Archaeometry **61**, 1 (2019) 211–225 doi: 10.1111/arcm.12409

A COMPARISON OF TWO METHODS FOR RECORDING ENTHESEAL CHANGE ON A POST-MEDIEVAL URBAN SKELETAL COLLECTION FROM AALST (BELGIUM)*

J. L. A. PALMER†

Faculty of Archaeology, Leiden University, Einsteinweg, 2333 CC Leiden, The Netherlands

K. QUINTELIER

Flanders Heritage Agency, Havenlaan 88 bus 5, 1000, Brussels, Belgium

S. INSKIP

McDonald Institute for Archaeological Research, University of Cambridge, Downing Street, Cambridge CB2 3ER, UK

and A. L. WATERS-RIST

Department of Anthropology, Faculty of Social Science, Western University, London, Ontario, Canada, N6A 3K7

This research compares two current methods for recording bony changes at muscle attachment sites, called entheseal changes (EC); the Mariotti method and the Coimbra method, to evaluate the concordance and comparability of results in a post-medieval skeletal collection from Aalst, Belgium (n = 116). For both methods, the EC scores produce broadly similar patterns, are symmetrical and differ between age groups. Statistical differences between the upper and lower limb and the lower limb of males and females only occur in the Mariotti method. With careful consideration of the influence of different EC score ranges, the results from the two methods can generally be compared.

KEYWORDS: HUMAN OSTEOARCHAEOLOGY, ACTIVITY MARKER, MUSCULOSKELETAL STRESS MARKERS, MARIOTTI METHOD, COIMBRA METHOD, FIBROCARTILAGINOUS ENTHESES

INTRODUCTION

Research into entheseal changes (EC) has evolved rapidly in the past few years as a means of reconstructing aspects of the activity patterns of past populations (Weiss 2015). Osteoarchaeologists have long noted the development and morphological variation of bony attachment sites of muscles and tendons (e.g., Angel 1945; Wells 1963); however, structured studies of EC only commenced in the 1980s (for an early example, see Dutour 1986) and gained momentum with the creation of a broadly applicable scoring method by Hawkey and Merbs (1995). Although several other methods have been created since (see, e.g., Robb 1998; Wilczak 1998), most studies have used the method suggested by Hawkey and Merbs (1995) (see Acosta *et al*. 2017) or the updated version thereof, designed by Mariotti *et al*. (2004, 2007). Entheses are divided into two types based on direct or indirect bony attachment, namely fibrous and

*Received 7 September 2016; accepted 30 April 2018
†Corresponding author: email j.l.a.palmer@arch.leidenuniv.nl

fibrocartilaginous entheses (Weiss 2015). While the methods listed above can be applied to all entheses, the focus of recent EC research has been on fibrocartilaginous entheses (Jurmain *et al*. 2012). The main reason for this is that a normal, 'no entheseal change' baseline for fibrous entheses has not been established (Villotte *et al*. 2016), whereas it has been ascertained for fibrocartilaginous entheses. This development has improved understanding of the nature of ECs and their usefulness in activity reconstruction. It has gone hand in hand with the creation of new scoring methods, such as the system of Villotte (2006), which incorporates the difference between fibrocartilaginous and fibrous entheses, or the later version focusing exclusively on fibrocartilaginous attachment sites (Villotte *et al*. 2010). More recently, a group of scientists, including Mariotti and Villotte, united in the Coimbra group and set out to create a standard scoring method, which, because of the limitations of fibrous entheses mentioned above, focused only on fibrocartilaginous entheses (Henderson *et al*. 2010, 2013, 2016b). Although a more standardized scoring method is certainly desirable to enable inter-study comparison (Henderson 2013), it does not address issues of comparability of previous and future studies using different methods. Given the considerable amount of research already conducted with different methods, and the likelihood that some researchers will continue to use different methods in the future, it is important to establish the extent to which results from different methods can and cannot be compared. Such an assessment must begin by examining the patterning of EC results generated by the different methods as applied to the same skeletons. If the older Mariotti *et al*. (2004, 2007) method and the new Coimbra method (Henderson *et al*. 2016b) can be shown to produce sufficiently similar result patterns, this can allow inter-study comparisons and thereby greatly enlarge the utilizable data set on entheseal change in different regions, contexts and time periods.

Therefore, this paper will compare EC results obtained with the Coimbra method (Henderson *et al*. 2016b) to those obtained with the Mariotti *et al*. (2004, 2007) method, to (1) evaluate the concordance of resulting EC patterns, (2) assess implications for inter-study comparison and (3) point out the practical advantages and limitations of each method.

MATERIALS AND METHODS

The sample of human skeletons for this methodological study was taken from the post-medieval Belgian Carmelite friary of Aalst, excavated in the current Hopmarkt. This site was excavated in two phases, the first in 2004–5 and the second in 2011. The collection was chosen for its good preservation, skeletal completeness and potential for further research. The cemetery population is recorded as containing monks as well as lay people, who were buried between AD 1497 and 1797 (De Groote *et al*. 2011). Both sexes and all age groups are represented. Sex was estimated through traits on the pelvis, cranium and mandible, following the Workshop for European Archaeologists guideline (Ferembach *et al*. 1980), the Phenice (1969) pubic traits, and metrics (Stewart 1979; McCormick *et al*. 1991; Steyn and Işcan 1999). For this research, only adults were selected, as entheses can develop and change differently in the growing skeleton (Villotte 2006; for an explorative study, see also Palmer *et al*. 2017). The 116 analysable individuals were divided into four age categories; early young adult (18–25 years), late young adult (26–35 years), middle adult (36–49 years) and old adult (50+ years) (Table 1). Age was estimated using the morphology of the pubic symphysis (Suchey and Brooks 1990), auricular surface (Buckberry and Chamberlain 2002) and sternal rib end (Işcan *et al*. 1984), as well as dental attrition (Maat 2001), cranial suture closure (Meindl and Lovejoy 1985) and the fusion state of late-fusing epiphyses (i.e., the sternal end of the clavicle, spheno-occipital synchondrosis, iliac crest and ischial tuberosities; Schaefer *et al*. 2009).

Table 1 *Age at death and sex distribution of the Aalst Hopmarkt sample*

	Sex		
Age category	*Female*	*Male*	*Total*
Early young adult	8	9	17
Late young adult	16	26	42
Middle adult	14	26	40
Old adult	6	11	17
	44	72	116

The most consistent result in EC research is the positive correlation between the development of changes at muscle attachment sites and age—for example, Weiss (2007), Molnar *et al*. (2011) and Michopoulou *et al*. (2017) on collections with osteologically assessed age at death; and Alves Cardoso and Henderson (2010), Henderson *et al*. (2012, 2017), Milella *et al*. (2012) and Villotte *et al*. (2010) on collections with known age at death). It is mainly new bone formation that is being related to ageing (Henderson *et al*. 2012, 2017). Although old adults have been shown to be less reliable for activity reconstruction (Niinimäki *et al*. 2013), they are included here to evaluate the differences between the results of the scoring methods. Statistical tests are run both with the entire sample (n = 116) and without the old adults (n = 99), to evaluate their influence on the sample.

Only the entheses that could be scored with both methods are used in this study. As such, only the fibrocartilaginous entheses (listed by Villotte *et al*. 2010) that were included in the Mariotti *et al*. (2004, 2007) method are used, namely the *M. triceps brachii* and *M. brachialis* attachments on the ulna, the *M. biceps brachii* on the radius, the *M. iliopsoas* on the femur, the *M. quadriceps femoris* and *M. popliteus* on the tibia, the *M. quadriceps femoris* on the patella and the *tendo calcaneus* (the Achilles tendon, which attaches the *M. plantaris*, *M. gastrocnemius* and *M. soleus* to the heel) on the calcaneus. An enthesis is only included if all traits included in both methods (described below) could be scored. The surface area of the enthesis was identified and delineated based on anatomical textbooks (Gray 1977; Paulsen and Waschke 2011) prior to EC scoring. Enthesis delineation can be challenging (more so for fibrous than fibrocartilaginous entheses), and thus a source of inter- and intra-observer error. Furthermore, it can be a cause for deviation between recording results not related to the actual method used. To limit error, all entheses were scored solely by the first author, whose recording process was to delineate the entheses and then score it using both methods, before moving on to the next entheses.

The two methods that this study compares use different criteria for observation and recording. The Mariotti *et al*. (2004, 2007) method uses three criteria for EC, namely robusticity, osteophyte formation and osteolytic lesions. The method was developed for 23 fibrous and fibrocartilaginous entheses, with a specific description and pictures for robusticity scores for each enthesis. These descriptions and pictures were added to incorporate the unique changes seen at each attachment site and to attempt to limit inter-observer error, a main problem in the Hawkey and Merbs (1995) method (Davis *et al*. 2013). Robusticity is scored between 1 and 3, with score 1 subdivided into three stages, a, b and c, whereas the other two markers are scored between 0 and 3. Per enthesis, three individual scores are obtained. Mariotti *et al*. (2007) allow researchers to decide how many categories to incorporate for robusticity based on the specifics of the collection under study (e.g., sample size, research question). For the current study, robusticity was scored as 1, 2 or 3, as recommended in the method paper (Mariotti *et al*. 2007), thus omitting the subdivisions that are possible for score 1, since including them would produce a data set that was mathematically difficult to synthesize.

The Coimbra group method (Henderson *et al*. 2016b) will be used it in the way it was presented at the 18th European Meeting of the Palaeopathology Association (Henderson *et al*. 2010). For this method, the fibrocartilaginous entheses are divided into two zones: zone 1 being the edge of the enthesis surface, at an obtuse angle to where the tendon fibres attach (i.e., the rim the furthest away from the direction in which the muscle/tendon pulls) and zone 2 being the rest of the surface (i.e., the majority of the entheseal surface). For zone 1, two traits are evaluated: bone formation and erosion. For zone 2, bone formation, erosion, fine porosity, macroporosity and cavitation are analysed. Thus, per enthesis, seven individual scores are obtained. The final version of the Coimbra method (2016) also includes the factor of textural change for zone 2. As data collection for this research was completed prior to that publication, this factor is not incorporated. The largest difference between the Coimbra method and the Mariotti method is that the Coimbra method does not include robusticity as a trait, but includes more possible changes occurring at the enthesis site.

Ten individuals were re-analysed using both methods 2 weeks after the original analysis to test for intra-observer variation. In all cases for both methods, the composite score deviation per enthesis was maximally one unit of measurement different from the original score. Per method, intra-observer error percentages were calculated for the composite score per individual. The percentages thus obtained from the 10 individuals were then combined into an average. The average error percentage for the 10 individuals was 6% difference for the Mariotti method and 4% difference for the Coimbra method. No specific enthesis was more sensitive to intra-observer error.

Neither method, as yet, includes a way to synthesize the data for statistical analysis. More research into the relative importance of the different entheseal changes (i.e., porosity, lytic lesions, new bone formation etc.) is necessary to provide the data to create an appropriate system of data synthesis, potentially giving different weights to different EC traits. For the purpose of this study, all scores are tallied per enthesis in the same manner for both methods, to create a composite score. For the Mariotti *et al*. (2004, 2007) method, this means that each entheses obtains a score between 1 and 9, whereas Coimbra method scores can range from 0 to 18. An average score is calculated for the left and right upper limb and lower limb, and these results are shown in bar charts. Thus, the Mariotti and Coimbra EC averages cannot be directly compared due to their different scoring parameters, but the overall trends and patterns in the results can be compared at a more general level. It must be noted that the creation of average scores based on ordinal data is, from a purely mathematical perspective, inappropriate. However, for EC scoring, the ordinal data system represents an artificial approach to organizing the underlying continuous spectrum of human variation at entheses. Therefore, reframing this type of ordinal data into averages for analysis is an acceptable way to elucidate patterns (as noted by Robb 1998, based on Weisberg 1992). For some individuals, not all entheses of the upper and lower limb could be scored. Individuals were included when more than half of the entheses under study were present and omitted when less than half were present.

Comparability of the methods is gauged both by assessing the similarity between their separate results and by testing their correlation. For each method separately, Wilcoxon signed-rank tests are used to assess the difference between the left and right side and the upper and lower limb. Similarly, for each method separately, the Kruskal–Wallis test is used to examine if there are significant differences in EC score by age and the Mann–Whitney U test is used for differences between the sexes. The comparability of the methods can then be gauged by the similarities and differences in statistical significance with the parameters of side (left vs. right), limb (upper vs. lower), age (EYA, LYA, MA, OA) and sex (male vs. female). Next, the results are assessed for correlation between the methods, with a Spearman's rho correlation test, based on the concordance of the two data sets per individual.

All statistical tests are performed using IBM SPSS 23 (IBM Corp., Armonk, NY: IBM SPSS Statistics for Windows, Version 23.0, released 2015). The statistical significance is set at $p \leq 0.05$.

RESULTS

An overview of all the statistical test results is provided in Table 2. This table provides specific *n*-values for each test done with each method, the test statistic value and the *p*-values.

Left and right side

There is no statistically significant evidence of asymmetry or side dominance in EC scores for either method (Wilcoxon signed-rank tests; Mariotti $p = 0.960$, Coimbra $p = 0.771$). This is true when all data are combined as well as when the upper and lower limbs are evaluated separately (Mariotti upper limb $p = 0.397$, lower limb $p = 0.448$, and Coimbra upper limb $p = 0.917$, lower limb $p = 0.991$). When the entheses data are plotted per specific muscle attachment site, the

Table 2 *All tests run for the two methods: statistically significant results are indicated in bold; for each method, the n-value is indicated in the far right column*

	Mariotti method			*Coimbra method*		
Variables	*p-value*	*Factor*	*n*	*p-value*	*Factor*	*n*
Upper vs. lower limb*	**$p = 0.044$**	**$Z = -2.010$**	97	$p = 0.878$	$Z = -0.153$	92
Age†	**$p < 0.001$**	**$\chi^{2(2)} = 34.381$**	116	**$p < 0.001$**	**$\chi^{2(2)} = 23.324$**	115
Upper limb age†	**$p < 0.001$**	**$\chi^{2(2)} = 24.893$**	111	**$p = 0.001$**	**$\chi^{2(2)} = 15.441$**	107
Lower limb age†	**$p < 0.001$**	**$\chi^{2(2)} = 27.881$**	102	**$p < 0.001$**	**$\chi^{2(2)} = 18.813$**	100
Age without old adults†	**$p < 0.001$**	**$\chi^{2(2)} = 24.860$**		**$p < 0.001$**	**$\chi^{2(2)} = 19.298$**	99
Upper limb age without old adults†	**$p < 0.001$**	**$\chi^{2(2)} = 16.527$**	95	**$p = 0.009$**	**$\chi^{2(2)} = 9.369$**	92
Lower limb age without old adults†	**$p < 0.001$**	**$\chi^{2(2)} = 21.379$**	89	**$p = 0.002$**	**$\chi^{2(2)} = 12.975$**	88
Sex‡	**$p = 0.007$**	**$U = 2099.00$**	116	**$p = 0.100$**	**$U = 1276.5$**	115
Upper limb sex‡	$p = 0.168$	$U = 1235.00$	111	**$p = 0.571$**	**$U = 1265.00$**	107
Lower limb sex‡	**$p = 0.006$**	**$U = 857.00$**	102	$p = 0.075$	$U = 956.00$	100
Left vs. right*	$p = 0.960$	$Z = -0.050$	104	$p = 0.771$	$Z = -0.292$	109
Left upper vs. right upper*	$p = 0.397$	$Z = -0.847$	83	$p = 0.917$	$Z = -0.104$	77
Left lower vs. right lower*	$p = 0.448$	$Z = -0.758$	84	$p = 0.991$	$Z = -0.011$	80
		p-value			Factor	n
Mariotti vs. Coimbra§		**$p < 0.001$**			**$\rho = 0.764$**	116
Mariotti vs. Coimbra upper limb§		**$p < 0.001$**			**$\rho = 0.651$**	111
Mariotti vs. Coimbra lower limb§		**$p < 0.001$**			**$\rho = 0.791$**	100

*Wilcoxon signed rank,
†Kruskal–Wallis,
‡Mann–Whitney *U*,
§Spearman's rho.

pattern of scores is similar for both methods for the *M. triceps brachii* and *M. brachialis* of the upper limb, and the *M. iliopsoas* and *M. quadriceps femoris* on the patella in the lower limb, and only slightly different for the other four entheses under study (see Fig. 1). In the *M. popliteus* entheses especially, differences in score patterns are very small. For this reason, left and right side data are pooled in subsequent tests, which has the benefit of increasing the sample sizes.

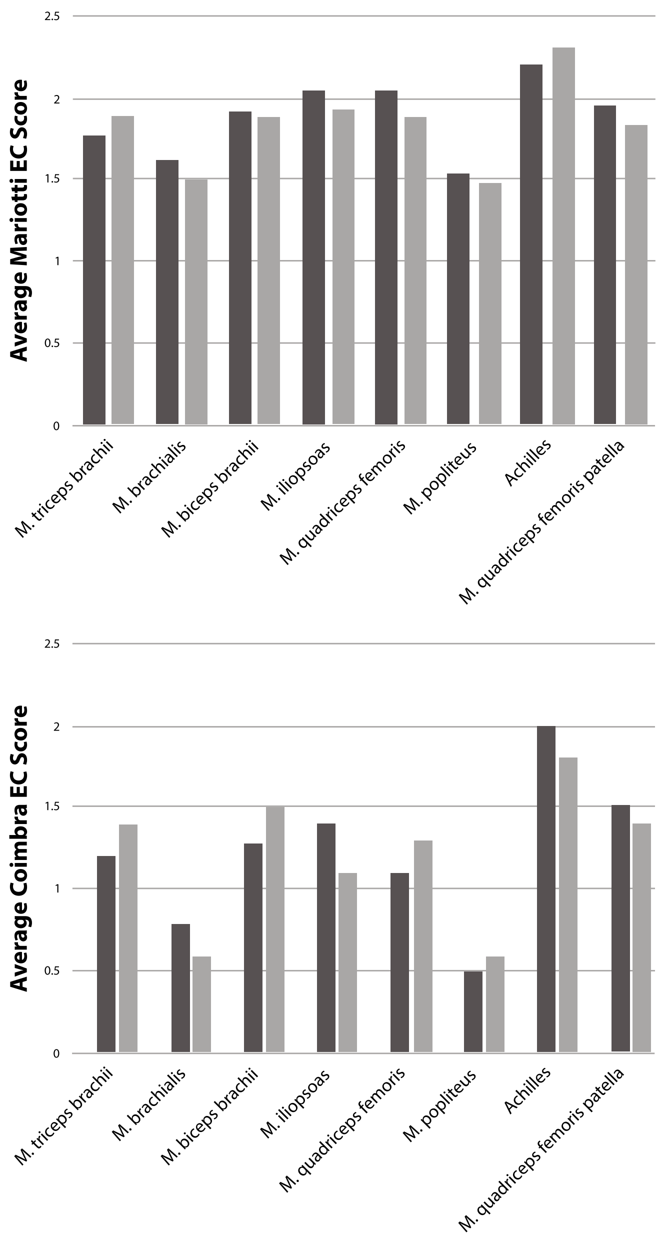


Figure 1 *The average EC score per muscle attachment site for the left and right sides: left side in dark grey, right side in light grey (n = 116). [Colour figure can be viewed at wileyonlinelibrary.com]*

Comparing the methods

A Spearman's rho correlation test shows a statistically significant positive correlation in scores between the methods ($p < 0.001$, $\rho = 0.764$) when all entheses are combined, as well as for the upper and lower limb individually (upper limb $p < 0.001$, $\rho = 0.651$; lower limb $p < 0.001$, $\rho = 0.791$). A scatterplot of all scored individuals shows that while results do cluster around the fit line, there is still a noteworthy amount of diversion (Fig. 2). The most diverging individuals have been labelled in the figure. Individuals 17 (an old adult male), 60 (also an old adult male) and 82 (a late young adult female) score notably higher in the Coimbra than the Mariotti method. Individuals 45 (an old adult female), 37 (a late young adult male) and 44 (an old adult female) score higher in the Mariotti method compared to the Coimbra method. Individuals 67 and 10 are both old adult males and score high in both methods, if more so in the Coimbra method.

Upper and lower limb

Wilcoxon signed-rank tests show a significant difference in EC scores between the upper and lower limbs in the Mariotti method ($p = 0.044$), with a mean EC score for the upper limb of 1.81 and for the lower limb of 1.95. For the Coimbra method, there is no significant difference between upper and lower limb ($p = 0.878$), with the mean EC score for the upper limb being 1.18, and 1.16 for the lower limb.

Age

Both methods show a significant difference in EC scores between age groups using the Kruskal–Wallis tests (Mariotti method $p < 0.001$, Coimbra method $p < 0.001$) for all EC combined as well as for the upper and lower limb separately (Mariotti, upper limb $p < 0.0001$, lower limb

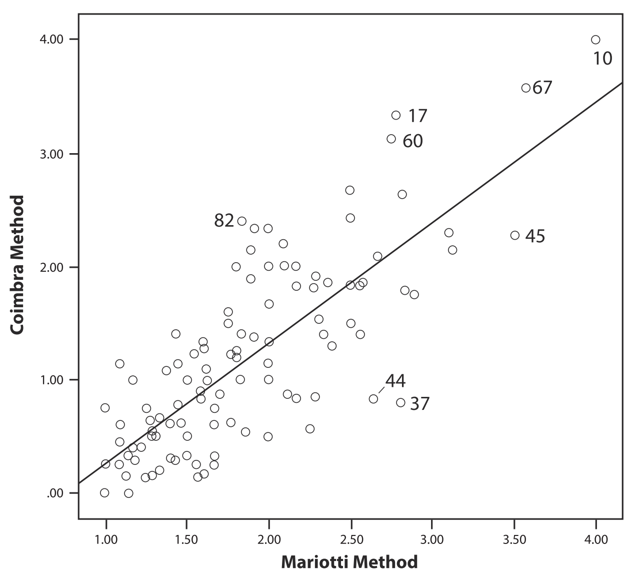


Figure 2 *A scatterplot showing the data from both methods with a fit line: outliers are labelled and discussed in the text* (*n* = 116, R^2 Linear = 0.615).

$p < 0.0001$; Coimbra: upper limb $p = 0.001$, lower limb $p < 0.001$) (for test statistics, see Table 2; see also Fig. 3).

Sex

According to a Mann–Whitney *U* test, when all EC scores are combined, the Mariotti method shows a significant difference between the EC scores of the sexes ($p = 0.007$), whereas this significant difference is absent in the Coimbra method ($p = 0.100$). When the upper- and lower-limb EC scores from the Mariotti method are analysed separately, the upper limb shows no statistically significant difference between the sexes ($p = 0.168$), whereas the lower limb does ($p = 0.006$). In the lower limb, the male EC scores are consistently higher than the female EC scores, whereas in the upper limb, males score higher for the *M. biceps brachii* attachments, but females score higher for the *M. triceps brachii*. For the Coimbra data, neither the upper nor the lower limb show a statistically significant difference between the sexes at the 0.05 level (upper limb $p = 0.571$, lower limb $p = 0.075$). However, when the muscle attachment score averages are regarded per muscle, even though statistical significance is not reached, the pattern is similar to that of the Mariotti method: greater lower-limb scores for males, and in the upper limb higher male scores for the *M. brachialis* and *M. biceps brachii* attachments, but higher female scores for the *M. triceps brachii* (see Fig. 4).

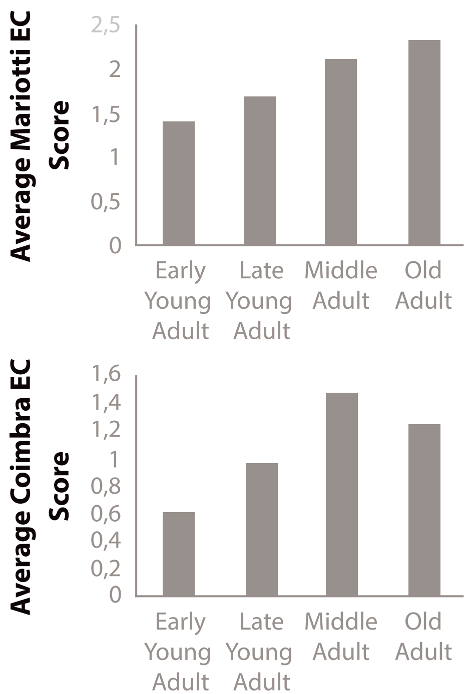
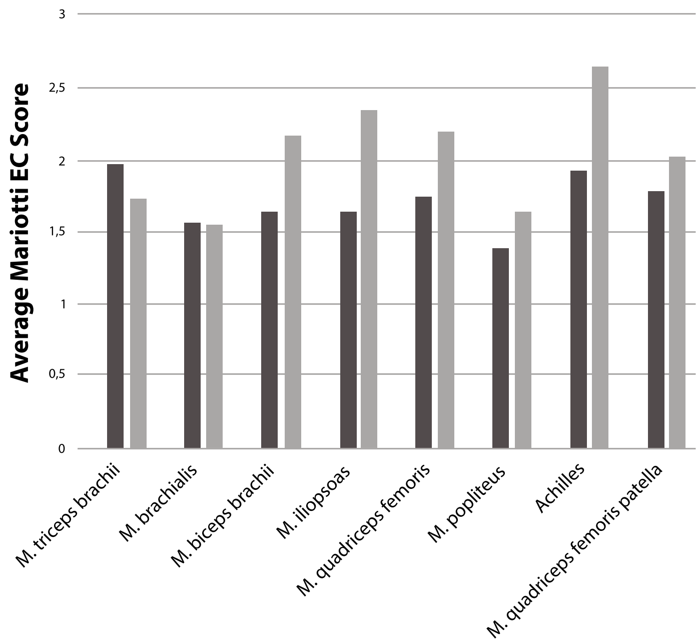


Figure 3 *The average EC score per age category for the upper and lower limbs combined (n = 116). [Colour figure can be viewed at wileyonlinelibrary.com]*



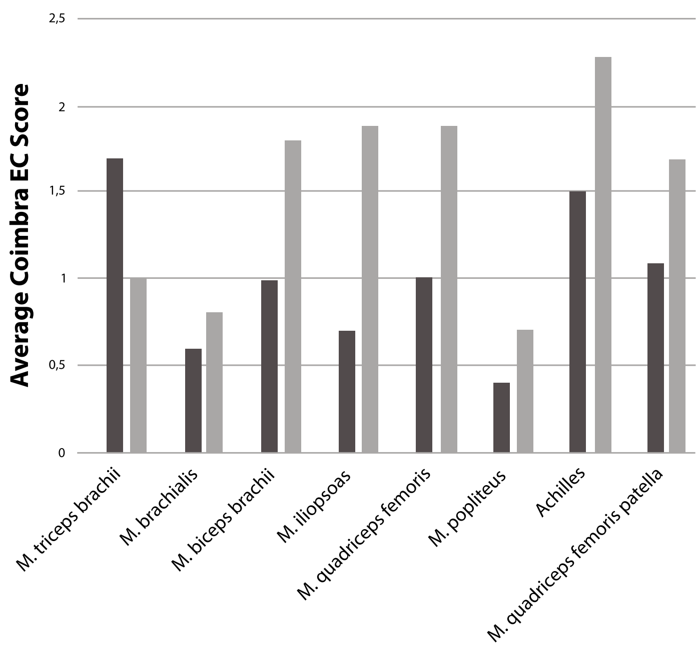


Figure 4 *The average EC score per muscle attachment site for males and females: females in dark grey, males in pale grey (n = 116); left- and right-side data pooled. [Colour figure can be viewed at wileyonlinelibrary.com]*

DISCUSSION

This paper aimed to assess whether the Coimbra method and the Mariotti method produce sufficiently similar results to permit the comparison of past (and possible future) research that used the different methods. Therefore, the concordance of results was analysed, and the implications for inter-study comparisons are now appraised along with consideration of the practical advantages and limitations of each method, together with some additional findings.

Concordance of results between methods

Given that both methods, for a large part, rely on the same traits (i.e., bone formation and lytic activity) as evidence of EC, it could be expected that they would deliver similar results. At a general level this is shown to be true; the methods give broadly concurring patterns of results, as seen in Figures 1 and 4. However, relative sex differences appear larger when using the Coimbra method (Fig. 4). The explanation for this is purely methodological: in the Mariotti method, all individuals have a minimum score of 1, whereas in the Coimbra method a score of 0 is possible, making differences between individuals with little or no EC and individuals with EC seem larger. Also, in Figure 1 the relative left–right differences are different between methods, although there is no clear explanation for this. There is a positive correlation between the methods, and this correlation reaches very high statistical significance, yet when all data are pooled in a scatterplot, the underlying variation within this overall picture of similarity between methods is illustrated (Fig. 2). No single explanation emerges to explain the outliers in this graph, and no single entheses, group of entheses or trait can be singled out as creating inter-method inconsistency. It is noteworthy that four of the individuals diverging the farthest from the fit line are old adults; and that of these four old adults, the higher Coimbra scores are males where the higher Mariotti scores are female. Future research into which entheseal change traits are more common in which sex and age group can perhaps explain this. It is unsurprising that the highest-scoring individuals are both old adults, given the correlation between age and EC (discussed above).

Both methods demonstrate no statistical difference between left and right and significant differences in EC among age groups. However, only the Mariotti data demonstrate a significant difference between the upper and lower limb, a difference driven in part by a statistically significant difference in EC between the lower limbs of males and females. Thus, there is perhaps more difference between the methods' results than expected. Some of these differences could lead researchers to different conclusions. For example, if the ECs were scored using the Mariotti method, one would conclude that there was a significant difference between the lower-limb ECs, and hence activity patterns, of males and females, whereas one would probably conclude that no such difference existed if the data were generated using the Coimbra method. In the former scenario, it might have been concluded that males were more mobile than females; and in the latter scenario, that they were not.

As with most archaeological studies, it must be borne in mind that limited sample size will impact the statistical results (for a discussion of the effect of sample size on EC, see Henderson and Nikita 2016); however, as the sample sizes are highly similar for both methods (Table 2), it can be assumed, at least for concordance between methods, that the impact of sample size was limited.

Implications for inter-study comparison

Comparison between studies and populations is key to achieving a better understanding of how changes in morphology of muscle attachment should be interpreted. As both the Mariotti *et al.* (2007) method and Coimbra method (Henderson *et al.* 2016b) aim to register the changes at muscle attachment sites to analyse physical activity, the differences between results obtained via the two methods are an important finding, further highlighting the complexity of entheseal change research. Michopoulou *et al*. (2017) tested the correlation between the Coimbra method and bone cross-sectional geometry, an activity marker with proven reliability. They found that EC can at least be partially caused by activity when observed via the Coimbra method. In an earlier study (Michopoulou *et al*. 2015), they conducted the same test upon the Mariotti *et al*. (2004) method and found no correlations between activity as construed via cross-sectional geometry and EC as observed via osteolytic and osteophytic lesions. However, as that study did not use the Mariotti *et al*. (2007) method, and nor did it incorporate robusticity as described in Mariotti *et al*. (2007), these results do not offer any conclusive outcomes for our current study. For pre-industrial populations, we are forced to rely mainly on archaeological materials to answer questions about if and to what extent ECs are related to activity. Further testing of both methods on skeletal collections of known activity can elucidate their respective validity; however, given the complicated nature of activity throughout a person's life, which is much wider than one's registered occupation, this still holds limitations.

The results from this research suggest that general usage of studies conducted with these two different EC recording methods is suitable only when comparing the same muscle attachment sites, and bearing in mind the intrinsic properties of each method, as these may sometimes lead to different statistical outcomes. The largest differences are seen in the lower limb, where sex differences are statistically significant in the Mariotti but not the Coimbra method, a difference that is also reflected in the statistically significant difference between the upper and lower limb in the Mariotti but not the Coimbra method. Researchers must remain aware of the larger range of scores possible in the Coimbra method, which can make intra- and inter-individual differences appear larger than in the Mariotti method simply as a result of the scoring system. Yet, the number of similarities in results between methods outweighs the number of differences, and we argue that when the discernment of broad patterns is the goal of using studies that have employed the different methods, this can be acceptably achieved with careful consideration of the results and the aforementioned factors.

Advantages and limitations of each method

For fibrocartilaginous entheses, the Coimbra method, which uses seven traits per EC, is the most detailed, giving a very accurate overview of the osseous changes occurring at an entheses (except for changes in robusticity). Although there has been anatomical research that has improved our understanding of EC, we are still not yet fully knowledgeable about the relationship between a muscle and its bony attachment site. Therefore, this level of detail in recording is desirable, to allow for in-depth analysis of the occurring osseous changes. Due to the more complex nature of this method, it ideally requires training by someone proficient in the field to learn to distinguish between the two zones and the different traits. For fibrous entheses, which cannot be analysed using the Coimbra method, the Mariotti method can be retained. The issue remains, however, that we do not yet have a normal, 'no entheseal change' baseline for fibrous entheses (Villotte *et al*. 2016). The Mariotti *et al*. (2004, 2007) method does not provide a solution to this, so as of now it is a factor to be kept in mind. However, as fibrous attachment sites change in a less complex manner (i.e., not as prone to porosity, no two distinguishable physiological zones) the Mariotti method, with its three scored traits, robusticity, osteolytic lesions and osteophyte formation, is also intrinsically well suited to the analysis of fibrous entheses. Given its more limited terminology, relative simplicity, detailed descriptions and pictures per enthesis, it can be learned more rapidly and autodidactically, if a sufficiently large reference collection with variation in entheseal changes is available for training. A recent study by Milella *et al.* (2015) used only the factor of robusticity to analyse activity, with considerable success. However, the recording of bone formation and bone resorption, two major types of bony change used in all areas of osteoarchaeology, adds extra information that any well-trained osteologist is capable of observing correctly. Therefore, it can be suggested that these two other traits scored in the Mariotti *et al.* (2004, 2007) method should be reserved for more general studies of EC. Based on all the above, it can be suggested that the best course of action is to implement the Coimbra method for fibrocartilaginous entheses while reserving the Mariotti method for fibrous entheses.

Further findings

In both methods, it is noteworthy that the visualization (Fig. 4) of the male and female *M. triceps brachii*, *M. brachialis* and *M. biceps brachii* shows a sex difference that is obscured when the three muscle scores are grouped. The *M. triceps brachii* score is higher in females and the other two entheses score higher in males. This highlights the necessity of observing not only muscle groups, but also each muscle separately.

With the factor osteophyte formation, researchers must take into account that some entheses, such as the Achilles on the calcaneus, the *M. quadriceps femoris* on the patella and the *M. triceps brachii* on the ulna, are potentially more prone to new bone formation than other entheses. This has been noted by Villotte (2006), who grouped these three entheses as 'group 2' in his scoring method, identifying them as entheses that show enthesophyte formation at their edges frequently, and osteolytic lesions infrequently and very rarely in combination with osteophyte formation. Figure 2 shows how average EC scores for the Achilles on the calcaneus and *the M. quadriceps femoris* on the patella are relatively higher than for other entheses, but this must not be interpreted solely as evidence of more extensive use of these muscles. Physiological factors such as the existent bursae, location on the bone and the angle of pull will influence the EC scores regardless of the method used. Further research is necessary to evaluate the extent of differences, establish if and how this can be corrected for, and how to incorporate this when interpreting data. Both the sex differences in EC scores and bone formation differences between different entheses indicate that comparison between studies is only possible when the same muscle attachment sites are analysed.

CONCLUSION

In the current study, while some differences in EC results exist, the Mariotti and Coimbra EC recording methods show results that are broadly sufficiently similar to allow for comparison of general EC patterns between studies on the same entheseal sites using the different methods, but only when the intrinsic differences in score range, and hence possible differences in statistical significance, are taken into account. We argue that with careful consideration, a trend seen in a given sample using one method can be applied to discuss trends seen in another sample using the other method. Although more studies on different populations are necessary to bolster these findings, these results are very promising, as inter-method comparability would facilitate continuity within EC research. The general comparability shown by the current study serves as an extra boost to the positive impulse for the field of EC research generated by the new Coimbra method. EC research has already delivered tantalizing results (see, e.g., Eshed *et al.* 2004; Lieverse *et al.* 2009; Havelkovà *et al.* 2013; Palmer *et al.* 2014), making it an important field for future osteoarchaeological research. However, given the remaining lacunae in our knowledge, fundamental research and larger data sets are necessary before solid conclusions about past activities can be made from EC. By striving towards a standardized system of observation and scoring, larger comparisons that can allow solid activity-related patterns to emerge become possible. Future research should develop a standardized method to synthesize the EC data for statistical analyses, with an ideal analytical tool being one that limits or removes the effect of the score range, thereby permitting a more valid comparison of results generated from different methods.

ACKNOWLEDGMENTS

We would like to thank Wouter De Maeyer and Stijn Van Cauwenbergh at Solva archaeology for granting permission to use this skeletal collection, and the anonymous reviewers for their valuable feedback.

REFERENCES

Acosta, M. A., Henderson, C. Y., and Cunha, E., 2017, The effect of terrain on entheseal changes in the lower limbs, *International Journal of Osteoarchaeology*, **27**, 828–38.

Alves Cardoso, F., and Henderson, C. Y., 2010, Enthesopathy formation in the humerus: data from known age-at-death and known occupation skeletal collections, *American Journal of Physical Anthropology*, **141**, 550–60.

Angel, J. L., 1945, Skeletal material from Attica, *Hesperia*, **14**, 279–363.

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**, 231–9.

Davis, C. B., Shuler, K. A., Danforth, M. E., and Herndon, K. E., 2013, Patterns of interobserver error in the scoring of entheseal changes, *International Journal of Osteoarchaeology*, **23**, 147–51.

De Groote, K., De Maeyer, W., Moens, J., Quintelier, K., Van Cleven, F., Vanden Berghe, I., and Vernaeve, W., 2011, Het karmelietenklooster van Aalst (prov. Oost-Vl.) (1497–1797): het gebouwenbestand, de begravingen en het fysisch-antropologische onderzoek, *Relicta*, **8**, 83–250.

Dutour, O., 1986, Enthesopathies (lesions of muscular insertions) as indicators of the activities of Neolithic Saharan populations, *American Journal of Physical Anthropology*, **71**, 221–4.

Eshed, V., Gopher, A., Galili, E., and Hershkovitz, I., 2004, Musculoskeletal stress markers in Natufian hunter–gatherers and Neolithic farmers in the Levant: the upper limb, *American Journal of Physical Anthropology*, **123**, 303–15.

Ferembach, D., Schwindezky, I., and Stoukal, M., 1980, Recommendation for age and sex diagnoses of skeletons, *Journal of Human Evolution*, **9**, 517–49.

Gray, H., 1977, *Gray's Anatomy*, Bounty Books, London.

Havelkovà, P., Hladík, M., and Velemínský, P., 2013, Entheseal changes: do they reflect socioeconomic status in the early medieval Central European population? (Mikulčice–Klášteřisko, Great Moravian Empire, 9th–10th century), *International Journal of Osteoarchaeology*, **23**, 237–51.

Hawkey, D. E., and Merbs, C. F., 1995, Activity-induced musculoskeletal stress markers (MSM) and subsistence strategy changes among ancient Hudson Bay eskimos, *International Journal of Osteoarchaeology*, **5**, 324–38.

Henderson, C. Y., 2013, Subsistence strategy changes: the evidence of entheseal changes, *HOMO—Journal of Comparative Human Biology*, **64**, 491–508.

Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Lopreno, G., Villotte, S., and Wilczak, C., 2012, The effect of age on entheseal changes at some fibrocartilaginous enthuses, *American Journal of Physical Anthropology*, **147**(S54), 163.

Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Perréard-Lopreno, G., Villotte, S., and Wilczak, C., 2010, Scoring entheseal changes: proposal of a new standardised method for fibrocartilaginous entheses [online], Poster presented at the 18th European Meeting of the Paleopathology Association, Vienna, 23–26 August 2010; https://www.uc.pt/en/cia/msm/Vienna2010.pdf (accessed 10 August 2013).

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

Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Villotte, S., and Wilczak, C., 2016b, The new 'Coimbra method': a biologically appropriate method for recording specific features of fibrocartilaginous entheseal changes, *International Journal of Osteoarchaeology*, **26**, 925–32.

Henderson, C. Y., Mariotti, V., Santos, F., Villotte, S., and Wilczak, A., 2017, The new Coimbra method for recording entheseal changes and the effect of age-at-death, *Bulletins et Mémoires de la Société d'anthropologie de Paris*, **29**, 140–9.

Henderson, C. Y., Mariotti, V., Wilczak, C., and Santos, F., 2016a, The 'new Coimbra method' and the effect of age, Presented at the conference 'Working your Fingers to the Bone: An Interdisciplinary Conference on Identifying Occupation from the Skeleton', 6–8 July 2016, Coimbra, Portugal.

Henderson, C. Y., and Nikita, E., 2016, Accounting for multiple effects and the problem of small sample sizes in osteology: a case study focusing on entheseal changes, *Archaeological and Anthropological Sciences*, **8**, 805–17.

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**, 147–56.

Jurmain, R., Alves Cardoso, F., Henderson, C., and Villotte, S., 2012, Bioarchaeology's holy grail: the reconstruction of activity, in *A companion to paleopathology* (ed. A. L. Grauer), 531–52, Wiley-Blackwell, Chichester.

Lieverse, A. R., Bazaliiskii, V. I., Goriunova, O. I., and Weber, A. W., 2009, Upper limb musculoskeletal markers among Middle Holocene foragers of Siberia's Cis-Baikal region, *American Journal of Physical Anthropology*, **138**, 458–72.

Maat, G. J., 2001, Diet and age-at-death determinations from molar attrition: a review related to the Low Countries, *Journal of Forensic Odontostomatology*, **19**(1), 18–21.

Mariotti, V., Facchini, F., and Belcastro, M. G., 2004, Enthesopathies: proposal of a standardized scoring method and applications, *Collegium Antropologicum*, **28**, 145–58.

Mariotti, V., Facchini, F., and Belcastro, M. G., 2007, The study of entheses: proposal of a standardised scoring method for twenty-three entheses of the postcranial skeleton, *Collegium Antropologicum*, **31**, 291–313.

McCormick, W. F., Stewart, J. H., and Greene, H., 1991, Sexing of human clavicles using length and circumference measurements, *American Journal of Forensic Medicine and Pathology*, **12**, 175–81.

Meindl, S., and Lovejoy, O., 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.

Michopoulou, E., Nikita, E., and Henderson, C. Y., 2017, A test of the effectiveness of the Coimbra method in capturing activity-induced entheseal changes, *International Journal of Osteoarchaeology*, **27**, 409–17.

Michopoulou, E., Nikita, E., and Valakos, E. D., 2015, Evaluating the efficiency of different recording protocols for entheseal changes in regards to expressing activity patterns using archival data and cross-sectional geometric properties, *American Journal of Physical Anthropology*, **158**, 557–68.

Milella, M., Alves Cardoso, F., Assis, S., Perreard Lopreno, G., and Speith, N., 2015, Exploring the relationship between entheseal changes and physical activity: a multivariate study, *American Journal of Physical Anthropology*, **156**, 215–23.

Milella, M., Belcastro, M. G., Zollikofer, C. P. E., and Mariotti, V., 2012, The effect of age, sex, and physical activity on entheseal morphology in a contemporary Italian skeletal collection, *American Journal of Physical Anthropology*, **148**, 379–88.

Molnar, P., Ahlstrom, T. J., and Leden, I., 2011, Osteoarthritis and activity—an analysis of the relationship between eburnation, musculoskeletal stress markers (MSM) and age in two Neolithic hunter–gatherer populations from Gotland, Sweden, *International Journal of Osteoarchaeology*, **21**, 283–91.

Niinimäki, S., Niskanen, M., Niinimäki, J., Nieminen, M., Tuukkanen, J., and Junno, J. A., 2013, Modeling skeletal traits and functions of the upper body: comparing archaeological and anthropological material, *Journal of Anthropological Archaeology*, **32**, 347–51.

Palmer, J. L. A., Hoogland, M. L. P., and Waters-Rist, A. L., 2014, Activity reconstruction of post-medieval Dutch rural villagers from upper limb osteoarthritis and entheseal changes, *International Journal of Osteoarchaeology*, **26**, 78–92.

Palmer, J. L. A., Waters-Rist, A. L., and Lieverse, A., 2017, A proposed method for scoring subadult entheseal morphology [online], Poster presented at the 86th Annual Meeting of the American Association of Physical Anthropology, New Orleans, 19–22 April 2017; https://www.researchgate.net/publication/319879136_A_proposed_method_for_scoring_subadult_entheseal_morphologyDOI:10.13140/RG.2.2.32815.64160 (accessed 18 September 2017).

Paulsen, F., and Waschke, J. (eds.), 2011, *Sobotta atlas of anatomy*, Elsevier, Munich.

Phenice, T. W., 1969, A newly developed visual method of sexing the os pubis, *American Journal of Physical Anthropology*, **30**, 297–302.

Robb, J., 1998, The interpretation of skeletal muscle sites: a statistical approach, *International Journal of Osteoarchaeology*, **8**, 363–77.

Schaefer, M., Black, S., and Scheuer, L., 2009, *Juvenile osteology: a laboratory and field manual*, Academic Press, San Diego, CA.

Stewart, T. D., 1979, *Essentials of forensic anthropology*, Charles C, Thomas, Springfield, IL.

Steyn, M. Y., and Işcan, M. Y., 1999, Osteometric variation in the humerus: sexual dimorphism in South Africans, *Forensic Science International*, **106**, 77–85.

Suchey, S., and Brooks, J. M., 1990, Skeletal age determination based on the os pubis: a comparison of the Acsádi–Nemeskéri and Suchey–Brooks methods, *Human Evolution*, **5**, 227–38.

Villotte, S., 2006, Connaissances médicales actuelles, cotation des enthésopathies: nouvelle méthode, *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, **18**, 65–85.

Villotte, S., Assis, S., Alves Cardoso, F., Henderson, C. Y., Mariotti, V., Milella, M., Pany-Kucera, D., Speith, N., Wilczak, C. A., and Jurmain, R., 2016, In search of consensus: terminology for entheseal changes (EC), *International Journal of Paleopathology*, **13**, 49–55.

Villotte, S., Castex, D., Couallier, V., Dutour, O., Knüsel, C. J., and Henry-Gambier, D., 2010, Enthesopathies as occupational stress markers: evidence from the upper limb, *American Journal of Physical Anthropology*, **142**, 224–34.

Weisberg, H., 1992, *Central tendency and variability*, SAGE Publications, Newbury Park, CA.

Weiss, E., 2007, Muscle markers revisited: activity pattern reconstruction with controls in a central California Amerind population, *American Journal of Physical Anthropology*, **133**, 931–40.

Weiss, E., 2015, The surface of bones: methods of recording entheseal changes, *Surface Topography: Metrology and Properties*, **3**, 1–1.

Wells, C., 1963, Hip disease in ancient man: report of three cases, *Journal of Bone and Joint Surgery*, **45B**, 790–1.

Wilczak, C. A., 1998, Consideration of sexual dimorphism, age, and asymmetry in quantitative measurements of muscle insertion sites, *International Journal of Osteoarchaeology*, **8**, 311–25.