Received: 3 February 2022 | Revised: 14 August 2022 | Accepted: 22 August 2022

DOI: 10.1002/oa.3157

RESEARCH ARTICLE

Age-at-death estimation in archaeological samples: Differences in population means resulting from different aging methods can be predicted from the mean ages of method-specific reference samples

Denise U. Navitainuck[1] | Werner Vach[1,2] | Sandra L. Pichler[1] | Kurt W. Alt[1,3]

[1]Integrative Prehistory and Archeological Science (IPAS), Department of Environmental Sciences, University of Basel, Basel, Switzerland

[2]Basel Academy for Quality and Research in Medicine, Basel, Switzerland

[3]Center of Natural and Cultural Human History, Danube Private University, Krems, Austria

Correspondence
Denise U. Navitainuck, Integrative Prehistory and Archeological Science (IPAS), Department of Environmental Sciences, University of Basel, Spalenring 145, CH-4055 Basel, Switzerland.
Email: denise.navitainuck@unibas.ch

Funding information
Studienstiftung des Deutschen Volkes

Abstract

Age mimicry is a well-known phenomenon in the application of osteological age-estimation methods. Age mimicry refers to the fact that predicting age-at-death from a specific trait (age indicator) based on the relation observed in a specific reference sample implies that age estimates to some degree reflect the age structure of the reference sample. In particular, the estimated population mean in a target population in which an age-estimation method is applied is shifted towards the mean in the method-specific reference sample. Consequently, differences in population means between different age-estimation methods in the same target population may be due to differences in mean age of the reference samples used to develop the age-estimation methods. We aim at quantifying the expected magnitude for such differences. Fifteen different traditional age-estimation methods were applied to a sample of 675 adult individuals from the early medieval cemetery of Mannheim-Seckenheim. The relation of the observed estimated population age means and the mean age in the reference samples was analyzed by linear regression. We find that up to 80% of the variation in the estimated population age means can be explained by the variation of the mean age in the reference samples. Furthermore, differences in the magnitude of 3 to 4 years in the mean age between two reference samples can imply a 1-year difference in estimated target population age means. Because large differences in mean age between reference samples used to develop different age-estimation methods are common, some care is needed in interpreting differences between individual age estimates or population mean age estimates in cases where different age-estimation techniques are used.

KEYWORDS

age mimicry, aging methods, population mean, reference sample

1 | INTRODUCTION

An individual's age-at-death is one of the most basic and important pieces of information in both forensic and bioarchaeological contexts and likewise at the population level. Many characteristics have been identified for age diagnosis and various age-estimation methods developed. The pubic symphysis, the facies auricularis of the os ilium and dental wear are prominent examples for such characteristics. In everyday osteological work, it is common usage to consider multiple traits and to come to a final judgment, which is also based on the observers' individual experience. The judgment is typically expressed as a numerical age interval or with respect to an age classification system. European anthropologists commonly follow the system introduced by Martin (1928) using the six classes Infans I (0–6), Infans II (7–12), Juvenis (13–18/20), Adultas (20–40), Maturitas (40–60), and Senilis (60+), although there are also variations (Herrmann, 1990; Lohrke, 2004). By contrast, American anthropologists often use the seven classes fetus, infant (0–3), child (3–12), adolescent (12–20), young adult (20–35), middle adult (35–50), and old adult (50+) (cf. Buikstra & Ubelaker, 1994). In cases of larger uncertainty, multiple categories are chosen in either classification system.

However, age-estimation methods can be also applied more formally, assigning point estimates of age instead of age classes, potentially accompanied by a quantification of the precision. In recent years, there has been an increasing interest in such more formal approaches to age estimation, which may reduce the subjective elements and hence increase comparability across studies. This was in particular expressed and stimulated by the so-called "Rostock Manifesto" (Hoppa & Vaupel, 2002). However, it is well known that using age estimates based on reference samples carries the risk to shift estimates of the age-at-death distribution in a given sample towards the distribution of the reference sample (e.g., Aykroyd et al., 1997; Bocquet-Appel & Masset, 1982; Konigsberg & Frankenberg, 1992). This phenomenon has been called "age mimicry" (Mensforth, 1990) and has led to discussions on the applicability of age-estimation methods in forensic or archaeological contexts. Buckberry (2015) states that mean age estimates in a new target population are a direct reflection of the age structure of the reference sample and should be avoided unless the target population has a similar age-at-death structure compared with the reference sample.

Traditional age-estimation methods report age-at-death as a function of a score derived from a trait, or as a regression equation. This means that they use the reference sample to predict age directly based on the observed trait. Typical examples are the Suchey–Brooks method, translating maturational changes in the pubic symphysis into six stages and assigning the sex-specific mean age observed in the reference sample to each stage, or the regression equation developed by Prince and Ubelaker (2002) to translate three quantitative markers related to dental root translucency into an age estimate. Age mimicry then has the following effect on age estimates: If two samples with different mean ages are drawn from the same reference population, age estimates based on the two samples will systematically differ; a formal derivation of the statement is found in Appendix S2. Age estimates based on the reference sample with the lower mean age will tend to give lower estimates than those based on the older reference sample, that is, the one with the higher mean age. Moreover, if individual age estimates are obtained by one specific age-estimation method in a target population, then the estimated population mean in the target population will be biased in the direction of the reference sample. Therefore, if the true (but unknown) mean age in the target population is below the mean age in the reference sample, target population mean age is overestimated, or underestimated vice versa.

Age mimicry has an obvious effect if different age-estimation methods are applied in the same target population. The methods will differ in the estimated population means if the reference samples used to develop the different age-estimation methods differ in their mean values. Indeed, application of different age-estimation methods in the same sample can lead to highly varying estimates of the population mean age (Clark et al., 2019; Wittwer-Backofen et al., 2008). However, it is not well explored to which degree such differences can actually be explained by differences in the mean age of the reference sample. This study aims to explicitly quantify the magnitude of the differences in estimated population age means we have to expect due to differences in mean age between the reference samples.

For this purpose, we applied 15 different age-estimation methods based on eight different osteological/dental traits in an early medieval archaeological population from Mannheim-Seckenheim, Germany. All selected methods are well established and allow translating a specific age indicator into an age estimate. We compare the estimated age distributions between the 15 methods and also describe correlations between the age-estimation methods. The observed population mean values from the different methods in the target population are then correlated with the mean age in the corresponding reference samples. The selected methods often also provide a quantification of the uncertainty of the specific age estimate in form of intervals or standard deviations, but this aspect does not play a role in our investigation.

In selecting age-estimation methods for this case study, we aimed at including different skeletal elements but also at comparing different methods on the same feature (e.g., the facies auricularis of the os ilium). Our focus was to select well-established and noninvasive methods in the field of paleodemographic studies (Buckberry, 2015; Falys & Lewis, 2011; Wittwer-Backofen et al., 2008), which can be applied in large archaeological samples by macroscopic inspection and without application of imaging techniques or further technical methods of investigation.

We did not include more recent age-estimation methods that do not suffer from age mimicry as these methods do not use the relation between trait expression and age observed in the reference sample to simply predict age based on the observed trait. In contrast, they use information on biological age development by predicting the state of the trait based on age in the reference sample. Transition analysis introduced by Boldsen et al. (2002) combines this information with a priori assumptions on the age distribution in the target sample to obtain individual age estimates. This approach has been studied by several authors (Bullock et al., 2013; Godde & Hens, 2015; Hens &

Godde, 2016; Jooste et al., 2016; Milner & Boldsen, 2012; Simon & Hubbe, 2021). Alternatively, the information can be combined with distributional assumptions on the age distribution to directly obtain an estimate of the latter (Konigsberg & Frankenberg, 1992, 2002; Müller et al., 2002). These approaches are outlined in Appendices S3 and S4.

2 | MATERIAL AND METHODS

2.1 | Age-estimation methods

Fifteen well-known and frequently used age-estimation methods were chosen for this study, all of which represent nondestructive procedures. The methods are based on assessing features on the pelvis (os pubis and os ilium), skull, dentition (tooth wear and root translucency), clavicle, or first rib. They are listed and described in Table 1. All traits were assessed according to the guidelines described in the original publications by one assessor (DN), who is a PhD student and familiar with the application of the methods. By realizing the age estimation by only one observer, we avoided interobserver errors. Additional scores or categories other than in the original publications are explained below.

Table 1 presents the 15 age-estimation methods considered in this study.

The term "age-estimation method" here refers to procedures generating a single age estimate for an individual. For most methods this is based on translating a score or a phase determined from the specific features of a given trait into an age value. The translation tables used in the paper are summarized in Table 2. Age-estimation according to Lovejoy's (1985) method of assessing dental wear is based on two translation tables for the maxilla and the mandibula, respectively. These are regarded as two separate methods in our investigation, although usually the two age estimates are averaged. The combined method of Nemeskéri et al. (1960) is based on translation tables for four different features of which only two (cranial sutures and pubic symphysis) were considered in our investigation. For the other two, the original publication required sawing open the proximal femur and humerus epiphyses, and we excluded such invasive measures in our analysis. As Brooks and Suchey (1990) reported sex-specific translation tables, this method was only applied to individuals with available sex estimations based on DSP2 (Bruzek et al., 2017). The two age-estimation methods based on dental root translucency were performed on a subsample of 100 adult individuals. This subsample consisted of 50 female and 50 male individuals over 25 years of age, with sex estimation using DSP2 and a first (preliminary) age estimation using tooth wear (Lovejoy, 1985 and Miles, 1963). If available, one single-rooted tooth from each quadrant was assessed and the age estimates averaged (tooth notation according to FDI ISO 3950; Alt & Türp, 1998). Overall, tooth root transparency was measured on the labial surface in 383 single-rooted teeth. The methods by Miles (1963) and Kunos et al. (1999) are not included in Table 2, as they require reading the age from a graph or from several tables, depicting the relation of the traits to age.

Table 2 is the translation table used in application of the age-estimation methods.

Five of the age-estimation methods translate scores to an interval and not a single age. Here, the midpoint of the interval was used as age estimate. The methods of Todd (1920) and Lovejoy et al. (1985) do not assign an upper bound to the interval for the oldest individuals. Here, we made an own choice for the age estimate. Szilvássy (1977) assigns an age between 26 and 30 to the oldest subjects. Here, the assessor (DN) assigned an age of 30+ for individuals in whom epiphysis closure was achieved some time since and showing first signs of osteoarthritis of the facies articularis sternalis. For three of the age-estimation methods, the assessor (DN) introduced intermediate score values to reflect that some individuals could not be placed in one of two predefined score values. For these intermediate values, age was interpolated from the two neighboring values.

2.2 | The target population: An archaeological sample

All 675 adult individuals from the early medieval cemetery of Mannheim-Seckenheim ("Hermsheimer Bösfeld"), Germany, were analyzed in this study (see also Navitainuck et al., 2021). Classification as an adult individual was based on complete epiphyseal closure (especially the spheno-basilar junction).

2.3 | Mean values of the reference samples

The original publications describing the development of each age-estimation method were screened for any information on the mean age of the reference sample used. Only in one publication was the mean value reported directly. However, with one exception, we were able to reconstruct the specific mean ages in an approximate manner at least. The details are described in Appendix S6. Hence, with the exception of Lovejoy's method based on the facies auricularis, for each age-estimation method, a corresponding reference sample mean value could be identified. Note that the study of Miles (1963) uses a pseudo-reference sample of quasi-age-known/age-determined individuals. In younger individuals, age was assessed based on determining the functional age of the molars by means of seriation and comparison with age-known individuals from other series. Miles states that the system is one of decreasing reliability with advancing age because age was extrapolated for older individuals due to antemortem tooth loss and so on.

2.4 | Statistical methods

Before investigating the relation between population mean age estimates and the mean age of the reference samples, we performed preliminary analyses with regard to the single age-estimation methods. The applicability of the methods is depicted by the number of

TABLE 1 Aging methods

Aging technique and trait by skeletal element	Abbreviation	Reference	Age assessment	Reference population
Os pubis				
Todd, pubic symphysis	PS-T	Todd (1920)	10 phases translated to age ranges of 2 to 5 years	Todd collection
Combined method, pubic symphysis	PS-N	Nemeskéri et al. (1960)	5 phases, translated to mean age and SD	Recent forensic sample, Budapest
Suchey–Brooks method, pubic symphysis	PS-BS	Brooks and Suchey (1990)	6 phases with sex-specific images, translated to sex-specific mean age and SD	Suchey collection 1977–1979 autopsy sample, Los Angeles
Os ilium				
Lovejoy et al., facies auricularis	FAu-L	Lovejoy et al. (1985)	8 phases translated into 5 year age ranges	Hamann–Todd collection
Buckberry amd Chamberlain, facies auricularis: Transverse organization, surface texture, apical changes, microporosity, and macroporosity	FAu-BC	Buckberry and Chamberlain (2002)	5 features assigned to 3 to 5 phases. Resulting composite score reduced to 7 stages, translated to mean age and SD	Preliminary test on Bedford et al. (1989) sample, blind test on Christ Church, Spitalfields, London, sample
Cranium				
Combined method, endocranial suture closure	EnS-N	Nemeskéri et al. (1960)	5 phases, translated to mean age and SD	Recent forensic sample, Budapest
Meindl and Lovejoy, ectocranial suture closure: Vault and lateral	EcSv-ML EcSa-ML	Meindl and Lovejoy (1985)	5/7 features assigned to 4 phases. Summary score translated to mean age and SD	Hamann–Todd collection
Clavicle				
Szilvássy, facies articularis sternalis	FAr-S	Szilvássy (1977)	3 phases translated to age ranges	Recent forensic sample, Vienna
First rib				
Kunos et al., sternal end, caput costae, tuberculum costae	RI-K	Kunos et al. (1999)	A time line indicating the occurrence of specific features of each trait allows to translate specific features into an age. Age estimates are then averaged over features and traits.	Hamann–Todd collection
Dentition				
Miles method, dental wear of lower molars	DW-M	Miles (1963)	Wear stages are plotted on a time line allowing to translate the wear stage of each molar into an age. The final age estimation is the average of the estimates of all three molars	Breedon-on-the-hill, Leicestershire, Anglo-Saxon burial site, 700–900 AD
Lovejoy, dental wear: Maxilla & mandibular	DW-L1 DW-L2	Lovejoy (1985)	9/10 functional attritional stage translated into 5 to 11 year ranges	Libben site population (hunter–gatherer diet)
Lamendin method, tooth root translucency	RT-L	Lamendin et al. (1992)	Explicit formula: $A = (0.18 * P) + (0.42 * T) + 25.53$	Recent French autopsy sample
Prince and Ubelaker, tooth root translucency	RT-PU	Prince and Ubelaker (2002)	Explicit sex-specific formulas: Males: $A = 1.10\ (RH) + 0.31\ (P) + 0.39\ (T) + 11.82$ Females: $A = 0.15\ (RH) + 0.29\ (P) + 0.39\ (T) + 23.17$	Terry collection

Note: The variables used in the formulas based on root translucency are abbreviated in the following manner: A = age, RH = root height, P = height of periodontium reduction, T = translucency height.

TABLE 2 Translation of phases/stages into ages

Aging method	Phase/stage		Mean age
PS-T		1	18.5
		1.5	*19.5*
		2	20.5
		3	23
		4	25.5
		5	28.5
		6	32.5
		7	37
		8	41.5
		9	47
		10	65†
PS-N		1	26.3
		2	46.5
		3	51.1
		3.5	*54.6*
		4	58.1
		5	68.5
PS-BS		1	18.5
Male		2	23.4
		3	28.7
		4	35.2
		5	45.6
		6	61.2
PS-BS		1	19.4
Female		2	25.0
		3	30.7
		4	38.2
		5	48.1
		6	60.0
FAu-L		1	22
		2	27
		3	32
		4	37
		5	42
		6	47
		7	55
		8	70†
FAu-BC	5, 6		17.33
	7, 8		29.3
	9, 10		37.9
	11, 12		51.4
	13, 14		59.9
	15, 16		66.7
	17, 18, 19		72.3
EnS-N		1	28.6
		2	43.7
		3	49.1
		4	60.0
		5	65.4
EcSv-ML	1, 2		30.5
	3, 4, 5, 6		34.7
	7, 8, 9, 10, 11		39.4
	12, 13, 14, 15		45.2
	16, 17,18		48.8
	19, 20		51.5
EcSa-ML		1	32.0
		2	36.2
	3, 4, 5		41.1
		6	43.4
	7, 8		45.5
	9, 10		51.9
		11, 12, 13, 14	56.2
FAr-S		1	19
		1.5	*21*
		2	23
		2.5	*25.5*
		3	*28*
		>3	
DW-L1		1	15
		2	18
		3	20
		4	22
		5	27
		6	32.5
		7	37.5
		8	45
DW-L2		1	15
		2	18
		3	20
		4	22
		5	27
		6	32.5
		7	37.5
		8	42.5
		9	50

Note: The methods PS-T, FAu-L, FAr-S, DW-L1, and DW-L2 translate the scores or phases to a small age range and not a single value. Here the middle of the interval was used. For PS-T and FAu-L, the original publications do not report an upper bound for the respective oldest stages. The values chosen by us are marked with an †. For FAr-S, the oldest stage was too wide in order to justify such a choice so no age value was chosen for stage >3. Additional (intermediate) score values chosen by the assessor are marked in italics. Not all values shown in this table were actually assigned to individuals in the sample analyzed in this study. In particular, this applied to EnS-N, Stage 5 of which was never assigned.

individuals for whom the methods could be applied. The distribution of the individual age estimates achieved by the single methods is visualized by dot plots. To explore systematic differences between the age-estimation methods, we considered all pairwise comparisons. For each pair of methods, we determined the mean difference between the two methods including only those individuals for whom both methods could be applied. A two-sided paired *t* test was used to assess the statistical significance for a deviation of the mean difference from 0, using a significance level of 5%. To depict the degree of association between the individual age estimates created by different age-estimation methods, we considered all pairwise Pearson correlation coefficients. The coefficients are displayed as a heat map, using the heatplot command provided by Jann (2019).

In the main analysis, the mean age estimates from the Bösfeld sample are related to the mean ages of the reference samples using linear regression. The variation explained by the differences in reference sample means is measured by the adjusted R^2 value (expressed as a percentage). To take into account that the target population mean estimates have a known imprecision, we also performed a meta-regression using the observed standard error of the mean as additional input. The slope of the regression is reported together with a 95% confidence interval. The slope is used to describe how differences in mean age between two reference samples translate into a difference between population mean age estimates.

All statistical computations have been performed using Stata 16.1. The computations and the data used in the figures are documented in Appendix S7.

TABLE 3 The number of subjects scored for each age-estimation method

Abbreviation	Aging method	*n*
PS-T	Pubic symphysis (Todd)	86
PS-N	Pubic symphysis (Nemeskeri)	87
PS-BS	Pubic symphysis (Brooks and Suchey)	65
FAu-L	Facies auricularis (Lovejoy)	272
FAu-BC	Facies auricularis (Buckberry and Chamberlain)	255
EnS-N	Endocranial suture closure (Nemeskeri)	83
EcS1-ML	Ectocranial suture closure vault (Meindl and Lovejoy)	102
EcS2-ML	Ectocranial suture closure anterior (Meindl and Lovejoy)	81
FAr-S	Facies articularis sternalis (Szilvássy)	60
FR-K	First rib (Kunos et al.)	94
DW-M	Tooth wear molars (Miles)	380
DW-L1	Tooth wear maxilla (Lovejoy)	385
DW-L2	Tooth wear mandibula (Lovejoy)	431
RT-L	Lamendin	100[a]
RT-PU	Prince and Ubelaker	100[a]

[a]Root translucency-based methods were applied to a random sample of 100 subjects.

3 | RESULTS

3.1 | Applicability of the age-estimation methods

Table 3 gives an overview of the applicability of the different age-estimation methods.

The methods based on tooth wear were most widely applicable, followed by those based on the facies auricularis, applicable in about 40% of individuals. All other age-estimation methods were applicable (or applied) in 102 or fewer individuals.

3.2 | Observed age distributions and systematic differences between age-estimation methods

Figure 1 depicts the distribution of the age estimates for each method. As some methods are based on only a few score values, they also produce few different age values, and hence their distributions are quite discrete. Substantial differences can be observed with respect to the mean values. These are, however, partially due to the fact that some of the methods mainly distinguish ages within a certain range. The assessment of the sternal clavicle of Szilvássy (1977), for example, produces values only in the range up to 30 years of age, whereas the methods of Meindl and Lovejoy (1985) for cranial suture closure can only produce age estimates above that age, and the two methods of Nemeskéri et al. (1960) each only produce a single value below the age of 43.

In general, we observe distinct differences in the age distribution and the mean age between the different age-estimation methods. This holds even for different methods based on the same trait. The three estimations based on the pubic symphysis give mean estimates in the range from 33 to 51 years, while the two methods based on the facies auricularis differ in the estimated mean age by 8 years.

The differences observed in Figure 1 may to some degree be explained by differences in the subpopulations to which each age-estimation method could be applied. However, statistically significant differences are also frequently observed in pairwise comparisons if considering only those subjects assessed by both methods (Figure 2). In particular, the two methods by Nemeskéri et al. (1960) produce on average estimates 10 to 15 years higher than all other methods. In contrast, the methods of Brooks and Suchey (1990), Szilvássy (1977), and Miles (1963) tend to produce values lower than most other age-estimation methods. In general, differences between the age-estimation methods are often 5 years or more, that is, of a relevant absolute magnitude.

3.3 | Correlations among age estimates

The pairwise correlations between the age estimates of the different methods are shown in Figure 3. As may have been expected,

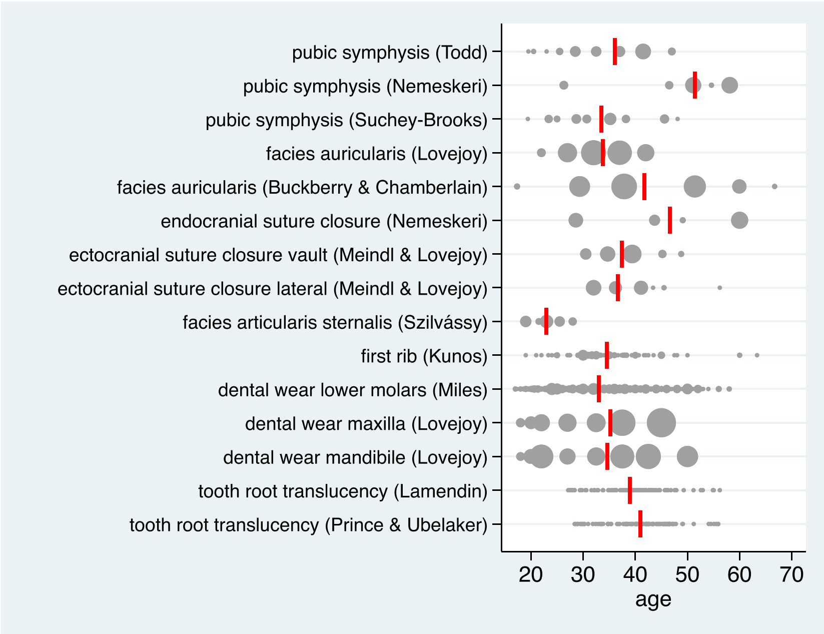

FIGURE 1 The distribution of the estimated age values for each method. The size of the dots represents the number of individuals from the target population assigned a specific age value. The red vertical lines indicate the mean age estimate for each method. [Colour figure can be viewed at wileyonlinelibrary.com]

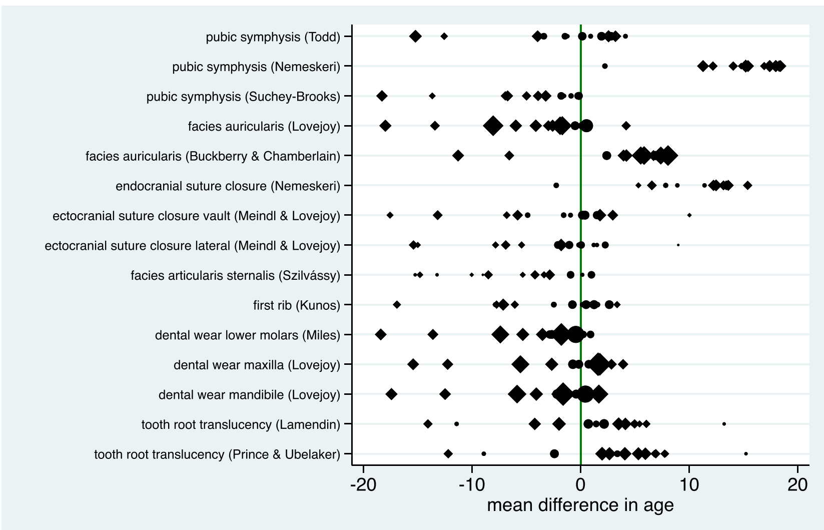

FIGURE 2 Mean differences in the pairwise comparison of age-estimation methods. Each dot refers to the mean difference in age estimation when comparing two age-estimation methods and including all individuals for whom both methods are applicable. The size of the plot symbol is relative to the number of individuals. Values with differences significant from 0 are marked by diamonds, and all others are marked by dots. [Colour figure can be viewed at wileyonlinelibrary.com]

we can observe high correlations for methods based on the same trait. An exception is the methods based on the cranial suture closure where correlations are less pronounced, in particular between the vault and the anterior suture complex. In general, at least moderate correlations among all age-estimation methods are to be expected, as all methods aim to estimate age. However, distinct differences in the degree of correlation are observed between methods assessing different traits or features. In particular, the estimates based on dental root translucency showed partially negative or rather low positive correlations with the majority of the other age-estimation methods. In addition, age estimation based on endocranial suture closure showed on average a higher correlation to the other age-estimation methods than age estimation based on the ectocranial sutures.

3.4 | Relation of target population means to mean age in the reference sample

Figure 4 depicts the relation between the estimated target population age means according to the different age-estimation methods and the mean age in the corresponding reference samples. Generally, a high correlation is observed, with the mean values in the reference sample explaining 68% of the variation in the estimated target population age means.

The age-estimation methods by Szilvássy (1977) and Nemeskéri et al. (1960) contribute markedly to the variation in the target population means. In our appraisal, this is not only due to the notably low or high age means in the corresponding reference samples but is also related to the limited number of possible values the age estimates for

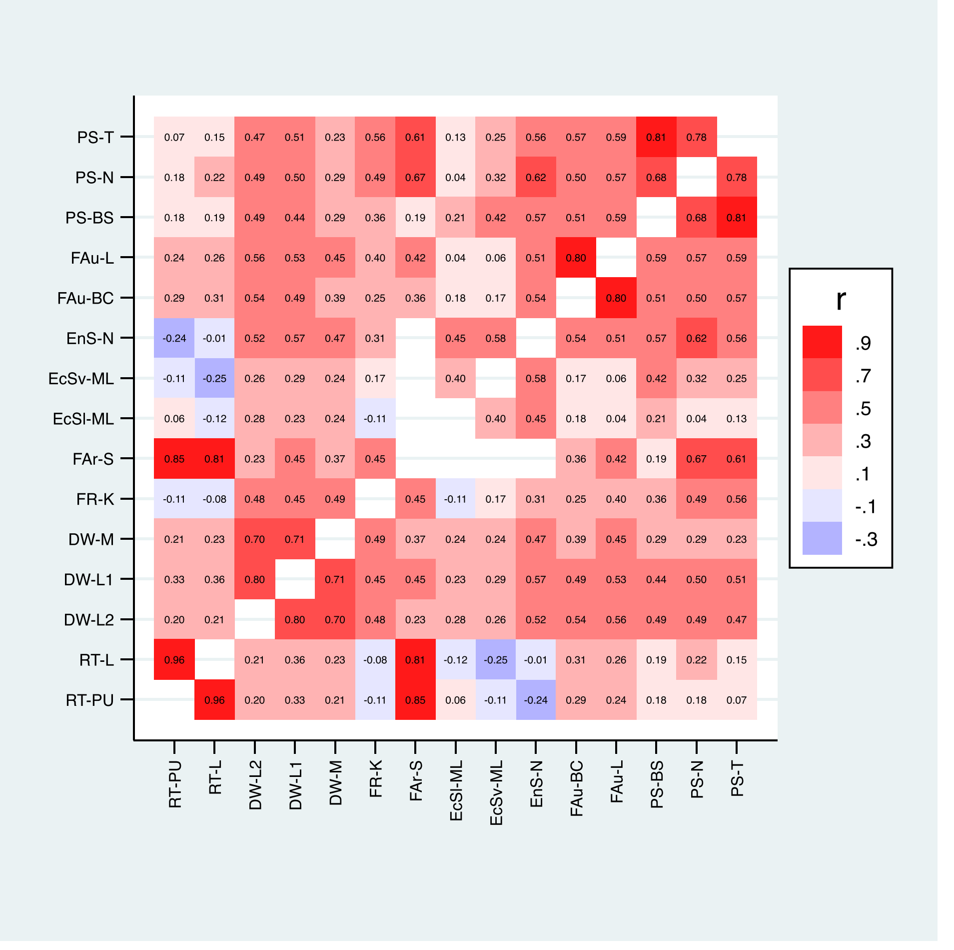

FIGURE 3 Pairwise correlations among the 15 age-estimation methods expressed in a heat map. Correlations are only computed if there are at least five individuals with age estimates for both methods. For method abbreviations, see Table 1. [Colour figure can be viewed at wileyonlinelibrary.com]

these methods can assume. This is particularly obvious for the method of Szilvássy (1977), allowing to assign an explicit maximum age of 30. Furthermore, the methods of Nemeskéri et al. (1960) assign only a single age value below the age of 43 or 46 years, respectively, and hence fail to differentiate between individuals below those ages.

Excluding the three corresponding data points from the statistical analysis, 80% of the remaining variation in the target population age means are explained by the variation in mean age in the reference samples. Taking into account the imprecision of the target population mean estimates, an even higher 86% of the variation is explained. Fitting a corresponding regression line (cf. Figure 4) obtains a slope of 0.28 (95% CI: 0.18–0.38). This means that a difference of 1 year between the mean ages in two reference samples predicts a difference of 0.28 years in the corresponding target population age means. We can rescale this statement also in the following way: We find that a difference of 3.6 years between the mean ages in two reference samples predicts a difference of 1 year in the corresponding target population means.

A concrete example may illustrate the relevance of this finding. The age-estimation methods based on dental wear considered in this paper use a reference sample with a mean age about 35 years. On the other hand, the methods based on root translucency use reference samples with a mean age of about 57 years. The difference in age means between these methods is 22 years, which implies that we can expect a difference of $22 * 0.28 = 6.2$ years in population mean age estimates between the methods. Similarly, we can state that a difference of 18 years in mean age between two reference samples implies a difference of 5 years in the target population age means, and a difference of 36 years implies a difference of 10 years.

4 | DISCUSSION

In this study, we applied 15 different age-estimation methods to 675 adult individuals from an archaeological population. In general, estimates of the mean age in the target population differed substantially between the different age-estimation methods used. This is also true for pairwise comparisons between methods, restricting the analysis to those individuals for whom both age-estimation methods were applicable. This is in line with a similar investigation presented by Wittwer-Backofen et al. (2008) based on 121 individuals from the Early Medieval skeletal sample of Lauchheim, Germany. In particular, they observed the highest target population mean values for the so-called "complex method" of Nemeskéri et al. (1960) and the auricular surface approach by Buckberry and Chamberlain (2002), as it was also the case in our study. Furthermore, Clark et al. (2019) applied five

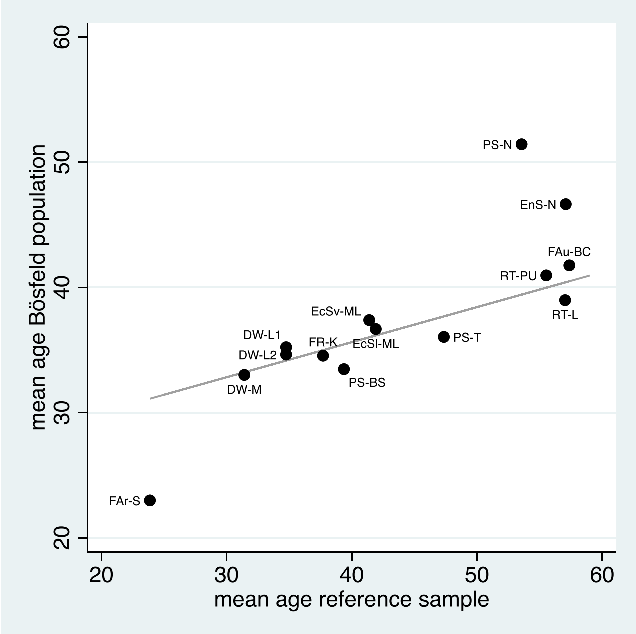

FIGURE 4 Scatterplot of the target population age means according to the different age-estimation methods and the mean age in the corresponding reference sample. The line shows the fitted regression after exclusion of three data points (PS-N, EnS-N, and FAr-S). A justification for excluding these data points is given in the text. [Colour figure can be viewed at wileyonlinelibrary.com]

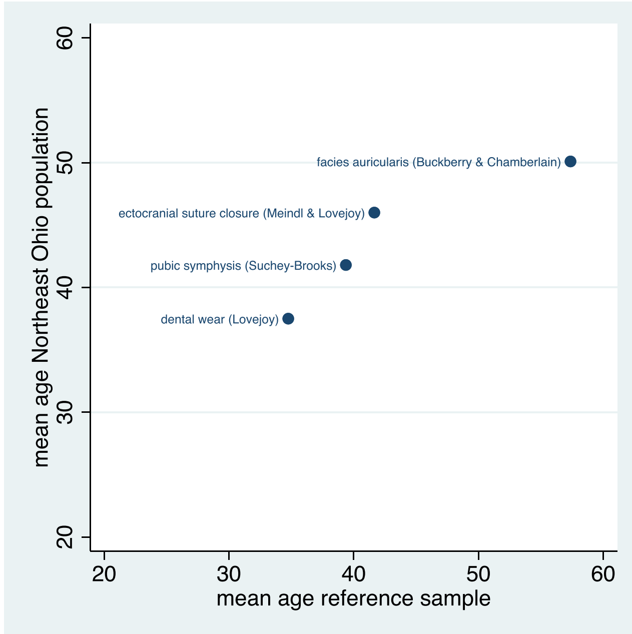

FIGURE 5 Scatterplot of the target population age means according to four different age-estimation methods in the study of Clark et al. (2019) and the mean age in the corresponding reference sample. Note that Clark et al. do not differentiate between vault and anterior in using the ectocranial suture closure and evaluated maxilla and mandibula together in evaluating dental wear. [Colour figure can be viewed at wileyonlinelibrary.com]

different age-estimation methods to 93 individuals from Late Archaic and Prehistoric periods in Northeast Ohio. Four of these methods were also considered in our study, returning nearly the same ranking in mean age in both their and our studies.

In addition, we investigated the variation in the target population mean age estimates with regard to the well-known phenomenon of age mimicry, which suggests that target population age estimates to a certain degree reflect the age structure in the reference samples used in developing the specific age-estimation methods. We were able to demonstrate that up to 86% of the variation in the target population mean estimates is explained by the variation in mean age in the reference samples. We found that differences of 18 years in the mean values between two reference samples explain a difference of 5 years in target population mean estimates. Differences of such a magnitude between mean values in the reference samples are not uncommon, as illustrated by our investigation. Consequently, some care is needed when comparing individual age estimates or population mean age estimates if different age-estimation methods are used.

A monotonic relation between the mean target population age and the mean age in the reference sample is also detectable in the results of Clark et al. (2019), as shown in Figure 5. Unfortunately, with four age-estimation methods and a small target population that study does not allow to estimate the regression line in a reliable manner.

Although the present study focused on population mean estimates, insights into the bias due to using different age-estimation methods also apply to individual age estimates. A bias in the population mean implies that individual age estimates are on average biased to the same degree. However, the variation of individual age estimates across different methods cannot be explained by the differences in reference sample age means alone. In addition, variation in individual age estimates also reflects different expressions of separate age indicators within a single individual.

In principle, the relation observed between the mean values in the reference samples and target population mean values for the different age-estimation methods can be used for some type of bias correction. Averaging among the 14 age-estimation methods included in the comparison, we obtain an average target population mean value of 37.4 years and an average reference sample mean value of 43.8. This difference of 5.4 years suggests that the estimate 37.4 overestimates the true target population mean by $0.28 * 5.4 = 1.5$ years. However, this correction changes the estimated target population mean age, and the argument has to be iterated. In Appendix S5, a corresponding bias correction is outlined, and a bias-corrected estimate of 33.9 years is obtained for the true target population mean.

However, instead of developing bias corrections, it would be more useful to use approaches that allow estimating the mean age in the target population (and even the whole age distribution) without suffering age mimicry. As mentioned above, such techniques do exist and have been applied successfully; the principle behind these techniques is outlined in Appendix S3, where also further references are given. There are, however, two challenges with such approaches.

First, they require that in the reference population the distribution of the age indicator is described as a function of age. This is in contrast to all age-estimation methods considered in this paper, which are based on considering (mean) age as a function of the age indicator. The reporting in the original publications usually does not allow to derive a description in the opposite direction. Second, these techniques require fitting specific types of models to the data from the target sample, which is beyond the scope of standard statistical packages.

Although age mimicry is a quantitative phenomenon, it is usually described by qualitative terms in the literature. For example, Clark et al. (2019) refer to age mimicry as "a phenomenon in which the age distribution of the target population mirrors the distribution of the reference population," while Liversidge et al. (2010) use the phrase "age estimation mirrors the data upon which a method is based." These descriptions refer to the original investigation of Bocquet-Appel and Masset (1982), who presented a graphical comparison of the age distributions in reference samples with the estimated age distribution in corresponding target populations. Our study serves as a direct extension of that investigation in considering such 14 comparisons and focusing on the mean estimates only. We have thus been able to quantify the effect of age mimicry in an empirical manner.

Such a quantification can be also helpful for expert age estimations based on long-term experience. In that scenario a final estimate should also consider that some age-estimation methods are well known to tend to either overestimate age-at-death or otherwise underestimate old age when using the given age estimates. This applies in particular to the complex method introduced by Nemeskéri et al. (cf. Jopp, 2007; Kemkes-Grottenthaler, 1996, 2002; Langenscheidt, 1985). Quantifications as provided by our investigation can be helpful for informing to which degree the suggested age estimates are to be corrected. However, this is always a tricky issue, as the bias depends on the difference in mean age between the target population and the reference sample, meaning that some prior guess about the mean age of the target population is necessary in order to deduce the correct value for the correction.

We were also able to confirm some correlations between the different age-estimation methods, reflecting that the features they evaluate are indeed related to age. The poorest association was found for the methods based on dental root translucency, which has been also observed by Wittwer-Backofen et al. (2008). Correlations between age estimations based on the same traits were rather high, with cranial suture closure as a notable exception. This result is in line with previous discussions about the potential differences between the three age-estimation approaches based on suture closure (Bindl, 2009; Key et al., 1994; Meindl & Lovejoy, 1985). In our sample, age estimation based on endocranial suture closure showed on average a higher correlation with the other age-estimation methods than ectocranial suture closure. In contrast, Bindl (2009) regarded ectocranial suture closure as superior to endocranial closure. Key et al. (1994) recommended combining both endocranial and ectocranial suture closure to enable age estimation for both young and older adults.

Our investigation is, however, subject to some limitations. All age-estimation methods were applied by a single observer, which may imply that systematic differences noted between the age-estimation methods may be unequal in a different observer. Similarly, we disregarded any differences in the sex distributions among the reference samples and between the reference samples and the target population, which may also affect differences in the target population mean values. However, because the unexplained variation in the target population mean estimates is small, we may conclude that the impact of these limitations is of negligible magnitude and our calculations are trustworthy.

5 | CONCLUSIONS

In this study, we were able to quantify the effect of age mimicry on target population age means in relation to the method-specific reference samples used for age estimation in archaeological samples. Our results suggest that differences of about 3.5 years in mean age between two reference samples imply a systematic difference in age estimates of 1 year. Because large differences in mean age between reference samples used to develop different age-estimation methods are common, some care is needed in interpreting differences between individual age estimates or population mean age estimates in cases where different age-estimation techniques are used.

ACKNOWLEDGMENT

The study was funded by a scholarship of Studienstiftung des Deutschen Volkes. Open access funding provided by Universitat Basel.

CONFLICT OF INTEREST

There is no conflict of interest.

DATA AVAILABILITY STATEMENT

Data are available on request from authors.

ORCID

Denise U. Navitainuck https://orcid.org/0000-0002-6082-0778

REFERENCES

Alt, K. W., & Türp, J. C. (1998). Roll call: Thirty-two white horses on a red field. The advantages of the FDI two-digit system of designating teeth. In K. W. Alt, F. Rösing, & M. Teschler-Nicola (Eds.), *Dental anthropology* (pp. 41–56). Springer.

Aykroyd, R. G., Lucy, D., Pollard, A. M., & Solheim, T. (1997). Regression analysis in adult age estimation. *American Journal of Physical Anthropology*, *104*, 259–265. https://doi.org/10.1002/(SICI)1096-8644(199710)104:2<259::AID-AJPA11>3.0.CO;2-Z

Bedford, M. E., Russell, K. F., & Lovejoy, C. O. (1989). *The auricular surface aging technique: 16 color photographs with descriptions*. Kent State University.

Bindl, R. (2009). *Analytische Sterbealtersbestimmung von Skelettfunden. Möglichkeiten und Grenzen bei der Bearbeitung von historischem und rezentem Skelettmaterial. Nat. Diss*. Universität Jena.

Bocquet-Appel, J. P., & Masset, C. (1982). Farewell to paleodemography. *Journal of Human Evolution*, *11*, 321–333. https://doi.org/10.1016/S0047-2484(82)80023-7

Boldsen, J., Milner, G., Konigsberg, L., & Wood, J. (2002). Transition analysis: A new method for estimating age from skeletons. In R. D. Hoppa & J. W. Vaupel (Eds.). *Paleodemography: Age distributions from skeletal samples* (pp. 73–106). Cambridge University Press.

Brooks, S., & Suchey, J. M. (1990). Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution*, *5*(3), 227–238. https://doi.org/10.1007/BF02437238

Bruzek, J., Santos, F., Dutailly, B., Murail, P., & Cunha, E. (2017). Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology. *American Journal of Physical Anthropology*, *164*, 440–449. https://doi.org/10.1002/ajpa.23282

Buckberry, J. (2015). The (mis)use of adult age estimates in osteology. *Annals of Human Biology*, *42*(4), 323–331. https://doi.org/10.3109/03014460.2015.1046926

Buckberry, J. L., & Chamberlain, A. T. (2002). Age estimation from the auricular surface of the ilium: A revised method. *American Journal of Physical Anthropology*, *119*, 231–239. https://doi.org/10.1002/ajpa.10130

Buikstra, J. E., & Ubelaker, D. H. (1994). Standards for data collection from human skeletal remains. In *Proceedings of a Seminar at the Field Museum of Natural History, Organized by Jonathan Haas* (p. 44). Arkansas Archeological Survey.

Bullock, M., Marquez, L., Hernandez, P., & Ruιz, F. (2013). Paleodemographic age-at-death distributions of two Mexican skeletal collections: A comparison of transition analysis and traditional aging methods. *American Journal of Physical Anthropology*, *1522*, 67–78.

Clark, M. A., Simon, A., & Hubbe, M. (2019). Aging methods and age-at-death distributions: Does transition analysis call for a re-examination of bioarchaeological data? *International Journal of Osteoarchaeology*, *30*, 206–217. https://doi.org/10.1002/oa.2848

Falys, C., & Lewis, M. E. (2011). Proposing a way forward: A review of standardisation in the use of age categories and ageing techniques in osteological analysis (2004–2009). *International Journal of Osteoarchaeology*, *21*, 704–716. https://doi.org/10.1002/oa.1179

Godde, K., & Hens, S. M. (2015). Modeling senescence changes of the pubic symphysis in historic Italian populations: A comparison of the Rostock and forensic approaches to aging using transition analysis. *American Journal of Physical Anthropology*, *156*, 466–473. https://doi.org/10.1002/ajpa.22671

Hens, S. M., & Godde, K. (2016). Auricular surface aging: Comparing two methods that assess morphological change in the ilium with Bayesian analyses. *Journal of Forensic Sciences*, *61*, S30–S38. https://doi.org/10.1111/1556-4029.12982

Herrmann, B. (Ed.) (1990). *Prähistorische Anthropologie. Leitfaden der Feld- und Labormethoden*. Springer. https://doi.org/10.1007/978-3-642-61514-6

Hoppa, R. D., & Vaupel, J. W. (2002). The Rostock Manifesto for paleodemography: The way from stage to age. In R. D. Hoppa & J. W. Vaupel (Eds.), *Paleodemography. Age distributions from skeletal samples* (pp. 1–8). Cambridge University Press. https://doi.org/10.1017/CBO9780511542428.001

Jann, B. (2019). *HEATPLOT: Stata module to create heat plots and hexagon plots. Statistical Software Components S458598*. Boston College Department of Economics. Revised 27 Aug 2021

Jooste, N., L'Abbé, E. N., Pretorius, S., & Steyn, M. (2016). Validation of transition analysis as a method of adult age estimation in a modern south African sample. *Forensic Science International*, *266*, 580.e1–580.e7. https://doi.org/10.1016/j.forsciint.2016.05.020

Jopp, E. (2007). *Methoden zur Alters- und Geschlechtsbestimmung auf dem Prüfstand. Eine rechtsmedizinische empirische Studie* (Vol. 8). Forschungsergebnisse aus dem Institut für Rechtsmedizin der Universität Hamburg. Kovač.

Kemkes-Grottenthaler, A. (1996). Critical evaluation of osteomorphognostic methods to estimate adult age at death: A test of the "complex method". *Homo*, *46*, 280–292.

Kemkes-Grottenthaler, A. (2002). Aging through the ages: Historical perspectives on age indicator methods. In R. D. Hoppa & J. W. Vaupel (Eds.), *Paleodemography. Age distributions from skeletal samples* (pp. 48–72). Cambridge University Press. https://doi.org/10.1017/CBO9780511542428.004

Key, C., Aiello, L., & Molleson, T. (1994). Cranial suture closure and its implications for age estimation. *International Journal of Osteoarchaeology*, *4*, 193–207. https://doi.org/10.1002/oa.1390040304

Konigsberg, L. W., & Frankenberg, S. R. (1992). Estimation of age structure in anthropological demography. *American Journal of Physical Anthropology*, *89*, 235–256. https://doi.org/10.1002/ajpa.1330890208

Konigsberg, L. W., & Frankenberg, S. R. (2002). Deconstructing death in paleodemography. *American Journal of Physical Anthropology*, *117*, 297–309. https://doi.org/10.1002/ajpa.10039

Kunos, C. A., Simpson, S. W., Russell, K. F., & Hershkovitz, I. (1999). First rib metamorphosis: Its possible utility for human age-at-death estimation. *American Journal of Physical Anthropology*, *110*, 303–323. https://doi.org/10.1002/(SICI)1096-8644(199911)110:3<303::AID-AJPA4>3.0.CO;2-O

Lamendin, H., Baccino, E., Humbert, J. F., Tavernier, J. C., Nossintchouk, R. M., & Zerilli, A. (1992). A simple technique for age estimation in adult corpses: The two criteria dental method. *Journal of Forensic Sciences*, *37*, 1373–1379.

Langenscheidt, F. (1985). *Methodenkritische Untersuchungen zur Paläodemographie am Beispiel zweier fränkischer Gräberfelder* (Vol. 2). Materialien Zur Bevölkerungswissenschaft. Bundesinst. f. Bevölkerungsforschung.

Liversidge, H. M., Smith, B. H., & Maber, M. (2010). Bias and accuracy of age estimation using developing teeth in 946 children. *American Journal of Physical Anthropology*, *143*, 545–554. https://doi.org/10.1002/ajpa.21349

Lohrke, B. (2004). *Kinder in der Merowingerzeit. Gräber von Mädchen und Jungen in der Alemania. Freiburger Beiträge zur Archäologie und Geschichte des ersten Jahrtausends* (Vol. 9). Leidorf.

Lovejoy, C. O. (1985). Dental wear in the Libben population: Its functional pattern and role in the determination of adult skeletal age at death. *American Journal of Physical Anthropology*, *68*, 47–56. https://doi.org/10.1002/ajpa.1330680105

Lovejoy, C. O., Meindl, R. S., Pryzbeck, T. R., & Mensforth, R. P. (1985). Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology*, *68*, 15–28. https://doi.org/10.1002/ajpa.1330680103

Martin, R. (1928). *Lehrbuch der Anthropologie* (2nd ed). Fischer, Stuttgart.

Meindl, R. S., & Lovejoy, C. O. (1985). Ectocranial suture closure: A revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American Journal of Physical Anthropology*, *68*, 57–66. https://doi.org/10.1002/ajpa.1330680106

Mensforth, R. P. (1990). Paleodemography of the Carlston Annis (Bt-5) late archaic skeletal population. *American Journal of Physical Anthropology*, *82*, 81–99. https://doi.org/10.1002/ajpa.1330820110

Miles, A. E. W. (1963). The Dentition in the Assessment of Individual Age in Skeletal Material. In D. Brothwell (Ed.), *Dental anthropology* (pp. 191–209). Pergamon Press. https://doi.org/10.1016/B978-0-08-009823-4.50015-X

Milner, G. R., & Boldsen, J. L. (2012). Transition analysis: A validation study with known-age modern American skeletons. *American Journal of Physical Anthropology*, *148*, 98–110. https://doi.org/10.1002/ajpa.22047
Müller, H. G., Love, B., & Hoppa, R. D. (2002). Semiparametric methods for estimating paleodemographic profiles from age indicator data. *American Journal of Physical Anthropology*, *117*, 1–14.
Navitainuck, D. U., Vach, W., Alt, K. W., & Schibler, J. (2021). Best practice for osteological sexing in forensics and bioarchaeology: The utility of combining metric and morphological traits from different anatomical regions. *International Journal of Osteoarchaeology*, *31*, 987–1000. https://doi.org/10.1002/oa.3014
Nemeskéri, J., Harsányi, L., & Acsádi, G. (1960). Methoden zur Diagnose des Lebensalters von Skelettfunden. *Anthropologischer Anzeiger*, *24*, 70–95.
Prince, D. A., & Ubelaker, D. H. (2002). Application of Lamendin's adult dental aging technique to a diverse skeletal sample. *Journal of Forensic Sciences*, *47*, 107–116. https://doi.org/10.1520/JFS15209J
Simon, A. M., & Hubbe, M. (2021). The accuracy of age estimation using transition analysis in the Hamann-Todd collection. *American Journal of Physical Anthropology*, 1–9. https://doi.org/10.1002/ajpa.24260
Szilvássy, J. (1977). Altersschätzung an den sternalen Gelenkflächen der Schlüsselbeine. *Beiträge zur Gerichtlichen Medizin*, *35*, 343–345.
Todd, T. W. (1920). Age changes in the pubic bone. I the male white pubis. *American Journal of Physical Anthropology*, *3*, 285–334. https://doi.org/10.1002/ajpa.1330030301
Wittwer-Backofen, U., Buckberry, J., Czarnetzki, A., Doppler, S., Grupe, G., Hotz, G., Kemkes, A., Larsen, C. S., Prince, D., Wahl, J., Fabig, A., & Weise, S. (2008). Basics in Paleodemography. A comparison of age indicators applied to the early medieval skeletal sample of Lauchheim. *American Journal of Physical Anthropology*, *137*, 384–396. https://doi.org/10.1002/ajpa.20881

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Navitainuck, D. U., Vach, W., Pichler, S. L., & Alt, K. W. (2022). Age-at-death estimation in archaeological samples: Differences in population means resulting from different aging methods can be predicted from the mean ages of method-specific reference samples. *International Journal of Osteoarchaeology*, *32*(6), 1226–1237. https://doi.org/10.1002/oa.3157