View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Bradford Scholars

UNIVERSITY of BRADFORD

Library

The University of Bradford Institutional Repository

http://bradscholars.brad.ac.uk

This work is made available online in accordance with publisher policies. Please refer to the repository record for this item and our Policy Document available from the repository home page for further information.

To see the final version of this work please visit the publisher's website. Available access to the published online version may require a subscription.

Link to Publisher's version: *http://dx.doi.org/10.3109/03014460.2015.1046926*

Citation: Buckberry J (2015) The (mis)use of adult age estimates in osteology. Annals of Human Biology, 42 (4): 323-331.

Copyright statement: © 2015 Taylor & Francis. The Version of Record of this manuscript has been published and is available in Annals of Human Biology, 2015, http://dx.doi.org/10.3109/03014460.2015.1046926

The (mis)use of adult age estimates in osteology

Jo Buckberry

Archaeological Sciences, University of Bradford, Bradford, BD7 1DP, UK

j.buckberry@bradford.ac.uk

phone: +44 (0)1274 234289

Keywords: age estimation; age range; mean age; age categories; biological age; social age

Abstract

Context: Adult age-at-death is presented in a number of different ways by anthropologists. Ordinal categories predominate in osteoarchaeology, but do not reflect individual variation in ageing, with too many adults being classified as 'middle adults'. In addition, mean ages (derived from reference samples) are overly-relied upon when developing and testing methods. In both cases, 'age mimicry' is not adequately accounted for.

Objectives: To highlight the many inherent biases created when developing, testing and applying age-estimation methods without fully considering the impact of 'age mimicry' and individual variation.

Methods: The paper draws on previously published research (Web of Science, Pub Med, Google Scholar) on age estimation methods and their use in anthropology.

Results and Conclusions: There is a lack of consistency in the methods used to estimate age, and for the mode of combining them. Ordinal categories are frequently used in osteoarchaeology, whereas forensic anthropologists are more likely to produce case-specific age ranges. Mean ages reflect the age structure of reference samples, and should not be used to estimate age for individuals from populations with a different age-at-death structure. Individual-specific age ranges and/or probability densities should be used to report individual age. Further research should be undertaken on how to create unbiased, combined method age estimates.

Age estimation of adults remains one of the most complex, yet essential, aspects of human skeletal analysis (Aykroyd *et al.* 1997; Bocquet-Appel and Masset 1982; Hoppa and Vaupel 2002a; Konigsberg and Frankenberg 1992; Milner *et al.* 2008). In a society ruled by numbers, we are eager to ascribe a specific age to skeletonised remains, either to facilitate the identification of unidentified remains in a forensic situation, or to better interpret the life, death and burial of archaeological humans. Yet many methods of age estimation, especially for adults but also for non-adults, are woefully inadequate at producing point age estimates. Over the last few decades, research into the development, refinement and testing of adult age-estimation methods has expanded, resulting in a plethora of techniques, and variants of techniques, available for application, with the aim of improving accuracy and precision.

Age estimation in non-adults, where dental and skeletal maturation are employed to estimate chronological age, typically produces estimates that are both accurate and precise. Dental development in particular has been shown to be minimally affected by external factors such as disease or malnutrition (Elamin and Liversidge 2013) and thus less variation in age for a specific stage of development. Once skeletal maturity has been reached, age is typically estimated using degenerative changes of the skeleton, dental wear and (less frequently) microscopic analysis of bone and cementum. All of these, except tooth cementum annulations, have only a broad relationship with chronological age (Gauthier and Schutkowski 2013; Grosskopf and McGlynn 2011; Naji *et al.* early view; Wittwer-Backofen *et al.* 2004). Most skeletal indicators of adult age have been shown to vary between individuals and between populations, reflecting the complexity of the ageing process and the myriad of factors that may influence it (see Mays, this volume and Jackes 2000).

A key requirement for large-scale research into past populations is the ability to combine and compare datasets. Yet it is widely known that researchers collect their data in a variety of ways, using different combinations of methods, making synthesis and comparison difficult (Roberts and Cox 2003). Even where the same basic methods are used, the final estimated age can be defined differently (Garvin and Passalacqua 2012) and is often reported using ordinal categories which also vary by researcher or research group (Falys and Lewis 2011). Surprisingly little attention is given to how age ranges from a suite of methods are combined, and how these resultant age estimates are then consistently placed within age categories. How age estimates are presented to other researchers is discussed even less.

This paper explores how age estimates from skeletonised remains are reported, and the underlying assumptions made by different approaches to age estimation. It will argue that we need to be more transparent in our reporting of age estimates, and advocates the use of individual-specific age estimates over pre-determined age categories.

A Standard Toolkit?

Despite various attempts to standardise methods in osteology (e.g. Brickley and McKinley 2004; Buikstra and Ubelaker 1994; Ferembach *et al.* 1980), individual researchers will commonly select their own preferred suite of methods for age estimation, and will combine resultant age estimates in different ways. These probably partially reflect the methods recommended in 'Standards' publications, and the methods taught within universities (and therefore perhaps the preferences of tutors). Less frequently used methods include recently published methods, perhaps due to decreased awareness of development in the first few years following publication, a lack of subject-specific continuing professional development, and a lower level of independent testing. However, individual researchers who have developed specific methods are more likely to use their own method (Falys and Lewis 2011). For large-scale osteological projects, complex and/or time-consuming methods, and those requiring specialist equipment (including computer software) are often unfeasible; for example, tooth cementum annulation methods require destructive sampling of the material, take up considerable staff time, and initial lab set up costs are high. Two recent surveys of methods used to estimate adult age have revealed interesting patterns in the methods employed by osteoarchaeologists and forensic anthropologists, indicating that while there is wide variation in the methods employed, certain skeletal regions (pelvic joints, ribs, cranial sutures, teeth) are most likely to be used for age estimation.

A review of adult age-estimation techniques used in osteological analysis was undertaken by Falys and Lewis (2011), utilising papers published in three journals between 2004 and 2009. They found that dental attrition, cranial suture closure, and changes to the pubic symphysis and auricular surface were used most frequently to estimate age, with sternal rib ends being utilised less often, perhaps due to poor preservation of ribs in archaeological contexts. However, they found that the specific method used (e.g. Suchey-Brooks (1990) versus Todd (1920; 1921) for the pubic symphysis) varies considerably. It was noticeable that population-specific methods were rarely employed, worryingly so for dental attrition (Falys and Lewis 2011). Continental European researchers tended to follow the recommendations of Ferembach and colleagues (1980), whereas North Americans followed Buikstra and Ubelaker (1994) and Bass (1995). Korean and Japanese researchers were most likely to use population-specific standards (Falys and Lewis 2011). The situation in the UK was not reported, and there was no strong indication in this paper that the BABAO standards (O'Connell 2004) were being extensively followed, but perhaps their publication in 2004 meant they were too recent to have made a huge impact for the period examined.

A similar study of forensic anthropologists, undertaken via a survey of members of the American Academy of Forensic Sciences Physical Anthropology section reveals similar patterns. The pubic symphysis, sternal rib ends and the auricular surface were the preferred skeletal regions for age estimation (Garvin and Passalacqua 2012); unsurprisingly dental wear was ranked as least preferred, probably due to the low levels of dental attrition and thus lower correlation with age in modern populations. For each skeletal region a wide variety of methods were employed, with the most commonly used specific methods the Suchey-Brooks (Brooks and Suchey 1990; Suchey *et al.* 1988) method for the pubic symphysis (95.3% of respondents); the Lovejoy *et al.* (1985b) method for the auricular surface (84.5%); the İşcan *et al.* (İşcan and Loth 1986; İşcan *et al.* 1984; İşcan *et al.* 1985; İşcan *et al.* 1987) method for sternal ribs (89.9%) and the Meindl and Lovejoy (1985) method for cranial suture closure (61.2%, although 38.8% of respondents reported that they typically did not use suture closure). Other methods, including dental wear, medial clavicle fusion, maxillary suture closure, bone histology, tooth cementum annulations, dentine translucency, arthritis and quality of bone were also reported as being used, but less frequently (Garvin and Passalacqua 2012).

Both of these studies indicate that while there is wide variation in practice, generally speaking many researchers tended to examine all or a combination of the pubic symphysis, auricular surface, ribs, cranial suture closure and (in osteoarchaeology) dental wear more frequently that other skeletal regions or methods.

Age Ranges, Precision and Accuracy

The greatest challenge in adult age estimation is the lack of methods which are both accurate and precise; where one of these key attributes is met, it is usually at the expense of the other. Ageing processes are highly variable between individuals; where this is fully recognised in an age-estimation method, the end result is usually a wide, imprecise age range (see Table I for examples). A common feature of these wide ranges is high degree of overlap between adjacent phases; thus a single individual could fall into a number of different phases. Other methods present narrow, precise age ranges. Here the likelihood is that an unknown individual may fall outside the age range given for the stage observed. In some cases there is no overlap between adjacent phases, implying a transformation from one phase to another almost overnight (see Figure 1). While wide age ranges are undoubtedly more accurate – Osborne *et al.* (2004) reported accuracies of 94-100% for each of their auricular surface stages – an age range of 29 to 89 years (Osborne stage 6) is arguably not particularly informative. In contrast, the precise age estimates we seek have much higher levels of inaccuracy when a binary system (falls / does not fall into specified range) is utilised. A comparison of the Suchey-Brooks (1990) pubic symphysis, Lovejoy *et al.* (1985b) auricular surface and Buckberry-Chamberlain (2002) revised auricular surface methods revealed accuracy rates of 71%, 27% and 86% respectively; the low level of accuracy for the Lovejoy *et al.* (1985b) auricular surface method is a direct result of the narrow 5 or 10 year age ranges given in this method, which was developed before 95% or 100% age ranges were advocated (Rissech *et al.* 2012).

TABLE I HERE

FIGURE 2 HERE

The Meaning of Means: Mean Ages, Inaccuracy and Bias

Most age-estimation methods are published with mean ages, standard deviations and either 95% or 100% age ranges for each stage described. It is usually this data that is used both to estimate the age of unknown individuals, and as the basis for tests on different populations. However, one thing that is sometimes overlooked is the inherent relationship between these descriptive statistics and the age distribution of the reference sample. If older adults are well represented in a reference sample, the mean ages for each stage are older than if a younger reference sample was used (Bocquet-Appel and Masset 1982; Jackes 1992). Thus, an age-estimation method developed on an older reference sample will have older mean ages for each stage than the same method would have if it had been developed on a younger reference sample. The age-at-death profile for the reference sample will have an impact on accuracy in two ways. First, if relatively few individuals represent several decades, then little information regarding age-related changes within these decades will be available. Second, if a phase crosses more than one decade, the higher number of individuals in (for example) the older decade will produce a bias, increasing the mean age-at-death for that stage. When tested on a different population, bias and inaccuracy (calculated using formulae 1 and 2, below) will be worse for the decades underrepresented in the reference sample. Logically, inaccuracy and bias will be lower when a method is tested on a population with a similar age-at-death structure to the reference sample.

$$\text{Inaccuracy} = \Sigma \mid \text{estimated} - \text{actual} \mid / \text{N}$$

Formula 1: Inaccuracy. For most tests of age-estimation techniques, 'estimated' age will be the mean age given for a given stage.

$$\text{Bias} = \Sigma\, (\text{estimated} - \text{actual}) / \text{N}$$

Formula 2: Bias. For most tests of age-estimation techniques, 'estimated' age will be the mean age given for a given stage.

This issue is illustrated well by the Buckberry-Chamberlain auricular surface method. The reference sample, Christ Church Spitalfields, has far more older adults than younger adults, and individuals under the age of 25 are particularly underrepresented (see Figure 2). The mean ages at death for the 6 phases (see Table I) are skewed towards the peak of middle to older adults. Due to the small number of young adults in the sample, very few individuals in the reference sample were classified into phases 1 and 2, and the age ranges for these two stages are narrow (see Figure 1b). This may mean that young auricular surfaces are typically similar to each other. However, the individuals present in the Spitalfields sample probably do not reflect the full variation that might have been seen in a population with many young adults; to investigate variation in age indicators in younger adults a different reference sample with many younger adults needs to be used.

Narrow age ranges are often associated with the earlier phases of many different age-estimation methods, even when young adults were well represented in the reference sample; another factor at play is the variation in the rate of ageing seen between individuals within a population. We know that individuals age at different rates (see Mays this volume), but this change is cumulative. With increasing age there has been longer for the difference in ageing rate to manifest in the skeletal regions being studied, thus age ranges for later phases are usually wider; in younger individuals there has been less time for different rates of ageing to make an impact, so the observed variation is likely to be smaller and age ranges are generally narrower.

FIGURE 2 HERE

The bias created by unevenly distributed reference samples has been discussed on many occasions (e.g. Bocquet-Appel and Masset 1982; Hoppa and Vaupel 2002b). The 'Rostock Manifesto' emphasised the need to avoid 'age mimicry' (Hoppa and Vaupel 2002b); one approach is to avoid using skewed reference populations, but rather use one with a uniform distribution of ages to develop methods (Usher 2002, and Konigsberg, this volume), an approach that is seen in many recent age-estimation studies. However, few target populations would have a uniform age-at-death distribution. It could be argued that methods developed on reference samples with a similar age-at-death structure to what is expected for the reference sample should be selected for use (see below).

The bias of the reference sample on the mean ages for stage for the revised auricular surface method were acknowledged in Buckberry and Chamberlain (2002, 235). Bayesian analysis was undertaken to minimise the influence of the reference sample, with posterior probabilities of age, given the auricular surface stage, for uniform priors provided. However, when the method has been tested on different populations, bias and inaccuracy, based on the mean ages derived from, and reflecting the structure of, the Spitalfields population are commonly used (Falys *et al.* 2006; Gocha *et al.* 2015; Hens and Belcastro 2012; Moraitis *et al.* 2014; Mulhern and Jones 2005; Rissech *et al.* 2012; San Millán *et al.* 2013). Unsurprisingly, given the old demography of the Spitalfields reference sample, the vast majority of these studies have found that the Buckberry-Chamberlain method performs better on older individuals (Falys *et al.* 2006; Gocha *et al.* 2015; Hens and Belcastro 2012; Mulhern and Jones 2005; San Millán *et al.* 2013). Indeed, the equal age-at-death distribution used in the study by Mulhern and Jones (2005) presumably will have reduced biases created by the uneven age-at-death distribution in the original paper. The descriptive statistics reported in this study are therefore potentially preferable to those reported in Buckberry and Chamberlain (2002), particularly when calculating inaccuracy and bias.

Combining and Presenting Age Estimates

Once methods have been selected, and age estimates obtained from each method employed, anthropologists have to combine them to give a meaningful overall age estimate for the individual analysed. Indeed, the use of multiple skeletal regions to estimate age has consistently argued to be more accurate than using just one single indicator (e.g. Bedford *et al.* 1993; Gocha *et al.* 2015; Kemkes-Grottenthaler 2001; Lovejoy *et al.* 1985a). Despite the focus on using standard methods, and on testing these for inaccuracy and bias (see above), comparatively little attention has been given to how individual age estimates should be combined, weighted or presented. This issue was highlighted by Garvin and Passalacqua, who noted that subjective 'experience' was used to combine age estimates and produce and appropriate age range more frequently than using a consistent approach utilising age ranges, areas of overlap, mean ages or standard deviations (Garvin and Passalacqua 2012). But we should not belittle the significance of 'experience' – during large scale studies of age estimation it has become apparent that experience-based age estimates may well be more accurate (Milner and Boldsen 2012) and current research is being undertaken to combine age-related data from across the whole skeleton, including anatomical features not typically used for age estimation, to replicate this experience-based approach (Milner and Boldsen 2014; Weise *et al.* 2009).

The ways in which different methods are combined are probably almost as numerous as the range of methods available for age estimation. Many tests of methods have shown that certain methods perform better for certain age groups, however age estimation for the over 50s remains a particular challenge (Milner and Boldsen 2012). While there is an emphasis on being consistent in the use of methods, particularly in post-Daubert era forensic anthropology, it is arguably better to combine methods in a manner best suited to the individual skeleton being analysed, placing more reliance on the skeletal regions which have been shown to be more reliable for the broad age group the skeleton belongs to. A common approach is to utilise late fusing epiphyses to support young age estimates; in these cases adding the broad ranges provided by the pubic symphysis or cranial suture closure are unlikely to usefully contribute to a more precise age estimate, although these would usually have also pointed towards a younger age. The two-step procedure (TSP) is a formalised selective approach which advocates the use of the Suchey-Brooks pubic symphysis and Lamendin dentine translucency methods (Brooks and Suchey 1990; Lamendin *et al.* 1992). Here, if the pubic symphysis (which performs best for younger individuals) is recorded as phases I, II or III, the age estimate is taken from just this element; however if scored as IV, V or VI, the dentine translucency method (which performs better on older individuals) is also applied (Baccino *et al.* 2014). Gocha and colleagues (2015) found that using different auricular surface methods alongside the Suchey-Brooks (Brooks and Suchey 1990) pubic symphysis method provided the most accurate results; they recommended using age estimates from Osborne *et al.* (2004) for individuals with a pubic symphysis in stages I to IV, but the Buckberry-Chamberlain (2002) auricular surface method for individuals with a Suchey-Brooks phase V or VI. Dental wear cannot reliably estimate the age of older adults; in one study, for individuals with dental wear ages of 45+, emphasis was placed on age estimates provided by the pubic symphysis and auricular surface (Gauthier and Schutkowski 2013). Of course a systematic selective approach to combining age estimates assumes that the same age indicators are available for all individuals, which is often not the case due to taphonomic damage (Gauthier and Schutkowski 2013; Kemkes-Grottenthaler 2001; Naji *et al.* early view; Wittwer-Backofen *et al.* 2008), and is much easier when there is a broad level of agreement between different age estimates (Milner and Boldsen 2012).

Use of Mean Ages

Researchers inclined towards precise age estimates may be more likely to favour mean age estimates. A minority of forensic practitioners reported taking an average of mean ages (then ascribing an 'appropriate' range), or using the range of mean ages produced by multiple methods as a final age estimate (Garvin and Passalacqua 2012). The use of mean age data to produce age estimates is problematic, as the mean age for any stage will be dependent on the overall age distribution of a population, an issue widely recognised when dealing with palaeodemographic analysis, but less so for individual age estimates (Bocquet-Appel and Masset 1982; see above). Unaltered mean ages should not be used to estimate age, unless the reference sample used to age an unknown individual has a very similar age-at-death profile to the population the unknown individual came from. As discussed above, this means methods need to be selected carefully based on what is known, or expected, of each individual case or population. Bayesian approaches offer an opportunity to avoid biases inherent in mean ages.

Ordinal Age Categories

Within osteoarchaeology, where age estimates are commonly used to compare populations, groups within populations or to assess patterns of pathology, individuals are often placed within an ordinal age category. Many different systems are in use, from a binary division into 'young' and 'mature' through to ascribing individuals to specific decades (Falys and Lewis 2011). Buikstra and Ubelaker (1994) recommend using three broad categories (see Table II) which are advocated by the BABAO standards (O'Connell 2004). Figure 13 in Buikstra and Ubelaker (1994) provides a helpful indication as to how the phases of different pelvic methods fit into the proposed three adult age categories; for Todd's (1920) pubic symphysis method and the Lovejoy *et al.* (1985b) auricular surface method (which both provide precise age ranges with little to no overlap between phases; see Table I), these correspond to the ordinal categories well, allowing for easy categorisation. For the Suchey-Brooks method, most phases cross between two categories, and phases III, IV and V for females and IV, V and VI for males cross from young adult, span the whole of the middle adult age range, and into the old adult category. Presumably here the mean ages would be used, along with the phases for the other two methods, to place an individual into a specific category. However, as noted above, caution should be applied when using any mean age as these reflect biases inherent in most methods (see also cautionary note by O'Connell 2004).

TABLE II HERE

The drawback with ordinal systems is that by forcing an *individual* (with their own unique set of age indicators) into one of these categories, data about that individual is potentially lost (Milner and Boldsen 2012). The use of late fusing epiphyses to identify young adults is commonplace; however there is an assumption that all young adults have open or partial fusion at some skeletal locations. Individuals who mature early are more likely to be over-aged and identified as middle adults using this approach (see Table III for example). The converse is true at the other end of the spectrum; individuals with final stage pubic symphysis, auricular surfaces and sternal ribs would be identified as 'old adults', but those with middle phase features, falling into the plateau of age-estimation methods, would be more likely to be categorised as middle adults, despite the wide ranges ascribed to phases in some methods (see above). This is often because mean ages for these stages fall within the 'middle adult' category, even though the range is much wider. This issue has been described as the "attraction of the middle" (Masset 1989, 81), whereby the age of young adults are typically overestimated and the age of older adults is typically underestimated, a phenomenon seen in a variety of studies (e.g. Molleson 1993; Saunders *et al.* 1992). Perhaps we need to be more open about what these ordinal groups reflect:

- Young adults: individuals with known youthful features (see Table III); individuals ageing biologically at a slow rate (see Figure 3)

- Old adults: individuals who have aged rapidly and have features usually seen in individuals older than 45, or 50, years
- Middle adults: everyone else. These individuals may fall within the middle decades, but this group will also include younger adults who have reached a level of maturity quickly (see Table III), or older adults who are ageing biologically at a slower rate (see Figure 3).

TABLE III HERE

FIGURE 3 HERE

Of course certain age methods (with more precise age ranges) make it far easier to place an individual skeleton into one of these ordinal categories. Personal experience and analysis of many skeletal recording forms completed by different researchers at the University of Bradford has shown that the age estimates from the Brothwell (1981) dental wear or from the Lovejoy *et al.* (1985b) auricular surface methods often correspond to the young / middle / older age categories an individual is placed in, suggesting the results of these precise, yet less accurate, methods have dominated age category selection. Given the inaccuracy of the Lovejoy auricular surface method (Osborne *et al.* 2004) and the indiscriminate use of uncalibrated dental wear methods (Falys and Lewis 2011), this is extremely worrying. The results of other methods typically cross category boundaries, yet this potential variability is rarely acknowledged in the resultant age range for a specific individual.

Despite this critique, ordinal systems do have some benefits. The nature of the ordinal categories makes them useful vectors for expressing biological age, particularly for age-related processes where we understand the underlying mechanisms and provide a simple method for comparing population age structure. Dental wear is an ideal method for creating ordinal age groups (Mays 2010), and, in populations with high wear rates, calibrated methods have been shown to have a high correlation with chronological age (Mays, this volume), but have the limitation of not being able to distinguish between older adults or to identify the oldest old. The Brothwell (1981) method, for example, has a final age category of 'about 45+'. Including evidence of ante-mortem tooth loss and the degree of alveolar resorption offers potential here (Mays 2014; Miles 2001), but if an individual's teeth were lost due to an exceptional pathological process this would result in over-ageing. Ordinal categories may also reflect some aspects of social age. The direct association between skeletal ageing and other biological ageing process has not been fully investigated, however, individuals who are skeletally immature may also appear youthful in life; it is possible that those who achieve end-phase pubic symphysis and auricular surfaces first may also be seeing and feeling the effects of increasing age in different ways, for example with grey hair and aching joints. Of course this is conjecture, but perhaps a more nuanced investigation in the ageing process, and not just age estimation, would be a fruitful area of research.

Overall, pre-determined ordinal age groups do not fully reflect the range of variation seen in the ageing process and are inadequate representations of estimated individual age. Simple comparisons of age structure can be made using them, but reducing age variation to three or four age categories can mask age-related patterns; most demographic studies would divide populations into five or ten year age groups before comparing populations.

Age Ranges and Probability Densities

Using the full age ranges provided for any given method is more likely to reflect the potential variability seen between individuals, particularly for methods where outliers were not removed. In these cases examining the full range obtained would give a wide, inclusive age range, whereas reporting the area of overlap (assuming there was an area of overlap for all methods employed) presents the area of consensus. Thus age estimates could be presented as 'most likely to be' and 'would not exclude' age ranges (as advocated by Lynnerup *et al.* 2008, e.5). In this case, the age range for each individual skeleton is likely to be different, but the age estimate given will better reflect the age and biological status of that individual. While this is an ideal approach for presenting the age of an individual, these age ranges will not easily permit the comparison of populations. We need to develop methods of combining multiple age estimates from a single population using these bespoke age ranges, by placing proportions of individuals into different (and ideally narrow) categories.

In palaeodemographic studies, the overrepresentation of middle adults can be overcome by using probability densities of age for each stage, and proportionally distributing the number of individuals within each stage across the different ages (Chamberlain 2000; Chamberlain 2006). Bayesian approaches allow prior probabilities of age to be incorporated into models, allowing for more nuanced investigation of past populations and comparison of mortality profiles (e.g. Gowland and Chamberlain 2005; Nagaoka and Hirata 2008; Redfern and Chamberlain 2011; Storey 2007). The same approach can be used for individual age estimates (Boldsen *et al.* 2002). By presenting age as a probability density, the full range of possible ages is represented, but the age range where an individual is most likely to fall is highlighted, with maximum likelihood being used to give a point estimate. This approach is exemplified by the ADBOU program for transition analysis, where probability densities for individual skeletal regions (cranial sutures, pubic symphysis, auricular surface) are shown alongside densities for the traits combined using a uniform prior, and for an informative prior (Boldsen *et al.* 2002). Unfortunately these probability density graphs are not easily calculated by hand and are not readily available for many 'standard' methods, but probability matrices of age, given stage, have been published (e.g. Buckberry and Chamberlain 2002; Chamberlain 2000; Gowland and Chamberlain 2005). For this approach to be adopted more widely, computer interfaces allowing for the creation of probability densities which do not require each user to undertake the complex mathematics need to be developed. Ideally these must include probability densities for popular and traditional methods, to facilitate reanalysis of data collected over the last 20-30 years (especially where skeletal material has been reburied) and to allow direct comparisons with newly analysed material. An easy-to-use method with free interface and accessible training is essential if we are to progress age estimates and palaeodemography widely within osteoarchaeology.

Conclusion

This paper has discussed how age is commonly estimated for skeletonised adult human remains, highlighting a reluctance to adopt new methods and a lack of a standardised methodology both in terms of methods applied and how different age estimates are combined. The tension between accuracy and precision remains an issue. Mean age estimates are a direct reflection of the age structure of the reference sample, and should be avoided unless the target population has a similar age-at-death structure. Equally, the calculation of inaccuracy and bias from these mean ages needs to be undertaken with care. Ordinal age groups are frequently used in osteoarchaeology and offer a reasonable mechanism for comparing populations and for investigating biological and potentially social age, but in many cases may not reflect the chronological age of an individual, particularly for individuals categorised as 'middle adults'. The use of combined, individual age ranges (e.g. most likely to be 29 to 57 years, but would not exclude 21 to 88), as per Lynnerup *et al.* (2008, e.5) is advocated, to ensure the data presented is accurate and transparent. We should strive towards developing easy-to-use interfaces allowing probability densities of age to be produced for a combination of 'standard' methods at a skeleton-by-skeleton level, in order to better represent the variation in age-related change seen within an individual and to create robust demographic profiles.

Acknowledgements

I thank the following people for many delightful and thoughtful discussions about age estimation: Sonia Zakrzewski, George Milner, Svenja Weise, Niels Lynnerup, Chiara Villa, Andrew Chamberlain, and the Lauchheim study group. Julia Beaumont and Hannah Koon kindly discussed aspects of this paper with me. I am indebted to students on the MSc Human Osteology and Palaeopathology and PhD students at the University of Bradford, who have challenged my perceptions of age estimation and inspired me to write this paper. I also thank the anonymous reviews for their insightful comments on an earlier draft of this paper. Any errors remain my own.

References

Aykroyd RG, Lucy D, Pollard AM, Solheim T. 1997. Regression analysis in adult age estimation. *American Journal of Physical Anthropology* 104:259-265.

Baccino E, Sinfield L, Colomb S, Baum TP, Martrille L. 2014. Technical note: The two step procedure (TSP) for the determination of age at death of adult human remains in forensic cases. *Forensic Science International* 244:247-251.

Bass WM. 1995. *Human osteology: a laboratory and field manual*. Colombia: Missouri Archaeological Society.

Bedford ME, Russell KF, Lovejoy CO, Meindl RS, Simpson SW, Stuart-MacAdam PL. 1993. Test of the multifactorial aging method using skeletons with known ages-at-death fron the Grant collection. *American Journal of Physical Anthropology* 91:287-297.

Bocquet-Appel JP, Masset C. 1982. Farewell to palaeodemography. *Journal of Human Evolution* 11:321-333.

Boldsen JL, Milner GR, Konigsberg LW, Wood JW. 2002. Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW, editors. *Palaeodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press. p 73-106.

Brickley M, McKinley JI, eds. 2004. *Guidelines to the standards for recording human remains*: BABAO/IFA.

Brooks ST, Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Ascádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution* 5:227-238.

Brothwell DR. 1981. *Digging up bones*. Oxford: Oxforn University Press.

Buckberry JL, Chamberlain AT. 2002. Age estimation from the auricular surface of the ilium: a revised method. *American Journal of Physical Anthropology* 119:231-239.

Buikstra J, Ubelaker DH. 1994. *Standards for data collection from human skeletal remains*. Arkansas: Arkansas Archeological Survey.

Chamberlain AT. 2000. Problems and prospects in palaeodemography. In: Cox M, Mays S, editors. *Human osteology in archaeology and forensic medicine*. London: Greenwich Medical Media Ltd. p 101-116.

Chamberlain AT. 2006. *Demography in archaeology*. Cambridge: Cambridge University Press.

Elamin F, Liversidge HM. 2013. Malnutrition has no effect on the timing of human tooth formation. *PloS one* 8(8):e72274.

Falys CG, Lewis ME. 2011. Proposing a way forward: a review of standardisation in the use of age categories and ageing techniques in osteological analyses (2004-2009). *International Journal of Osteoarchaeology* 21:704-716.

Falys CG, Schutkowski H, Weston D. 2006. Auricular surface aging: worse than expected? A test of the revised method on a documented historic skeletal sample. *American Journal of Physical Anthropology*.

Ferembach D, Schwidetzky I, Stoukal M. 1980. Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution* 9:517-549.

Garvin HM, Passalacqua NV. 2012. Current practices by forensic anthropologists in adult skeletal age estimation. *Journal of Forensic Sciences* 57(2):427-433.

Gauthier J, Schutkowski H. 2013. Assessing the application of tooth cementum annulation relative to macroscopic aging techniques in an archeological sample. *HOMO - Journal of Comparative Human Biology* 64:42-57.

Gocha TP, Ingvoldstad ME, Kolatorowicz A, Cosgriff-Hernandez M-TJ, Sciulli PW. 2015. Testing the applicability of six macroscopic skeletal aging techniques on a modern Southeast Asian sample. *Forensic Science International* doi: 10.1016/j.forsciint.2014.12.015.

Gowland RL, Chamberlain AT. 2005. Detecting plague: palaeodemographic characterisation of a catastrophic death assemblage. *Antiquity* 79:146-157.

Grosskopf B, McGlynn G. 2011. Age diagnosis based on incremental lines in dental cementum: A critical reflection. *Anthropologischer Anzeiger* 68(3):275-289.

Hens SM, Belcastro MG. 2012. Auricular surface aging: a blind test of the revised method on historic Italians from Sardinia. *Forensic Science International* 214:209.e201-209.e205.

Hoppa RD, Vaupel J, eds. 2002a. *Palaeodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press.

Hoppa RD, Vaupel JW. 2002b. The Rostock Manifesto for paleodemography: The way from stage to age. In: Hoppa RD, Vaupel JW, editors. *Palaeodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press. p 1-8.

İşcan MY, Loth SR. 1986. Determination of age from the sternal rib in white females: a test of the phase model. *Journal of Forensic Sciences* 31:990-999.

İşcan MY, Loth SR, Wright RK. 1984. Age estimation from the rib by phase analysis: white males. *Journal of Forensic Sciences* 29:1094-1104.

İşcan MY, Loth SR, Wright RK. 1985. Age estimation from the rib by phase analysis: white females. *Journal of Forensic Sciences* 30:853-863.

İşcan MY, Loth SR, Wright RK. 1987. Racial variation in the sternal extremity of the rib and its effect on age determination. *Journal of Forensic Sciences* 32:452-466.

Jackes M. 1992. Palaeodemography: problems and techniques. In: Saunders SR, Katzenberg MA, editors. *Skeletal biology of past peoples: research methods*. New York: Wiley Liss Inc. p 189-224.

Jackes MK. 2000. Building the bases for paleodemographic analysis: adult age determination. In: Katzenberg MA, Saunders SD, editors. *Biological Anthropology of the Human Skeleton*. New York: Wiley-Liss. p 417-466.

Kemkes-Grottenthaler A. 2001. The reliability of forensic osteology - a case in point. *Forensic Science International* 117:65-72.

Konigsberg LW, Frankenberg SR. 1992. Estimation of age structure in anthropological demography. *American Journal of Physical Anthropology* 89:235-256.

Lamendin H, Baccino E, Humbert JF, Tavernier JC, Zerilli A. 1992. A simple technique for age estimation in adult corpses: the two criteria dental method. *Journal of Forensic Sciences* 37:1373-1379.

Lovejoy CO, Meindl RS, Mensforth RP, Barton TJ. 1985a. Multifactorial determination of skeletal age at death: a method and blind tests of its accuracy. *American Journal of Physical Anthropology* 68:1-14.

Lovejoy CO, Meindl RS, Pryzbeck TR, Mensforth RP. 1985b. Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology* 68:15-28.

Lynnerup N, Belard E, Buch-Olsen K, Sejrsen B, Damgaard-Pedersen K. 2008. Intra- and interobserver error of the Greulich–Pyle method as used on a Danish forensic sample. *Forensic Science International* 179: 242.e241–242.e246.

Masset C. 1989. Age estimation on the basis of cranial sutures. In: İşcan MY, editor. *Age markers in the human skeleton*. Springfield: Charles C. Thomas. p 71-103.

Mays S. 2010. *The archaeology of human bones*. London: Routledge.

Mays S. 2014. Resorption of mandibular alveolar bone following loss of molar teeth and its relationship to age at death in a human skeletal population. *American Journal of Physical Anthropology* 153(4):643-652.

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. *American Journal of Physical Anthropology* 68:57-66.

Miles AEW. 2001. The Miles method of assessing age from tooth wear revisited. *Journal of Archaeological Science* 28:973-982.

Milner GR, Boldsen JL. 2012. Transition analysis: a validation study with known-age modern American skeletons. *American Journal of Physical Anthropology* 148:98-110.

Milner GR, Boldsen JL. 2014. Evaluating old and new skeletal age indicators – which ages of transition pass the "box test"? [abstract]. *American Journal of Physical Anthropology* S58:186.

Milner GR, Wood JW, Boldsen JL. 2008. Advances in paleodemography. In: Katzenberg MA, Saunders SR, editors. *Biological anthropology of the human skeleton*. New York: Wiley-Liss. p 561-600.

Molleson T. 1993. Growing old. In: Molleson T, Cox M, editors. *The Spitalfields project volume 2: the anthropology The middling sort*. York: Council for British Archaeology.

Moraitis K, Zorba E, Eliopoulos C, Fox SC. 2014. A test of the revised auricular surface aging method on a modern European population. *Journal of Forensic Sciences* 59(1):188-194.

Mulhern DB, Jones EB. 2005. Test of revised method of age estimation from the auricular surface of the ilium. *American Journal of Physical Anthropology* 126:61-65.

Nagaoka T, Hirata K. 2008. Demographic structure of skeletal populations in historic Japan: a new estimation of adult age-at-death distributions based on the auricular surface of the ilium. *Journal of Archaeological Science* 35:1370-1377.

Naji S, Colard T, Blondiaux J, Bertrand B, d'Incau E, Bocquet-Appel J-P. early view. Cementochronology, to cut or not to cut? *International Journal of Paleopathology*:doi.org/10.1016/j.ijpp.2014.1005.1003.

O'Connell L. 2004. Guidance on recording age at death in adults. In: Brickley M, McKinley JI, editors. *Guidelines to the standards for recording human remains*. Southampton: BABAO and the IFA. p 14-17.

Osborne DL, Simmons TL, Nawrocki SP. 2004. Reconsidering the auricular surface as an indcator of age at death. *Journal of Forensic Sciences* 49(5):905-911.

Redfern RC, Chamberlain AT. 2011. A demographic analysis of Maiden Castle hillfort: Evidence for conflict in the late Iron Age and early Roman period. *Journal of Paleopathology* 1(1):68-73.

Rissech C, Wilson J, Winburn AP, Turbón D, Steadman D. 2012. A comparison of three established age estimation methods on an adult Spanish sample. *International Journal of Legal Medicine* 126:145-155.

Roberts CA, Cox M. 2003. *Health and disease in Britain from Prehistory to the present day*. Stroud: Sutton.

San Millán M, Rissech C, Turbón D. 2013. A test of Suchey-Brooks (pubic symphysis) and Buckberry-Chamberlain (auricular surface) methods on an identified Spanish sample: paleodemographic implications. *Journal of Archaeological Science* 40:1743-1751.

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 Sciences Journal* 25:97-118.

Storey R. 2007. An elusive palaeodemography? A comparison of two methods for estimating the adult age distribution of deaths at Late Classic Copan, Honduras. *American Journal of Physical Anthropology* 132:40-47.

Suchey JM, Brooks ST, Katz D. 1988. *Instructions for use of the Suchey-Brooks system for age determination of the female os pubis. Instructional materials accompanying female pubic symphyseal models of the Suchey-Brooks system*. Colorado: Distributed by France Casting.

Todd TW. 1920. Age changes in the pubic bone I: the male white pubis. *American Journal of Physical Anthropology* 3:285-334.

Todd TW. 1921. Age changes in the pubic bone. II: The pubic of the male negro-white hybrid. III: The pubis of the white female. IV: The pubis of the female white-negro hybrid. *American Journal of Physical Anthropology* 4:1-70.

Usher BM. 2002. Reference samples: the first step in linking biology and age in the human skeleton. In: Hoppa RD, Vaupel JW, editors. *Palaeodemography: age distributions from skeletal samples*. Cambridge: Cambridge University Press. p 29-47.

Weise S, Boldsen JL, Gampe J, Milner GR. 2009. Calibrated expert inference and the construction of unbiased paleodemographic mortality profiles [abstract]. *American Journal of Physical Anthropology* S48:269.

Wittwer-Backofen U, Buckberry J, Czarnetzki A, Doppler S, Grupe G, Hotz G, Kemkes A, Larsen CS, Prince D, Wahl J and others. 2008. Basics in Paleodemography: A Comparison of Age Indicators Applied to the Early Medieval Skeletal Sample of Lauchheim. *American Journal of Physical Anthropology*.

Wittwer-Backofen U, Gampe J, Vaupel JW. 2004. Tooth cementum annulation for age estimation: Results from a large known-age validation study. *American Journal of Physical Anthropology* 123:119-129.

Figure captions:

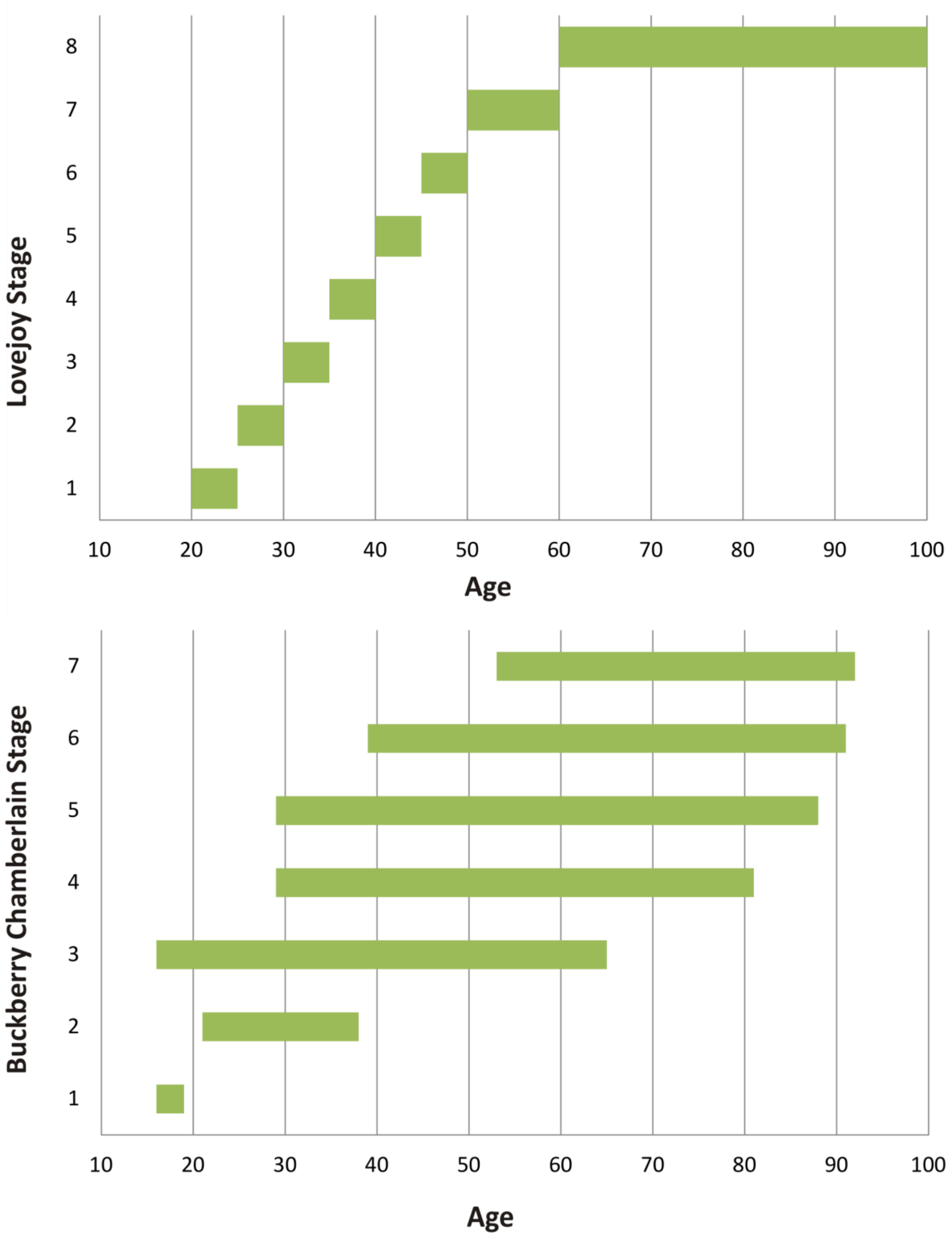

Figure 1: Comparison of age ranges reported by the a) Lovejoy *et al.* (1985b) and b) Buckberry-Chamberlain (2002) auricular surface methods. The Lovejoy *et al.* system implies a rapid transformation between adjacent stages, with no overlap in age ranges. There is a high degree of overlap for the age ranges between adjacent phases for the Buckberry-Chamberlain method, and an individual of a certain age could fall into several different stages; the resultant age ranges are, arguably, too wide to be of use in isolation, particularly for the middle phases. The narrow age ranges for stages 1 and 2 may reflect the low number of younger individuals in the Christ Church Spitalfields sample.

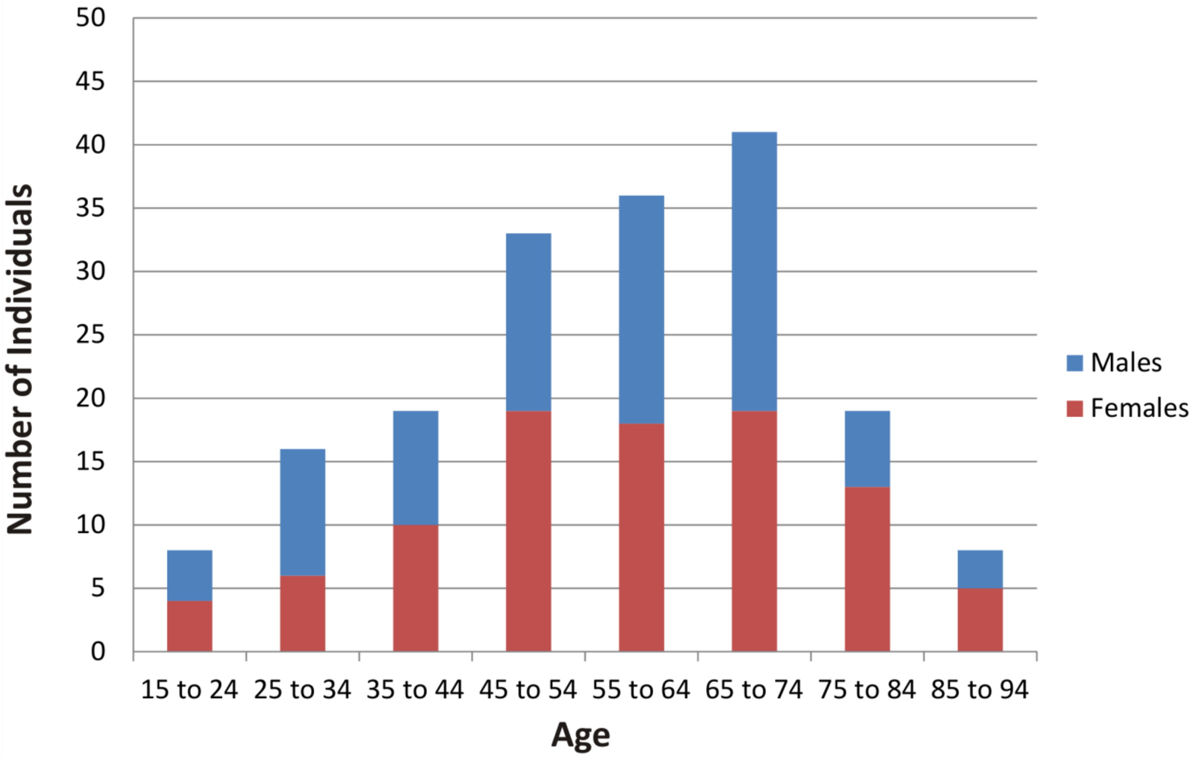

Figure 2: Demography of the Christ Church Spitalfields sample utilised to develop the Buckberry-Chamberlain auricular surface method.

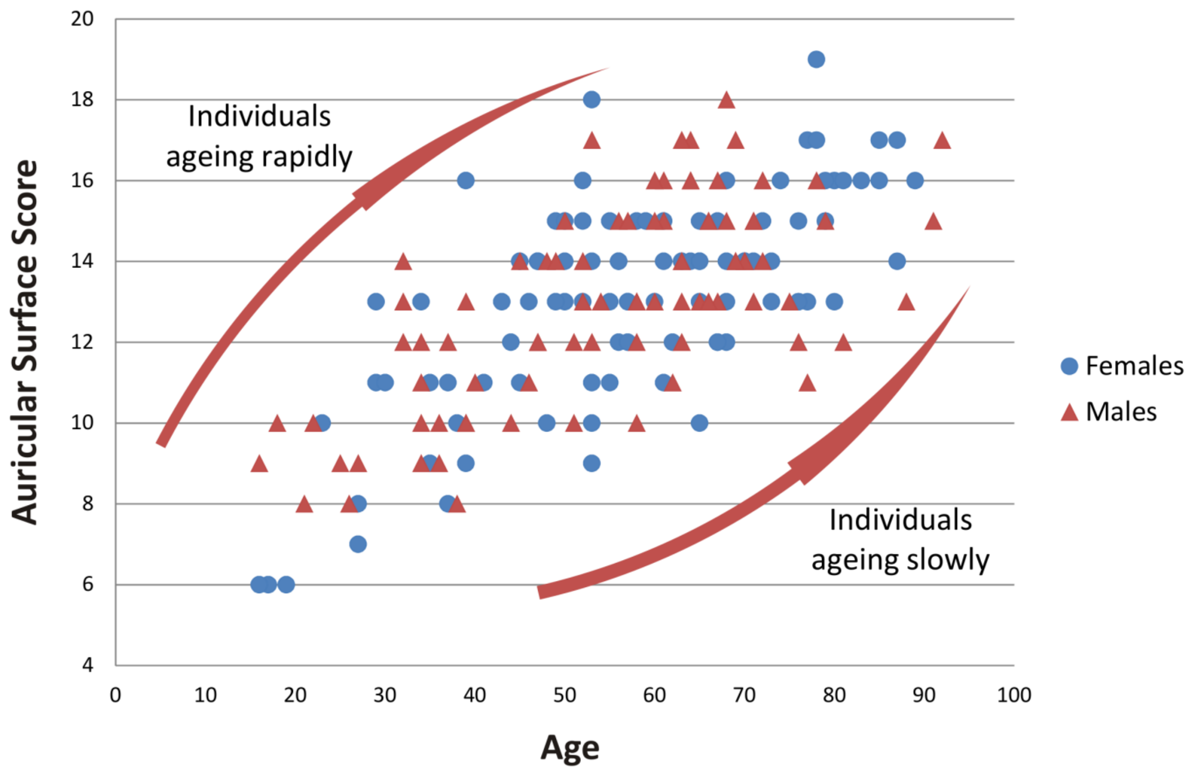

Figure 3: Scatter plot of age against composite score for the Buckberry-Chamberlain auricular surface method. The range of ages for each score probably reflects differences in the rate of aging for different individuals within the population.

Phase	Todd pubic symphysis	Suchey-Brooks pubic symphysis (females)	Suchey-Brooks pubic symphysis (males)	Lovejoy *et al* auricular surface	Buckberry-Chamberlain auricular surface	Osborne *et al* auricular surface
1	18-19	15-24	15-23	20-24	16-19	≤27
2	20-21	19-40	19-34	25-29	21-38	≤46
3	22-24	21-53	21-46	30-34	16-65	≤69
4	25-26	26-70	23-57	35-39	29-81	20-75
5	27-30	25-83	27-66	40-44	29-88	24-82
6	30-35	42-87	34-86	45-49	39-91	29-89
7	35-39			50-59	53-92	
8	40-45			60+		
9	45-49					
10	50+					

Table I: Age ranges reported for stages for some pelvic methods of age estimation. The earlier publications on specific skeletal regions (Lovejoy *et al.* 1985b; Todd 1920) provide narrower age ranges than more recent revisions which typically provide wide age ranges, which may be more reflective of human variation. Age ranges taken from Todd (1920), Brooks and Suchey (1990), Lovejoy *et al* (1985b) Buckberry and Chamberlain (2002) and Osborne *et al.* (2004). The stages do not correspond between different methods.

Age category	Age range (years)
Young adult	20-34
Middle adult	35-49
Old adult	50+

Table II: Age categories suggested by Buikstra and Ubelaker (1994)

Stages of union	Age data (males)	Age data (females)	Probable interpretation
Nonunion	≤25 years	≤23 years	Non-adult (when found with other areas of non-union or incomplete third molars) or young adult
Partial union	17-30 years	17-30 years	Young adult? But some may be as old as 30 years, which often falls within a 'middle adult' category in an ordinal system
Union	≥ 21 years	≥ 21 years	Middle or older adults (other indicators will probably be used to assess age). Some of these individuals may be as young as 22 years

Table III: Age-related data for the sternal end of the clavicle from Webb and Suchey (1985). Final column indicates how these stages may be interpreted in a three-category ordinal system (young adult, 17-25; middle adult, 26-45; old adult, 46+). Interpretation would be different if other ordinal categories were used, such as those given in Table II.