An evaluation of Bayesian age estimation using the auricular surface in modern Greek material

An evaluation of Bayesian age estimation using the auricular surface in modern Greek material

Efthymia Nikita[a], Panagiota Xanthopoulou[b], Elena F. Kranioti[c]

[a]Science and Technology in Archaeology Research Center, The Cyprus Institute, 2121 Aglantzia, Nicosia, Cyprus; e.nikita@cyi.ac.cy

[b]Department of Biology, Section of Animal and Human Physiology, National and Kapodistrian University of Athens, Greece; xanthopop@bio.auth.gr

[c]School of History, Classics & Archaeology, The University of Edinburgh, Edinburgh EH8 9AG, United Kingdom; elena.kranioti@ed.ac.uk

Submitted: 20/5/2018

Accepted:27/07/2018

Abstract

Pelvic morphology is highly reflective of both sex and age changes in humans, making it a popular research focus in forensic anthropology. Relevant studies range from traditional descriptive to more complicated approaches involving statistical modeling, with the latter having become excessively popular in the last decades. The present study examines the performance of Bayesian statistics in age estimation based on the morphological changes observed on the iliac auricular surface. The aim is two-fold: a) to test whether a Bayesian approach can improve age-at-death estimation compared to the original Lovejoy et al. (1985a) and Buckberry and Chamberlain (2002) methods, and b) to explore the impact of adopting different samples as informative priors as well as for obtaining the transition analysis parameters. For this purpose, two modern Greek documented collections have been used, the Athens and the Cretan Collection. Our results found no clear improvement in age prediction when adopting Bayesian age estimation, with only one exception: Athenian males for the Buckberry and Chamberlain (2002) method. The choice of samples for transition analysis and as informative priors affected the results but this effect was statistically non-significant.

Keywords: Forensic Anthropology; age estimation; auricular surface; Bayesian statistics; transition analysis

Introduction

Age-at-death estimation is a key parameter in forensic anthropological analysis as age is one of the three main variables in establishing an individual's biological profile, along with sex and stature (Byers 2017; Dirkmaat 2012). Despite its importance, age-at-death estimation for adults is problematic due to the fact that skeletal age and chronological age become more and more disassociated as the individual grows older due to the cumulative impact of multiple factors on skeletal degeneration, such as pathology, activity and diet (Nawrocki 2010). A large number of methods focused on different anatomical parts of the skeleton have been proposed for age estimation in adult remains (e.g. Brooks and Suchey 1990; Meindl and Lovejoy 1985). Validation studies of these methods in different assemblages have produced mixed and often discouraging results, stressing the high level of inter-population variation in the ageing process and the need for different approaches (Falys et al. 2006; Hershkovitz et al. 1997; Michopoulou et al. 2017; Saunders et al. 1992; Xanthopoulou et al. 2018).

Additionally to the limitations of each ageing method, an important general problem highlighted by Bocquet-Appel and Masset (1982) is 'age mimicry', that is, the fact that the ages estimated in a target sample will be biased by the demographic profile of the reference sample used in order to develop each skeletal ageing method. In particular, traditional ageing methods associate specific morphological skeletal changes with a mean age in the reference population. However, this mean age largely depends on the age structure of the reference population; therefore, the age distribution of the reference sample will largely affect the age distribution of the target sample in which the method is applied (Bocquet-Appel and Masset 1982, 1995; Konigsberg and Frankenberg 1992).

Another serious limitation of traditional ageing methods is that few of them allow the combination of skeletal ageing markers from different anatomical structures. This is a serious issue since, as highlighted by Milner and Boldsen (2012), looking simultaneously at anatomical structures across the skeleton can improve age estimates because each structure will be informative about a single transition at some point in adulthood. Among the proposed approaches for combining traditional age indicators are the Complex and Multifactorial Methods (Acsádi and Nemeskéri 1970; Ferembach et al. 1980; Lovejoy et al. 1985b); however, these do not appear to provide better results than the individual age indicators they are based on (Milner and Boldsen 2012).

Bayesian age estimation, that is, transition analysis (TA) coupled with Bayesian statistics, has been proposed as a potential solution to the above problems. TA estimates the age of transition between successive stages of an age marker and, when combined with Bayesian analysis, the age intervals estimated for each stage of transition control for biased age-at-death estimates (Boldsen et al. 2002; Konigsberg et al. 2008). The past years have witnessed an increase in the number of papers using Bayesian age estimation, while the popularity of this approach is further attested by the imminent integration of the ADBOU software for age estimation using Bayesian statistics (Boldsen et al. 2002) into Fordisc, a computer program that employs statistical methods for sex, ancestry, and stature estimation, widely used by forensic anthropologists. Among the studies examining the performance of Bayesian age estimation, some found it to outperform traditional approaches by providing more realistic mortality profiles (Bullock et al. 2013) or higher accuracy of the generated age ranges (Godde and Hens 2012; Hens and Godde 2016). In addition, a number of scholars have proposed new ageing methods or adaptations of existing ones using Bayesian analysis. Methods for age estimation from the clavicle (Langley-Shirley and Jantz 2010), the first rib (DiGangi et al. 2009), dental wear (Prince et al. 2008), the development of the third molars (Tangmose et al. 2015), or a revised version of the Suchey-Brooks method (Berg 2008) are some examples of this trend. Even though most studies acknowledge the importance of Bayesian age estimation and many note its superior performance compared to traditional methods for age-at-death estimation, serious concerns have also been expressed with regard to the suitability of the reference populations used in the estimation of the TA parameters (e.g. Schmitt et al. 2002), the effectiveness of the anatomical structures used in capturing age-at-death changes (Milner and Boldsen 2012) and the overall potential of this approach to provide accurate age estimates (Lottering et al. 2013).

The aim of the present paper is to examine the extent to which Bayesian age estimation can improve age estimation based on the auricular surface of the ilium in two different modern Greek assemblages, one from Athens and the other from Crete. Previous validation studies of traditional ageing methods on these two assemblages have produced contradictory results. Xanthopoulou et al. (2018) found the auricular surface of the ilium to produce the most accurate age estimates in the Athens Collection compared to the pubic symphysis and cranial sutures. In particular, the authors noted that the Lovejoy et al. (1985a) method gives the highest rates of correct age classifications when adopting expanded age ranges (78% to 83%),

while the Buckberry and Chamberlain (2002) method gave the most accurate results compared to all methods in the older age groups (56% in the total sample, 80% to 82% in individuals over 51 years). Similarly, Moraitis et al. (2014) found the Buckberry and Chamberlain (2002) method to work well in the Athens Collection and in their study the good performance of this method was identified also among younger adults. In contrast, Michopoulou et al. (2017) found the Buckberry and Chamberlain (2002) method to perform very poorly in the Cretan Collection (19% to 23% correct age classifications). We should stress that in the Xanthopoulou et al. (2018) study, the use of Bayesian age estimation by means of the ADBOU software produced even lower correct age classifications (28% to 44% for the auricular area of the ilium) compared to the original Lovejoy et al. (1985a) and Buckberry and Chamberlain (2002) methods.

In order to examine whether the performance of the auricular surface of the ilium as an age marker can improve further in these two Greek collections, most notably in the Cretan one, we applied Bayesian analysis on the Lovejoy et al. (1985a) and Buckberry and Chamberlain (2002) methods and experimented with different informative priors and populations for the derivation of transition analysis parameters. In this context, we provide a complete package of R functions with detailed instructions for their use so that even a novice scholar can easily apply this technique in a customized manner, using his/her own TA samples and informative priors. This code along with detailed instructions for its use are given as supplementary material (R_Instructions_for_Bayesian_age_estimation and R_BayesianAgeEstimationFunctions).

Bayesian age estimation

The Bayesian approach for age estimation computes the conditional probability Pr(a|c) that the skeletal remains are from an individual who died at age = a given that the observed age marker is c via *Bayes' Theorem* (Boldsen et al. 2002; Konigsberg et al. 2008):

$$\Pr(a|\mathrm{c}) = \frac{\Pr(\mathrm{c}|a)\mathrm{f}(a)}{\int_{shift}^{\infty} \Pr(\mathrm{c}|\mathrm{x})\mathrm{f}(\mathrm{x})\mathrm{dx}} \quad (1)$$

where Pr(c|*a*) is the probability of observing an age marker c given that the individual died at age = *a*, f(*a*) is the age-at-death distribution of a reference population called *informative prior*, and *shift* is the minimum age in the informative prior. It is seen that the application of a Bayesian approach for age estimation presumes the estimation of the age-at-death distribution function f(*a*) of the informative prior as well as the conditional probabilities Pr(c|*a*). The function f(*a*) is estimated using *survival analysis*, whereas the conditional probabilities Pr(c|*a*) may be obtained using *transition analysis*. Then Equation (1) is adopted to produce probability density functions that can be used to estimate the highest posterior densities (HPD) and highest posterior density ranges (HPDR), which approximate confidence ranges.

According to the conventions outlined in the "Rostock Manifesto" (Boldsen et al. 2002), f(*a*) must be estimated before Pr(c|*a*) can be assessed and, ideally, f(*a*) is the probability distribution of lifespans in the target population. Since it is not always possible to estimate f(*a*) in the target population, we select an informative prior that reflects a similar age distribution to the target sample. f(*a*) may be estimated from the function (Konigsberg et al. 2008; Prince et al. 2008; Wood 2011):

$$f(x) = h(x)S(x) \quad (2)$$

where h(x) is the *Gompertz model*:

$$h(x) = a_3 e^{b_3 x} \quad (3)$$

and S(x) the *survival function*:

$$S(x) = \exp[\frac{a_3}{b_3}(1 - e^{b_3 x})] \quad (4)$$

i.e. the function that describes the *survivorship rate* in the informative prior. Here, a_3 is known as the *baseline mortality*, whereas b_3 is the *senescent component*. These parameters may be estimated as follows: From the sample of the ages of the informative prior we compute the survivorship rate, for example using the *survival* library of R, and fit the model of Equation (4) to these data. This option involves a non-linear fitting. Alternatively, we may maximize the sum of all log(f(x)) values computed over all ages (Konigsberg et al. 2008).

Note that in all these computations as well as when f(*a*) is applied in Equation (1), the minimum age in the informative prior (*shift*) is subtracted from the x values.

As already pointed out above, the conditional probabilities Pr(c|*a*) necessary for the application of Equation (1) are obtained using TA. TA is a statistical method for estimating the age (age-at-transition) at which an individual moves from one age stage to the next higher stage. To estimate these ages-at-transition, we may use *probit regression models for ordinal responses* to fit the ordinal variable c that codes the age stages (dependent variable) to the log of the real ages (Konigsberg et al. 2008 and http://faculty.las.illinois.edu/lylek/TransAna/TransAna.htm). The relationships that express these models are:

$$\text{probit}(P_i) = a_i + b*\log(\text{age}) \quad (5)$$

or

$$\text{probit}(P_i) = a_i + b_i*\log(\text{age}). \quad (6)$$

depending on whether the fitting models exhibit a common slope, b, or different slopes, b_i. In the above equations i = 1, 2, … (J-1), where J is the number of factors of the ordinal variable. It is seen that the fitting procedure yields (J-1) intercepts, a_1, a_2, …, $a_{(J-1)}$ and, depending on the researcher choice, either (J-1) slopes, b_1, b_2, …, $b_{(J-1)}$, or one common slope, b. At the age-at-transition from stage i to (i+1) on average half of the observations will fall at stage i and the other half at stage i + 1. Therefore, at the mean age-at-transition, μ_i, between stages i and (i+1), $P_i = 0.5$, $\text{probit}(P_i) = \text{probit}(0.5) = 0$ and from Equation (5), we obtain $0 = a_i + b*\log(\mu_i)$, which yields $\log(\mu_i) = |a_i/b|$. Similarly, from Equation (6), we obtain $\log(\mu_i) = |a_i/b_i|$. In addition, from the properties of the probit function, it arises that when $\log(\text{age}) = \log(\mu_i) + \sigma$, then $P_i = 0.841$. Therefore, probit(0.841) = 1 and eventually $\sigma = 1/b$. For σ_i, we similarly have $\sigma_i = 1/b_i$.

It is seen that the ratio (absolute value) $|a_i/b_i|$ or $|a_i/b|$ gives the mean age-at-transition, μ_i, between stages i and (i+1), and $1/b_i$ or $1/b$ is the standard deviation, σ_i or σ, of the distribution curve of the age-at-transition. Thus, based on the above equations, TA estimates the log of

ages-at-transition and their common (or not) standard deviation, σ. These quantities may be used to compute the conditional probability Pr(c|*a*) in Equation (1). According to Konigsberg et al. (2008), the conditional probability Pr(c|*a*) can be estimated from:

$$\Pr(c=i|a) = \Phi(\log(age), \mu_{i-1}, \sigma) - \Phi(\log(age), \mu_i, \sigma), \quad i = 2, 3, \ldots, J-1 \tag{7}$$

where Φ is the cumulative distribution function of the normal distribution, age is measured in years, and μ_i is the age-at-transition from phase c = i to c = i+1. When c = 1 or c = J, Pr(c|*a*) can be estimated from the following expressions:

$$\Pr(c=1|a) = 1 - \Phi(\log(age), \mu_1, \sigma) \text{ and } \Pr(c=J|a) = \Phi(\log(age), \mu_{J-1}, \sigma) \tag{8}$$

The above analysis concerns just one trait. If the ageing method involves m traits, Boldsen et al. (2002) proposed the following expression:

$$Pr(a|c) = \prod_{k=1}^{m} Pr(a|c_k) \tag{9}$$

where Pr(*a*|c) is the conditional probability that the skeletal remains are from an individual who died at age = *a* given that the observed age marker c is in fact c_1 for trait 1, c_2 for trait 2 … c_m for trait m. Thus, we multiply all the distributions of the individual traits and divide by the integral of this product to create a proper probability density function.

The Bayesian age estimation described above is a standard approach developed in order to improve age prediction (Boldsen et al. 2002; Konigsberg and Frankenberg 1992; Konigsberg et al. 2008) and constitutes a basic tool for age-at-death estimation for the last two decades. However, as any statistical technique, the accuracy of its predictions depends upon the fulfilment of the various assumptions on which it is based. The first source of problems comes from the Bayes theorem, Equation (1), and refers to the selection of the informative prior. In particular, the informative prior and the target sample should reflect similar age distributions (Hoppa and Vaupel 2002). However, the age distribution of the target sample is not known in bioarchaeological assemblages, and this is a serious issue in palaeodemography as the population age structure is precisely what one wants to discover when looking at

skeletons. This issue may be addressed by using a uniform, or uninformative, prior distribution. This is a conservative approach in the sense that a specific population's age-at-death pattern is not imposed onto the ages being estimated, but it may not yield particularly good results (Milner and Boldsen 2012). In modern assemblages the age distribution of the target sample may be known but there is a problem when the demographic profile of the target sample is markedly different from that of the parent population.

Another source of problems originates in the assumptions of the transition analysis. The validity of the obtained results depends on whether the probit regression models with or without a constant slope, Equations (5) and (6), do describe the c vs. log(age) data of the training sample. This requirement is satisfied and a good fitting is achieved if the training sample is relatively large so that all stages of the skeletal ageing marker under study are sufficiently represented. A second assumption is that the distribution of log(age) in the training sample follows the normal distribution, Equations (7) and (8).

If the above preconditions are fulfilled, Bayesian age estimation is expected to give better predictions than traditional ageing methods (e.g. Brooks and Suchey 1990; Lovejoy et al. 1985a; Meindl and Lovejoy 1985) since it should reduce the age mimicry effect, it can make predictions based on multiple traits, and can provide confidence intervals for each age estimate. The main limitation of the Bayesian approach is that it is rather sensitive to the violation of its basic assumptions, as stated above, and this is likely the reason why in certain cases this method has given poor results (see Introduction). In addition, it is more laborious than traditional methods and requires the use of specialized software.

Materials and Methods

Materials

Two modern Greek documented skeletal collections were studied. The first is housed at the Department of Animal and Human Physiology at the National and Kapodistrian University in Athens. The sample examined consisted of 59 females and 81 males and represents individuals who lived mainly in the second half of the twentieth century (Eliopoulos 2007; Xanthopoulou et al. 2018). The second collection is housed at the facilities of the Forensic

Pathology Division of the Hellenic Ministry of Justice and Human Rights in Crete. It consists of 54 females and 52 males, exhumed from St Konstantinos and Pateles cemeteries in Heraklion, Crete, who died between 1963 and 1997 (Kranioti et al. 2008). Table 1 depicts the age profile of the two Greek collections used in this study.

For the derivation of the transition analysis parameters, we additionally used the data provided in Hens and Godde (2016) for modern American males (n = 372) curated in the William M. Bass Donated Skeletal Collection. This collection includes a variety of donors from United States and the world and represents a sample largely ethnically distinct from the Greek collections under study. Finally, among the informative priors for males, we used the Balkan sample (n = 212) from Konigsberg et al. (2008) in order to draw the demographic parameters of our models from a geographically proximal group to Athens and Crete. The only information provided in Konigsberg et al. (2008) about this sample is that it consists of Balkan genocide victims but no further indication as to the exact origin of the individuals is given. In addition, Konigsberg et al. (2008) stress that the ages-at-death of this sample are those reported by relatives, therefore, they may not always be exact.

Table 1 Profile of the Greek material under study

Age group	**Athenian Male**	**Athenian Female**	**Cretan Male**	**Cretan Female**
18-30	14	4	0	3
31-40	8	6	1	3
41-50	14	11	1	1
51-60	16	9	14	6
61-70	14	7	11	15
71-80	8	12	15	10
81-90	7	10	10	16
Total	***81***	***59***	***52***	***54***

Methods

The effectiveness of Bayesian statistics in age estimation was explored in the present paper using the Lovejoy et al. (1985a) and the Buckberry and Chamberlain (2002) methods. Both methods focus on the morphology of the auricular surface of the ilium and they were used in this paper both in their original form as well as using Bayesian analysis. For the remainder of this paper, the Lovejoy et al. (1985a) method in its original and Bayesian form will be denoted as L and the Buckberry and Chamberlain (2002) method as BC. For the Buckberry and Chamberlain (2002) method, when Bayesian analysis is applied on the composite score, the method will be denoted as BCc.

To test the performance of Bayesian statistics, a combination of target samples, TA samples and informative priors was adopted. Initially we examined the case where the sample from Athens was used as target, TA and informative prior in order to test how Bayesian age estimation performs in an 'ideal' case. Subsequently, we used the Athens Collection as the target sample and informative prior, and used the TA parameters from the American assemblage in order to explore how much age predictions change when the analysis parameters are obtained from a population ethnically distinct from the target one. Finally, we used the Cretan sample as target, the Cretan, Athenian and American samples to obtain the TA parameters and the Athenian, Cretan and Balkan samples as informative priors successively.

Method performance

We used three approaches in order to test how different models perform in age-at-death estimation:

1. We counted the number of cases where the documented age-at-death fell within different intervals (±5, ±10, ±20 years) around the predicted mean age-at-death. For L the midpoint of each interval of the original method was used as the mean estimated age, while for the last phase, the estimated age was arbitrarily set at 70 years since Lovejoy et al. (1985a) only provide a broad age estimate over 60 years for individuals belonging to this stage. Subsequently, the percentage of successful predictions of the Bayesian methods was compared to that of the original methods using the same intervals around the estimated mean age-at-death. Fisher's test was used to identify statistically significant differences between methods. We must note that other authors (e.g. Hens and Godde 2016) performed this

comparison using different confidence intervals for the Bayesian analysis. For example, for L, the original version of which uses 5-year intervals, Hens and Godde (2016) used as a comparison the 50% Confidence Interval (CI) for the Bayesian version. This approach has, however, the serious limitation that often the intervals being compared are very different in range as the 50% CI for the Bayesian method results in intervals ranging from 5 to 18 years (Table 9 in Hens and Godde 2016). 95% CIs for the difference in the percentage successes between original and Bayesian methods were also estimated and used as a criterion.

2. We have estimated the Pearson correlation coefficients between the mean ages estimated using different methods (original and Bayesian). Statistically significant differences between the correlation coefficients were also explored.

3. We plotted the documented age-at-death for all individuals in the target samples as well as the age-at-death predicted using the original methods and their Bayesian implementation after all cases had been organized in increasing documented age-at-death.

The first two approaches should be used in conjunction since each one independently may result in misleading conclusions. It is possible to get a high correlation between methods but poor predictive accuracy, for example if one method produces systematically higher or lower values than the other. Nonetheless, a small correlation coefficient indeed suggests that a method has a poor performance. In addition, a high predictive accuracy may also be misleading because, for example, age predictions that fluctuate around the middle of the sample age distribution may show satisfactory classification rates but small correlation with the documented age-at-death of the individuals, indicating a poor method performance. As for the correlations, a small percentage of success shows again that a method is problematic.

Software

For the implementation of the Bayesian age estimation, a complete package of R functions has been written and is given as Supplementary material along with step-by-step instructions for its use. The functions can estimate the Gompertz model parameters, calculate the survivorship rate and plot Kaplan-Meier survivorship curves, run transition analysis (TA), estimate highest posterior densities (HPD) and highest posterior density ranges (HPDR) per individual or per sample and calculate the frequencies of correct age estimations within a certain age range.

In addition, the age-at-death distribution of different samples was compared by means of Kolmogorov-Smirnov tests using the ks.test. Differences between the correlation coefficients were explored using the R.test. The significance of the difference in the success rate of different ageing methods was tested using the prop.test. All tests were run in R.

Results

Demographic profiles

As pointed out above, Bayesian age estimation is sensitive to the selection of the informative prior as the demographic profile of this prior provides important information for the age prediction of the target sample. Therefore, optimal results are expected when the target sample is also the informative prior or when target and informative prior reflect similar age distributions. This similarity may be examined by means of Kolmogorov–Smirnov tests.

In our study, Kolmogorov–Smirnov tests showed that the Cretan, Athenian and Balkan samples have different age-at-death distributions (Athenian females vs. Cretan females p-value = 0.033; Athenian males vs. Cretan males p-value < 0.001; Athenian males vs. Balkan males p-value = 0.002; Cretan males vs. Balkan males p-value < 0.001). When looking at the Gompertz plots (Figure 1), we observe that the age-at-death distribution of Cretan males is indeed very different from the Athenian sample and even more different from the Balkan one. The distribution of the Athenian and Balkan male samples is also different but not as pronouncedly. Differences in the age-at-death distribution of the Athenian and Cretan females are also noted. Due to these dissimilarities in the demographic profiles of the samples under study, we examined below, among others, the case where the target sample is also the informative prior.

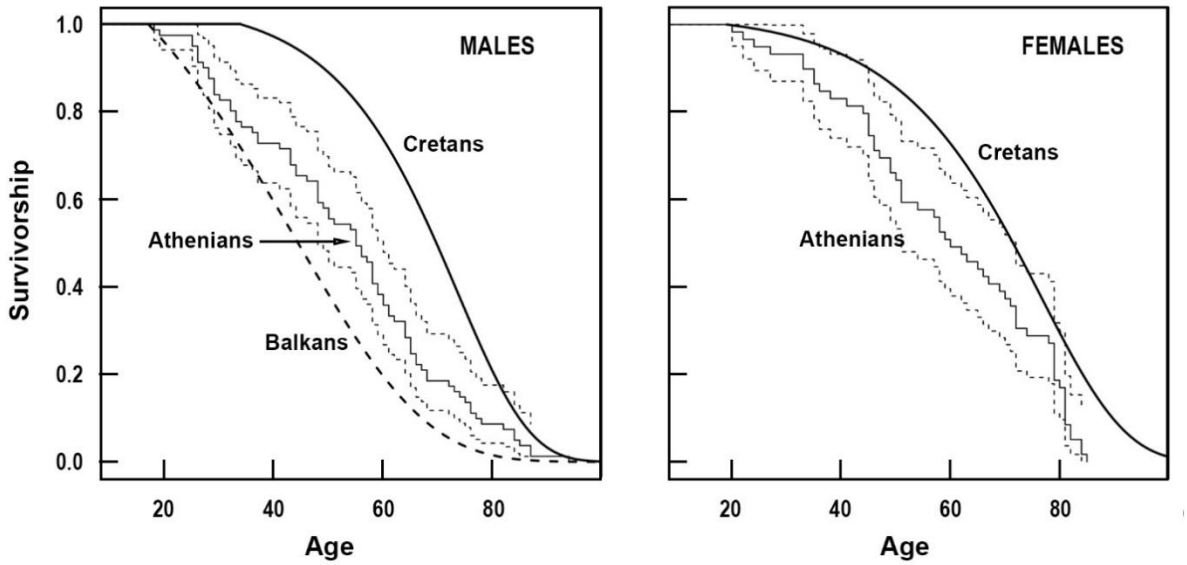

Figure 1. Kaplan-Meier survivorship curves (dotted lines) derived from Athenian males and females with superimposed Gompertz survivorship curves for Cretan males and females (solid lines) and Balkan males (broken line).

At this point we should stress that the Athenian and Cretan assemblages are temporally largely contemporary and geographically very proximal. Therefore, there is no reason for the populations from which they have been drawn to reflect a significantly different survivorship profile. Instead, the observed significant difference is likely due to the way these two modern reference collections have been put together through donations. In what concerns the Balkan sample, as stated in the Materials section, its documentation may be more problematic, but again this assemblage is geographically and temporally proximal to the two Greek assemblages used in this study.

Initial assessment

Before examining the relative performance of Bayesian analysis, we tested how the original L and BC methods work in order to determine how to proceed. In Table 2 it is seen that L works better than BC, and both methods give better results in females than males with few exceptions. In addition, both methods work notably poorly for Cretan males, as is particularly clear from the low correlation coefficient between mean estimated age-at-death and documented age-at-death. These results raise the question of whether the performance of these methods can be improved by adopting Bayesian approaches, especially with regard to

the performance of the BC method in the Athenian sample and both methods in the Cretan material.

Table 2 Percentage of correctly classified individuals in different year intervals and Pearson's correlation coefficients between estimated and documented age-at-death

Original	Athenian Female	Athenian Male	Cretan Female	Cretan Male
Percentage of correctly classified individuals				
L5	44.07	35.80	33.33	11.54
L10	69.49	67.90	57.41	34.62
L20	93.22	87.65	90.74	69.23
BC5	16.95	13.58	18.52	25.00
BC10	50.85	39.51	51.85	50.00
BC20	76.27	71.60	85.19	88.46
Pearson's correlation coefficients				
L	0.798	0.760	0.700	0.451
BC	0.620	0.740	0.641	0.327

Key: L=Lovejoy et al. (1985a), BC=Buckberry and Chamberlain (2002), numbers 5, 10, 20 indicate age ranges used to evaluate the performance of each approach.

Athenians as target sample, TA sample, and informative prior

Figure 2 shows the differences in the percentage performance of the Bayesian approaches compared to the original L and BC methods. Positive values in this figure suggest that the Bayesian approach improves the original methods and negative values show the opposite. When the zero value is outside the 95% CI, the difference between the performance of these methods is statistically significant. It can be seen that in females no statistically significant difference is found between the performance of the original versus Bayesian methods since the zero value is always inside the 95% CI. In contrast, in males the Bayesian version of BC gives statistically significantly better results than the original method in all year intervals and

also irrespective of whether the Bayesian analysis is applied on the composite score or on each individual age marker (the zero value is never inside the 95% CI).

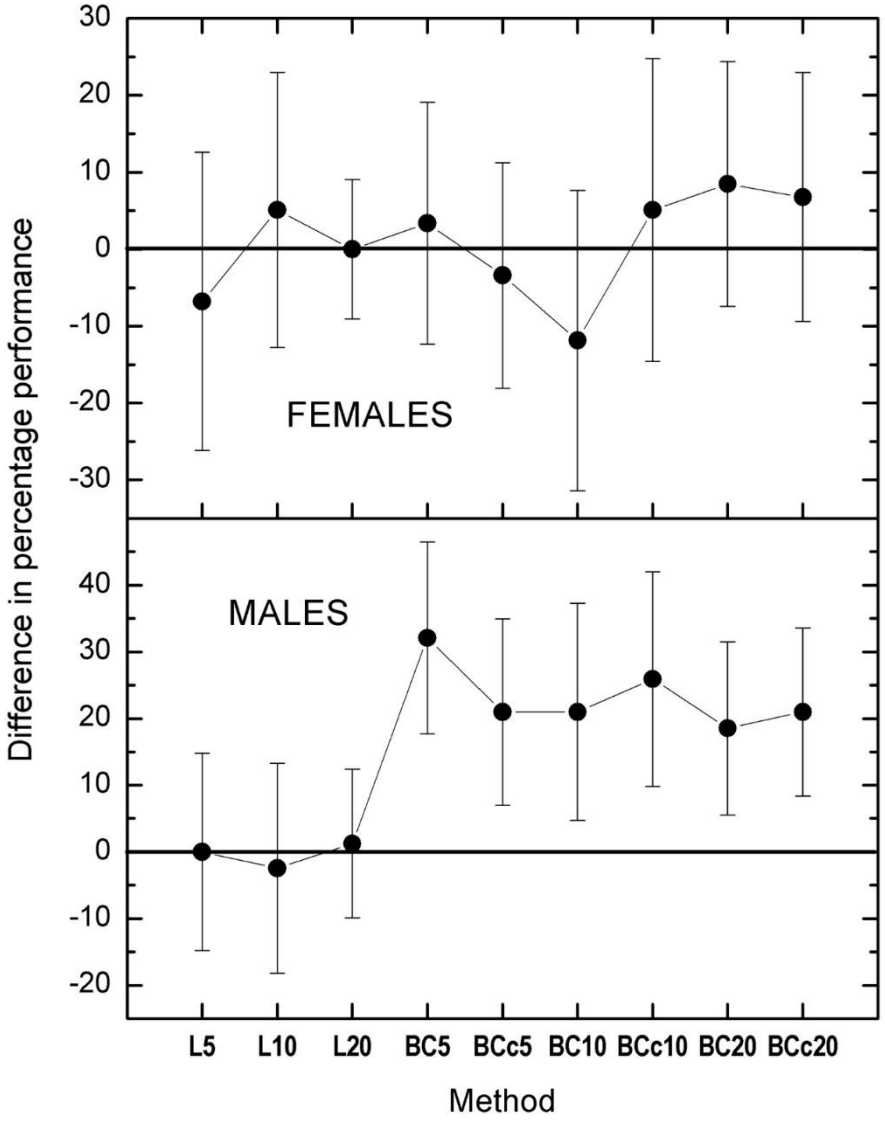

Figure 2. Differences in the percentage performance of the Bayesian approaches studied with respect to the original L and BC methods for the Athens sample using the same sample for TA and informative prior. Numbers 5, 10, 20 indicate age ranges used to evaluate the performance of each approach.

Figure 3 visualizes the documented and estimated ages-at-death using the original and Bayesian versions of L. In females it can be seen that with the exception of a few outliers, there is satisfactory agreement between the estimated and the documented age-at-death, while the implementation of Bayesian analysis has a minimal impact on the results. This has as a consequence a rather high Pearson's correlation coefficient (r = 0.8). Among males, the

picture is similar, though the Bayesian method is slightly better than the original one, as attested by the Pearson correlation coefficient of 0.76 for the original method and 0.79 for the Bayesian one.

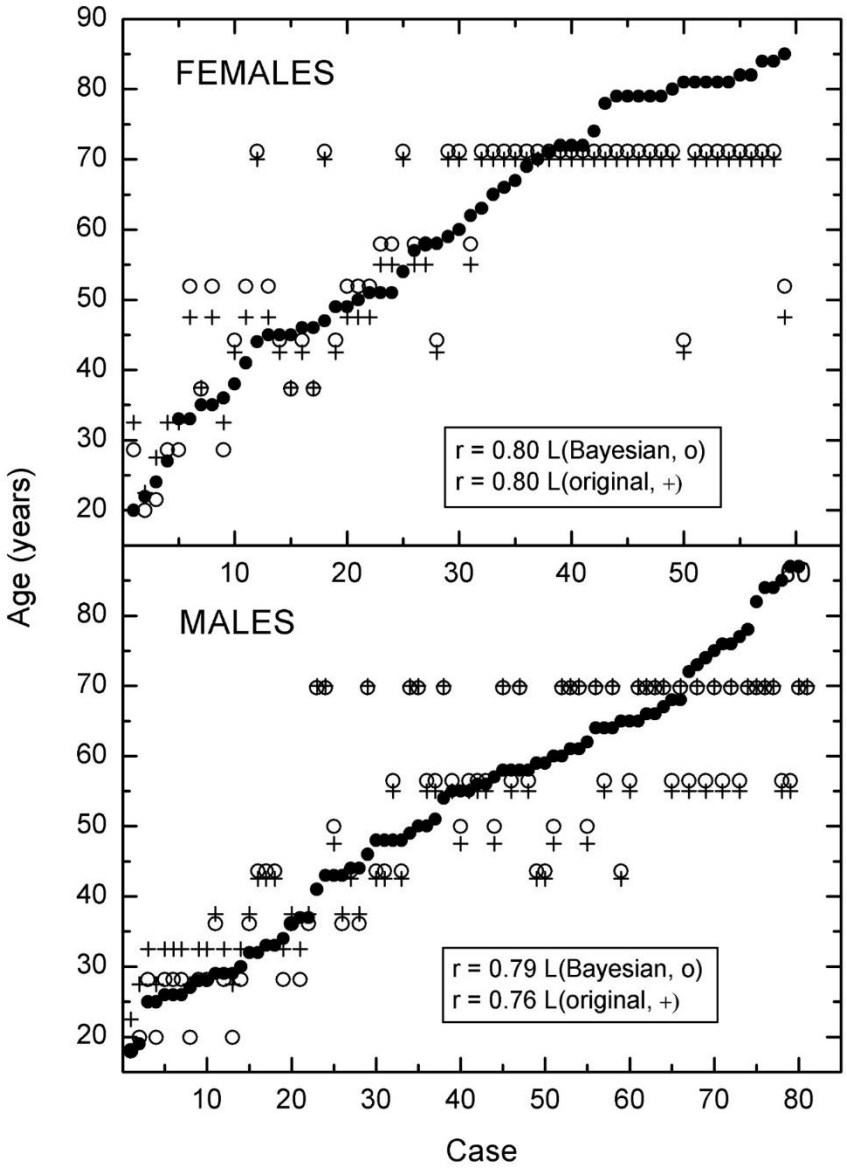

Figure 3. Documented ages of males and females for the Athens sample and their estimation by means of L original (+) and L Bayesian (o) using the same sample for TA and informative prior. r values are Pearson's correlation coefficients of the estimated vs. documented ages.

For BC (Figure 4), there is a notable overestimation of age in females younger than 50. Bayesian analysis tends to improve age prediction in younger adults but the original method is better for older adults. As a result, the Pearson correlation coefficient is almost the same for the Bayesian and traditional version of the method (r = 0.61-0.62) and smaller than the

corresponding one for L. In males, the original method again overestimates age in younger adults but now the use of Bayesian statistics clearly improves the results. The Pearson correlation coefficient is 0.74 for the original method and raises to 0.80 when using Bayesian analysis. Note that, as highlighted above (Figure 2), the improvement in the percentage performance of BC for males is statistically significant; however, the increase of the Pearson correlation coefficient from 0.74 to 0.80 is not significant. Finally, Figure 5 visualizes the comparative performance of the Bayesian implementation of BC when using the individual scores of each marker or the composite score. Despite the differences observed, both methods provide similar age estimates.

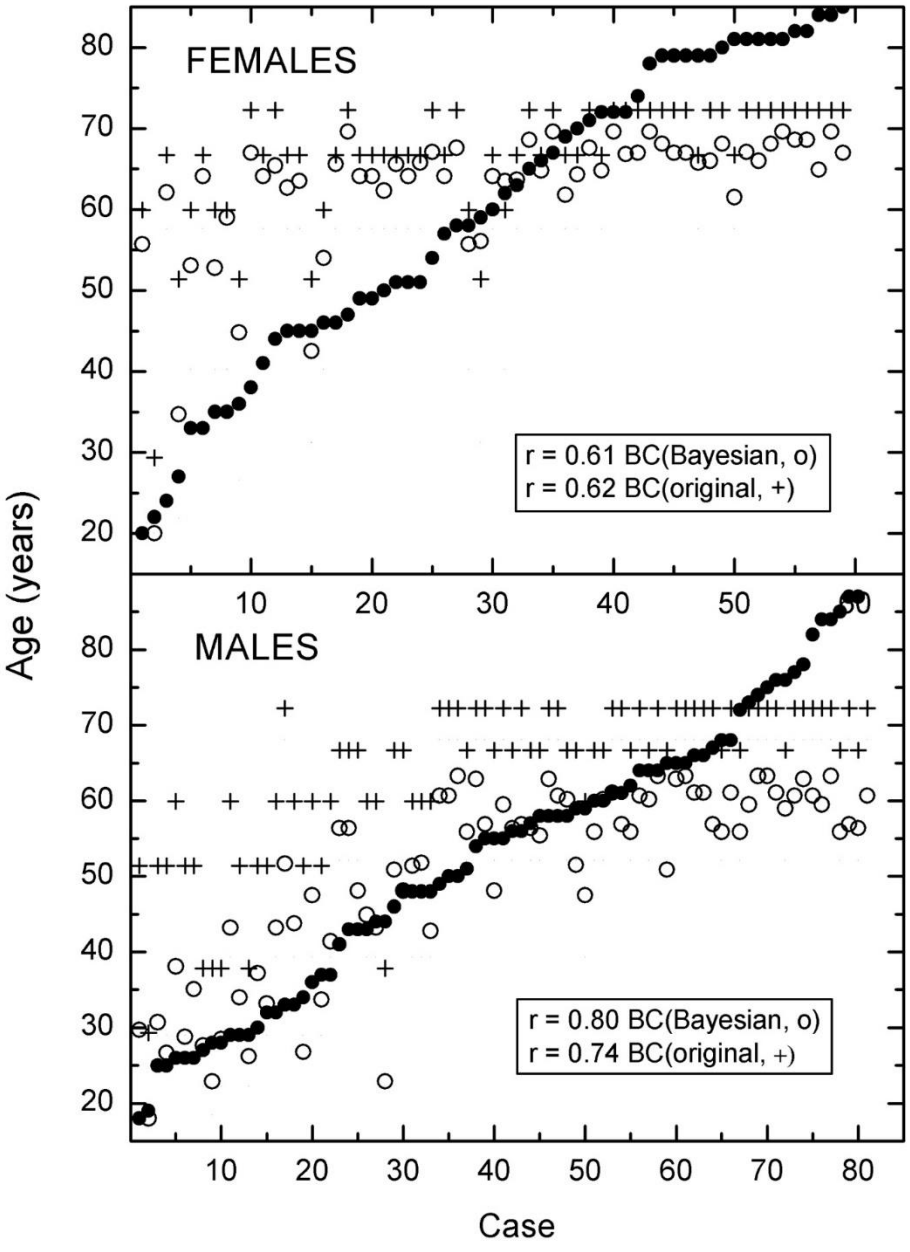

Figure 4. Documented ages of males and females for the Athens sample and their estimation by means of BC original (+) and BC Bayesian (o) using the same sample for TA and informative prior. r values are Pearson's correlation coefficients of the estimated vs. documented ages.

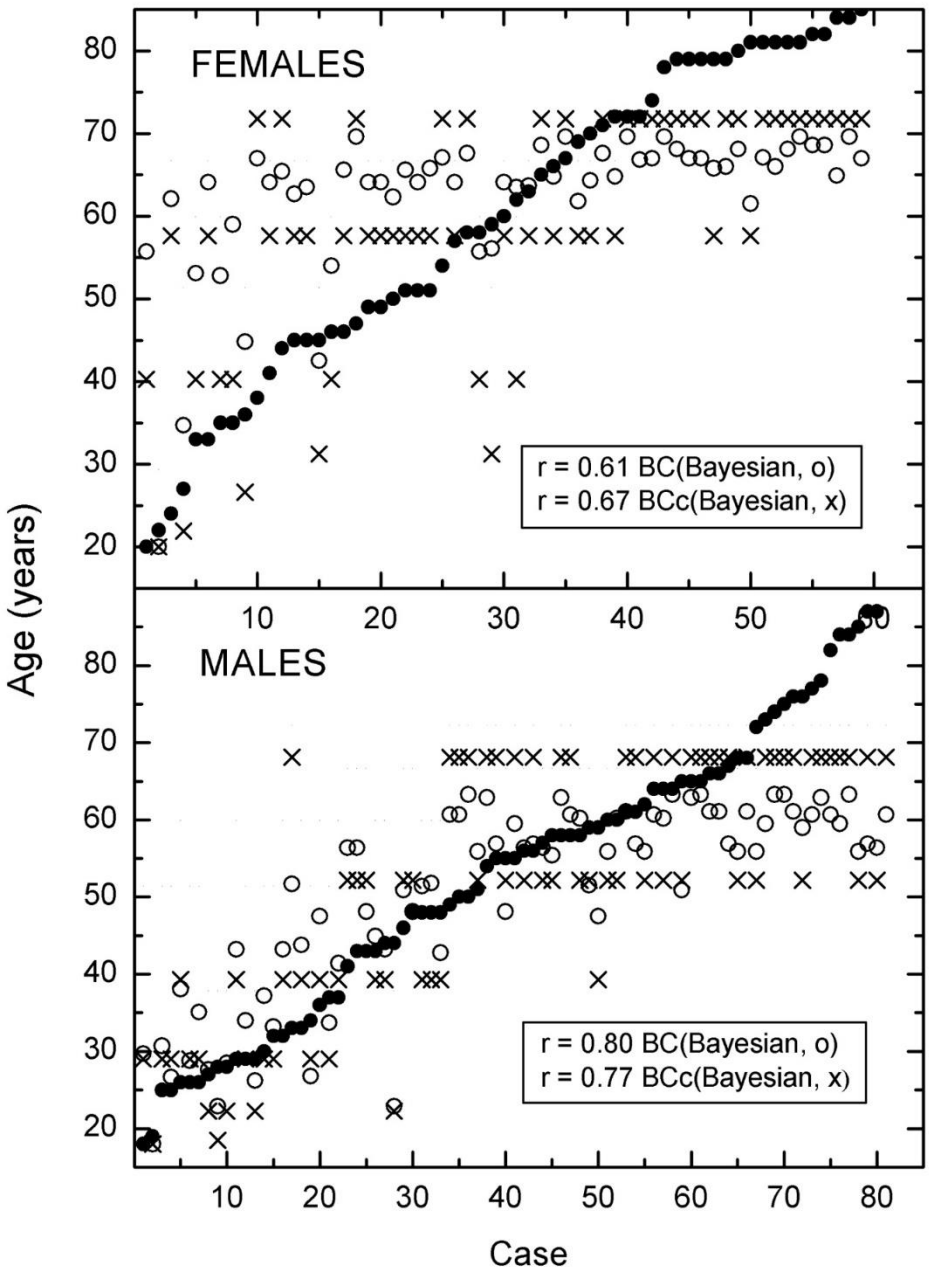

Figure 5. Documented ages of males and females for the Athens sample and their estimation by means of the Bayesian methods BC (o) and BCc (×)using the same sample for TA and informative prior. r values are Pearson's correlation coefficients of the estimated vs. documented ages.

Athenians as target sample and informative prior, Americans as TA sample

When the TA parameters for males are obtained from the American assemblage (Hens and Godde 2016), we observe that the success rate in age-at-death prediction generally decreases but this decrease is overall rather small and statistically non-significant (Figure 6). The correlation between predicted and documented age-at-death is very similar for L and BC whether in their original or in their Bayesian version, ranging from 0.76 to 0.81.

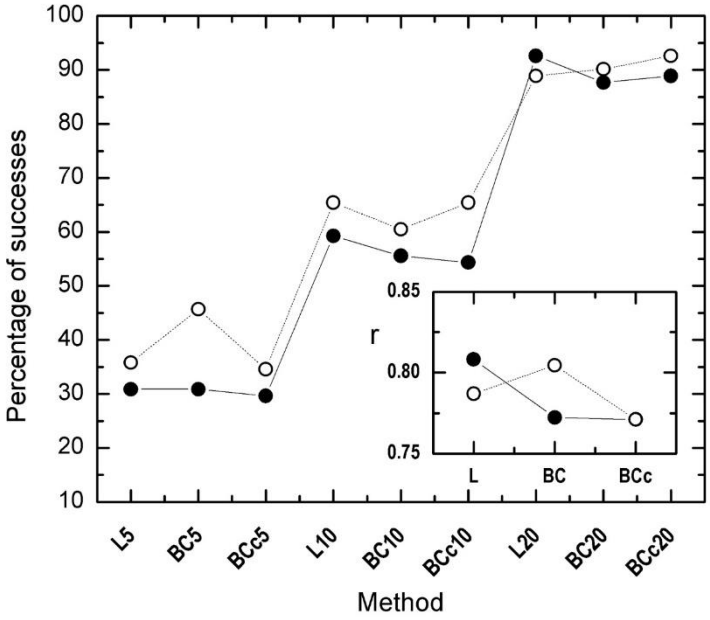

Figure 6. Percentage of successful age estimations in the Athens sample of males when the TA sample is the Athens sample (o) and the American sample (•) for the Bayesian methods under study at three age ranges, 5, 10 and 20. The corresponding Pearson's correlation coefficients of the estimated vs. documented ages are given in the inset.

Cretan females as target sample, Athenians as TA sample

In females (Figure 7), L gives better results than BC, both in its original form and its Bayesian versions but the difference is small and never statistically significant. In addition, the difference in the success rate between original and Bayesian methods is more pronounced in BC than in L but again this difference is not statistically significant. Finally, the use of an informative prior from Athens or Crete has a very small, and again non-significant, impact on the results. In Figure 8 it can be seen that when adopting an Athenian prior and TA sample, age estimates for adults up to 50 years are closer to documented ages when using L, both in

its original and Bayesian version. In contrast, age estimation deteriorates when adopting BC, either in its original or Bayesian version.

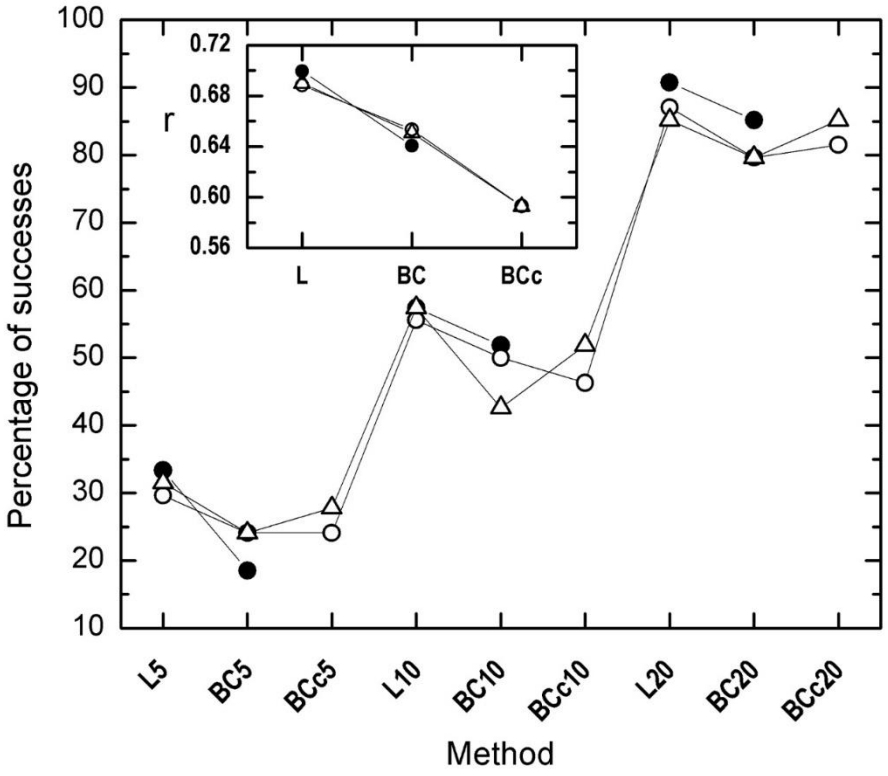

Figure 7. Percentage of success age estimations in the Cretan sample of females using the original L and BC methods (•) as well as the corresponding Bayesian approaches when the TA sample is the Athens sample and the informative prior is the Athens (o) and the Cretan (Δ) sample at three age ranges, 5, 10 and 20. The corresponding Pearson's correlation coefficients of the estimated vs. documented ages are given in the inset.

The Pearson correlation coefficient between the mean age-at-death estimated using L and the documented age-at-death of Cretan females is 0.69-0.70 and minimally affected by the choice of prior or the adoption of a Bayesian approach or not. The same correlation coefficient ranges from 0.59 to 0.65 for BC. We observe that these coefficients are lower than the corresponding ones when the Athenian sample is used as target. No statistically significant difference between methods is traced using tests for the significance of correlations.

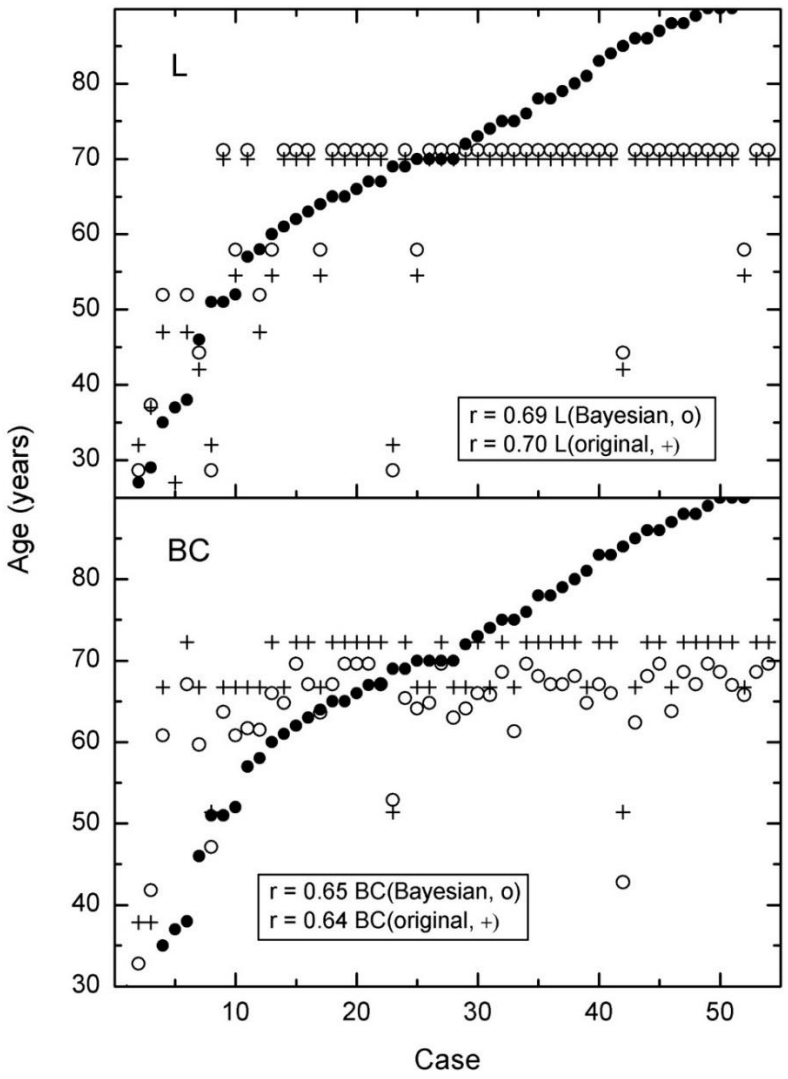

Figure 8. Documented ages of females for the Cretan sample and their estimation by means of L and BC, original (+) and Bayesian (o), using for TA and informative prior the Athens sample. Pearson's correlation coefficients of the estimated vs. documented ages are shown in the plot.

Cretan males as target sample, Athenians and Americans as TA samples

In Figure 9 it is seen that the success rates of BC and L largely overlap and fluctuate considerably depending on whether the original methods or their Bayesian versions have been used, as well as on the informative prior adopted. Bayesian statistics generally improve the success rate of L but this improvement is not statistically significant. In contrast, the original version of BC generally produces better or equally good results as its Bayesian versions when the Balkan and Athenian priors are used. Nevertheless, the small Pearson's correlation coefficients between the mean estimated age-at-death and the documented age-at-

death (0.38 to 0.45 for L and 0.28 to 0.38 for BC), highlight the fact that for Cretan males no method, original or Bayesian, actually provides satisfactory age estimations (Table 3).

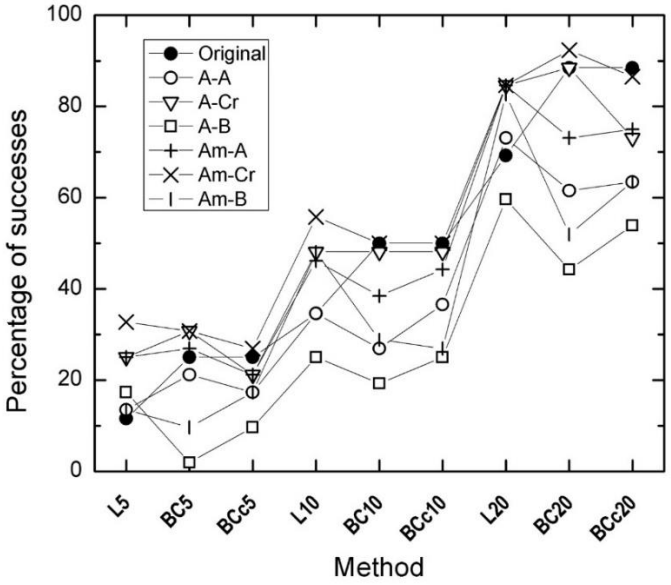

Figure 9. Percentage of successful age estimations in the Cretan sample of males using the original L and BC methods (•) as well as the corresponding Bayesian approaches when the TA sample is Athenian (A) and American (Am) and the informative prior from Athens (A), Crete (Cr) and the Balkans (B) at three age ranges, 5, 10 and 20.

Table 3 Pearson's correlation coefficients of the estimated vs. documented ages-at-death related to the data presented in Figure 9

Method	**L**	**BC**	**BCc**
Original	0.451	0.327	0.327
A-Ath	0.431	0.325	0.377
A-Cr	0.431	0.276	0.358
A-B	0.422	0.327	0.372
Am-Ath	0.408	0.338	0.382
Am-Cr	0.374	0.320	0.359
Am-B	0.404	0.342	0.379

Key: A = Athenian TA sample, Am = American TA sample, Ath = Athenian informative prior, Cr = Cretan informative prior, B = Balkan informative prior

Figure 10 shows that when using an American sample to obtain the TA parameters, even in the optimum case where a Cretan prior has been adopted (see Figure 9), there is no improvement in age estimation compared to the original methods, both for BC and L. This is additionally highlighted by the notably low Pearson's correlation coefficients between estimated and documented age-at-death, which ranges from 0.32 to 0.33 for BC and from 0.37 to 0.45 for L, depending upon whether the original or the Bayesian version is used (Table 3).

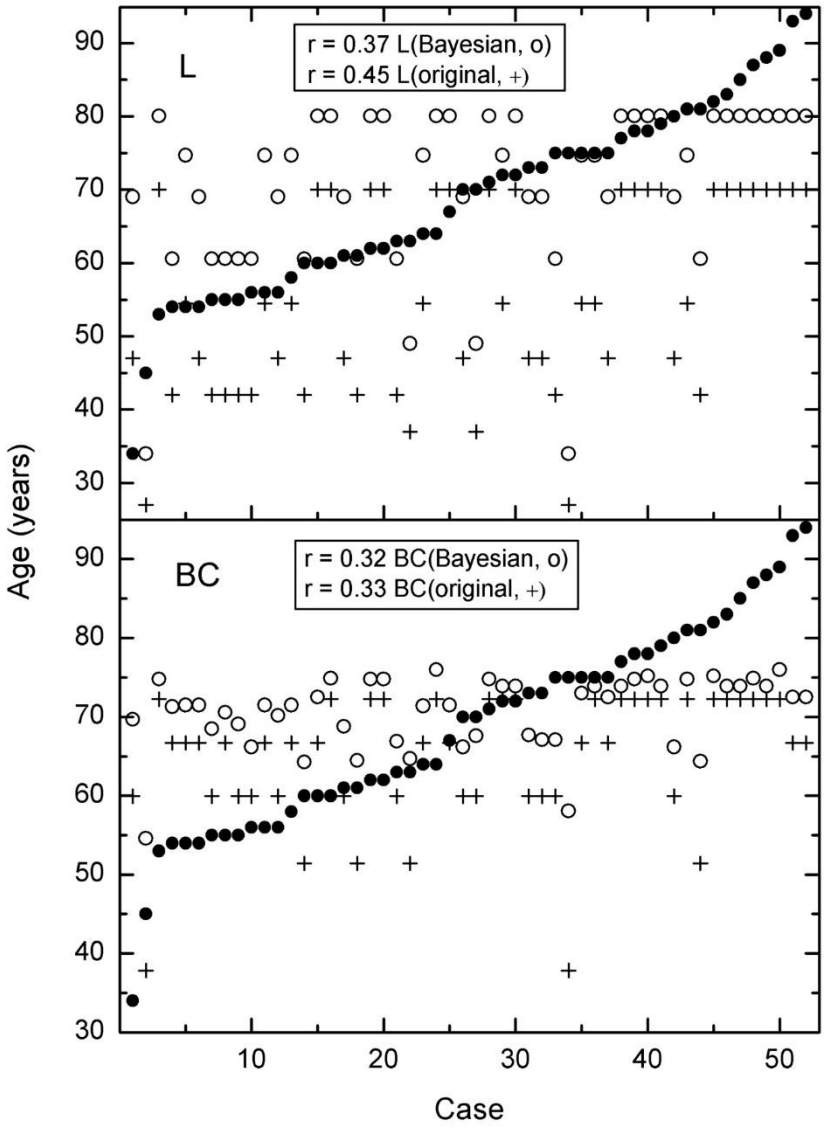

Figure 10. Documented ages of males for the Cretan sample and their estimation by means of L and BC, original (+) and Bayesian (o), using for TA the American sample and informative prior the Cretan sample. Pearson's correlation coefficients of the estimated vs. documented ages are shown in the plot.

Pooled Cretan sample as target, TA sample, and informative prior

In order to explore whether the poor results for the Cretan sample are due to the biased demographic nature of this collection (Table 1), we have explored the case where Cretans are used as target, TA sample, and informative prior. Note that in this instance we had to pool the male and female sample because the very small number of individuals younger than 50 years did not allow us to obtain transition analysis parameters per sex. In Figure 11 it is seen that in L and BC the performance of both the original and the Bayesian methods is still poor, as attested by the disassociation of the predicted age-at-death estimates from the documented ones. In no case do we observe an improvement in age-at-death prediction through the adoption of Bayesian methods compared to the traditional ones.

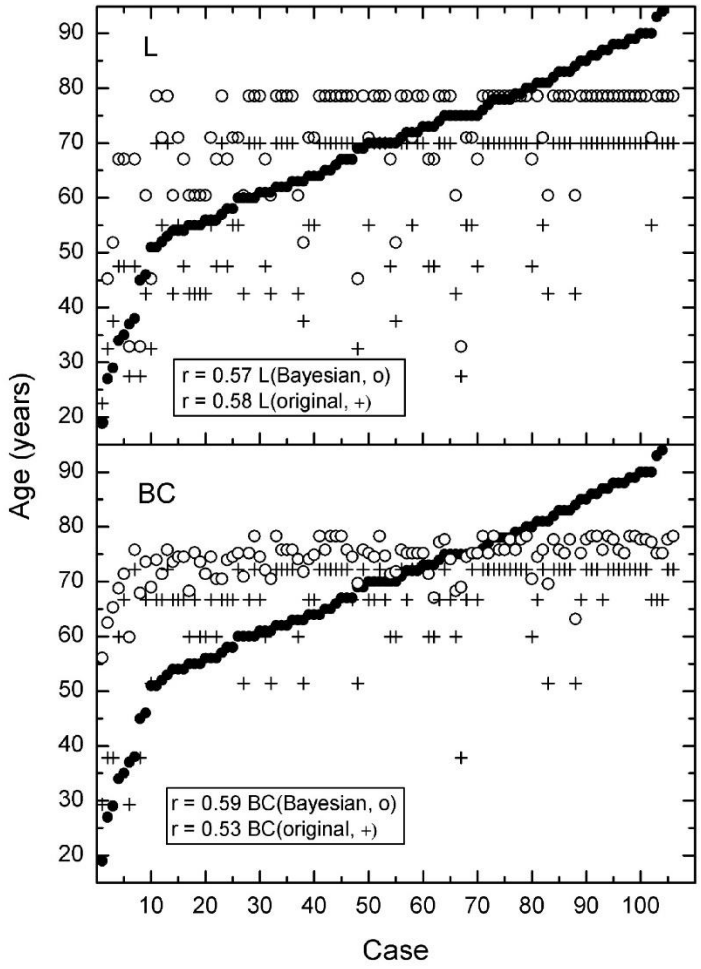

Figure 11. Documented ages-at-death for pooled sexes in the Cretan sample and estimated ages-at-death by means of L and BC, original (+) and Bayesian (o), using the Cretan sample

as TA sample and informative prior. Pearson's correlation coefficients of the estimated vs. documented ages are shown in the plot.

Discussion

The past years have witnessed a generalization in the use of Bayesian age estimation from skeletal remains in an attempt to overcome issues of 'age mimicry' (Bocquet-Appel and Masset 1982). The results of relevant studies are contradictory with many scholars reporting that Bayesian statistics improve age estimation compared to the original ageing methods (e.g. Godde and Hens 2012; Hens and Godde 2016) and others stressing the limitations of this approach and the need for further progress (e.g. Lottering et al. 2013; Milner and Boldsen 2012). The Bayesian approach to age-at-death estimation, developed by Lyle Konigsberg, Jesper Lier Boldsen and their colleagues in a number of key publications (e.g. Boldsen et al. 2002; Konigsberg and Frankenberg 1992; Konigsberg et al. 2008) has constituted a standard method during the past years, as explained above in section *Bayesian age estimation*. However, to achieve the optimum performance for a certain target sample, it is important to experiment with different reference samples and informative priors. In the current paper we explore the performance of the Bayesian method for two modern Greek assemblages, using different reference samples and informative priors from Greece, America and the Balkans. The morphology of the auricular surface of the ilium was our focus as an ageing marker by means of the Lovejoy et al. (1985a) and Buckberry and Chamberlain (2002) methods.

In summary, our results showed that when the Athenian sample is used as target, L produces satisfactory age-at-death estimates both in its original and Bayesian version in males and females. In the same sample, BC gives lower success rates but still satisfactory, while its performance improves significantly among males when adopting a Bayesian approach. Among Cretan females, L works satisfactorily for individuals younger than 50 years and again, no significant difference in its performance is observed when using the original or the Bayesian version, while BC provides less satisfactory age-at-death predictions. Finally, for Cretan males both L and BC provide particularly poor results irrespective of whether we implement them in their original or Bayesian version.

The above results first of all highlight that the implementation of Bayesian analysis did not generally lead to a significant improvement of age-at-death estimates in either Greek sample

used as a target. The only exception, as mentioned, was the Athenian male sample where the Bayesian implementation of BC indeed resulted in a significantly better age prediction. However, this result was not replicated among Athenian females, for the L method in either sex or for the Cretan target sample. It should also be stressed that the Bayesian version of BC among Athenian males, although improved compared to the original method, still did not produce more accurate age estimates than L. It is striking that this overall lack of better performance for Bayesian analysis in our samples was traced even when using the same assemblages (Athens and Cretan Collection, respectively) as target samples, TA samples and informative priors. These results highlight the lack of a clear pattern with regard to the improvement achieved by the Bayesian method, while the observed differences between males and females stress the importance of intra-sample variation in the senescence process.

A second important finding of the present study pertains to the impact of the assemblages used to derive the TA parameters. In all our tests, although the samples used as TA samples are important and we should aim at using assemblages temporally, geographically and ethnically proximal to the target sample, the actual impact this has on the estimated age-at-death is small. This is because the transition analysis used to compute $Pr(c \mid a)$ is based on the regression of c on *a*, which is not particularly sensitive to the demographic profile of the reference sample. In what concerns the informative prior, our study showed that its impact on the results is rather small when the overall performance of the Bayesian analysis is moderate (e.g. for Cretan females). In contrast, the informative prior has a significant impact on the results in cases where age prediction is very poor (e.g. for Cretan males), though even the selection of the optimum prior does not really improve age estimation in our data.

Despite the limitations of the Bayesian method highlighted in this paper, there are useful aspects to this approach that are not available with other current ageing methods. In particular, Bayesian analysis allows the combination of different skeletal age markers, the omission of markers that exhibit a low correlation with age-at-death and it can handle missing values for individual markers. This is very clear in the freely available ADBOU software, but it is also feasible with the R scripts written for the current paper. In the current paper we have made limited use of this functionality by combining L and BC in order to see if their joint use could improve the results of Bayesian analysis. Considering that both these methods focus on morphological changes in the same anatomical area, the iliac auricular surface, their combination mostly provided success rates in between those of each method used

independently. This is the reason why we did not report the relevant results. However, our scripts provide great flexibility with regard to the age markers used and the TA samples and informative priors selected. Our recommendation would be for more scholars to use the provided scripts in different target samples, with different combinations of reference populations and informative priors and adopting different skeletal markers in order to explore more systematically the effect of the demographic parameters of the models as well as the relative usefulness of different skeletal age markers and their combination.

References

Acsádi G, Nemeskéri J (1970) History of Human Life span and Mortality. Akadémiai Kiadó, Budapest.

Berg GE (2008) Pubic bone age estimation in adult women. J Forensic Sci 53:569-577. doi: 10.1111/j.1556-4029.2008.00712.x

Bocquet-Appel J, Masset C (1982) Farewell to paleodemography. J Hum Evol 11:321-333.

Bocquet-Appel JP, Masset C (1995) L'âge au décès dans les populations inhumées inhumées: comparaison de méthodes et de résultats. Antrop Port 13:39–48.

Boldsen JL, Milner GR, Konigsberg LW, Wood JM (2002) Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW (eds) Paleodemography: Age Distributions from Skeletal Samples. Cambridge University Press, New York, NY, pp 73-106.

Brooks S, Suchey JM (1990) Skeletal age determination based on the os pubis: a comparison of the Acsadi-Nemeskeri and Suchey-Brooks methods. Human Evolution 5:227-238.

Buckberry JL, Chamberlain AT (2002) Age estimation from the auricular surface of the ilium: a revised method. Am J Phys Anthropol 119:213-239. doi: 10.1002/ajpa.10130

Bullock M, Márquez L, Hernández P, Ruíz F (2013) Paleodemographic age-at-death distributions of two Mexican skeletal collections: comparison of transition analysis and traditional aging methods. Am J Phys Anthropol 152:67–78. doi: 10.1002/ajpa.22329

Byers SN (2017) Introduction to Forensic Anthropology, 5th edition. Routledge, London.

DiGangi EA, Bethard JD, Kimmerle EH, Konigsberg LW (2009) A new method for estimating age-at-death from the first rib. Am J Phys Anthropol 138:164-76. doi: 10.1002/ajpa.20916

Dirkmaat DC (2012) A Companion to Forensic Anthropology. Wiley-Blackwell, Chichester.

Eliopoulos C, Lagia A, Manolis S (2007) A modern, documented human skeletal collection from Greece. Homo 58:221-228. doi: 10.1016/j.jchb.2006.10.003

Falys CG, Schutkowski H, Weston DA (2006) Auricular surface aging: worse than expected? A test of the revised method on a documented historic skeletal assemblage. Am J Phys Anthropol 130:508-513. doi: 10.1002/ajpa.20382

Ferembach D, Schwidetzky I, Stloukal M. 1980. Recommendations for age and sex diagnoses of skeletons. J Hum Evol 9:517–549

Godde K, Hens S (2012) Age-at-death estimation in an Italian historic sample: a test of the Suchey-Brooks and transition analysis methods. Am J Phys Anthropol 149:259–65. doi: 10.1002/ajpa.22126

Hens SM, Godde K (2016) Auricular surface aging: Comparing two methods that assess morphological change in the ilium with Bayesian analyses. J Forensic Sci 61, Suppl 1: S30-38. doi: 10.1111/1556-4029.12982.

Hershkovitz I, Latimer B, Dutour O, Jellema LM, Wish-Baratz S, Rothschild C, Rothschild BM (1997) Why do we fail in ageing the skull from the sagittal suture? Am J Phys Anthropol 103:393-399. doi: 10.1002/(SICI)1096-8644(199707)103:3<393::AID-AJPA8>3.0.CO;2-R

Hoppa RD, Vaupel JW (2002) Paleo-demography: Age Distributions from Skeletal Samples. Cambridge University Press, Cambridge.

Konigsberg LW, Frankenberg SR (1992) Estimation of age structure in anthropological demography. Am J Phys Anthropol 89:235-256.

Konigsberg L, Herrmann N, Wescott D, Kimmerle E (2008) Estimation and evidence in forensic anthropology: age-at-death. J Forensic Sci 53:541-557. doi: 10.1111/j.1556-4029.2008.00710.x

Kranioti EF, İşcan MY, Michalodimitrakis M (2008) Craniometric analysis of the modern Cretan population. Forensic Sci Int 180: 110.e1-5. doi: 10.1016/j.forsciint.2008.06.018

Langley-Shirley N, Jantz R (2010) A Bayesian approach to age estimation in modern Americans from the clavicle. J Forensic Sci 55:571-583. doi: 10.1111/j.1556-4029.2010.01089.x

Lottering N, MacGregor DM, Meredith M, Alston CL, Gregory LS (2013) Evaluation of the suchey–brooks method of age estimation in an Australian subpopulation using computed tomography of the pubic symphyseal surface. Am J Phys Anthropol 150:386-399. doi: 10.1002/ajpa.22213

Lovejoy CO, Meindl RS, Pryzbeck TR, Mensforth RP (1985a) Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. Am J Phys Anthropol 68:15-28. doi: 10.1002/ajpa.1330680103

Lovejoy CO, Meindl RS, Mensforth RP, Barton TJ (1985b) Multifactorial determination of skeletal age at death: a method and blind tests of its accuracy. Am J Phys Anthropol 68:1–14. doi: 10.1002/ajpa.1330680102

Meindl RS, Lovejoy CO (1985) Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures. Am J Phys Anthropol 68:57-66.

Michopoulou E, Negre P, Nikita E, Kranioti EF (2017) The auricular surface as age indicator in a modern Greek sample: A test of two qualitative methods. Forensic Sci Int 280:246.e1-246.e7. doi: 10.1016/j.forsciint.2017.08.004

Milner GR, Boldsen JL (2012) Transition analysis: a validation study with known-age modern American skeletons. Am J Phys Anthropol 148:98-110. doi: 10.1002/ajpa.22047

Moraitis K, Zorba E, Eliopoulos C, Fox SC (2014) A test of the revised auricular surface aging method on a modern European population. J Forensic Sci 59:188-194. doi: 10.1111/1556-4029.12303

Nawrocki SP (2010) The nature and sources of error in the estimation of age at death from the human skeleton. In: Latham KE, Finnegan M (eds) Age estimation of the human skeleton. Charles C. Thomas, Springfield, pp 79-101.

Prince DA, Kimmerle EH, Konigsberg L (2008) A Bayesian approach to estimate skeletal age-at-death utilizing dental wear. J Forensic Sci 53:588-593. doi: 10.1111/j.1556-4029.2008.00714.x

Saunders SR, Fitzgerald C, Rogers T, Dudar C, McKillop H (1992) A test of several methods of skeletal age estimation using a documented archaeological sample. Canadian Society of Forensic Science Journal 25:97-118. doi: 10.1080/00085030.1992.10757005

Schmitt A, Murail P, Cunha E, Rouge D (2002) Variability of the pattern of aging on the human skeleton: evidence from bone indicators and implications on age at death estimation. J Forensic Sci 47:1203-1209.

Tangmose S, Thevissen P, Lynnerup N, Willems G, Boldsen J (2015) Age estimation in the living: Transition analysis on developing third molars. Forensic Sci Int 257:512.e1-512.e7. doi: 10.1016/j.forsciint.2015.07.049

Wood B (2011) Wiley-Blackwell Encyclopedia of Human Evolution. Wiley-Blackwell, Chichester.

Xanthopoulou P, Valakos E, Youlatos D, Nikita E (2018) Assessing the accuracy of cranial and pelvic ageing methods on human skeletal remains from a modern Greek assemblage. Forensic Sci Int 286: 266.e1-266.e8. doi: 10.1016/j.forsciint.2018.03.005