Estimation of adult skeletal age-at-death: statistical assumptions and applications

Richard Samworth and Rebecca Gowland

Statistical Laboratory

Centre for Mathematical Sciences

Wilberforce Road

Cambridge, UK

CB3 0WB

and

St John's College

Cambridge, UK

CB2 1TP.

Short title: Estimation of adult skeletal age-at-death

Corresponding author: Richard Samworth, Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge, UK. CB3 0WB. Tel: +44 1223 337950; Fax: +44 1223 337956; Email: r.samworth@statslab.cam.ac.uk.

Abstract

We examine the statistical assumptions underlying different techniques of estimating the age-at-death of a skeleton from one or more age indicators. The preferred method depends on which property of the distribution of the data in the reference sample is preserved in the skeleton to be aged. In cases where the conditional distribution of age given indicator is preserved, we provide 'look-up' tables giving essentially unbiased age estimates and prediction intervals, using a large reference sample and the auricular surface and pubic symphysis age indicators. Where this assumption is violated, but the conditional distribution of indicator given age is preserved, we find that an alternative model which attempts to capture the biological process of development of an individual has some attractive features, which may make it suitable for further study.

Key Words: Biological model; Conditional distribution; Linear regression; Look-up table

1 Introduction

The most appropriate method for estimating age-at-death from human skeletal remains has been a source of vigorous debate among anthropologists over the last two decades. See, for instance, Bouquet-Appel and Masset (1982, 1996); Konigsberg and Frankenberg (1992, 1994); Lucy et al. (1996); Konigsberg et al. (1997); Aykroyd et al. (1997, 1999); Hoppa and Vaupel (2002b). Several authors have suggested that regression techniques result in systematic under- or over-aging of skeletons (Konigsberg and Frankenberg, 1994; Konigsberg et al., 1997; Aykroyd et al., 1997, 1999). Consequently, a variety of other statistical methods have been explored with the aim of eliminating this apparent bias (e.g. Konigsberg and Frankenberg (1992, 1994, 2002); Bouquet-Appel and Masset (1996); Lucy et al. (1996); Konigsberg et al. (1997); Aykroyd et al. (1997, 1999); Boldsen et al. (2002); Konigsberg and Herrman (2002)). This research has culminated in the book 'Paleodemography', edited by Hoppa and Vaupel (2002b), in which the authors conclude that aging of skeletal remains should utilise methods based on an application of Bayes' theorem in order to obtain age-at-death estimates.

In this article, we re-examine the statistical assumptions on which the different proposed methods are based, and address some misconceptions which have arisen during the course of the debate. We use a large sample of known-age skeletal material to illustrate our discussion and to derive easy-to-use 'look-up' tables for estimating age-at-death from the auricular surface and pubic symphysis, singly and in combination. The known-age data were recorded from two skeletal collections: the 18th century sample from Spitalfields, London and a sample of the early 20th century skeletons from the Coimbra Identified Skeletal Collection (CISC), Portugal. The combination of two temporally and geographically diverse skeletal samples will go some way towards encompassing the range of variation observed between different skeletal populations (cf. Murray and Murray (1991); Hoppa (2000)). The auricular surfaces from 448 individuals and pubic symphyses from 376 individuals were scored according to

the eight phase system of Lovejoy et al. (1985) and the six phase system of Brooks and Suchey (1990) respectively. A single observer (the second author) recorded all of the data in order to eliminate inter-observer error. Since, in the majority of cases, the ages-at-death cited in the reference sample will have been rounded down to the nearest whole number of years, we added 0.5 years to all of the ages.

2 Statistical assumptions in age-at-death estimation

The statistical methods used to produce skeletal ageing techniques from known-age skeletal data have been a source of considerable debate and contention. Different techniques of age estimation make different assumptions; it is important to be precise about their nature, and this is most easily done in mathematical notation. We may assume that our reference sample consists of pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, where X_i represents the indicator measurement for the ith individual and Y_i represents the age-at-death of that individual. For simplicity, we assume for the moment that there is a single indicator available; it may be recorded as a continuous variable, as is the case for the transparency of dentin in teeth, or discrete, as for the auricular surface and pubic symphysis measurements studied below. It is natural to assume that the different pairs are independent (the indicator–age pair for one individual says nothing about the corresponding pair for another individual) and that they are identically distributed. To specify a model for the data, then, it suffices to prescribe the joint distribution of a generic pair (X_i, Y_i). In situations where it is the relationship between the variables which is of primary interest, as is the case in age estimation, this is usually done by proposing a model for the conditional distribution of one variable given the other. (The marginal distribution of the second variable, which is required to complete the specification of the joint distribution, is rarely of interest and mention of it is often

suppressed.)

One very simple model, which we do not propose as realistic, is that

$$Y_i = \alpha + \beta X_i + \epsilon_i, \tag{2.1}$$

where ϵ_i is independent of X_i and has a normal distribution with mean 0 and variance σ^2. We write $\epsilon_i \sim N(0, \sigma^2)$. This model specifies that the conditional distribution of Y_i given X_i is $N(\alpha + \beta X_i, \sigma^2)$. For completeness, we may say that the marginal distribution of X_i is $f_X(x)$. Using the data in our reference sample, we are able find estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$ for the unknown parameters in the model. We now suppose that a new individual (which we refer to as a target case, though it could be one individual from a target sample) has indicator–age pair (X^*, Y^*). We observe X^*, but not Y^*, of course. If, as in (2.1) above, we have specified the conditional distribution of Y_i given X_i, it is because we believe that the relationship between the variables for the target case is the same as for the reference sample, i.e. that

$$Y^* = \alpha + \beta X^* + \epsilon^*,$$

where ϵ^* is independent of X^* and has a $N(0, \sigma^2)$ distribution. We may then estimate Y^* by $\hat{Y} = \hat{\alpha} + \hat{\beta} X^*$. Thus the fundamental assumption in the use of regression of age on indicator for age estimation is that the conditional distribution of age given indicator is the same for the target case as for the reference sample. Notice, however, that there is no assumption that the marginal distribution $f_{X^*}(x)$ of X^* is the same as $f_X(x)$. In general, altering the marginal distribution of the indicator while preserving the conditional distribution of age given indicator will mean that the marginal distribution $f_Y(y)$ of ages in the reference sample is not the same as the corresponding marginal distribution $f_{Y^*}(y)$ for the target case. Hence regression of age on indicator does not require the assumption that the target case comes from the same marginal age-at-death distribution as does the reference sample. Incidentally, this last statement is in conflict with the argument in Box 1 on p. 95 of Konigsberg and Frankenberg (1994). In their application of Bayes' theorem in their Equation (1),

5

however, the first term in the numerator should refer to the target case rather than the reference sample. See Equation (11.1) on p. 566 of Davison (2003) for the appropriate version of Bayes' theorem in this context. We also note that the argument in the second paragraph of the same Box 1 is circular: an assumption is made that $f_Y(y)$ is equal to $f_{Y^*}(y)$, and the conclusion is that the regression of age on indicator makes this assumption.

The alternative way of specifying the joint distribution of a generic pair (X_i, Y_i) is to propose a model for the conditional distribution of indicator given age. (Mention of the marginal distribution $f_Y(y)$ of age will usually be suppressed.) In this instance, the fundamental assumption is that the conditional distribution of indicator given age is preserved in the target case.

But which is the more reasonable assumption to make? In practice, of course, each assumption will only be an approximation to the truth. In cases where assuming that the conditional distribution of age given indicator is preserved is an adequate approximation, we will see in Section 3 that making this assumption has several advantages. As well as the ease of modelling this distribution and the desirable features of the estimates and prediction intervals, we find that the methods of Section 3 adapt well to the multiple indicator setting, and do not require any assumption about the form of the marginal distribution $f_{Y^*}(y)$ of age in the target case.

However, there may also be situations where making the assumption that the conditional distribution of age given indicator is preserved is inappropriate. One instructive (though admittedly extreme) such scenario would be where the target sample is censored according to its age-at-death. Consider the hypothetical situation where an oracle who knew the true ages-at-death in the target sample were to remove those who were over 50 years old when they died. Except in unusual circumstances (such as if the indicator were perfectly correlated with age) this would mean that the conditional distribution of age given indicator would not be the same in the reference

6

and target samples. Estimates from the methods based on this assumption would therefore tend to be too high, particularly for later indicator stages. Of course, this censoring does not happen in practice, but it may be regarded as an extreme case of what might have happened at certain war graves, for example, where very few casualties might be expected to have been over 50 years old when they died.

On the other hand, censoring according to the age-at-death does not change the conditional distribution of indicator given age. This fact is demonstrated graphically in Figure 4 of Aykroyd et al. (1997); see also Davison (2003), pp. 203–206 for an excellent discussion of the effect of different types of censoring. In such circumstances, then, it would be more reasonable to assume that the conditional distribution of indicator given age is the same for the reference sample and target case. We discuss approaches to modelling this distribution in Section 4, as well as presenting a new model which attempts to capture the biological process of development of an individual.

After modelling the conditional distribution of indicator given age, however, there remains the issue of how to use this distribution to derive age estimates. There have been several proposals, which proceed via an application of Bayes' theorem, if one is willing to assume a form for the marginal age-at-death distribution in the target case – indeed much of the edited volume of Hoppa and Vaupel (2002b) is devoted to such methods. Where prior information on the marginal age-at-death distribution is available, which may be the case for certain burial sites, this may not be a problem. However, in the absence of such information, as may well be the case if we have a single skeleton to age, for example, it would seem prudent not to rely on such a model. We should mention in this regard that it has been suggested (e.g. Konigsberg and Frankenberg (1994), pp. 95-96) that a uniform marginal age distribution (with some choice of upper limit, which also has to be made) is appropriate because it is weak and equivalent to absence of information. Though this is a commonly-held belief, the two statements are different: absence of information about the distribution of Y^* is

7

equivalent to absence of information about the distribution of $\exp(Y^*)$; but a uniform distribution for Y^* does not give a uniform distribution for $\exp(Y^*)$. In general, we would not regard a uniform marginal age-at-death distribution as a realistic model for most age estimation problems. An attractive feature of our model in Section 4 is that we can obtain age estimates without assuming a particular form for the marginal distribution of ages in the target case.

3   Age estimation when the conditional distribution of age given indicator is preserved

Throughout this section, we make the assumption that the conditional distribution of age given indicator is the same for the reference sample and the target case. Based on our reference sample mentioned in Section 1, we construct age estimates and prediction intervals where either an auricular surface measurement or a pubic symphysis measurement (or both) are available. In the Appendix we show that our estimates are essentially unbiased.

3.1   Auricular surface present

Figure 1(a) gives a box and whisker plot of the ages at death of the individuals in the reference sample for which an auricular surface measurement was available. The numbers of individuals in stages 1 to 8 were 42, 45, 38, 74, 77, 75, 67 and 30. We assume that the data consist of independent and identically distributed pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, where X_i denotes the stage of the auricular surface of the ith individual (and so takes values between 1 and 8), and Y_i denotes the age-at-death of that individual. The aim is to model the conditional distribution of Y_i given X_i, and then to use this model to predict the age-at-death, Y^*, of another individual whose

age indicator measurement is X^*.

[Figure 1 to go here]

One striking feature of Figure 1(a) is that the spread of the data increases as the indicator increases. In fact, Figure 1(b) indicates that the standard deviation of the ages within a class is approximately proportional to the mean age in that class. It is therefore convenient to seek a model which relates $\log(Y_i)$ and X_i, as this transformation ensures that the conditional variance of $\log(Y_i)$ given X_i is approximately constant (cf. Brown (1993), p. 35). This transformation is also common practice when the response (i.e. the y-variable) is necessarily positive, as is the case for age.

The next decision to be made is whether to regard the x-variable as an unordered factor or as a numerical value. In the former case, the x-values are thought of as class labels, which could equally well be eight different colours, or eight different types of fruit. Our prediction for Y^* would then depend only on the data in the class to which X^* belongs. While this may at first sight seem attractive, it has the drawback that there is no guarantee that the predictions will increase as the age indicator increases. This is undesirable in view of the fact that the auricular surface measurement can only have reached a given stage after having passed through the previous ones. It could also lead to highly variable estimates when there is little data in a particular class, which becomes a more significant problem when the observations are further subdivided according to the pubic symphysis measurement.

For these reasons, we prefer to regard the auricular surface measurement as a numerical value, though we should note that this requires the implicit assumption that the age indicator stages are equally spaced (though see also the discussion in the last paragraph of this subsection). As a very basic model, we first consider the relationship

$$\log(Y_i) = \alpha + \beta X_i + \epsilon_i, \tag{3.1}$$

9

where α and β are unknown parameters to be estimated from the reference sample, and $\epsilon_i \sim N(0, \sigma^2)$. Thus in this model, the conditional distribution of $\log(Y_i)$ given X_i is $N(\alpha + \beta X_i, \sigma^2)$. In particular, in this model the conditional expectation of $\log(Y_i)$ given X_i is assumed to increase linearly with X_i, and the conditional variance of $\log(Y_i)$ given X_i is assumed to be constant.

The parameters α and β may be estimated by the method of least squares, which is equivalent in this instance (i.e. with normally distributed errors) to maximum likelihood estimation. In other words, we seek the values of α and β which minimise the residual sum of squares, given by

$$\sum_{i=1}^{n}(\log Y_i - \alpha - \beta X_i)^2.$$

Writing $Z_i = \log Y_i$, it is well-known (e.g. Montgomery and Peck (1992), pp. 8–9) that the least squares estimates of α and β are given by

$$\tilde{\beta} = \frac{S_{XZ}}{S_{XX}}, \quad \tilde{\alpha} = \bar{Z} - \tilde{\beta}\bar{X}, \tag{3.2}$$

where $S_{XZ} = \sum(X_i - \bar{X})(Z_i - \bar{Z})$, $S_{XX} = \sum(X_i - \bar{X})^2$ and $\bar{X} = n^{-1}\sum X_i$, $\bar{Z} = n^{-1}\sum Z_i$. Applying these formulae to our reference sample, we obtain the estimates $\tilde{\alpha} = 3.08$, and $\tilde{\beta} = 0.153$, so our predictions for $\log(Y^*)$ lie on the straight line shown in Figure 2.

[Figure 2 to go here]

A closer examination of Figure 2 suggests, however, that the assumption that the conditional expectation of $\log(Y_i)$ given X_i increases linearly with X_i may not be appropriate. We therefore consider the expanded model

$$\log(Y_i) = \alpha + \beta X_i + \gamma X_i^2 + \epsilon_i. \tag{3.3}$$

10

Since the model remains linear in the parameters α, β and γ, the parameters may again be estimated by least squares, though we note that there is no reason why our estimates of α and β should remain the same as in the model (3.1). Formulae for the least squares estimates in this instance are given in the Appendix, and we obtain $\hat{\alpha} = 2.79$, $\hat{\beta} = 0.321$ and $\hat{\gamma} = -0.0190$. An analysis of variance (ANOVA) test indicates that the slightly more complicated model (3.3) is preferable to the basic model (3.1), and diagnostic plots suggest that the fit of the model with the quadratic term is good.

Under our fundamental assumption on the property of the reference sample which is preserved in the target case, we have that

$$\log Y^* = \alpha + \beta X^* + \epsilon^*,$$

where $\epsilon^* \sim N(0, \sigma^2)$ and is independent of X^*. We therefore estimate Y^* by $\hat{Y} = \exp\{\hat{\alpha} + \hat{\beta}X^* + \hat{\gamma}(X^*)^2\}$. We may also derive an expression for a 90% prediction interval for Y^*, though since this requires some familiarity with properties of the multivariate normal distribution, we defer this derivation to the Appendix. Table 1 gives our results.

[Table 1 to go here]

For comparison, we did also examine the predictions from the model in which the auricular surface is regarded as an unordered factor. In this model, we would estimate $\log Y^*$ by the arithmetic mean of logarithm of the ages in the class to which X^* belongs, and this corresponds to estimating Y^* by the geometric mean of the ages in the appropriate class. The predictions from this model were similar to those presented in Table 1, in that all but two of the corresponding estimates from the two different models were within two years of each other, and the biggest discrepancy was an estimate of 52.6 years for Class 6 under the unordered factor model. A further ANOVA test confirmed the preference for the model (3.3).

11

3.2 Pubic symphysis present

Our analysis of the data for which only the pubic symphysis measurement was available was very similar to that in the previous subsection. The numbers of individuals in each of the six stages were 35, 28, 39, 94, 81 and 99. Again, we are led to make a logarithmic transformation of the ages, and the model

$$\log Y_i = \alpha + \beta X_i + \gamma X_i^2 + \epsilon_i \tag{3.4}$$

(where now X_i represents the pubic symphysis stage) was found to be most successful. Our parameter estimates were $\hat{\alpha} = 2.725$, $\hat{\beta} = 0.332$ and $\hat{\gamma} = -0.0159$, giving estimates for $\log Y^*$ which lie on the line in Figure 3. The predictions for Y^* are given in Table 2.

[Figure 3 to go here]

[Table 2 to go here]

3.3 Both indicators present

In this subsection, we write X_i^A and X_i^P respectively for the auricular surface and pubic symphysis measurements of the ith individual. The presence of both indicators greatly increases the number of plausible models to be considered, particularly due to possible interaction effects between the two predictors, as in the fifth and sixth terms on the right-hand side of (3.5) below. ANOVA tests were used as a basis for comparison of different models, and suggested that the model

$$\log Y_i = \beta_1 + \beta_2 X_i^A + \beta_3 (X_i^A)^2 + \beta_4 X_i^P + \beta_5 X_i^A X_i^P + \beta_6 (X_i^A)^2 X_i^P + \epsilon_i \tag{3.5}$$

was the most plausible. The presence of a quadratic term in the auricular surface but not in the pubic symphysis is perhaps a little surprising at first, though the quadratic

component does appear to be more significant in Figure 1 than in Figure 3. The least squares estimators of the parameters were $\hat{\beta}_1 = 2.54$, $\hat{\beta}_2 = 0.323$, $\hat{\beta}_3 = -0.0315$, $\hat{\beta}_4 = 0.194$, $\hat{\beta}_5 = -0.0453$ and $\hat{\beta}_6 = 0.00576$, yielding the predictions in Table 3. Certain predicted ages and associated intervals in the top-right of Table 3 have been left blank, as there is insufficient data in these categories to justify a prediction via this model. Of course, since these pairs of age indicator measurements are very rare, this is unlikely to be a problem when using the table in practice.

[Table 3 to go here]

Table 3 has several interesting features. First, notice that moving down or to the right in the table increases both the estimated age and the endpoints of the prediction interval. Second, if the estimated ages from the individual indicator models are in broad agreement, then the prediction interval from the combined model tends to be a little narrower than either interval from the separate models. Thus, if an individual has an auricular surface measurement in stage 5, then from Table 1, we estimate the age-at-death to be 50.3 with a 90% prediction interval of width 47.0 years, while a pubic symphysis measurement of 5 gives an estimated age of 53.8 with an interval of width 57.8 years, from Table 2. However, from the combined model (3.5), we obtain a predicted age of 50.7 with an interval of width 41.4 years.

Another important point to note is that it is not necessarily the case that the predicted ages from the combined model lie between the two predictions from the separate models. For instance, an auricular surface measurement of 8 gives an estimated age of 62.8, while a pubic symphysis measurement of 6 gives an estimated age of 62.9. However, the combined prediction is 74.2. In fact we regard this as a desirable characteristic: the second piece of evidence strengthens the argument that the individual died in old age. Interestingly, the mean age-at-death of the 13 individuals having the age indicator pair $(8, 6)$ was 70.0, but this may be considered a little lower

13

than might be expected, as the mean of the 28 individuals having the indicator pair $(7, 6)$ was 71.3. Thus the predicted age from the combined model (3.5) seems entirely reasonable.

4 Age estimation when the conditional distribution of indicator given age is preserved

The techniques of Section 3 give very satisfactory answers to the problem of age estimation when the conditional distribution of age given indicator is the same in the reference sample as for the target case. It was argued in Section 2, though, that there may be circumstances where this assumption would be a poor approximation, and it would be more reasonable to assume that the conditional distribution of indicator given age is preserved. But how should one model the conditional distribution of indicator given age? And how should one use this model to estimate the age-at-death of a target skeleton? The first obvious possibility is to use a similar regression technique to that employed in Section 3. This has several drawbacks: an examination of Figure 1(a) suggests that the conditional variance of indicator given age is certainly not constant (it appears to increase up to ages around 50 before decreasing again) and there does not appear to be an obvious transformation to correct this. Moreover, when using the method of least squares in a prediction problem, it is natural to place the unobserved variable to be predicted (age) on the y-axis. The variable (indicator) which is on the x-axis is (by assumption) observed without error, and the error in the prediction is in the y-direction. Regression of y on x then minimises the errors in the direction in which they occur, but this is not the case for the regression of x on y. Additional practical problems are mentioned in Aykroyd et al. (1997): the variance of the predicted ages is larger for the regression of x on y than y on x, and the confidence intervals are of a more complicated form.

14

We also argued in Section 2 that there may well be situations where prior information on the marginal age-at-death distribution of the target case is not available. In the remainder of this section, therefore, we investigate a new model for the conditional distribution of indicator given age and with the aim of producing age estimates which do not rely on a model for the marginal age-at-death distribution in the target case.

The idea of this model is to try to capture the underlying biological process by which an individual moves between stages over time. In order to describe this model, it is convenient to alter our notation from what we have used previously. We can consider that at time t after the birth of an individual, its indicator (either auricular surface or pubic symphysis) is in stage X_t. Although the aging process is continuous, the pubic symphysis and auricular surface measurements are recorded as whole numbers. We can therefore think of the individual 'jumping' up a stage at particular times. As an initial model, then, we may say that the individual spends an exponential length of time in stage 1, before jumping to stage 2, where the individual spends another exponential length of time (independent of the time spent in the first stage, and possibly with a different parameter for the exponential distribution) before jumping to stage 3, and so on. The exponential family of distributions provides a natural model for the waiting time before jumping up a stage, because it has a memoryless property. That is, the probability that an exponential random variable is greater than $s + t$, given that it is greater than s, is equal to the (unconditional) probability that it is greater than t. In fact, no other family of continuous distributions has this property, so it is often used to model failure times for components (e.g. light bulbs), waiting times for calls at a telephone exchange etc..

At some random time T (with marginal distribution $f_T(t)$, say), the individual dies, and we observe X_T, the indicator stage at that time. In mathematical notation, then, we may consider the data to be independent and identically distributed pairs $(T_1, X_{T_1}), \ldots, (T_n, X_{T_n})$, where T_i represents the age-at-death of the ith individual,

15

and X_{T_i} represents its indicator stage at that time. Given the age-at-death T_i for the ith individual, we have

$$X_{T_i} = 1 + \max\left\{m \leq M : \sum_{j=1}^{m} E_j^i \leq T_i\right\},$$

where M is the maximum possible indicator stage and $E_1^i, \ldots, E_M^i$ are independent random variables with E_j^i having an exponential distribution with parameter λ_j (that is, mean $1/\lambda_j$). We set $\lambda_M = 0$ to ensure that X_T cannot be greater than M. In fact, we found that the slight modification of the model obtained by preventing a jump to the second stage before a years resulted in a much better fit to the data. Thus we now have

$$X_{T_i} = 1 + \max\left\{m \leq M : \sum_{j=1}^{m} E_j^i \leq T_i - a\right\}.$$

The parameter a must be estimated from the data in the reference sample, in addition to $\lambda_1, \ldots, \lambda_{M-1}$, though since one individual died aged 20.5 years in auricular surface stage 2, we constrained a to lie between 0 and 20 years. The parameters may be estimated by the method of maximum likelihood. The log-likelihood for the parameters (excluding the term involving $f_T(t)$, which does not depend on the parameters) is

$$\ell(a, \lambda_1, \ldots, \lambda_{M-1}) = \sum_{i=1}^{n} \log\left\{P\left(\sum_{j=1}^{X_{T_i}-1} E_j^i \leq T_i - a\right) - P\left(\sum_{j=1}^{X_{T_i}} E_j^i \leq T_i - a\right)\right\}.$$

It may be proved that

$$P\left(\sum_{j=1}^{m} E_j^i \leq y\right) = 1 - \sum_{j=1}^{m} e^{-\lambda_j y} \prod_{\substack{1 \leq k \leq m \\ k \neq j}} \frac{\lambda_k}{\lambda_k - \lambda_j}, \tag{4.1}$$

so the log-likelihood is

$$\ell(\lambda_1, \ldots, \lambda_{M-1}, a)$$

$$= \sum_{i=1}^{n} \log\left\{\sum_{j=1}^{X_{T_i}} e^{-\lambda_j(T_i-a)} \prod_{\substack{1 \leq k \leq X_{T_i} \\ k \neq j}} \frac{\lambda_k}{\lambda_k - \lambda_j} - \sum_{j=1}^{X_{T_i}-1} e^{-\lambda_j(T_i-a)} \prod_{\substack{1 \leq k \leq X_{T_i}-1 \\ k \neq j}} \frac{\lambda_k}{\lambda_k - \lambda_j}\right\}.$$

16

This function may be maximised numerically to yield maximum likelihood estimates $\hat{a}, \hat{\lambda}_1, \ldots, \hat{\lambda}_{M-1}$ of the parameters, for example, using the 'optim' function in the statistical programming language 'R'. Given X_{T^*}, the age indicator measurement of a new individual at its time of death, we propose to estimate T^* by

$$\hat{T} = \hat{a} + \frac{1}{\hat{\lambda}_1} + \ldots + \frac{1}{\hat{\lambda}_{X_{T^*}-1}} + \frac{1}{2\hat{\lambda}_{X_{T^*}}}.$$

In other words, after the initial period of $\hat{a}$ years, we estimate the time spent in each of the earlier indicator stages by their estimated means, and add on one half of the estimated mean length of time spent in the final stage. The rationale for the factor of one half here is that if death had not occurred and we were to observe that the next jump happened at time E after the final jump which did occur, it would be natural to estimate that the death occurred at time $E/2$ after this final jump. As we do not observe E, we estimate it by estimating its mean.

For the auricular surface data, the maximum likelihood estimates for the parameters were $\hat{a} = 20$, $\hat{\lambda}_1 = 0.365$, $\hat{\lambda}_2 = 0.132$, $\hat{\lambda}_3 = 0.193$, $\hat{\lambda}_4 = 0.0959$, $\hat{\lambda}_5 = 0.0928$, $\hat{\lambda}_6 = 0.0730$, $\hat{\lambda}_7 = 0.0297$. The resulting age estimates for the first five indicator stages were 21.4, 26.5, 32.9, 40.7 and 51.3, in close agreement with those presented in Table 1. Estimates for later stages were unreliably high. There appear to be two problems: firstly, while each data point gives information about the parameter λ_1, only individuals who died in stages 7 and 8 yield any information about λ_7, so the parameter estimates will be more variable for later parameters. It also appears, however, that the assumption of the exponential distribution may be too restrictive in later stages. The exponential distribution has its standard deviation equal to its mean, and the small parameter estimates for λ_6 and λ_7 may be more a reflection of the increased standard deviation for the waiting time for a jump (a fact noted by Boldsen et al. (2002), p. 83) rather than the increased mean. Ideally, we would like to be able to use a more flexible family (such as the Weibull family, of which the exponential distribution is a special case). In order to apply this in practice, we would need to be able to compute the analogue of (4.1) for a Weibull family – in other

17

words, to compute the distribution function of a sum of independent Weibull random variables having different parameters. This does not appear to be straightforward, though Monte-Carlo methods may help in this regard.

For the pubic symphysis data, the maximum likelihood estimates were $\hat{a} = 20$, $\hat{\lambda}_1 = 0.669$, $\hat{\lambda}_2 = 0.141$, $\hat{\lambda}_3 = 0.157$, $\hat{\lambda}_4 = 0.0642$, $\hat{\lambda}_5 = 0.0505$, yielding estimates for the first four indicator stages of 20.7, 25.0, 31.8 and 42.8, again in very close agreement with the estimates in Table 2. The later estimates are again unreliably high.

The main attractions of this model over those proposed in Section 3 and elsewhere in the literature are that it represents an attempt to model the underlying biological process and that the estimates do not rely on a model for the marginal age-at-death distribution. However, in addition to the lack-of-fit problem in the later stages mentioned above, there are other drawbacks. Firstly, even if the model were correct, we would still have the problem of how to estimate the age in the final indicator stage. One way to do this would be to model the time spent in the final indicator stage before death. It could also be argued that the method is not as easily adapted to the multiple indicator setting. Assuming that the processes X_t^A and X_t^P, representing the auricular surface and pubic symphysis measurements at time t, evolved independently, we would estimate the age-at-death by the average of the individual estimates. Thus the combined prediction would always lie between the two separate predictions, which may not always be desirable (cf. the discussion at the end of Section 3.3). Lastly, while the model does have a theoretical, biological basis, ideally we would also like to be able to verify the model empirically.

To obtain prediction intervals for the age-at-death in each indicator stage, we do have to assume a form for the marginal age-at-death distribution for the target case. The following bootstrap algorithm could then be applied:

1. Generate T^* from the assumed marginal age-at-death distribution $f_{T^*}(t)$ for the target case

2. Generate independent exponential random variables $E_1, \ldots, E_M$ where $E_j \sim \mathrm{Exp}(\hat{\lambda}_j)$

3. Set $X_{T^*} = 1 + \max\left\{ m \leq M : \sum_{j=1}^{m} E_j \leq T^* - \hat{a} \right\}$

4. Repeat steps 1–3 a total of 100,000 times, recording each value of T^* and X_{T^*}

5. Sort into ascending order the values of T^* for which the corresponding value of X_{T^*} was equal to 1. If there are n_1 such values of T^*, the $(0.05n_1)$th and $(0.95n_1)$th values in the sorted list give lower and upper endpoints for a 90% prediction interval (round $0.05n_1$ and $0.95n_1$ to the nearest whole number if necessary). Repeat for the other possible values of X_{T^*}.

To illustrate this method, we applied it to our reference sample with two different marginal age-at-death distributions for the target case. In the first, we took $f_{T^*}(t)$ to be the empirical distribution of the ages in the reference sample (so step 1 of the algorithm above corresponded to sampling T^* uniformly at random from the ages in the reference sample). Thus, in this example, the marginal age-at-death distributions were virtually identical in the reference sample and the target case. For the auricular surface indicator, the lower and upper endpoint pairs in the first five stages were $(16.5, 27.5)$, $(22.5, 46.5)$, $(24.5, 53.5)$, $(27.5, 65.5)$ and $(32.5, 72.5)$ respectively, while for the pubic symphysis, the corresponding pairs in the first four indicator stages were $(16.5, 22.5)$, $(21.5, 43.5)$, $(23.5, 53.5)$ and $(28.5, 70.5)$. These values are generally in close agreement with those presented in Tables 1 and 2.

On the other hand, in the second example we chose $T^* = 15 + \Gamma$, where Γ had a gamma distribution with parameters 1.5 and 0.1 (so T^* has mean 30 years and standard deviation 12.2 years). This distribution was chosen to represent a crude

approximation to the marginal age-at-death distribution at a war grave. The interval endpoints for first five stages of the auricular surface indicator were $(15.8, 24.9)$, $(21.2, 37.7)$, $(23.0, 43.1)$, $(25.6, 52.0)$ and $(28.8, 59.7)$, while the endpoints for the first four stages for the pubic symphysis indicator were $(15.7, 22.3)$, $(21.0, 35.9)$, $(22.8, 42.1)$ and $(25.5, 54.1)$. We see that the upper endpoints of the intervals in particular are dramatically reduced. It is not a surprise that these intervals are markedly different from those in Tables 1 and 2 since the conditional distributions of age given indicator are dissimilar in the reference and target cases. One might also modify the age estimates in this situation, for instance by estimating the age-at-death in stage 1 by the average of the values of T^* for which X_{T^*} was equal to 1. The revised estimates from the auricular surface indicator were 19.6, 27.4, 31.0, 36.5 and 42.0, while for the pubic symphysis indicator, they were 18.7, 26.2, 30.4 and 37.1. Finally, we note that it is not necessarily clear how best to combine the two intervals if both indicators are available.

5   Conclusions

In the last twenty years, a wide variety of different techniques have been proposed for the aging of adult skeletons. We have found that different methods rely on different assumptions, and the question of which is the most appropriate procedure depends on the assumptions that the investigator is prepared to make. In particular, one must decide which property of the distribution of the data in the reference sample is preserved in the target sample.

In situations where the assumption that the conditional distribution of age given indicator is preserved is adequate, we recommend the methodology of Section 3. We have applied our methodology to a large reference sample and produced look-up tables which give anthropologists straightforward estimates, together with prediction

intervals, in cases where an auricular surface or pubic symphysis measurement (or both) are available.

However, we have also seen that there may be other situations in which it may be more reasonable to assume that the conditional distribution of indicator given age is preserved. Such circumstances are more problematic, particularly if prior information on the marginal age-at-death distribution of the target skeleton is not available. We have examined a simple model of the biological process of indicator development. The model appears to have some attractive features, although further research into more general versions of the model presented in Section 4 may yield more reliable estimates, particularly at later indicator stages. Even though the estimates in Section 3 should not be regarded as 'correct' if the conditional distribution of age given indicator is not preserved, they still serve as a useful basis for comparison with estimates from other models. Moreover, we find it reassuring that the estimates agree very closely with those from the model in Section 3, at least for the early indicator stages.

A   Appendix

A.1   Least squares regression and unbiasedness

In this subsection, we discuss the statistical theory underlying the predictions in Section 3. This is important because of the current criticism levelled at the use of regression in skeletal age estimation. For simplicity, we will again treat the case

where only a single age indicator is available, though the argument easily extends to the multiple predictor setting. Recall that the data in the reference sample consists of pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, where X_i denotes the age indicator measurement of the ith individual and Y_i denotes the age-at-death of that individual. We assume that the pairs are independent and identically distributed and attempt to model the conditional distribution of age given indicator under the assumption that this distribution is preserved in the target case.

Assume for the moment that it is through the raw ages themselves, rather than their logarithms, that we model this distribution. One simple possibility is that

$$Y_i = \alpha + \beta X_i + \epsilon_i, \tag{A.1}$$

where ϵ_i is independent of X_i and has a $N(0, \sigma^2)$ distribution. Notice that if we write $Y = (Y_1, \ldots, Y_n)^T$, $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^T$ and $\theta = (\alpha, \beta)^T$, then we can write the model in vector form as

$$Y = \mathbf{X}\theta + \epsilon, \tag{A.2}$$

where $\mathbf{X}$ is the $n \times 2$ matrix whose first column consists entirely of ones, and whose second column has entries $X_1, \ldots, X_n$. Similarly, the model

$$Y_i = \alpha + \beta X_i + \gamma X_i^2 + \epsilon_i, \tag{A.3}$$

can be written in the same form (A.2), where now $\theta = (\alpha, \beta, \gamma)^T$, and $\mathbf{X}$ has the same first two columns as before, as well as an additional third column with entries $X_1^2, \ldots, X_n^2$. In general, we let p denote the number of columns of $\mathbf{X}$. Apart from the fact that we are now discussing raw ages rather than logarithms of ages, all of the models considered in Section 3 are of the form (A.3). The least squares estimator $\hat{\theta}$ of θ minimises

$$\sum_{i=1}^{n} \{Y_i - (\mathbf{X}\theta)_i\}^2,$$

where $(\mathbf{X}\theta)_i$ denotes the ith component of the vector $\mathbf{X}\theta$. The famous formula (Brown (1993), p. 40) for the least squares estimator is

$$\hat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^TY, \tag{A.4}$$

and it is well-known that $\hat{\theta}$ has a multivariate normal distribution with mean vector θ and covariance matrix $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$. For the model (3.1), it may be verified that (A.4) yields the same estimates of the parameters as does (3.2), once Y has been replaced by $\log Y$ in (A.4).

Now suppose that the age indicator and age-at-death of a new individual are given by the pair (X^*, Y^*). By assumption, we may write

$$Y^* = \mathbf{X}^*\theta + \epsilon^*,$$

where $\mathbf{X}^*$ is equal to the first row of $\mathbf{X}$, but with X_1 replaced with X^*, and ϵ^* has a $N(0, \sigma^2)$ distribution and is independent of X^*. Notice that we make no assumption that the marginal distribution $f_{X^*}(x)$ of X^* is equal to $f_X(x)$.

In this situation, we would estimate Y^* by $\hat{Y} = \mathbf{X}^*\hat{\theta}$. Thus, in the model (A.3), we would have $\mathbf{X}^* = \begin{pmatrix} 1 & X^* & (X^*)^2 \end{pmatrix}$ and $\hat{Y} = \mathbf{X}^*\hat{\theta} = \hat{\alpha} + \hat{\beta}X^* + \hat{\gamma}(X^*)^2$. It then follows from a standard property of the multivariate normal distribution that $\mathbf{X}^*\hat{\theta} \sim N(\mathbf{X}^*\theta, \sigma^2\tau^2)$, where $\tau^2 = \mathbf{X}^*(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^*)^T$.

The crucial point is this: the expected value of $\hat{Y}$ is $\mathbf{X}^*\theta$, which is equal to the expected value of Y^*. In this natural sense, then, our predicted age is unbiased. (As is conventional in regression problems, we have treated $X_1, \ldots, X_n$ and X^* as fixed at their observed values; in mathematical terminology, the expectations are taken conditional on these values.)

The unbiasedness described above is in contrast with the argument of Aykroyd et al. (1997) and Aykroyd et al. (1999), which says that the least squares regression technique overestimates the age of young individuals and underestimates the age of

old individuals. Their justification is that if, after fitting the basic linear model (A.2), one plots the points $(Y_i, \hat{\epsilon}_i)$ for $i = 1, \ldots, n$, where $\hat{\epsilon}_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$ is the ith residual, then the least squares regression line of $\hat{\epsilon}_i$ on Y_i has positive gradient. While this is true, we do not regard the predicted ages as biased as a result, as this argument says nothing about the predicted age for a new individual. The real question of interest is: given that we observe the age indicator measurement X^* for an individual not in the reference sample, do we on average over- or under-predict their age? This is the question addressed in the previous paragraph.

A minor complication arises when we first take a logarithmic transformation of the y-values, as in Section 3. In this case, our model is

$$\log Y = \mathbf{X}\theta + \epsilon,$$

and we estimate Y^* by $\hat{Y} = \exp(\mathbf{X}^*\hat{\theta})$. Now $\hat{Y}$ has a lognormal distribution, and it may be shown that the expected value of $\hat{Y}$ is $\exp(\mathbf{X}^*\theta + \sigma^2\tau^2/2)$. On the other hand, the expected value of Y^* is $\exp(\mathbf{X}^*\theta + \sigma^2/2)$. The value of τ^2 depends on X^*, but for each of the data sets in Section 3 and each possible value of X^*, it is less than 1. If we define the multiplicative bias of our estimate to be the ratio of these expected values, then the multiplicative bias is $\exp\{\sigma^2(\tau^2 - 1)/2\}$. To gain an idea of its magnitude, we estimate the multiplicative bias by $\exp\{\hat{\sigma}^2(\tau^2 - 1)/2\}$, where

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \{Y_i - (\mathbf{X}\theta)_i\}^2$$

is an unbiased estimate of σ^2. For each of the data sets in Section 3 and each possible value of X^*, our estimate of the multiplicative bias of $\hat{Y}$ was between 1 and 4 per cent. The bias is therefore relatively small, though we could modify our estimate of Y^* as $\tilde{Y} = \exp\{\mathbf{X}^*\hat{\theta} + \hat{\sigma}^2(1 - \tau^2)/2\}$ to obtain a virtually unbiased estimate. This has the drawback that the variance of $\tilde{Y}$ is greater than that of $\hat{Y}$, so we did not implement this in Section 3. It should be emphasised that this small bias arises purely as a result of the logarithmic transformation, and is not related to the points raised in,

for example, Aykroyd et al. (1997), Konigsberg et al. (1997), Aykroyd et al. (1999) and Konigsberg and Frankenberg (1994).

The next problem is to derive a $100(1 - \alpha)\%$ prediction interval for Y^*, where $0 < \alpha < 1$. Now $\log \hat{Y} - \log Y^* = \mathbf{X}^*(\hat{\theta} - \theta) - \epsilon^*$, which has a normal distribution with mean 0 and variance $\sigma^2(\tau^2 + 1)$. It follows from an elementary property of the family of t-distributions (cf. Casella and Berger (1990), pp. 225–226), that

$$\frac{\log \hat{Y} - \log Y^*}{\hat{\sigma}\sqrt{\tau^2 + 1}}$$

has a t-distribution with $n - p$ degrees of freedom. If we denote the upper α-point of this symmetric distribution by $t_{n-p}(\alpha)$, then

$$
\begin{aligned}
1 - \alpha &= P\left(-t_{n-p}(\alpha/2) \leq \frac{\log \hat{Y} - \log Y^*}{\hat{\sigma}\sqrt{\tau^2 + 1}} \leq t_{n-p}(\alpha/2)\right) \\
&= P\left(\hat{Y}\exp\{-\hat{\sigma}\sqrt{\tau^2 + 1}\,t_{n-p}(\alpha/2)\} \leq Y^* \leq \hat{Y}\exp\{\hat{\sigma}\sqrt{\tau^2 + 1}\,t_{n-p}(\alpha/2)\}\right),
\end{aligned}
$$

so a $100(1 - \alpha)\%$ interval for Y^* has endpoints at $\hat{Y}\exp\{-\hat{\sigma}\sqrt{\tau^2 + 1}\,t_{n-p}(\alpha/2)\}$ and $\hat{Y}\exp\{\hat{\sigma}\sqrt{\tau^2 + 1}\,t_{n-p}(\alpha/2)\}$. These were the expressions which were used for the construction of prediction intervals throughout Section 3.

[Table 4 to go here]

[Table 5 to go here]

[Table 6 to go here]

25

References

Aykroyd RG, Lucy D, Pollard AM and Roberts CA. 1999. Nasty, brutish, but not necessarily short: a reconsideration of the statistical methods used to calculate age at death from adult human skeletal and dental age indicators. American Antiquity 64; 55–70.

Aykroyd RG, Lucy D, Pollard AM and Solheim T. 1997. Technical note: regression analysis in adult age estimation. Am. J. Physical Anthropology 104; 259–265.

Boddington A. 1987. From bones to population: the problem of numbers. In Death, Decay and Reconstruction., Boddington A, Garland A, Janaway AN (ed.). Manchester Univ. Press; 180–197.

Boldsen JL, Milner GR, Konigsberg LW and Wood JW. 2002. Transition analysis: a new method for estimating age from skeletons. In Paleodemography., Hoppa RD and Vaupel JW (ed.). Cambridge Univ. Press; 73–106.

Bouquet-Appel J-P and Masset C. 1982. Farewell to Paleodemography. J. Human Evolution 11; 321–333.

Bouquet-Appel J-P and Masset C. 1996. Palaeodemography: expectancy and false hope. Am. J. Physical Anthropology 89; 235–256.

Brooks ST and Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Asca'di–Nemeskéri and Suchey–Brooks methods. Human Evolution 5; 227–238.

Brown PJ. 1993. Measurement, Regression, and Calibration. Oxford Univ. Press.

Buikstra JE and Konigsberg LW. 1985. Palaeodemography: critiques and controversies. Am. Anthropologist 87; 316–333.

Casella G and Berger RL. 1990. Statistical Inference. Duxbury Press, California.

Chamberlain AT. 2000. Problems and prospects in palaeodemography. In Human Osteology in Archaeology and Forensic Science., Cox M and Mays S (ed.). Greenwich Medical Media, London; 101–116.

Davison AC. 2003. Statistical Models. Cambridge Univ. Press.

Hoppa RD. 2000. Population variation in osteological aging criteria: an example from the pubic symphysis. Am. J. Physical Anthropology 111; 185–191.

Hoppa RD and Vaupel JW. 2002. The Rostock manifesto for paleodemography. In Paleodemography., Hoppa RD and Vaupel JW (ed.). Cambridge Univ. Press; 1–8.

Hoppa RD and Vaupel JW (ed.). 2002. Paleodemography. Cambridge Univ. Press.

Howell N. 1982. Village composition implied by a paleodemographic life table: the Libben site. Am. J. Physical Anthropology 59; 263–269.

Konigsberg LW and Frankenberg SR. 1992. Estimation of age structure in anthropological demography. Am. J. Physical Anthropology 89; 235–256.

Konigsberg LW and Frankenberg SR. 1994. Paleodemography: "not quite dead". Evolutionary Anthropology 3; 92–105.

Konigsberg LW and Frankenberg SR. 2002. Deconstructing death in paleodemography. Am. J. Physical Anthropology 117; 297–309.

Konigsberg LW, Frankenberg SR and Walker RB. 1997. Regress what on what? Paleodemographic age estimation as a calibration problem. In Integrating Archaeological Demography., Paine RR (ed.); 64–88.

Konigsberg LW and Herrman NP. 2002. Markov Chain Monte Carlo estimation of hazard model parameters in paleodemography. In Paleodemography., Hoppa RD and Vaupel JW (ed.). Cambridge Univ. Press; 222–242.

Lovejoy CO, Meindl RS, Pryzbeck TR and Mensforth RP. 1985. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. Am. J. Physical Anthropology 68; 15–28.

Lucy D, Aykroyd RG, Pollard AM and Solheim T. 1996. A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. J. Forensic Sciences 41; 189–194.

Montgomery DC and Peck EA. 1992. Introduction to Linear Regression Analysis. Wiley, New York.

Murray K and Murray T. 1991. A test of the auricular surface ageing technique. J. Forensic Sciences 36; 1162–1169.

Paine RR and Harpending HC. 1998. Effect of sample bias on palaeodemographic fertility estimates. Am. J. Physical Anthropology 105; 231–240.

Van Gervan D and Armelagos G. 1983. Farewell to paleodemography? Rumours of its death have been greatly exaggerated. J. Human Evolution 12; 353–360.

Walker PL, Johnson JR and Lambert PM. 1988. Age and sex biases in the preservation of human skeletal remains. Am. J. Physical Anthropology 76; 183–188.

Wood JW, Holman DL, O'Connor KA and Ferrell RJ. 2002. Mortality models in paleodemography. In Paleodemography., Hoppa RD and Vaupel JW (ed.). Cambridge Univ. Press; 129–168.

Auricular surface	1	2	3	4	5	6	7	8
Lower 90% endpoint	14.0	18.3	22.9	27.6	32.1	35.9	38.6	39.9
Estimated age	22.0	28.7	35.9	43.3	50.3	56.3	60.6	62.8
Upper 90% endpoint	34.7	45.0	56.4	68.1	79.1	88.4	95.1	98.9

Table 1: The middle row shows the estimate $\hat{Y}$ of Y^*, when the auricular surface measurement X^* takes each of the values 1 to 8. The other two rows give the endpoints of a 90% prediction interval for Y^*. Endpoints for a 68% interval are given in Table 4.

Pubic symphysis	1	2	3	4	5	6
Lower 90% endpoint	13.6	18.1	23.2	29.0	34.5	40.9
Estimated age	20.9	27.8	35.7	44.5	53.8	62.9
Upper 90% endpoint	32.3	42.7	55.0	68.5	82.7	96.8

Table 2: The middle row shows the estimate $\hat{Y}$ of Y^*, when the pubic symphysis measurement X^* takes each of the values 1 to 6. The other two rows give the endpoints of a 90% prediction interval for Y^*. Endpoints for a 68% interval are given in Table 5.

Lower 90% endpoint		Auricular surface							
		1	2	3	4	5	6	7	8
Pubic symphysis	1	13.3	16.3	18.8	20.6	21.6	-	-	-
	2	15.5	18.5	21.0	23.0	24.3	-	-	-
	3	18.1	21.0	23.5	25.7	27.2	28.0	-	-
	4	20.9	23.7	26.3	28.5	30.5	31.9	32.8	-
	5	24.1	26.8	29.3	31.7	34.0	36.4	38.6	40.7
	6	27.7	30.1	32.5	35.1	38.0	41.4	45.2	49.6

Estimated age		Auricular surface							
		1	2	3	4	5	6	7	8
Pubic symphysis	1	19.9	24.2	28.2	31.0	32.5	-	-	-
	2	23.2	27.5	31.4	34.5	36.3	-	-	-
	3	27.1	31.2	35.1	38.3	40.6	41.8	-	-
	4	31.6	35.4	39.1	42.5	45.3	47.8	49.1	-
	5	36.8	40.2	43.7	47.2	50.7	54.1	57.5	60.8
	6	43.0	45.6	48.7	52.3	56.6	61.6	67.4	74.2

Upper 90% endpoint		Auricular surface							
		1	2	3	4	5	6	7	8
Pubic symphysis	1	29.7	36.3	42.3	46.7	48.8	-	-	-
	2	34.6	41.0	46.9	51.5	54.3	-	-	-
	3	40.5	46.6	52.2	57.0	60.5	62.4	-	-
	4	47.6	53.0	58.3	63.2	67.5	70.8	73.3	-
	5	56.2	60.4	65.1	70.2	75.4	80.5	85.6	90.8
	6	66.6	69.0	72.9	78.1	84.4	91.8	100.5	111.2

Table 3: The middle block of entries give estimates $\hat{Y}$ of Y^* when both auricular surface and pubic symphysis measurements are available. The other two blocks give the endpoints of the corresponding lower and upper 90% prediction intervals. Endpoints for a 68% interval are given in Table 6.

Auricular surface	1	2	3	4	5	6	7	8
Lower 68% endpoint	16.7	21.8	27.4	33.0	38.4	42.9	46.1	47.7
Upper 68% endpoint	29.0	37.7	47.2	56.9	66.1	73.9	79.6	82.6

Table 4: Endpoints of a 68% prediction interval for Y^*, for each possible auricular surface measurement.

Pubic symphysis	1	2	3	4	5	6
Lower 68% endpoint	16.1	21.4	27.6	34.4	41.5	48.5
Upper 68% endpoint	27.2	36.0	46.4	57.8	69.7	81.6

Table 5: Endpoints of a 68% prediction interval for Y^*, for each possible pubic symphysis measurement.

Lower 68% endpoint		Auricular surface							
		1	2	3	4	5	6	7	8
Pubic symphysis	1	15.6	19.1	22.1	24.3	25.4	-	-	-
	2	18.2	21.6	24.7	27.0	28.5	-	-	-
	3	21.2	24.6	27.6	30.1	31.9	32.8	-	-
	4	24.6	27.8	30.8	33.4	35.7	37.4	38.5	-
	5	28.5	31.4	34.3	37.1	39.9	42.6	45.2	47.7
	6	32.9	35.5	38.2	41.1	44.5	48.4	53.0	58.2
Upper 68% endpoint		Auricular surface							
		1	2	3	4	5	6	7	8
Pubic symphysis	1	25.3	30.9	36.0	39.7	41.5	-	-	-
	2	29.5	35.0	40.0	43.9	46.3	-	-	-
	3	34.5	39.8	44.6	48.7	51.6	53.3	-	-
	4	40.5	45.2	49.8	54.0	57.6	60.5	62.5	-
	5	47.6	51.4	55.6	60.0	64.4	68.8	73.1	77.4
	6	56.0	58.6	62.2	66.7	72.1	78.4	85.8	94.7

Table 6: Endpoints of a 68% prediction interval for Y^* when both auricular surface and pubic symphysis measurements are available.
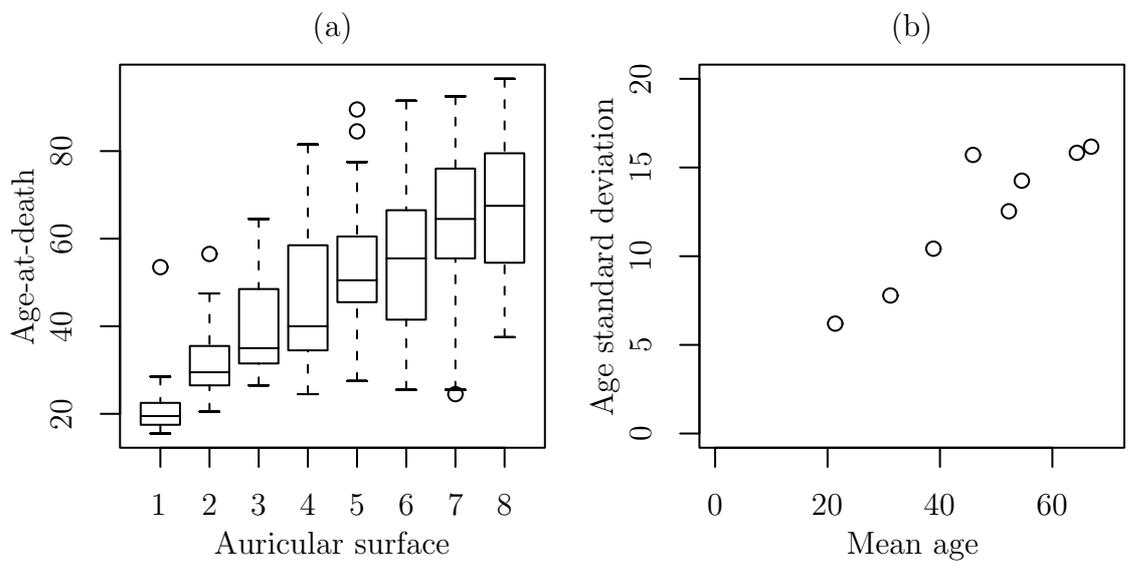
Figure 1: Panel (a) is a box and whisker plot of the ages-at-death within each auricular surface class for the reference sample. Panel (b) is a plot of the mean age-at-death within each class against the corresponding standard deviation.
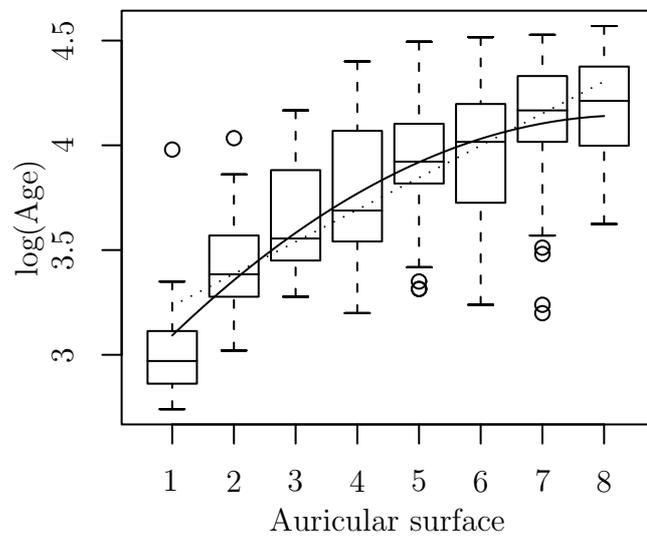
Figure 2: A plot of the logarithm of the ages at death against the auricular surface stage. The dotted straight and solid curved lines are those obtained from fitting the models (3.1) and (3.3) respectively.
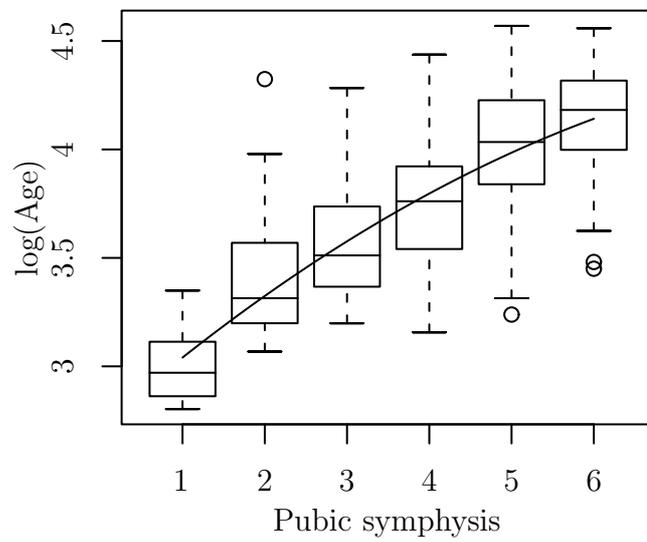
Figure 3: A plot of the logarithm of the ages at death against the pubic symphysis stage. The line is obtained from fitting the model (3.4) respectively.