THE UNIVERSITY of EDINBURGH

The auricular surface as age indicator in a modern Greek sample

Citation for published version:
Michopoulou, E, Negre, P, Nikita, E & Kranioti, E 2017, 'The auricular surface as age indicator in a modern Greek sample: A test of two qualitative methods' Forensic Science International, vol. 280, no. 246, pp. 246.e1-246.e7. DOI: 10.1016/j.forsciint.2017.08.004

Digital Object Identifier (DOI):
10.1016/j.forsciint.2017.08.004

Published In:
Forensic Science International

Title: The auricular surface as age indicator in a modern Greek sample: a test of two qualitative methods

Authors: Effrosyni Michopoulou, Pierrick Negre, Efthymia Nikita, Elena F. Kranioti

PII:	S0379-0738(17)30303-1
DOI:	http://dx.doi.org/10.1016/j.forsciint.2017.08.004
Reference:	FSI 8947

To appear in:	*FSI*

Received date:	2-12-2016
Revised date:	3-8-2017
Accepted date:	4-8-2017

Please cite this article as: Effrosyni Michopoulou, Pierrick Negre, Efthymia Nikita, Elena F.Kranioti, The auricular surface as age indicator in a modern Greek sample: a test of two qualitative methods, Forensic Science Internationalhttp://dx.doi.org/10.1016/j.forsciint.2017.08.004

Technical note: The auricular surface as age indicator in a modern Greek sample: a test of two qualitative methods.

Effrosyni Michopoulou[1], Pierrick Negre, MSc[2], Efthymia Nikita, PhD[3,4], Elena F. Kranioti, MD, Phd[1,5]

[1]Edinburgh Unit for Forensic Anthropology, School of History Classics and Archaeology, University of Edinburgh, 4 Teviot place, EH8 9AG, Edinburgh, UK.

[2]Institut médico-légal , Centre Hospitalier Régional Universitaire, 191 Avenue du Doyen Gaston Giraud, 34295 Montpellier, Cedex 5, France.

[3]Science and Technology in Archaeology Research Center, The Cyprus Institute, Nicosia, Cyprus

[4]Department of Archaeology, University of Sheffield, Sheffield, UK.

[5]Forensic Pathology Division Crete, Hellenic Republic Ministry of Justice, Transparency and Human Rights, Heraklion, Crete, Greece.

Author for correspondence and reprint requests:

Elena F. Kranioti

Edinburgh Unit for Forensic Anthropology,

School of History, Classics and Archaeology, University of Edinburgh

William Robertson Wing, Old Medical School, Teviot Place,

Edinburgh, EH8 9AG

Tel. +44 (0)131 650 2368

Fax. +44 (0)131 650 2378,

E-mail: elena.kranioti@ed.ac.uk

Highlights

- This paper tests the suitability of two qualitative methods of age estimation
- The methods are based on the morphology of the auricular surface
- Error estimates are provided for direct and indirect observation
- Both methods present limited value for accurate age estimates in Greeks
- Alternative age markers are more appropriate for skeletonised remains In Greece

ABSTRACT

The auricular surface is often found very well preserved, thus age-related changes in this anatomical area can be important for any set of human remains that require identification under different taphonomic conditions. This study tests the Buckberry & Chamberlain (2002) and Schmitt et al. (2005) methods in predicting the age of individuals in a documented sample from Crete, Greece. Both methods were used to record changes on the auricular surface in a mixed-sex sample of 74 individuals, directly as well as through photographs, by two independent observers. Cohen's kappa and intra class correlation coefficients (ICC) were used in order to assess inter-observer and intra-observer agreement. Results showed than none of the methods predicted age with sufficient accuracy, as high error rates were recorded. The Schmitt et al. (2005) method performed better, mainly because the age ranges it uses are broader. Scoring through photographs does not seem to introduce bias in predicting age, as demonstrated by the high intra-observer agreement rates. Inter-observer agreement was also high. The low intra- and inter-observer error rates suggest that the poor performance of both methods in the Cretan sample is not due to a lack of clarity in the description of the morphological changes recorded on the auricular surface; rather it should be attributed to a poor correlation between these changes and age at death in our material.

KEYWORDS

Forensic population data; auricular surface; age estimation; Cretan collection

INTRODUCTION

Age has always been an issue of major importance in the study of unidentified skeletal remains. During the past decades, great effort has been devoted to the study of the age at death of individuals based on skeletal evidence, using different schemes, as well as different anatomical parts of the body. The literature on age estimation methods shows a variety of approaches, macroscopic, microscopic and analytical. Cranial suture closure [1], pubic symphyseal surface morphology [2], sternal rib end morphology [3], and auricular surface morphology [4] constitute typical examples of macroscopic morphological methods for age estimation. On a microscopic and analytical level, bone histomorphometry methods rely on the quantification of age-related remodelling changes of the bone on different skeletal elements [5, 6]. Chemical/analytical methods range from aspartic amino-acid racemisation [7, 8] and bone material properties quantification [9, 10] to DNA methylation techniques [11, 12]. Each technique presents strengths and weaknesses related to reliability, applicability, cost- and resource-effectiveness and despite the general guidelines of all professional bodies (to "use a combination of all available methods"), this is not always feasible, especially in forensic facilities with few non-medical specialists. From this point of view morphological methods are more popular as they do not require sophisticated equipment and expertise. Yet, these techniques are sensitive to observer and population bias, thus, their reliability must be demonstrated before application in forensic settings.

The current paper focuses on the morphology of the auricular surface of the os coxae as an age marker. The auricular surface is often found very well preserved, thus age-related changes in this anatomical area can be important for any set of human remains under different taphonomic conditions. The first to propose aging standards based on morphological changes on the auricular area were Lovejoy et al. [4]. Lovejoy's study uses eight modal age stages of 5 years each, starting from juveniles and ending in 60+ years, and describes in detail the changes in the appearance of the auricular surface in every age group. The features recorded include grain and density, macroporosity, billowing, striations, apex lipping, activity in the retroauricular area and transverse organization. Several scholars applied Lovejoy's method in different samples and noted that older individuals tended to be underaged, whereas the age of younger individuals was overestimated [13-17]. In addition, many studies found that the 5-year intervals proposed by Lovejoy et al. [4] are unrealistically narrow [14, 18].

More recently, Buckberry and Chamberlain [18] revised this method (hereafter B-C) to improve age predictions and developed a scoring system where the features recorded on the auricular surface include transverse organization, surface texture, porosity and apex sharpness. The revised method examines each feature as an independent variable and suggests that it is scored separately using an ordinal scale. Subsequently, all individual scores are summed per element and age is estimated according to the total score. Similarly, Schmitt et al. [19] (hereafter SC) developed a method that focuses on the same overall attributes as the Buckberry and Chamberlain [18] method, as well as the entheseal changes on the iliac tuberosity. In this case, all variables are recorded in a binary scale, except for granulation and porosity (SSPIB), for which an ordinal system is used. Finally, Igarashi et al.'s [17] approach is based on the binary scoring (presence – absence) of 13 variables per individual, mainly concerning surface texture, porosity, granularity and hypertrophy.

The aforementioned methods have been tested in different populations and the results showed a poor to fair performance. In specific, Falys et al. [20] tested the B-C [18] revised method in a sample dating from the late 17th to the early 19th centuries. Results showed that the method was not reliable due to the extensive variation in the morphological changes of the auricular surface and suggested that only very broad age assessments can be reached by adopting it. Moraitis et al. [21] applied the B-C [18] method on a modern Greek sample, namely the Athens Collection, and their results showed a variation of 5.5-12.6 years across all age stages. Stage VII was the one that appeared to correspond better with the known ages of the individuals (56-78 years). However, Moraitis et al.'s [21] study agrees with previous ones that indicated age underestimation for older individuals along with overestimation for younger ones.

The present work tests the B-C [18] and SC [19] methods using a modern Cretan sample, with documented age and sex. The aim is to examine both how accurately these methods can predict the age at death of the individuals and to quantify the inter- and intra-observer error. A secondary goal of the study is to investigate whether high quality photographs can provide reliable enough information to be used in place of direct observations as it is a common phenomenon for forensic practitioners to provide consultation from images rather than direct examination of the body.

MATERIALS AND METHODS

The sample used for this study consists of 74 Cretan individuals of known sex and age at death from the Cretan collection, housed at the facilities of the Forensic Pathology Division of the Hellenic Ministry of Justice and Human Rights in Crete. The skeletal remains were exhumed from St Konstantinos and Pateles cemeteries in Heraklion, Crete. The individuals examined here died between 1963 and 1997. More information on the collection can be found in Kranioti et al. [22] and Kranioti and Michalodimitrakis [23]. The age range of the material is shown in Table 1.

For consistency, the right auricular surface was observed when possible, whereas data from the left side were only collected when the right one was unobservable. We have chosen the right innominate due to the better preservation of this skeletal element in our sample. Bones showing deformities of a pathological nature that could distort the auricular surface were excluded from the sample. Pictures of each innominate bone analyzed were taken with a Nikon D31000 camera in order to generate a digital image bank. Data collection was performed by two observers and involved both the direct recording of auricular surface morphological traits and the indirect recording through photographs. Observer 1 (E.M.) recorded the auricular traits both directly and indirectly, while Observer 2 (P.N.) recorded them only directly.

The methods described by Buckberry and Chamberlain [18] and Schmitt et al. [19] were applied to all material without prior knowledge of the age at death of the remains. According to the B-C method [18], five features were recorded for each auricular surface: transverse organization (TO), surface texture (ST), microporosity (MI), macroporosity (MA) and apical changes (AP). According to the SC method [19], four features were observed on the auricular surface: transverse organization (SSPIA), granulation and porosity (SSPIB), articular surface modification and apical modification (SSPIC), and iliac tuberosity changes (SSPID).

Correlation coefficients were calculated to examine the relationship between individual auricular surface features and actual age for the SC and B-C methods as well as between the composite score and actual age for the B-C method. For this purpose, Spearman's rho (rs) was used.

The reliability of the B-C method was tested by measures of bias and inaccuracy, following Hens and Belcastro [24]. A similar approach could not be adopted for the SC method because the latter does not produce mean age estimates. The accuracy of each method was also assessed by counting the number of individuals for whom a correct age estimate was attained for each observer, that is, the age at death of the individuals fell within the predicted age interval. The measure traditionally used to test for inter-observer error is Cohen's kappa for nominal data and weighted kappa for ordinal data [25]. This measure was adopted here to compare the results obtained by observers 1 and 2 from the direct recording of auricular morphology using each method. In addition, this measure was used to assess intra-observer error for observer 1 when recording auricular surface morphology directly and indirectly, as well as directly on two separate occasions with a three-month interval between each recording session. An intraclass correlation coefficient (ICC) test was carried out as well in order to examine the intra- and inter-observer error rates [26, 27]. The main limitation of this test compared to weighted kappa is that only absolute agreement between observers is taken into account, thus any difference between scores, no matter how small, is considered a disagreement.

RESULTS

Correlation with age

Correlation coefficients between individual auricular surface features and documented age at death for the SC and B-C methods as well as between the composite score and age at death for the B-C method are illustrated in Table 2. Compared with the correlation coefficients given in the original Buckberry and Chamberlain [18] paper, which ranged from 0.319 for apical changes to 0.533 for macroporosity and were equal to 0.609 for the composite score, ours are smaller for both the individual traits and the composite score. The Schmitt et al. [19] paper does not provide any correlation coefficients, so no comparison between our results and the ones based on which this method was developed can be made. However, it is seen that Spearman's rho is particularly low for SSPIA (-0.080 to 0.099) and low to moderate for the remaining features (0.199 to 0.460).

Method Accuracy

Following Hens and Belcastro [24], the reliability of the B-C method was tested by measures of bias and inaccuracy (Table 3). These measures were estimated both for the entire sample as well as separately for individuals younger than 50 years, between 51 and 70 years and over 71 years in order to explore age-related biases. We observe that the B-C method overall underestimates the age of the individuals in the Greek assemblage as the bias is negative in both observers and for both the direct and indirect scoring methods. This underestimation is observed in individuals over 51 years old and is more pronounced in individuals over 70 years old. In contrast, the age of individuals younger than 50 years old is overestimated. The inaccuracy of the method is greater for individuals over 70 years old while for individuals younger than 50 and between 51 and 70 years of age the relative inaccuracy depends upon the observer and the method used to record the age markers (direct or indirect). The latter finding is very likely due to the small sample size for individuals younger than 50 years in our sample. As mentioned above, a similar approach for testing bias and inaccuracy could not be used for the Schmitt et al. [19] method, since the latter does not provide mean age estimates.

As seen in Table 4, the application of the SC [19] method assigned 25 to 53 out of the 74 individuals to the correct age stage. Observer 2 outperformed Observer 1 by achieving correct age stage classification in 46 or 53 out of 74 cases. For Observer 1, using the method directly or indirectly did not appear to have a particular effect on the results. Note that in the SC [19] original publication, two tables with score-to-age correspondence are presented. The first one is based on "populations whose age distribution corresponds to a life expectancy of 30 years". The second table is based on "a reference population whose age distribution is uniform". In our sample, age estimates based on the former distribution were more accurate than based on the latter in all cases and for both observers. This is a surprising result given the age distribution of our material, which is characterized by a heavy bias toward elderly individuals. However, it must be stressed than correct age estimation using the SC [19] method was mostly achieved in cases where age was generally assessed as being over a certain threshold (e.g. >60) and not within a specific range; therefore, overall the results presented in Table 4 for this method should be treated cautiously.

When applying the B-C [18] method, age was correctly predicted in merely 14 to 17 out of 74 individuals based on the mean ± standard deviation values associated with each composite score of the original publication. Both observers achieved comparable results, with Observer 1 having a slight precedence over Observer 2. Direct vs indirect scoring does not seem to have a particular effect here either. In addition, the correlation between the mean age estimated using this method and the age at death of the individuals was rather small (Spearman's rho ≈ 0.300) and a Wilcoxon test showed that in all cases and for both observers the difference between the estimated ages and the age at death was statistically significant (p-value always < 0.05). To explore further the accuracy of this method, we plotted the mean predicted age against the age at death of the individuals, as seen in Figures 1 and 2.

The regression models are:

Observer 1 - Direct scoring: $y = 0.1272x + 50.381$, with $R^2 = 0.0683$

Observer 2 - Direct scoring: $y = 0.3598x + 26.985$, with $R^2 = 0.1679$

Observer 1 - Indirect scoring: $y = 0.1839x + 44.032$, with $R^2 = 0.1193$

In the ideal case, the slopes (0.3598, 0.1839, and 0.1272) should have been close to 1 and the intercepts (26.985, 44.032, and 50.38) close to zero. It is clear that the B-C [18] method shows a poor performance for age prediction in our dataset.

In Figure 1 it can also be seen that all individuals over 70 years of age (y axis) are under-aged and this issue becomes even more pronounced among older individuals. Subsequently, we removed all individuals over 70 from our sample and calculated the percentage of individuals assigned to the correct age range when adopting ±10 year-intervals around the mean age provided by Buckberry and Chamberlain [18]. This percentage was merely 55 to 60% depending upon the observer. When increasing the interval to ±15 years, this percentage raised to 60 to 80% depending upon the observer. These results suggest that even in the age ranges where the method seems to work (below 70 years), and even when adopting particularly large age intervals (±15 years), the performance of the method is poor for our dataset.

Intra- and Inter-observer error

As can be seen in Table 5, when applying the B-C [18] method, ST is the trait with the lowest inter-observer agreement (17.57%). MI, MA and AP show the highest agreement between observers (50-52.7%). When one-stage differences between observers are taken into account, the agreement for ST raises to 63.52% while that for MI, MA and AP is around 90%, suggesting that the overall agreement between observers is rather high. TO lies in between the remaining traits with regard to inter-observer agreement: 37.84% of the scores were identical between observers and 79.73% of the scores fell within a one-stage difference. Recording traits directly and through photographs produced different results, as seen in the comparisons between scores from photographs and direct observation obtained by Observer 1. The percentage of scores belonging to the same stage in both cases ranges from merely 27.03% for ST to 60.81% for TO. However, it is also seen that, in the majority of cases, the difference between the two scores laid within one stage. Finally, with regard to intra-observer error rates when direct scoring is employed on separate occasions, the percentage agreement between scores ranges from 58.07% for ST to 90.32% for AP.

As shown in Table 6, the SC [19] method produces lower error rates than the B-C [18] one, but this is to be expected since three out of the four traits of this method have only two categories. In general, the agreement between observers was over 60%, in some cases even over 70%, for the binary characters. For SSPIB, the inter-observer error rates are largely comparable to those of the B-C [16] method, with agreement close to 40%, but when within one-stage differences are taken into account, agreement between observers is over 80%. The intra-observer agreement when the auricular surface traits were recorded through photographs and directly is rather high (~80%), suggesting that indirect recording of these traits has a small impact on the results. Even more strikingly, intra-observer agreement when the method is used directly on two separate occasions ranges from 83.87 to 96.77%.

When the differences between observers are statistically tested (Table 7) the only trait that appears to exhibit a non-significant agreement between observers for the B-C [18] method, is ST. All other traits for both methods showed a statistically significant agreement between observers (p-value always < 0.05). It must be noted that the kappa values for the significant results range from 0.279 to 0.455 for Cohen's kappa and 0.461 to 0.716 for ICC, which suggests a fair to substantial agreement between observers, according to the interpretative scores by Landis and Koch [28]. Regarding intra-observer error, the agreement between the scores of the same observer was statistically significant whether direct or indirect recording was used and it ranged from fair to almost perfect [28] (direct vs. indirect scoring: $\kappa = 0.302$ to 0.676, ICC = 0.448 to 0.807, $p<0.05$; direct scoring on two separate occasions: $\kappa = 0.362$ to 0.931, ICC = 0.547 to 0.965, $p<0.05$).

DISCUSSION

Age estimation is unquestionably critical in forensic identification of decomposed human remains. A variety of macroscopic, microscopic and analytical methods are available to the forensic practitioner variable in cost, time, feasibility, expertise, accuracy and reliability. The primary criterion for the selection of the method is dictated by the specifics of the case, yet, in all occasions one seeks a method that combines little effort, speed in the assessment and accurate results. In that aspect, morphological methods tend to be quicker and easier to apply compared to microscopic or analytical methods and involve zero cost, which makes them popular amongst professionals lacking lab resources. Amongst the common morphological methods, authors agree that cranial sutures are not performing well as suture ossification seems to have also a genetic influence [29, 30] while pelvic morphology performs more consistently amongst populations. The current paper tested the applicability of two age estimation methods based on pelvic morphology for the first time in a sample from Crete, Greece.

More specifically, the current study explored the accuracy and reliability of the SC [19] and B-C [18] methods for age estimation based on the morphology of the auricular surface. Regarding the accuracy of these methods, it was shown that both perform poorly in the modern Cretan sample and particularly broad age ranges should be adopted in order to achieve correct age classifications, which renders these approaches of limited use to our material. There are several possible etiologies for this result: 1. an insufficient description of the adopted traits resulting in a general difficulty in the correct identification and recording of these traits, 2. inherent limitations pertaining to the strength of correlation between age at death and the recorded morphological changes, 3. different age distribution of the original and the current sample, or 4. secular change.

If the first etiology were correct, then high inter- and intra-observer error rates would have been obtained in our study. As expected, certain traits were more difficult to record consistently than others. In specific, surface texture, as presented in B-C [18], had the lowest agreement for the majority of intra- and inter-observer tests. On the other hand, microporosity, macroporosity and apical changes showed the highest agreement, while scoring using the SC [19] method provided overall low error rates for all features, even though this is attributed to the fact that three out of the four traits have a binary scoring system. Overall inter- and intra-observer agreement in our study was fair to exceptionally high, suggesting that the description of the recorded morphological changes is sufficiently clear. It is noteworthy that intra-observer agreement when the auricular surface traits were recorded through photographs and directly was rather high (~80%), suggesting that the traits under study are easy enough to identify and record correctly even indirectly; thus high quality photographs may substitute direct recording in cases where a collection is not easily accessible while photographs are available.

Regarding the second etiology, it appears that indeed in our material morphological changes in the auricular surface, as captured by the B-C [18] and SC [19] methods, are not correlated strongly enough with the age at death of the individuals in order to be used as accurate age markers. Our results are in agreement with those of Falys et al. [20], who, even though found a positive association between age and composite scores when applying the B-C [18] method on a British sample, concluded that only very broad developmental stages may be identified by means of this method, all of which exhibited substantial intra-stage individual heterogeneity in age. The results of the Falys et al. paper [20] are in agreement with these of Hens and Belcastro [24], who also concluded that due to the pronounced variation in the age ranges that are associated with each composite score, fewer stages with wider age ranges should be adopted when age estimation from the auricular surface is attempted. Moraitis et al. [21], on the other hand, applied the B-C [18] method on a modern Greek sample, namely the Athens Collection, and found a deviation of 5.5-12.6 years on all age stages. However, this study did not provide information on the actual number of individuals from the Athens Collection for whom age could be correctly estimated using the B-C [18] method, rather the authors reported only average age estimates per composite score; therefore, their results are not directly comparable to those of the present study.

Thirdly, our results regarding the bias and inaccuracy of the B-C method potentially highlight the limitations of applying methods developed based on assemblages with a specific age distribution to samples with a different (skewed) age profile. Our Cretan assemblage exhibited an over-representation of individuals older than 70 years and an under-representation of individuals younger than 50 years. Even though Buckberry and Chamberlain [18] highlight that their method was developed using the Spitalfields Collection where young individuals were under-represented, from the Appendix of their paper where all individuals under examination are given, it is clear that the age distribution of their material is more homogenous than ours.

An additional possible confounding factor for the poor results of our study is secular change. The B-C method was developed based on individuals from the crypt of Christ Church, Spitalfields, dating between 1646 and 1859, while our Cretan material dates between 1963 and 1996. If secular change is indeed the primary factor underlying the poor performance of the B-C method in our sample, then this casts even further doubt as to the applicability of this method to archaeological remains which will exhibit much more pronounced temporal and lifestyle differences to the Spitalfields material than our sample. As the authors of the original publication suggested themselves, their method "needs to be tested and redefined, using large, multiracial, and known-age modern and, if possible, archaeological populations" [18, p. 236]. It is possible that the use of transition analysis and Bayesian statistics would improve the performance of this method on different assemblages but the application of such approaches in a Portuguese sample found that "the accuracy of the revised method only slightly increased with application of Bayesian modeling" (31, p. S35).

More research is required in different materials before the aforementioned conclusions can be generalized. However, the low correlations identified between the different auricular age markers and the actual age of the individuals, both for the B-C and the SC methods, suggests that in our assemblage there is limited potential for these methods to result in improved age estimates and alternative age markers should be preferred.

Conclusions

The results of our study have shown that both the Buckberry and Chamberlain [18] and the Schmitt et al. [19] methods for age estimation based on auricular surface morphology are not accurate in our modern Cretan sample, consisting of 74 adult individuals, of known age at death. In specific, both methods correctly predicted age in exceptionally few cases. The Schmitt et al. [19] method assigned individuals to the correct age range mostly for composite scores that corresponded to particularly broad intervals, which are of limited use in forensic analysis. Similarly, the number of individuals classified in the correct age interval using the Buckberry and Chamberlain [18] method, using the mean and standard deviation values provided by the authors in the original paper, was particularly low. Using this method, age misestimation was most pronounced in individuals over 70 years old, but even among younger groups, the required age intervals in order to achieve correct age estimates were too large to be useful. Thus, we conclude that both the B-C and SC methods present limited value for accurate age estimates in our assemblage and alternative age markers should be preferred for application in forensic cases.

Acknowledgments

EK would like to thank the District Attorney of Heraklion and Mr. C. Kavalos, Director of Cemeteries, Crete, for permitting the creation of the Cretan collection, Dr. D. Nathena for assistance with the curation of the collection; and Dr. A. Papadomanolakis, Head of the Forensic Division, of the Ministry of Justice, Transparency and Human Rights for providing facilities for the study and storage of the collection. Efthymia Nikita's contribution to this paper was supported by a Marie Skłodowska-Curie Individual Fellowship [Programme/Call: H2020 — H2020-MSCA-IF-2015, Proposal: 702991 — HumAn]. Special thanks to the anonymous reviewers for taking the time to read our work and provide valuable comments

References

[1] R.S. Meindl, C.O. Lovejoy, Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures, Am J of Phys Anthropol. 68 (1985) 57-66.
[2] J.M. Suchey, S.T. Brooks, Skeletal age determination based on the male os pubis [manuscript handout], 12th International Congress of Anthropological and Ethnological Sciences. (1988) Zagreb, Yugoslavia.
[3] M.Y. Işcan, S.R. Loth, R.K. Wright. Metamorphosis at the sternal rib end: a new method to estimate age at death in white males, Am J of Phys Anthropol. 65 (1984) 147-56.
[4] C.O. Lovejoy, R.S. Meindl, T.R. Pryzbeck, R.P. Mensforth, Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death, Am J of Phys Anthropol. 68 (1985) 15-28.
[5] S.D. Stout, R.R. Paine, Brief communication: histological age estimation using rib and clavicle. Am J of Phys Anthropol. 87 (1992) 111-115.
[6] C. Crowder, H. Jarred , S.D. Stout, Rib Histomorphometry for Adult Age Estimation, Forensic Microscopy for Skeletal Tissues. 915 (2012) 109-127
[7] L.J. Bada, Amino Acid Racemization Dating of Fossil Bones, Annual Review of Earth and Planetary Sciences. 13 (1985) 241-268.
[8] L. Canoira, M.J. Garcia-Martinez, J.F. Llamas, J.E. Ortiz, D.T. Trinidad, Kinetics of amino acid racemization (epimerization) in the dentine of fossil and modern bear teeth, International Journal of Chemical Kinetics. 35 (1997) 576–591.
[9] P. Zioupos, A. Williams, G. Christodoulou, R. Giles, Determining 'age at death' for forensic purposes using human bone by a laboratory-based biomechanical analytical method, Journal of the Mechanical Behavior of Biomedical Materials. 33 (2014) 109-123.
[10] Bonicelli, X. Bledar, E.F. Kranioti, P. Zioupos. Rib biomechanical properties exhibit diagnostic potential for accurate ageing in forensic investigations. PLoS ONE 12(5): e0176785. https://doi.org/10.1371/journal.pone.0176785
[11] X. Xu, DNA methylation and cognitive aging, Oncotarget. 6 (2015) 13922-13932.
[12] C. Giuliani, M. Sazzini, M.G. Bacalini, C. Pirazzini, E. Marasco, E. Fontanesi, C. Franceschi, D. Luiselli, P. Garagnani, Epigenetic Variability across Human Populations: A Focus on DNA Methylation Profiles of the KRTCAP3, MAD1L1 and BRSK2 Genes, Genome Biology and Evolution. 8 (2016) 2760-2773.
[13] D.L. Osborne, T.L. Simons, S.P. Nawrocki, Reconsidering the Auricular Surface as an Indicator of Age at Death, J Forensic Sci. 49 (2004) 1–7.
[14] K.A. Murray & T. Murray, A test of the auricular surface aging technique, J Forensic Sci. 36 (1991) 1162-1169.
[15] S.R. Saunders, C. Fitzgerald, T. Rogers, C. Dudar, H. McKillop, A Test of Several Methods of Skeletal Age Estimation using a Documented Archaeological Sample, Canadian Society of Forensic Sci J. 25 (1992) 97-118.

[16] M.E. Bedford, K.F. Russel, C.O. Lovejoy, R.S. Meindl, S.W. Simpson, P.L. Stuart-MacAdam, Test of the Multifactorial Aging Method Using Skeletons With Known Ages-at-Death From the Grant Collection, Am J Phys Anthropol. 91 (1993) 287-297.

[17] Y. Igarashi, U. Kagumi, W. Tetsuaki, K. Eisaku, New Method for Estimation of Adult Skeletal Age at Death From the Morphology of the Auricular Surface of the Ilium, Am J Phys Anthropol. 128 (2005) 324–339.

[18] J.L. Buckberry, A.T. Chamberlain, Age Estimation from the Auricular Surface of the Ilium: A Revised Method. Am J Phys Anthropol. 119 (2002) 231–239

[19] A. Schmitt, Une nouvelle méthode pour estimer l'âge au décès des adultes à partir de la surface sacro-pelvienne iliaque. A new method to assess adult age at death from the iliac sacro-pelvic surface, Bulletins et Memoires de la Societe d' Anthropologie de Paris. 17 (2005) 89-101.

[20] C.G. Falys, S. Holger, A.W. Darlene, Auricular Surface Aging: Worse Than Expected? A Test of the Revised Method on a Documented Historic Skeletal Assemblage. Am J Phys Anthropol. 130 (2006) 508–513.

[21] K. Moraitis, E. Zorba, C. Eliopoulos, S.C. Fox, A test of the revised auricular surface aging method on a modern European population. J Forensic Sci. 59 (2014) 188-94.

[22] E.F. Kranioti, M.Y. Işcan, M. Michalodimitrakis, Craniometric analysis of the modern Cretan population, Forensic Sci Int. 180 (2008) 110.e1-5.

[23] E.F. Kranioti, M. Michalodimitrakis, Sexual dimorphism of the humerus in contemporary Cretans - a population-specific study and a review of the literature. J Forensic Sci. 54 (2009) 996-1000.

[24] S.M. Hens, M.G. Belcastro, Auricular surface aging: a blind test of the revised method on historic Italians from Sardinia, Forensic Sci Int. 214 (2012) 209.e1-5.

[25] J. Cohen, Weighed kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin. 70 (1968) 213–220.

[26] M.J. Haber, H.X. Barnhart, Coefficients of agreement between two observers or methods of measurements from quantitative data with replicative measurements. Satist. Methods Med. Res. 17 (2006) 151-169.

[27] J. Cheyenne, M.S. Lewis, M. Heather Garvin, Reliability of the Walker Cranial Nonmetric Method and Implications for Sex Estimation. J Forensic Sci. 61 (2016) 743-751.

[28] J.R. Landis, G.G. Koch, The Measurement of Observer Agreement for Categorical Data. Biometrics. 33 (1977) 159-174.

[29] K. Wolff, E. Hadadi, Z. Vas, A Novel Multidisciplinary Approach Toward a Better Understanding of Cranial Suture Closure: The First Evidence of Genetic Effects in Adulthood. Am J Hum Biol. 25 (2013):835-43.

[30] K. Wolff, Z. Vas, P. Sotonyi, L.G. Magyar, Skeletal age estimation in Hungarian population of known age and sex, Forensic Sci Int. 223 (2012) 374.e1-8.

[31] S.M. Hens, K. Godde, Auricular surface aging: Comparing two methods that assess morphological change in the ilium with Bayesian analyses, J Forensic Sci. 61 (2016) S30-S38.

Figure 1. Chronological versus estimated age plot and the corresponding regression lines for Observer 1 and 2 after direct scoring and Observer 1 after scoring in photographs.

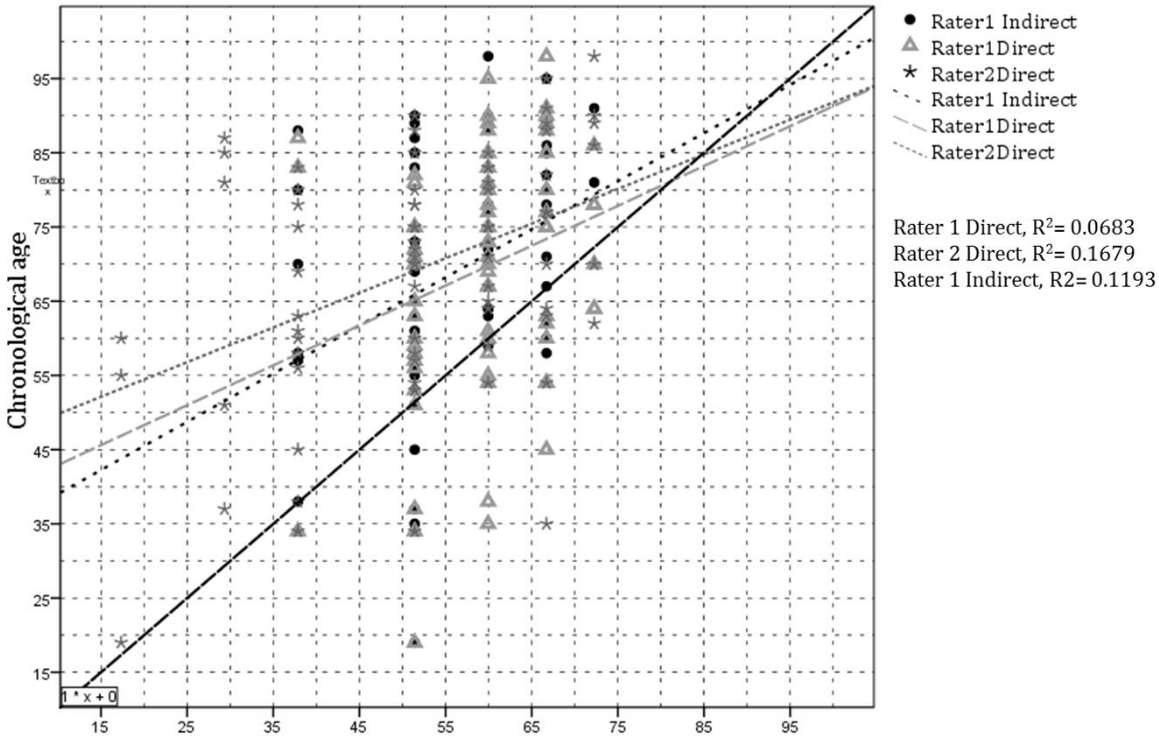

Figure 2. Graph showing chronological versus estimated age for Observer 1 and 2 after direct scoring and Observer 1 after scoring in photographs.

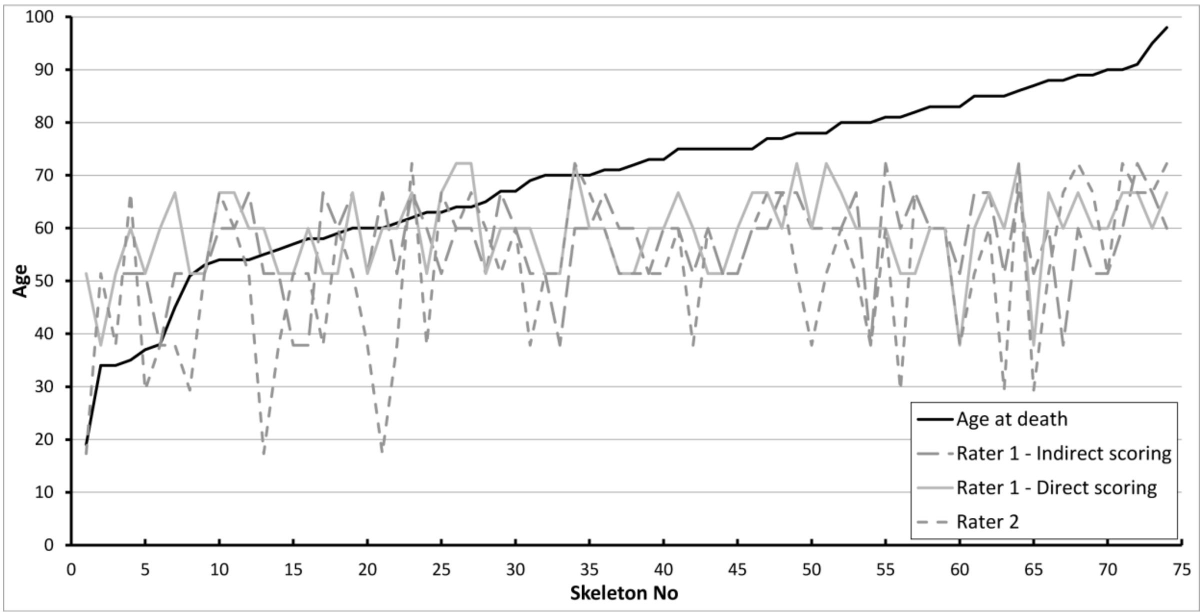

Table 1. Sample age distribution

Age range	**No of individuals**
19-29	1
30-39	5
40-49	1
50-59	11
60-69	12
70-79	20
80-89	18
90-99	5
>100	1

Table 2. Correlation coefficients (r_s) and corresponding p-values between auricular surface features and documented age at death for different raters and scoring methods.

Feature	r_s (p-value)
Rater 1 –Direct scoring	
SSPIA	-0.080 (0.497)
SSPIB	0.363 (0.001)
SSPIC	0.285 (0.014)
SSPID	0.206 (0.078)
TO	0.017 (0.885)
ST	-0.131 (0.264)
MI	0.123 (0.296)
MA	0.194 (0.097)
AP	0.431 (0.000)
Composite Score	0.253 (0.030)
Rater 1 –Indirect scoring	
SSPIA	-0.057 (0.627)
SSPIB	0.336 (0.003)
SSPIC	0.199 (0.089)
SSPID	0.304 (0.008)
TO	0.036 (0.758)
ST	-0.135 (0.253)
MI	0.399 (0.000)
MA	0.232 (0.047)
AP	0.341 (0.003)
Composite Score	0.307 (0.008)
Rater 2	
SSPIA	0.099 (0.399)
SSPIB	0.427 (0.000)
SSPIC	0.460 (0.000)
SSPID	0.421 (0.000)
TO	0.147 (0.210)
ST	0.160 (0.172)
MI	0.259 (0.026)
MA	0.408 (0.000)
AP	0.430 (0.000)
Composite Score	0.359 (0.002)

Key: TO = transverse organization, ST = surface texture, MI = microporosity, MA = macroporosity, AP = apical changes, SSPIA = transverse organization, SSPIB = granulation and porosity, SSPIC = articular surface modification and apical modification, SSPID = iliac tuberosity changes

Table 3. Bias and inaccuracy of estimated age to documented age at death for the B-C method

Actual age	N	Bias	Inaccuracy	Bias	Inaccuracy	Bias	Inaccuracy
		Rater 1 - direct		Rater 1 - indirect		Rater 2	
50>	7	5.19	9.94	13	13	19.53	19.53
51-70	28	-10.73	13.48	-5.24	9.01	-2.10	7.28
71<	39	-26.29	26.29	-22.43	22.43	-21.28	21.28
Overall	74	-17.92	18.96	-13.80	15.23	-12.01	13.97

Table 4. No of individuals assigned to the correct age class using the B-C and SC methods

	B-C	SC	
Rater 1 - direct scoring	16/74	47/74*	25/74**
Rater 1 - indirect scoring	17/74	45/74*	31/74**
Rater 2 - direct scoring	14/74	53/74*	46/74**

* based on a reference population whose age distribution corresponds to a life expectancy of 30 years

** based on a reference population whose age distribution is uniform

Table 5. Inter-and intra-rater comparisons – B-C method

Comparison	TO		ST		MI		MA		AP	
	Same stage (%)	1-stage difference (%)	Same stage (%)	1-stage difference (%)	Same stage (%)	1-stage difference (%)	Same stage (%)	1-stage difference (%)	Same stage (%)	1-stage difference (%)
Rater 1 vs Rater 2 (direct scoring)	28/74 -37.84	31/74 -41.89	13/74 -17.57	34/74 -45.95	39/74 -52.7	28/74 -37.84	37/74 -50	33/74 -44.6	39/74 -52.7	30/74 -40.54
Rater 1 (direct vs indirect scoring)	45/74 -60.81	18/74 -24.32	20/74 -27.03	33/74 -44.6	42/74 -56.76	26/74 -35.14	37/74 -50	31/74 -41.89	34/74 -45.96	33/74 -44.6
Rater 1 (direct scoring on two separate occasions)	27/31 -87.1	Mar-31 -9.68	18/31 -58.07	Dec-31 -38.71	23/31 -74.19	Jul-31 -22.58	26/31 -83.87	Apr-31 -12.9	28/31 -90.32	Feb-31 -6.45

Table 6. Inter-and intra-rater comparisons - SC method

Comparison	SSPIA*	SSPIB		SSPIC*	SSPID*
	Same stage (%)	Same stage (%)	1-stage difference (%)	Same stage (%)	Same stage (%)
Rater 1 vs Rater 2 (direct scoring)	48/74(64.87)	30/74(40.54)	34/74(45.95)	50/74(67.57)	59/74(79.73)
Rater 1 (direct vs indirect scoring)	58/74(78.38)	49/74(66.22)	19/74(25.68)	62/74(83.78)	61/74(82.43)
Rater 1 (direct scoring on two separate occasions)	30/31(96.77)	26/31(83.87)	5/31(16.13)	29/31(93.55)	26/31(83.87)

*SSPIA, C and D only have 2 categories, therefore there is no point in presenting the number of 1-stage differences.

Table 7. (Weighted) Cohen's kappa/p-value and Intraclass correlation coefficients /p-value

	Test	B-C					SC			
Comparison		TO	ST	MI	MA	AP	SSPI A	SSPI B	SSPI C	SSPI D
Rater 1 vs Rater 2 (direct scoring)	Cohen's (weighted) kappa	0.322/0.001	0.004/0.972	0.279/0.012	0.455/0.000	0.421/0.000	0.293/0.010	0.405/0.000	0.352/0.002	0.358/0.002
	ICC	0.557/0.000	0.008/0.486	0.478/0.003	0.716/0.000	0.583/0.000	0.461/0.004	0.626/0.000	0.521/0.001	0.530/0.001
Rater 1 (direct vs indirect scoring)	Cohen's (weighted) kappa	0.493/0.000	0.310/0.006	0.302/0.007	0.326/0.002	0.402/0.000	0.520/0.000	0.642/0.000	0.676/0.000	0.444/0.000
	ICC	0.661/0.000	0.479/0.003	0.448/0.006	0.526/0.001	0.595/0.000	0.691/0.000	0.787/0.000	0.807/0.000	0.617/0.000
Rater 1 (direct scoring on two separate occasions)	Cohen's (weighted) kappa	0.852/0.000	0.807/0.000	0.649/0.000	0.730/0.000	0.823/0.000	0.931/0.000	0.894/0.000	0.871/0.000	0.362/0.029
	ICC	0.929/0.000	0.900/0.000	0.791/0.000	0.853/0.000	0.757/0.000	0.965/0.000	0.944/0.000	0.931/0.000	0.547/0.017