ensures presence of these traditionally underrepresented subgroups in the final cohort. Here, stratified sampling uses information about age at death or documented biological sex as important criteria for sample selection. This approach counters known biases in skeletal collections by sampling a wide range of variation in the reference sample used to develop the method. However, even if documented sex is used to sample human variation, it does not mean that skeletal aging is specific to sex. Neither does documented sex necessarily reflect the social and personal aging experiences of the person accompanied with this notation. In addition to biases in skeletal variation, the quality and accuracy of documentary data are crucial to sample selection.

It is worthwhile to consider how the documented records and identities available today in skeletal collections constitute another bias in and of itself. The process by which this information was obtained and preserved is embedded in systemic structures that are inherently selective. As a response, this level of selectivity should be mirrored when reviewing biographic information available for individuals in reference samples [8,11]. The quality of the documenting materials can vary according to the specific context of the collection itself. In the case of the Terry Collection, even though age was reported on the death record by local facilities, it was not uncommon that they gave an estimate when the exact age was unknown [9]. Other points of concern could potentially include "age heaping" (approximation of age at death to the nearest age that is a multiple of five) or inflating the elderly's age as a sign of respect [9]. In practice, someone who died at 69 years old would be indistinguishable from someone who died at 70. Yet, when accounting for the mechanisms above a person's documented age of 69 is likely correct, whereas a person's documented age of 70 could be either correct, incorrect by as little as one year, or incorrect by as much as 15 to 20 years, and it is not possible to determine which is the case [9,11]. Selecting the 69-year-old individual effectively samples the age-related variation while avoiding possible erroneously documented ages.

Still, documenting and record-keeping demonstrate a level of respect and acknowledgement of individuality by the original amassers [33]. In other contexts, said respect and acknowledgement did not happen and identities were systematically not preserved, such as for anatomical remains from red markets in South Asia [39,44,45] or the graves of enslaved African American people and their descendants [46]. Although documented skeletal collections enable the development and evaluation of osteological methods, the inherent biases in their creation and curation can create flawed knowledge. This stands in contrast to one of the primary goals of forensic anthropology and bioarchaeology, which is reconstructing personhood [37].

Limitations and Future Research

It is worth noting how more morphological data from the auricular surface slightly increases accuracy already obtained by the retroauricular area alone [5]. The previous method presented a classification into age groups based on two scores compiled into one predictor variable. In the current method, scores required increase to 17 in total. The substantial increment in time and effort does increase accuracy of the classifications, but not by the same magnitude. In this study, obtaining 15 additional scores translated into a 6% increase in correct classifications. Although obtaining more variables does introduce more nuance to the results, this could have other practical implications such as more chances for observer error or for damage. In the case of damage to other areas of the auricular surface, the approach using scores only from the retroauricular area [5] could be performed.

Additional research will necessarily evaluate other aspects of this approach, such as interobserver reproducibility. On the other hand, intraobserver error provided moderately accurate results. Further research is still needed to investigate whether the magnitude and direction of scoring error may actually impact the age result [47]. In addition, the strategies for method validation performed (observing two skeletal collections to create the data-collection protocol, generating equations using only two age groups at a time but testing and reporting their accuracy in three age groups, or assigning final age classification by an observer that did not obtain the *p*-values) provided promising results. However, further research with a fully independent group of individuals is also needed and could investigate the potential presence of age mimicry in the results with more detail.

Other potential pitfalls of this procedure are shared with the retroauricular area method [5], including the inherently subjective selection of boundaries for age groups, and the production of relatively broad intervals of age.

5. Conclusions

This paper describes a systematic re-assessment and preliminary evaluation of how age-related changes in the auricular surface are observed and analyzed. Firstly, the reference sample of individuals who make up the foundation of this research was chosen through stratified sampling to maximize the range of variation included. Secondly, three multiple logistic regressions that classified individuals into three age groups were selected to balance ease of use with utility and accessibility of the method. Lastly, the protocol provided in the Supplementary Materials aims to clarify and standardize the scoring procedures for data collection. The results reflect an improved classification for younger individuals compared with the retroauricular area approach [5], while the previous high accuracy in identifying older individuals is maintained. Overall, the method classified individuals into their correct age group 66% of the time and provided partly correct information a further 26% of the time. Therefore, this method can establish accurately whether an individual is below 40 years old, or above 65 years old.

Still, age estimation attempts to imperfectly model the complex relationship between age at death and certain skeletal features, whose morphology is affected by age and also by a myriad of other biological, social, and cultural factors. The aging process inherently carries significant variability among different individuals as well. Whether due to those factors or the variability of the aging process itself, obtaining an age estimate that is both accurate and precise is a matter yet to be solved. This research strives to investigate whether increased clarity and transparency of a method will aid interpretation and transform into more repeatable results. Balancing accuracy of the results with an accessible methodology that can be used consistently has the potential to address the multifaceted nature of age estimation when navigating this challenging field forward.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/forensicsci6030067/s1>, Protocol S1: Data Collection Protocol for Age Estimation based on the Auricular Surface of the Ilium; Sheet S1: Data Collection Sheet for Age Estimation based on the Auricular Surface of the Ilium.

Author Contributions: Conceptualization, E.S.-S. and J.A.; data curation, E.S.-S. and J.A.; software, E.S.-S. and J.A.; formal analysis, E.S.-S. and J.A.; investigation, E.S.-S. and J.A.; methodology, E.S.-S. and J.A.; validation, E.S.-S. and J.A.; writing—original draft preparation, E.S.-S.; writing—review and editing, E.S.-S. and J.A.; visualization, E.S.-S.; supervision, J.A.; project administration, J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This research project was marked exempt from the requirements of Research Ethics Review at Simon Fraser University. According to the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2), research relying exclusively on secondary use of anonymous information is exempt from ethical review so long as the research process does not generate identifiable information.

Informed Consent Statement: Not applicable; no personal information was exposed. Individuals included in this research were collected under the Anatomy Act of Missouri (Terry Collection at the Smithsonian Institution) and the Anatomy Act of Ontario (Grant Collection at the University of Toronto) for anatomical instruction and research.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. They may be made available upon project completion.

Acknowledgments: The authors would like to acknowledge the help and insights from the Collections Manager in accessing and navigating the Terry Collection at the Smithsonian National Museum of Natural History. Similarly, the authors would like to acknowledge the Department of Anthropology at the University of Toronto for providing access and guidance at the Grant Collection.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Buckberry, J.L.; Chamberlain, A.T. Age Estimation from the Auricular Surface of the Ilium: A Revised Method. *Am. J. Phys. Anthropol.* **2002**, *119*, 231–239. [[CrossRef](#)] [[PubMed](#)]
2. Lovejoy, C.O.; Meindl, R.S.; Pryzbeck, T.R.; Mensforth, R.P. Chronological Metamorphosis of the Auricular Surface of the Ilium: A New Method for the Determination of Adult Skeletal Age at Death. *Am. J. Phys. Anthropol.* **1985**, *68*, 15–28. [[CrossRef](#)] [[PubMed](#)]
3. Osborne, D.L.; Simmons, T.L.; Nawrocki, S.P. Reconsidering the Auricular Surface as an Indicator of Age at Death. *J. Forensic Sci.* **2004**, *49*, JFS2003348-7. [[CrossRef](#)]
4. Osborne, D. Reconsidering the Auricular Surface as an Indicator of Age at Death. Master's Thesis, Western Michigan University, Kalamazoo, MI, USA, 2000.
5. Sierra-Serrano, E. Alternative Approaches for Capturing Age-Related Changes on the Auricular Area of the Ilium and Their Use for Adult Age Estimation. Ph.D. Thesis, Simon Fraser University, Burnaby, BC, Canada, 2026.

6. Bocquet-Appel, J.P.; Masset, C. Farewell to Paleodemography. *J. Hum. Evol.* **1982**, *11*, 321–333. [\[CrossRef\]](#)
7. Masset, C. Age Estimation on the Basis of Cranial Sutures. In *Age Markers in the Human Skeleton*; İşcan, M.Y., Ed.; Springfield: Irvine, CA, USA, 1989; pp. 71–103.
8. Albanese, J. Strategies for Dealing with Bias in Identified Reference Collections and Implications for Research in the 21st Century. In *Identified Skeletal Collections: The Testing Ground of Anthropology?* Henderson, C.Y., Cardoso, F.A., Eds.; Archaeopress Publishing Ltd.: Bicester, UK, 2018; pp. 59–82.
9. Sharman, J.; Albanese, J. Bioarchaeology and Identified Skeletal Collections: Problems and Potential Solutions. In *Identified Skeletal Collections: The Testing Ground of Anthropology?* Henderson, C.Y., Cardoso, F.A., Eds.; Archaeopress Publishing Ltd.: Bicester, UK, 2018; pp. 83–114.
10. Albanese, J.; Dagdag, A.; Skalic, C.; Osley, S.; Cardoso, H. The Fall and Rise of Identified Reference Collection: It Is Possible and Necessary to Transition from a Typological Conceptualization of Variation to Effective Utilization of Collections. *Forensic Sci.* **2022**, *2*, 438–454. [\[CrossRef\]](#)
11. Hunt, D.R.; Albanese, J. History and Demographic Composition of the Robert J. Terry Anatomical Collection. *Am. J. Phys. Anthropol.* **2005**, *127*, 406–417. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Maaranen, N.; Buckberry, J. Exploring Age—Transition Analysis as a Tool for Detecting the Elderly. In *Across the Generations: The Old and the Young in Past Societies*; Lillehammer, G., Murphy, E.M., Eds.; AM-UiS Publisher: Stavanger, Norway, 2018; Volume 12, pp. 129–130.
13. Navitainuck, D.U.; Vach, W.; Alt, K.W.; Pichler, S.L. Osteological Age-at-Death Estimation in an Archaeological Sample Avoiding Age-Mimicry: A Six-Step Approach. *Archaeol. Anthropol. Sci.* **2024**, *16*, 126. [\[CrossRef\]](#)
14. Falys, C.G.; Lewis, M.E. Proposing a Way Forward: A Review of Standardisation in the Use of Age Categories and Ageing Techniques in Osteological Analysis (2004–2009). *Int. J. Osteoarchaeol.* **2011**, *21*, 704–716. [\[CrossRef\]](#)
15. Aykroyd, R.G.; Lucy, D.; Pollard, A.M.; Solheim, T. Technical Note: Regression Analysis in Adult Age Estimation. *Am. J. Phys. Anthropol.* **1997**, *104*, 259–265. [\[CrossRef\]](#)
16. Aykroyd, R.G.; Lucy, D.; Pollard, A.M.; Roberts, C.A. Nasty, Brutish, but Not Necessarily Short: A Reconsideration of the Statistical Methods Used to Calculate Age at Death from Adult Human Skeletal and Dental Age Indicators. *Am. Antiq.* **1999**, *64*, 55–70. [\[CrossRef\]](#)
17. Hoppa, R.D.; Vaupel, J.W. (Eds.) The Rostock Manifesto for Paleodemography: The Way from Stage to Age. In *Paleodemography*; Cambridge University Press: Cambridge, UK, 2002; pp. 1–8.
18. Miranker, M. A Comparison of Different Age Estimation Methods of the Adult Pelvis. *J. Forensic Sci.* **2016**, *61*, 1173–1179. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Pagano, M.; Gauvreau, K. *Principles of Biostatistics*, 2nd ed.; Duxbury: Pacific Grove, CA, USA, 2000.
20. Boldsen, J.L.; Milner, G.R.; Konigsberg, L.W.; Wood, J.W. Transition Analysis: A New Method for Estimating Age from Skeletons. In *Paleodemography*; Hoppa, R.D., Vaupel, J.W., Eds.; Cambridge University Press: Cambridge, UK, 2002; pp. 73–106.
21. Garvin, H.M.; Passalacqua, N.V. Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation. *J. Forensic Sci.* **2012**, *57*, 427–433. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Li, C.; Ji, G. Age Estimation from the Permanent Molar in Northeast China by the Method of Average Stage of Attrition. *Forensic Sci. Int.* **1995**, *75*, 189–196. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Gustafson, G. Age Determinations on Teeth. *J. Am. Dent. Assoc.* **1950**, *41*, 45–54. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Lu, C.K.; Yee, M.C.S.; Ravi, S.B.; Pandurangappa, R. Forensic Age Estimation of Chinese Malaysian Adults by Evaluating Occlusal Tooth Wear Using Modified Kim's Index. *Int. J. Dent.* **2017**, *2017*, 4265753. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Albanese, J. A Metric Method for Sex Determination Using the Hipbone and the Femur. *J. Forensic Sci.* **2003**, *48*, 1–11. [\[CrossRef\]](#)
26. Albanese, J. A Method for Estimating Sex Using the Clavicle, Humerus, Radius, and Ulna. *J. Forensic Sci.* **2013**, *58*, 1413–1419. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Albanese, J. Approaches for Constructing and Using Reference Samples to Maximize the Utility of New Forensic Methods. *Eur. J. Anat.* **2021**, *25*, 63–82.
28. Albanese, J.; Tuck, A.; Gomes, J.; Cardoso, H.F.V. An Alternative Approach for Estimating Stature from Long Bones That Is Not Population- or Group-Specific. *Forensic Sci. Int.* **2016**, *259*, 59–68. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Albanese, J. The Grant Human Skeletal Collection and Other Contributions of J. C. B. Grant to Anatomy, Osteology, and Forensic Anthropology. In *Identified Skeletal Collections: The Testing Ground of Anthropology?* Henderson, C.Y., Cardoso, F.A., Eds.; Archaeopress Publishing Ltd.: Bicester, UK, 2018; pp. 35–57.
30. White, T.D.; Black, M.T.; Folkens, P.A. *Human Osteology*; Academic Press: Cambridge, MA, USA, 2011.
31. Squires, K.; Roberts, C.A.; Márquez-Grant, N. Ethical Considerations and Publishing in Human Bioarchaeology. *Am. J. Biol. Anthropol.* **2022**, *177*, 615–619. [\[CrossRef\]](#)
32. Terry, R.J. On Measuring and Photographing the Cadaver. *Am. J. Phys. Anthropol.* **1940**, *26*, 433–447. [\[CrossRef\]](#)

33. de la Cova, C. Making Silenced Voices Speak. In *Theoretical Approaches in Bioarchaeology*; Cheverko, C.M., Prince-Buitenhuis, J.R., Hubbe, M., Eds.; Routledge: London, UK; Taylor & Francis Group: New York, NY, USA, 2020; pp. 150–169.
34. Adams, D.M.; Goldstein, J.Z.; Isa, M.; Kim, J.; Moore, M.K.; Pilloud, M.A.; Tallman, S.D.; Winburn, A.P. A Conversation on Redefining Ethical Considerations in Forensic Anthropology. *Am. Anthropol.* **2022**, *124*, 597–612. [[CrossRef](#)]
35. Blakey, M.L. Discussion of the Symposium on Ethics in the Curation and Use of Human Skeletal Remains. *Am. J. Biol. Anthropol.* **2025**, *186*, e24895. [[CrossRef](#)] [[PubMed](#)]
36. de la Cova, C. Ethical Considerations for Paleopathology. In *The Routledge Handbook of Paleopathology*; Routledge: London, UK, 2022; pp. 381–396.
37. Robbins Schug, G.; Halcrow, S.E.; de la Cova, C. They Are People Too: The Ethics of Curation and Use of Human Skeletal Remains for Teaching and Research. *Am. J. Biol. Anthropol.* **2025**, *186*, e70013. [[CrossRef](#)] [[PubMed](#)]
38. Go, M.C. The Decolonial Turn in Forensic Anthropology. *Am. J. Biol. Anthropol.* **2025**, *186*, e70003. [[CrossRef](#)] [[PubMed](#)]
39. Agarwal, S.C. The Bioethics of Skeletal Anatomy Collections from India. *Nat. Commun.* **2024**, *15*, 1692. [[CrossRef](#)] [[PubMed](#)]
40. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)]
41. Buckberry, J. The (Mis)Use of Adult Age Estimates in Osteology. *Ann. Hum. Biol.* **2015**, *42*, 323–331. [[CrossRef](#)] [[PubMed](#)]
42. Nawrocki, S.P. The Nature and Sources of Error in the Estimation of Age at Death from the Skeleton. In *Age Estimation of the Human Skeleton*; Latham, K.E., Finnegan, M., Eds.; Charles C. Thomas: Springfield, IL, USA, 2010; pp. 79–101.
43. de la Cova, C. Race, Health, and Disease in 19th-century-born Males. *Am. J. Phys. Anthropol.* **2011**, *144*, 526–537. [[CrossRef](#)] [[PubMed](#)]
44. Agarwal, S.C. The Disposability and Inclusion of Brown Bodies. *Am. J. Biol. Anthropol.* **2025**, *186*, e25003. [[CrossRef](#)] [[PubMed](#)]
45. Carney, S. *Wired*. 27 November 2007.
46. Dunnavant, J.; Justinvil, D.; Colwell, C. Craft an African American Graves Protection and Repatriation Act. *Nature* **2021**, *593*, 337–340. [[CrossRef](#)] [[PubMed](#)]
47. Mullins, R.A.; Albanese, J. Estimating Biological Characteristics with Virtual Laser Data. *J. Forensic Sci.* **2018**, *63*, 815–823. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.