Testing the Sternal Clavicle Ageing Method on a Post-Medieval Dutch Skeletal Collection

Blom, A.A.; Inskip, S.A.; Baetsen, W.A.; Hoogland, M.L.P.

Citation

Blom, A. A., Inskip, S. A., Baetsen, W. A., & Hoogland, M. L. P. (2018). Testing the Sternal Clavicle Ageing Method on a Post-Medieval Dutch Skeletal Collection. *Archaeometry*, *60*(6), 1391-1402. doi:10.1111/arcm.12402 archaeometry

Archaeometry **60**, 6 (2018) 1391–1402 doi: 10.1111/arcm.12402

TESTING THE STERNAL CLAVICLE AGEING METHOD ON A POST-MEDIEVAL DUTCH SKELETAL COLLECTION*

A. A. BLOM,[1]† S. A. INSKIP,[2] W. A. BAETSEN[1] and M. L. P. HOOGLAND[1]
[1]*Laboratory for Human Osteoarchaeology and Funerary Archaeology, Faculty of Archaeology, Leiden University, Einsteinweg 2, NL -2333 CCLeiden, The Netherlands*
[2]*McDonald Institute for Archaeological Research, University of Cambridge, Downing Street, Cambridge CB2 3ER, UK*

When a new method is produced, it should be assessed for accuracy and reliability before widespread use. A new method by Falys and Prangle for estimating age from degeneration of the sternal end of the clavicle is tested in this paper on an 18th–19th-century known-age Dutch population (80 clavicles). Intra- and inter-observer tests (Cohen's kappa) revealed slight to moderate agreement (k = 0.100–0.425 and 0.106–0.534 respectively). However, owing to large age ranges, 87% accuracy was achieved. Methodological improvements are suggested that could help the method further to improve age estimates for the elderly in past populations.

KEYWORDS: AGEING METHOD, TESTING, CLAVICLE, OLD ADULTS, POST-MEDIEVAL, MIDDENBEEMSTER (THE NETHERLANDS)

INTRODUCTION

Archaeology aims to reconstruct and understand past populations. When attempting to do so, it is essential to have an understanding of the characteristics of individuals in order to build up a picture of the wider aspects of a society as a whole (Cox 2000). In order to achieve this, biological profiles need to be established, requiring an assessment of, if possible, age-at-death, sex, stature, pathology and sometimes ancestry (White and Folkens 2005, 1; DiGangi and Moore 2013, 18). However, many age-estimation methods have been the subject of discussion in recent years over their usefulness, accuracy and ability to estimate age in older adults. For example, important skeletal elements used for age estimation, including the pubic symphysis (Todd 1920, 288) and sternal rib end (İşcan *et al*. 1984, 155; Nikita 2013, 324), are often damaged or missing in archaeological samples. There are also issues of high intra-observer error for some methods (e.g. ectocranial suture closure; Key *et al*. 1994, 195–6; Dorandeu *et al*. 2008, 50), and dental wear depends on a uniform diet in a population, and is of limited use on younger individuals who retain teeth (Mays 2002, 861). An especially significant problem has been the poor correlation between true age and ageing traits due to varying personal life histories of analysed skeletons (Buckberry and Chamberlain 2002, 231). This variation in the pace of degeneration results in the use of wide age ranges to maintain method accuracy. This problem is especially acute for individuals dying after their fifth decade of life, and has meant that within the field generally the uppermost age that can be given in 46 or more years. Although all methods have proven to be reliable and accurate to some extent, the nature of archaeological assemblages required methods not only to be based on more durable bones (Cox 2000) but also they should be developed so they can accurately estimate the age at death of individuals aged over 40 years.

---

*Received 7 August 2017; accepted 27 February 2018
†Correspondence: email a.a.blom@umail.leidenuniv.nl

The inability to identify individuals in their sixth decade of life or after not only has made it difficult to assess the relationship between biological and social variables and advancing age, but also has resulted in a skewed perception of old age in the past, with people assuming that living beyond the sixth decade was a rare occurrence. This has resulted in a phenomenon known as the 'invisible elderly' (Cave and Oxenham 2014). However, we know from historical sources and funerary inscriptions, particularly in Classical, Egyptian and other societies, that people did reach these ages.

In an attempt to improve the resolution of current age categories used in bioarchaeology, and to move beyond using the 46 or more years as the uppermost limit, Falys and Prangle (2015) presented a method that estimates age based on degenerative changes that take place at the sternal end of the clavicle. The rationale behind this is that the clavicle is one of the last elements to complete development, and therefore starts degenerating slightly later than other areas. Using the known-age-at-death samples from the Hamann-Todd, Pretoria, St. Bride's and Coimbra collections, this method scores surface topography, porosity (micro and macro) and osteophyte formation and estimates age-at-death based on a composite score. In addition, from the same scores, regression models were established based on linear regression analyses of trait scores to estimate age-at-death. A test of this method on the Church Christ Spitalfield Collection revealed a 96.4% accuracy for both approaches. In addition, for the composite score approach, the uppermost age categories (4 and 5), had lower limits which were approximately 12 years or more in advance of 46 years, demonstrating the potential to identify individuals who are more advanced in years.

While this shows promise, the authors highlighted a number of issues, including reproducibility, user ease and description clarity, that needed further attention before widespread use of the method (Falys and Prangle 2015). Additionally, a method always requires testing for objectivity, reproducibility and to check whether prior knowledge is needed to apply it properly (Ubelaker 2008, 50). In addition, as genetic and environmental conditions can greatly influence the biological aspects of ageing, it is possible for populations to have distinct ageing patterns. As such, the method should be tested for accuracy on geographically and temporally diverse known-age individuals.

The main aim of this study is to test the method developed by Falys and Prangle (henceforth called the 'clavicle method') to assess its accuracy and reproducibility. This is achieved by applying the method to a collection with known-age-at-death individuals from a late 19th-century skeletal collection from the rural settlement of Middenbeemster, The Netherlands. To assess reproducibility, intra- and inter-observer tests by individuals with differing levels of experience in the field are performed and assessed; and to assess accuracy, the estimated age stages using the method are compared with the known age-at-death of those individuals. This will indicate whether the method can be applied to Dutch post-medieval skeletal material, and help test whether the method is more widely applicable to material outside those from which the method was originally developed and tested. In addition, if the method is easy to use and provides reliable results, it could be applied to skeletal material of unknown age.

MATERIALS AND METHODS

Middenbeemster is a rural village in the Beemster—a polder reclaimed from the sea several centuries ago—in north Holland, The Netherlands (Griffioen 2011, 11) (Fig. 1). The individuals making up the Middenbeemster collection come from a dairy farmer and labourer community, and the majority of the graves date to AD 1829–66 (Griffioen 2011, 11; Hakvoort 2013, 11).

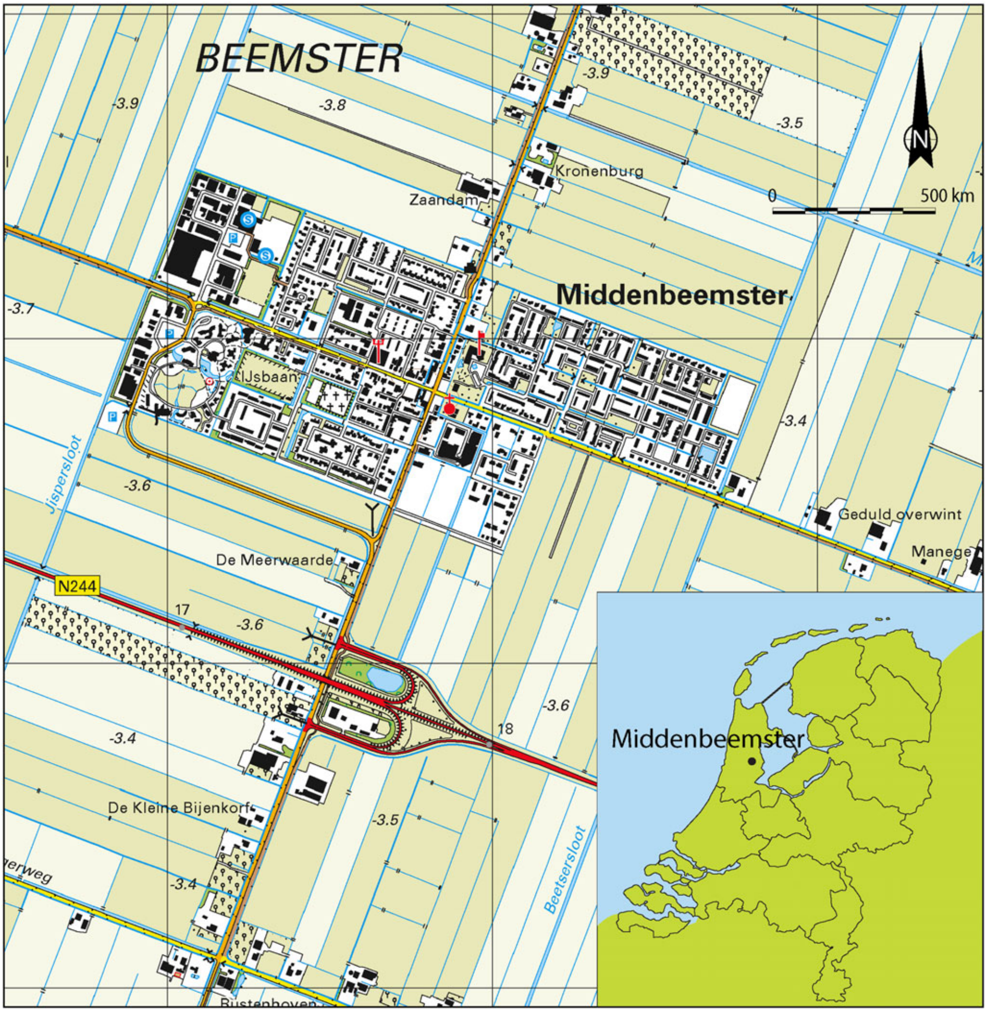

Figure 1 *Location of the excavation (red churchyard symbol †) on a crop of the topographical map of the Netherlands. Figure by Prof. M.L.P. Hoogland. [Colour figure can be viewed at wileyonlinelibrary.com]*

Approximately 450 primary graves were excavated in 2011 (Hakvoort 2013, 35; Waters-Rist and Hoogland 2013, 244).

The archival records consist of two sources: a cemetery ledger and a military document. The ledger contained names, age at death, lineages, marriages and occupations of a majority of the individuals between 1829 and 1866. This information was complemented by post-medieval military archives, detailing height, diseases and other abnormalities for most males (Waters-Rist and Hoogland 2013, 244). Mainly due to good preservation, known information and comparability of Middenbeemster with other Dutch collections, this collection is well suited to test the new clavicle method of Falys and Prangle in order to establish its accuracy for populations in the Low Countries.

A list of known age individuals was drawn up by a person not involved in the testing of the method. As the method should only be applied to individuals over 36 years of age (Falys and

Prangle 2015), adults known to be younger were excluded from the list. From this remaining list, clavicles with damage that obscured scoring to any part of the clavicle were also excluded. In total, 46 individuals were suitable for analysis, amounting to 41 left and 39 right clavicles. Of these 80 clavicles, 34 pairs could be analysed. A double-blind assessment—meaning an individual's age or sex was unknown before or during assessment—was performed to prevent biases towards sex and age.

For intra-observer error testing, a subset of 47 clavicles was scored twice by 'AB', a bachelor's student, with three weeks between observations. The first observation was denoted A_1 and the second A_2. For inter-observer error, values from A_1, who practised the method, were compared with results, denoted W_1, from a master's student who had no prior method experience; and one postdoctoral researcher, denoted S_1, also with no prior experience of the method. All observers obtained their degrees in the field of osteoarchaeology. For the inter-observer error test, 61 of the original clavicles were assessed.

Falys and Prangle (2015) evaluated the degeneration of three distinct features of the sternal end of the clavicle: surface topography, porosity and osteophyte formation. Surface topography is determined by two features: the general texture of the joint surface and the relief (undulation). As the individual ages, the surface becomes increasingly rough, the relief increases in undulation and the outline of the surface becomes increasingly irregular. The surface topography is scored on a scale from 1 (no element present) to 6 (complete surface breakdown). Porosity is also scored from 1 (no porosity present) to 6 (complete surface breakdown). In the clavicle method, a distinction is made between microporosity, in which the pores are less than 1 mm in diameter, and macroporosity, in which the pores are larger than 1 mm in diameter (Falys and Prangle 2015, 205). The scoring records the amount of surface that is subject to porosity rather than the amount of porosity present. A score of 6 entails both the amount of porosity present and the overlap of macropores leading to an undulated surface. With old age, the loss of bone is counteracted by depositing new bone, which is not laid down in neatly organized layers, but rather as 'bony outgrowths' especially on the edges of the joint. These bony outgrows are termed osteophytes. The formation of this trait is scored on a scale from 1 (no osteophytes present) to 4 (severe osteophyte formation). The presence of osteophytes is scored based on the length of the lipping (a consideration, not measurement) and on the amount of rim that shows outgrows. These features are scored individually using pictures and tables, after which the three feature scores are added to form a composite score. This composite score is used to ascribe an age stage to an individual, which is then compared with the known age at death obtained from the archives.

To test the intra- and inter-observer errors, and ultimately the reproducibility of the method, Cohen's kappa coefficient was calculated using IBM SPSS Statistics 23. Spearman's rank correlations were calculated to explore the relationship between male and female scores, left and right scores, and to quantify the relationship between surface topography, porosity and osteophyte formation with documented ages at death. Additionally, a distinction was made between the similarity in the assigned composite score and the similarity in the age stage to which the individual was assigned. A paired *t*-test was performed to assess whether there was any asymmetry in scores between left and right elements that may relate to handedness. For all tests, statistical significance was set at $p = 0.05$.

RESULTS

The intra- and inter-observer errors of the method were tested based on the agreement of assigned scores by the different observations/observers, and the age stages that go with these scores.

Table 1 shows Cohen's kappa's for both the intra- and inter-observer tests. When age stage is considered, the level of agreement for all tests—both intra- and inter-observer—is better than the results for the composite score alone. However, the strength of agreement remains fairly low, ranging from fair to moderate for the age stages, and from poor to substantial for individual features. The results in Table 1 imply that the method cannot be deemed reproducible in terms of scores, not by one observer in different attempts, and not between different observers. Additionally, there appears to be no significant difference between the accuracy of the method using the left or right clavicle, which was supported by a paired t-test ($t = –0.813$; d.f. (degrees of freedom) = 43; $p = 0.0420$). Although there is a significant difference between the left and right scores at the 5% level, this significance is only slight. Since there was a larger number of available left clavicles, this side was used to make all figures.

The goal of the new method is to estimate the age of older individuals more accurately than is possible with already existing methods. Figure 2 shows the correlation between composite score (from test A_1[1]) and the documented ages at death. Spearman's rank correlation coefficient for this sample ($r_s = 0.597$) was lower than Falys and Prangle's (2015) r_s, although age was still statistically correlated with composite score ($p < 0.01$) independent of sex, meaning that if composite scores are not grouped into stages, the chance of estimating age correctly is only moderate (see Fig. 2 for this relation). This result, however, does indicate that the higher composite scores are more frequently associated with increasing age, as is also concluded by Falys and Prangle, although their r_s was higher ($r_s = 0.755$; $p < 0.001$ for males and $r_s = 0.644$; $p < 0.001$ for females; Falys and Prangle 2015, 207).

The graphs for surface topography, porosity, osteophyte formation and documented age at death are provided as Figure 3, which shows the relatively low positive correlation between all features and documented age at death. As is also concluded by Falys and Prangle (2015, 207), Spearman's rank correlation coefficients revealed that surface topography ($r_s = 0.394$, $p < 0.05$) is more strongly correlated to increasing age than porosity ($r_s = 0.287$, $p < 0.05$) and osteophyte formation ($r_s = 0.392$, $p < 0.05$). The Spearman Rank order coefficient for surface topography ($r_s = 0.394$, $p < 0.05$) is similar to the result for the correlation between the composite scores and documented ages at death.

Subsequently, Spearman's rank tests (r_s) were performed separately for males and females to assess whether the composite scores of one group have a stronger correlation with documented age than the other. This revealed an $r_s = 0.726$, $p < 0.01$, for females and an $r_s = 0.503$, $p < 0.05$, for males. Figure 4 shows this correlation. It is clearly visible that the female line is slightly more inclined than the male line, which visualizes the already established increased accuracy for female age estimations for this method.

A paired t-test showed that there is a slight significant difference between the age estimates derived from the left and right sides at the 5% level, but this should be regarded as only weak evidence for handedness.

Overall, 92.5% of the individuals could be placed in the correct age stage based on the left clavicle and 87.18% based on the right clavicle.

DISCUSSION

Age-at-death estimation in archaeological samples is an important step towards the establishment of demographic profiles. These form a basis for the analysis of cultural perceptions of, and

---

[1] Test A_1 was used in Figure 2 since inter-observer error test A_2 did not comprise all individuals.

Table 1 *Weighted Cohen kappa statistics (k at significance level p) for the intra- and inter-observer reliability tests based on composite scores and age stages. Categories of strength of agreement were derived from Landis and Koch (1977, 165)*

		Individual scores									
		Surface topography		*Porosity*		*Osteophyte formation*		*Composite score*		*Age stage*	
Tests		*k*	*p*	*k*	*p*	*k*	*p*	*k*	*p*	*k*	*p*
Intra-observer agreement	A_1 to A_2 L	0.453	< 0.0001	0.330	0.019	–0.094	0.588	0.352	< 0.0001	0.388	0.0050
	A_1 to A_2 R	0.516	< 0.0001	0.326	0.008	0.347	0.016	0.100	0.2590	0.425	0.0030
Inter-observer agreement	A_1 to W_1 L	0.287	0.002	0.203	0.053	0.344	0.016	0.253	< 0.0001	0.405	< 0.0001
	A_1 to W_1 R	0.060	0.581	0.255	0.006	0.564	0.001	0.203	0.0030	0.359	0.0010
	A_1 to S_1 L	0.039	0.664	0.159	0.152	0.473	0.004	0.123	< 0.0001	0.240	0.0410
	A_1 to S_1 R	0.274	0.005	0.129	0.171	0.754	< 0.0001	0.106	0.1050	0.252	0.0180
	W_1 to S_1 L	0.248	0.008	0.550	< 0.0001	0.539	0.001	0.291	< 0.0001	0.333	0.0070
	W_1 to S_1 R	0.170	0.095	0.636	< 0.0001	0.590	< 0.0001	0.145	0.0450	0.534	< 0.0001
		Kappa statistic		< 0.00	0.00–0.20	0.21–0.40	0.41–0.60	0.61–0.80	0.81–1.00		
		Strength of agreement		Poor	Slight	Fair	Moderate	Substantial	Almost perfect		

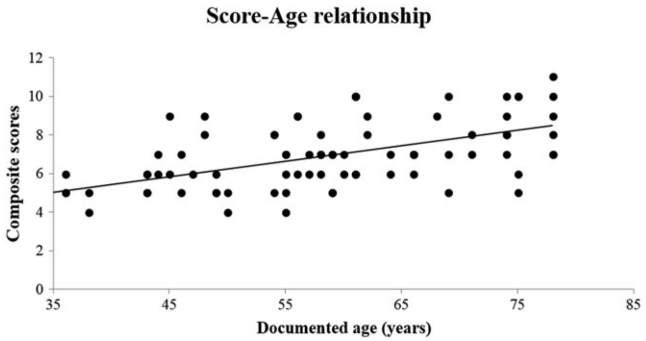

Figure 2 *Scatterplot of the relationship between documented age and the composite score (n = 80).*

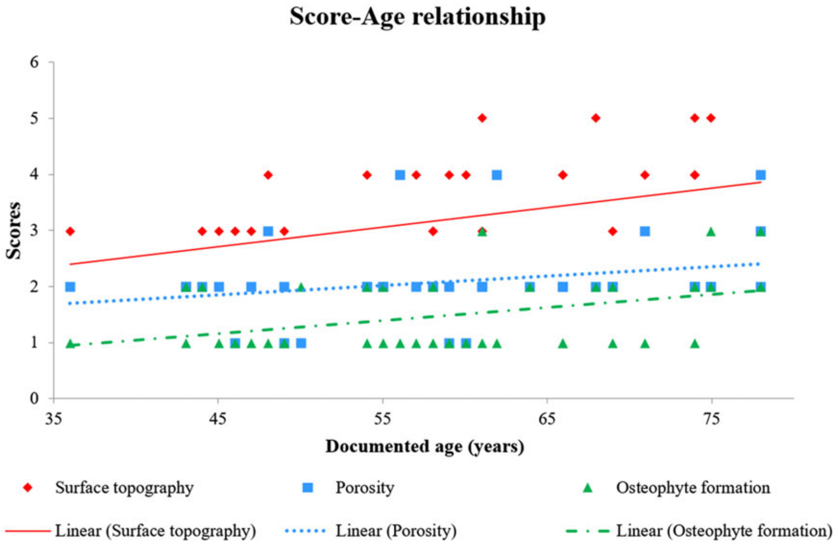

Figure 3 *Scatterplot of the relationship between documented age at death and individual scores, using the left clavicles (n = 41). [Colour figure can be viewed at wileyonlinelibrary.com]*

practices based on, age. However, at present our ability to learn about the relationship between old age and social and biological variables in the past is limited by the poor resolution of methods when dealing with the oldest individuals. The aim of this study was to assess intra- and inter-observer errors and the accuracy of the new clavicle ageing method developed by Falys and Prangle (2015) to improve age estimates when applied to a known age and sex skeletal collection from the post-medieval Netherlands. This was in order to assess the wider application of the method outside the original deriving study as well as to assess its accuracy when used by other researchers. Overall, the results corroborate that the ageing method based on the sternal end of the clavicle is in fact an accurate method with which to age adult individuals and have shown the method to be applicable to a post-medieval Dutch population.

However, this study has also shown that the method suffers from some difficulty in reproduction in repeat tests or by different observers, possibly partially solved by practice with this method. This problem is likely due to slight ambiguities in the descriptions and images provided

Figure 4 *Scatterplot of the relationship between documented age at death and the composite score divided into male and female results (based on the left clavicle). [Colour figure can be viewed at wileyonlinelibrary.com]*

as guidelines, along with lack of practice with this method and/or general experience in osteological analysis. The Cohen's kappa scores reflect this ambiguity well, despite the low number of observers, which makes the tests less definitive in quantitative terms. Nonetheless, the results do support the findings in this study qualitatively. Because the application of the methods immediately met with some discussion among the observers, the individual trait scores were submitted for statistical analysis to try to find the underlying cause of ambiguity (see Cohen's kappa scores in Table 1). The strength of agreement between the observers for each individual trait did not exceed a score of 'fair' in this study, whereas Falys and Prangle (2015, 205) found agreement strengths ranging from 'moderate' to 'substantial'. Falys and Prangle (2015, 212) already mention the possible inadequacy of using merely images and description in relation to inter-observer error data. Improvements could be made on two descriptions: (1) the description for score 4 in surface typography—'at least one round lump of bone is present on the generally flat surface' (i.e., nodule formation; Falys and Prangle 2015, 205)—has proven susceptible to variable interpretation between the observers. Each placed an emphasis on different aspects (e.g., either the 'at least one round lump of bone' or the 'generally flat surface'). Although other factors could have contributed, this impediment is possibly one contributing factor for the lower correlation between scores and documented age at death, especially in this study, as well as the relatively low degree of agreement within and between the different observers. It is possible that such confusion could be prevented by providing three-dimensional casts of 'prototypes' per stage, as has been done for sternal rib end and pubic symphysis methods.

(2) Another ambiguity lies in the limited amount of available knowledge about the causes, types and severity of bone porosity. Although it is widely accepted within osteology that porosity increases with age (Dequeker 1975, 100; Nelson and Weiss 1999, 3), Falys and Prangle (2015, 205) distinguish between micro- and macroporosity, and their method implies that they represent successive stages of severity (macroporosity appearing only after 50% of the surface is covered by microporosity). Reports and studies on the causes and consequences of porosity are not widely available, but one of the few studies discussing the difference between micro- and macroporosity (Henderson *et al.* 2013) does not discuss whether either of the two is a successive stage to the other. When assessing the order of appearance of porosity in the Middenbeemster collection, macropores

did appear before 50% of the surface was covered with micropores, or even before micropores had formed at all. If the presence of macroporosity were scored without consideration of the presence of microporosity, younger individuals might be aged significantly older than when first taking the presence of microporosity into account (e.g., Fig. 5). To avoid this problem, macropores were only considered when microporosity was present. Future studies might focus on quantifying the relationship between micro- and macroporosity for this method.

This issue of succession in micro- and macroporosity is an underlying problem of our poor understanding of the skeletal changes associated with the ageing process generally. Although general patterns of degeneration can be recognized, a true understanding of the different stages, the reasons for certain types of degeneration, and their succession and relation to one another are not clearly resolved and more studies are necessary to understand these processes.

Another point of discussion is whether the features used by Falys and Prangle to score stages of degeneration are truly associated with one another, as the method suggests. A study into the aetiology and expression of spinal osteoarthritis, for example, showed that some of the different features most likely have a different aetiology (Deverenski 2000, 352). A similar pattern cannot be excluded for clavicles without further research. Falys and Prangle (2015) suggest there might be a fixed relationship between the different traits by combining all the individual feature scores into one composite score, and ascribing an age (range) to these scores. However, the author found no definitive indication in the literature suggesting that the topography of the sternal end of the clavicle is related to the increase of porosity or to the formation of osteophytes, other than general age-related degeneration. Therefore, agreeing with Deverenski (2000, 352), systematic research should be carried out to study the relationship between different age- and activity-related degenerative symptoms, such as pitting, osteophyte formation, sclerosis, facet remodelling and porosity. The statistics performed in the original study also show that changes in surface topography are more strongly correlated to documented age at death than porosity, osteophyte formation and even the composite score. This study can largely corroborate these conclusions, except that here the correlations between composite scores and documented age at death turn out higher than for the other features (surface topography: $r_s = 0.394$, $p < 0.05$; porosity: $r_s = 0.287$, $p < 0.05$; osteophyte formation: $r_s = 0.392$, $p < 0.05$; versus composite score: $r_s = 0.597$, $p < 0.01$). However, as Falys and Prangle (2015) found the strongest correlation between age and surface topography, one might wonder if there was a need to combine the different, potentially

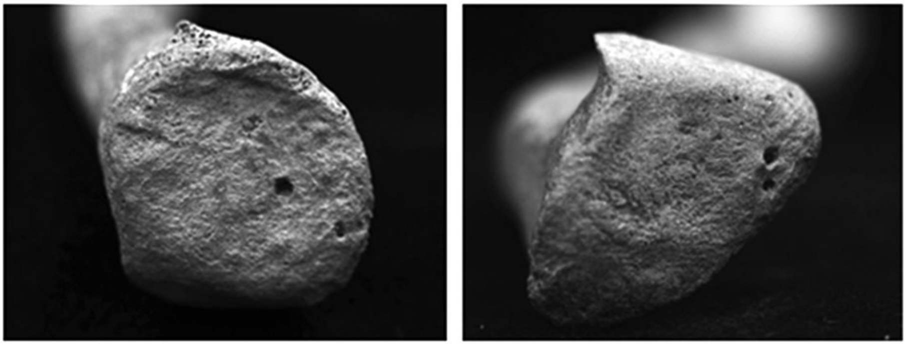

Figure 5 *Two examples of macroporosity without significant microporosity. The example of the left is individual MB11V422S0962 (aged 29 years); the example on the right is individual MB11V476S1054 (aged 31 years). If macroporosity were scored without taking lack of microporosity into account, these individuals would be classed as 42–53 and 43–84 years old respectively. Photos: author.*

unrelated features to form a composite score. Possibly the method could be improved upon if it would be based solely on surface topography, wherein porosity and osteophyte formation influence the flat topography of a young clavicle, or should be given more weight in a method that includes all three features. Further testing is required to ascertain this possibility.

The statistical analyses in this study have shown that the method is accurate for the Middenbeemster collection, placing at least 87% of the individuals in the appropriate age class. Despite the 'slight' to 'moderate' agreement in composite between the different observers on several levels, nearly all individuals were placed in the correct age category. The slightly lower accuracy rate may well improve with additional practice of the method. However, the already mentioned wide age ranges that make up the stages create results that cannot be deemed to be precise (a fact also concluded by Falys and Prangle 2015, 213). Based on the actual ages of sample individuals, 100% of all females could fall into age stage II for females (41–88 years) and 78% of males into the male age stage III (47–85 years), and the results would still be accurate since these age stages encompass such large age cohorts. This shows that age ranges as wide as 47 (41–88) years cannot be precise enough to improve upon the estimated age derived from other methods. Despite this lack of precision, the method does recognize older aged individuals in skeletal samples, especially if they fall into categories 4 and 5, which have age ranges that commence in the late sixth and seventh decades of life. This precision could be improved if smaller age ranges were identified. Unfortunately, the method, as it currently exists, cannot do so. The wide age ranges do, however, benefit the agreement scores: despite the low repeatabilities found for allocated scores in this study, accurate ageing was accomplished due to the large overlap between the age stages. However, their 'sufficiently broad age ranges', as mentioned by the authors, is something that some might not deem as being particularly useful because they decrease precision rates (Falys and Prangle 2015, 204). The age ranges are too large to improve upon currently existing ageing methods greatly, regardless if it provides a useful addition to other ageing methods, including the pubic symphysis and auricular surface, because this method could verify and support the age estimate achieved, and could indicate when someone was significantly older than 46 years. It is also important to note that the method provides an alternative area for age assessment when auricular surface and pubic symphysis are unobservable. The actual relationship between the clavicle method and other methods should, however, be researched in more detail to obtain a better understanding of the advantages and disadvantages of using this method over, or in combination with, other methods.

Although this study only uses descriptive statistics to estimate age, Falys and Prangle (2015) also provide three regression formulae for assessing age: one for females, one for males and one for individuals for which their sex cannot be determined. In the archaeological record, where it is likely that some individuals cannot have sex estimates ascribed to them, it is a great advantage to have a method that can ascertain age without knowing the sex. These regression formulae would benefit from independent testing in future.

CONCLUSIONS

This study aimed to test the ageing method developed for the sternal end of the clavicle by Falys and Prangle (2015). The method was assessed for reproducibility and applicability to collections other than those used for the development of the method or which were assessed by the original authors. The results of the intra- and inter-observer tests revealed that composite scores are not very reproducible in multiple tests or by different observers. However, the resulting age classes largely corresponded between observers due to their wide range. Spearman's rank correlation revealed that there is a weak correlation between the different features scored and composite score, and documented ages at death, but have shown that the composite score is more strongly related to age than the individual scores taken separately. When testing the method on males and females separately, the results showed the method to be slightly more accurate for females. A paired *t*-test revealed that there is no significant difference between left and right scores and the resultant age class, potentially indicating no significant difference relating to handedness.

The age stages assigned to individuals are too broad to deem the method precise, and may therefore not always divide the 46-or-more-years stage into smaller categories, perhaps with the exception of individuals falling into the last clavicle-age category. The method is, however, accurate in over 87% of cases, with the left clavicle slightly outperforming the right. The method can, therefore, be used to identify accurately the older individuals in a post-medieval Dutch skeletal collection, but is not necessarily much more precise than currently available methods.

The Falys and Prangle clavicle method provides an age-estimation technique based on a bone that is more durable in the archaeological record than the pubic symphysis or sternal rib ends, and the relationship between these methods should be assessed to understand their advantages and disadvantages compared with one another. Although the method could be improved upon by better descriptions and potentially by providing three-dimensional casts of examples, it proves valuable for ageing adult individuals in an archaeological record, and the subject should be researched in more detail.

ACKNOWLEDGMENTS

The authors thank the people of Middenbeemster and the Historische Genootschap Beemster (Historic Society Beemster) for their support. They also express their gratitude to the anonymous reviewers for their useful feedback.

REFERENCES

Buckberry, J. L., and Chamberlain, A. T., 2002, Age estimation from the auricular surface of the ilium: a revised method, *American Journal of Physical Anthropology*, **119**(3), 231–9.

Cave, C., and Oxenham, M., 2014, Identification of the archaeological ‘invisible elderly’: an approach illustrated with an Anglo-Saxon example, *International Journal of Osteoarchaeology*, **26**(1), 163–75.

Cox, M., 2000, Ageing adults from the skeleton, in *Human osteology in archaeology and forensic science* (eds. M. Cox and S. Mays), 61–81, Cambridge University Press, Cambridge.

Dequeker, J., 1975, Bone and ageing, *Annals of the Rheumatic Diseases*, **34**(1), 100–15.

Deverenski, J. R. S., 2000, Sex differences in activity-relates osseous change in the spine and the gendered division of labor at Ensay and Wharram Percy, UK, *American Journal of Physical Anthropology*, **111**(3), 333–54.

DiGangi, E. A., and Moore, M. K., 2013, Introduction to skeletal biology, in *Research methods in human skeletal biology* (eds. E. A. DiGangi and M. K. Moore), 3–28, Elsevier, Oxford.

Dorandeu, A., Coulibaly, B., Piercecchi-Marti, M., Bartoli, C., Gaudart, J., Baccino, R., and Leonetti, G., 2008, Age-at-death estimation based on the study of frontosphenoidal sutures, *Forensic Science International*, **177**(1), 47–51.

Falys, C. G., and Prangle, D., 2015, Estimating age of mature adults from the degeneration of the sternal end of the clavicle, *American Journal of Physical Anthropology*, **156**(2), 203–14.

Griffioen, A., 2011, *Inventariserend Veldonderzoek te Middenbeemster, Middenweg 148, Hollandia reeks 346*, Hollandia archeologen, Zaandijk.

Hakvoort, A., 2013, *De begravingen bij de Keyserkerk te Middenbeemster, Hollandia reeks 464*, Hollandia archeologen, Zaandijk.

Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Vilotte, S., and Wilczak, C., 2013, Recording specific entheseal changes of fibrocartilaginous entheses: initial tests using the Coimbra method, *International Journal of Osteoarchaeology*, **23**(2), 152–62.

İşcan, M. Y., Loth, S. R., and Wright, R. K., 1984, Metamorphosis at the sternal rib end: a new method to estimate age at death in white males, *American Journal of Physical Anthropology*, **65**(2), 147–56.

Key, C. A., Aiello, L. C., and Mollenson, T., 1994, Cranial suture closure and its implications for age estimation, *International Journal of Osteoarchaeology*, **4**, 192–207.

Landis, J. R., and Koch, G. C., 1977, The measurement of observer agreement for categorical data, *International Biometric Society*, **33**(1), 159–74.

Mays, S., 2002, The relationship between molar wear and age in an early 19th century AD archaeological human skeletal series of documented age at death, *Journal of Archaeological Science*, **29**(8), 861–71.

Nelson, D. A., and Weiss, M. L., 1999, Aging through the ages, in *The aging skeleton* (eds. C. J. Rosen, J. Glowacki, and J. P. Bilezikian), 3–10, Academic Press, San Diego, CA.

Nikita, E., 2013, Quantitative assessment of the sternal rib end morphology and implications for its application in aging human remains, *Journal of Forensic Sciences*, **58**(2), 324–9.

Todd, T. W., 1920, Age changes in the pubic bone. I. The male white pubis, *American Journal of Physical Anthropology*, **3**(3), 285–339.

Ubelaker, D. H., 2008, Forensic anthropology: methodology and diversity of applications, in *Biological anthropology of the human skeleton* (eds. M. A. Katzenberg and S. R. Saunders), 41–69, Wiley, Hoboken, NJ.

Waters-Rist, A. L., and Hoogland, M. L. P., 2013, Osteological evidence of short-limbed dwarfism in a nineteenth century Dutch family: achondroplasia or hypochondroplasia, *International Journal of Paleopathology*, **3**(4), 243–56.

White, T. D., and Folkens, P. A. (eds.), 2005, *The human bone manual*, Elsevier, London.