A sense of society: entheseal change as an indicator of physical activity in the Post-Medieval Low Countries: potential and limitations

Palmer, J.L.A.

Author: Palmer, J.L.A.
Title: A sense of society: entheseal change as an indicator of physical activity in the Post-Medieval Low Countries: potential and limitations
Issue Date: 2019-03-20

4

A recording method for sixteen nonadult muscle entheses

HOMO Journal of Comparative Human Biology

Under review

Authors:
J. L. A. Palmer
A. R. Lieverse
M. L. P. Hoogland
A. L. Waters-Rist

Abstract

Bioarchaeology lacks a system for recording the morphology of muscle and ligament attachment sites, called entheses, in growing individuals. Therefore, a new method is necessary to analyze this in nonadult individuals. Such information could eventually be useful in investigating factors that affect bone growth and development, including sex, age, puberty, pathology, and activity. To begin the creation of such a recording system this pilot study on 29 individuals of known age and sex (a) assesses the range of osseous changes of 16 entheses in the upper and lower limbs of growing individuals, and (b) presents a scoring method based on these findings that can be used by other researchers. The pilot results show no unidirectional relation between entheseal change (EC) and age, nor a clear relationship with sex. Ultimately, this method will allow researchers to evaluate if and how EC correlates with growth and development and factors that may affect those processes in nonadult remains from any archaeological or forensic context.

4.1 Introduction

Children provide a unique window into past cultures and societies. The way a society treats those who are, to a certain extent, dependent upon help, offers invaluable insights into the workings of humanity in a certain era and area. Bioarcheology can add unique data to our understanding of the life of growing individuals by analyzing the lesions and anatomical variations occurring in their skeleton. Currently, bioarchaeological research into individuals who have not yet reached skeletal maturity is often limited to the estimation of age-at-death, growth patterns, diet, and the assessment of pathology. The reasons for this are largely methodological. As the growing, developing skeleton has distinct physiological properties, many osteological methods, developed on adult human remains, cannot be applied (Ruff *et al.*, 2013). Additionally, osseous changes which take a long time to develop, or that emerge because of puberty, will not commonly occur in younger individuals. Furthermore, there has been a persistent misconception in osteoarchaeology that nonadult (<18 years) skeletal material is always more susceptible to taphonomic damage and less likely to preserve than adult material (e.g. Gordon and Buikstra 1981, Walker *et al* 1988). While this is true in some cases and soil conditions, in many cases it is not a problem. In past osteoarchaeology research this perception has led to an underrepresentation of studies of nonadults.

In the last two decades, there has been a major impetus in the study of nonadult remains, supported by the creation of well-structured manuals and books (i.e., Baker *et al.*, 2005; Scheuer and Black 2000; Schaefer *et al.*, 2009, Lewis 2017). Methods for estimating age-at-death (e.g., Primeau *et al.*, 2016) and sex (e.g., Stull *et al.*, 2017) of nonadults are being actively refined. In addition, aspects of childhood pathology and diet are becoming increasingly well-understood (e.g., Novak *et al.*, 2017, Stark 2014, Reitsema *et al.*, 2016, Tsutaja *et al.*, 2015). Less has been done in regard to investigating patterns of nonadult activity, in contrast to the vast amount of research on adult activity patterns (e.g., Hawkey and Merbs 1995, Henderson *et al.*, 2016, Eshed *et al.*, 2004, Milella *et al.*, 2015). Such knowledge will be important to gain for an improved understanding of factors related to nonadult growth but also of the development of entheseal changes (EC) in adults.

The current research aims to add to our growing osteological understanding of the lives of nonadults by analyzing and creating a method to assess variation in entheses within growing individuals. This is done using 29 known age and sex nonadult remains from a Dutch post-medieval cemetery, ranging in age from two to 17 years, free from macroscopic signs of skeletal pathology or abnormality. Assessment of intra- and inter-observer consistency in scoring is conducted. These results are critical in demonstrating this method that can be adopted by untrained osteoarchaeology researchers working on skeletal material from any spatiotemporal context. The EC results will also be assessed for correlation with age and sex. Our primary aim is to provide a standardized recording system which can be used by osteoarchaeologists to study how nonadult EC relates to growth and development, which in turn may be affected by variables such as sex, age, puberty, pathology, and physical activity. Until a standardized method for recording nonadult EC is established for osteoarchaeology we are missing out on a potentially valuable data source that could improve our understanding of biological and sociocultural variation during nonadulthood.

This article is structured to follow the method creation process as it occurred. This is done to allow readers to assess all decisions made in the method creation process and provide maximum transparency on the steps that led to the final proposed method.

4.2 Materials

The method is developed on the skeletal collection from the post-Medieval village of Middenbeemster, curated at the Human Osteology Laboratory of Leiden University. The collection is chosen for its good preservation, well understood historical (Falger *et al.*, 2012) and osteoarchaeological context (Griffioen *et al.*, 2012, Waters-Rist and Hoogland 2013, Lemmers *et al.*, 2013, Veselka *et al.*, 2015, Palmer *et al.*, 2016, Ziesemer *et al.*, 2015, Caroll *et al.*, 2016, Vikatou *et al.*, 2017), and the availability of individuals of known sex and age-at-death. Nonadults aged two to 17 years are used. Individuals younger than two years are omitted as the visible ECs are too subtle to be included in the same method. Individuals over 17 years are omitted because, even though full skeletal maturity is not yet reached, patterns of EC in these individuals better fit within existing adult methods (e.g. Hawkey and Merbs 1995, Henderson *et al.*, 2016, Mariotti *et al.*, 2004, Vilotte 2006). A total of 29 archivally identified individuals from the established age bracket, of sufficient completeness and good to excellent preservation, were available for analysis to assess the spectrum of EC (table1). This sample consists of seventeen girls and eleven boys.

	Archival age in years													
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**16**	**n**
Sex?			1											1
Female	1		4	1	1	1	2	3	1	1	1	1		17
Male		2	2	3	1		1					1	1	11
Total														29

Table 1: Individuals used for method creation, with their age and sex (n=29). For one 4-year old individual, sex is unknown.

4.3 Anatomy

An enthesis is a complex anatomical area where two different plastic and mobile tissues are made to connect to each other (Benjamin *et al.*, 2006). This junction poses inherent challenges to the body, which are further increased by force loading and force concentration (Benjamin *et al.*, 2006). The body accommodates these connections by creating a discrete site of attachment with a specific topography (Thomopoulos *et al.*, 2013), in the shape of tuberosities, processes, spines, ridges, crests, lines, fossæ, pits, depressions, grooves, furrows, fissures, or notches (Gray and Lewis 1918). Development of the entheses starts in utero in the embryonic stage (Zelzer *et al.*, 2014) and is completed when skeletal maturity is reached. Once an individual is fully grown, the entheses remains a plastic and changeable area, which can then be analyzed with the abovementioned adult methods.

Prior to skeletal maturation, bone and soft tissue interact in a specific fashion to allow for growth. Long bones in human nonadults, as well in the animals from which many bone growth models are created, increase in size and length primarily at the epiphyseal growth plates. Muscle and ligament attachment sites get dragged along this growing bone by the periosteum (Matyas *et al.*, 1990), thus ensuring that the attachment always stays at the same relative location, rather than being 'left behind' (Dörfl 1980a, b). The growing body maintains a fine balance of growth to allow for this precise maintenance of the relative location of tendon and ligament insertions (Stern *et al.*, 2015). The mechanical pull of muscles does not influence this isometric scaling system (Dörfl 1980b). This implies that, except in cases of premature closure of either the distal or proximal growth plate, or other developmental issues, nonadult muscles attach at the same locations on growing bone as they do in an adult skeleton (e.g., M. pectoralis major will always attach on the antero-lateral upper 1/3 of the humeral diaphysis).

Thus, although the location of an entheses is possible to find using adult anatomy handbooks, the actual appearance and variation visible at the entheseal site differs significantly between adults and nonadults. Both the underlying developmental process outlined above and the macroscopically visible enthesis on dry bone support this. Several osteoarchaeological researchers have observed how muscle and ligament attachments present differently in growing individuals; Castex (1980) noticed a high frequency of individuals aged 11-18 with attachments for the M. Pectoralis Major and M. Teres Major presenting in the shape of a fossa, as well as a higher frequency of hypotrochanteric fossa and rhomboid fossa in growing individuals from 5th to 6th Century AD Chalon-sur-Saône in France. These fossae have been interpreted as a result of an active phase of growth (Saunders 1978), and, in the case of the hypotrochanteric fossa and rhomboid fossa, recorded as non-metric traits when they are retained in adulthood (Finnegan 1978). Nonadults also frequently presented with fossae at entheseal sites in studies by Stirland (1996) and Mariotti and colleagues (2004). Villotte and Knüsel (2013: 141) stated that these fossae are "very common in immature human skeletons, in frequency but also in their distribution in the body". The current method aims to allow standardized observation of a selection of attachment sites for ligaments, fibrous muscle attachments, and fibrocartilaginous muscle attachment sites (see section 4.1), to further explore this attested phenomenon and to begin documenting the pattern of entheseal development at attachment sites.

4.4 Method creation

The initial development of this method occurred in four phases; selection of suitable entheses, observational analysis of entheses, draft method creation, and revised method creation. These phases are outlined in this section. The thus created method was then tested (section 6), and the results of these tests (section 7) were used to create the final method version (section 8).

4.4.1 Entheseal selection

Entheses were selected that would be most suited to analysis of their potential changes, based on the ability to distinguish and delineate them macroscopically on nonadult remains in the 2-17 year age range. Fibrous entheses predominately fulfilled these criteria, as fibrocartilaginous entheses are commonly close to the epiphyses of the bone and therefore obscured and complicated by the presence of the growth plates in a developing individual. This led to the selection of sixteen muscle and ligament attachment sites on seven bones (table 2).

Bone	**Muscle/ligament**	**Tissue type at site of attachment**
Scapula	Long head of M. triceps brachii	Fibrocartilaginous muscle attachment
Clavicle	Costoclavicular	Ligament attachment
	Conoid	Ligament attachment
	Trapezoid	Ligament attachment
Humerus	M. pectoralis major	Fibrous muscle attachment site
	M. latissimus dorsii/M. teres major	Fibrous muscle attachment site
	M. deltoideus	Fibrous muscle attachment site
Radius	M. pronator teres	Fibrous muscle attachment site
Ulna	M. pronator quadratus	Fibrous muscle attachment site
	M. brachialis	Fibrocartilaginous muscle attachment
Femur	M. gluteus maximus	Fibrous muscle attachment site
	M. vastus medialis	Fibrous muscle attachment site
	M. pectineus	Fibrous muscle attachment site
	M. gastrocnemius medial head	Fibrocartilaginous muscle attachment
	M. gastrocnemius lateral head	Fibrocartilaginous muscle attachment
Tibia	M. soleus	Fibrocartilaginous muscle attachment

Table 2: overview of the ligament and muscle entheses used, the bone they are located on and their anatomical attachment type.

4.4.2 Observation

The selected entheses were described in detail in ten individuals to assess which osseous changes can occur at each enthesis and distill the range of possible traits at nonadult entheses. Then, the observable traits per enthesis from all ten individuals were compiled into a list to form a reflection of the spectrum of EC. Room was left for expansion of this list if dictated by the subsequent stages.

4.4.3 Draft method creation

From this list of possible EC, an ordinal scoring scale was created for each enthesis. The reasoning behind this was that, because each muscle has its own distinctive size and shape, each enthesis can change in its own unique manner due to intrinsic factors such as its size and location on the skeleton, the angle of the bone, the angle of the muscle, and whether it is an origin or insertion site. Thus, separate descriptions were deemed necessary per enthesis incorporated in this method. A graded score was created for each enthesis, ranging from stage zero, meaning no observable EC, to stage three, which describes maximal EC. The first version of this nonadult EC scoring method ascribes a stage zero to entheses where the attachment site either cannot be delineated, or where delineation is just possible, depending on the enthesis under study (see appendix B1).

When testing this first version of the scoring method, it became clear that it often caused the observer to be torn between two possible scores (e.g. score 1 or 2, or score 2 or 3), thus allowing for a lot of potential intra- and interobserver error (Palmer *et al.*, 2017). Scoring in this manner also does not allow researchers to record and assess the specific osseous characteristics of each enthesis. It was therefore decided to deconstruct the method back to the list of potential osseous characteristics. From this list, a scoring form was created allowing the researcher to score, per observed enthesis in an individual, all the occurring osseous traits. Thus, seven characteristics were retained: enthesis delineation, porosity, surface irregularity/rugosity, ridges, depression/sulcus, presence of a distinct tubercle, and enthesophytes (defined below).

4.4.4 Revised method creation

For the next step of the method creation process, for the seven retained characteristics, score categories were created. Some osseous characteristics were simply scored as present (score of 1) or absent (score of 0). Other characteristics were scored in tripartite, ranging between no expression of the trait (score 0), medium expression of the trait (score 1), to maximal expression of the trait (score 2). Distinct thresholds were described for each score category to ensure maximal reproducibility by other scholars and to decrease intra-observer error. Section five defines the seven entheseal osseous characteristics, with section 5.8 explaining how their range of variability was categorized in scores zero to two. The method was then tested in this form to evaluate which ECs occur per specific enthesis, and to explore the presence of correlations with age and sex presented itself (section 7). Following the intra-and inter-observer agreement testing (section 8), the method was re-revised to improve utility and repeatability and presented it in its final form (section 9).

4.5 Terminology and scoring of traits

This section will first define the traits as they are used throughout this paper and in all stages of method creation. Section 5.8 will then present the scoring system as it was used for testing of trait expression and inter-and intra-observer agreement.

4.5.1 Entheseal delineation

Entheseal delineation is defined as whether an entheses that can be delineated, by visual identification, bone contour, palpation, or any of the traits below. Note: if entheseal delineation is not possible, all other traits score zero by default.

4.5.2 Porosity

Porosity indicates that the cortical bone surface appears permeable rather than solid and compact. In adult bone, a distinction is usually made between macro- and microporosity when evaluating this trait (e.g. Buckberry and Chamberlain (2002) for auricular surface age estimation, Henderson *et al.*, 2016 for EC). For nonadult bone, which is consistently more porous than adult bone, this distinction is not useful. Therefore, porosity is identified as any texture change that causes the entheseal area to appear porous, be it as a small patch of pinprick holes, diffuse holes, or larger pores.

4.5.3 Surface irregularity/rugosity

Any rough bone formation visible or palpable on the entheseal surface that does not present distinct enthesophytes, ridges, or tubercles (see below).

4.5.4 Ridges

This is the formation of a longitudinal line of elevated bone, creating a distinct ridge. The ridge can present a sharp or rugose 'peak', or a blunt/rounded elevated line.

4.5.5 Depression/sulcus

This is any indentation or cavity in the entheseal area, of any shape, which causes the enthesis (or a portion of it) to present a lower plane than the surrounding bone.

4.5.6 Presence of a distinct tubercle

This is any observable distinct tubercle, independent of size but larger than an enthesophyte (i.e. ca. >2mm). The surface of the tubercle can be scored for the other characteristics.

4.5.7 Enthesophytes

Any new bone formation that is a distinct nodule, spur, or bony projection <2mm.

4.5.8 Scoring classification

Table 3 represents the scoring classification as it was implemented for inter-and intra-observer agreement testing, prior to the final method creation.

Entheseal delineation	0 = entheses cannot be delineated 1 = entheses can be delineated
Porosity	0 = porosity covers less than half of the surface 1 = covers more than half of the delineated entheses
Surface irregularity/rugosity	0 = absent 1 = covers less than half of the delineated entheses 2 = covers more than half of the entheses
Ridges	0 = absent 1 = ridge presents a blunt/rounded elevated line 2 = ridge presents a sharp or rugose 'peak'
Depression/Sulcus	0 = absent 1 = edges of the depression are rounded and smooth 2 = edges of the depression/sulcus are sharp
Tubercle	0 = absent 1 = present
Enthesophytes	0 = absent 1 = only 1 enthesophyte observed 2 = 2+ enthesophytes observed

Table 3: initial scoring classification

Note that porosity was scored for EC only if it is not inherent to the anatomical area. For some entheses, the attachment site is always porous, due to the intrinsic properties of the bone. Of the sixteen entheses included in the method, only the lateral and medial heads of the gastrocnemius are always porous. Therefore, for these two entheses, porosity was not included in the scored traits for method testing.

4.6 Method testing

As a first test of the method, the scorable EC traits were structured into a recording form and the method tested on the 29 individuals of known age and sex by the first author. When scoring was impossible due to missing elements or taphonomic damage, a dash was put on the form. All observations were done near a window using natural light, without magnification. This testing of the method was done to evaluate the expression of the defined traits in the individual entheses and to explore how the EC interacted with the archival age and sex data.

As a second test of the method, the data were re-analyzed by the same observer for intra-observer agreement assessment, and by two other researchers to assess inter-observer agreement. The researchers also provided feedback on their experience with the method prototype. These tests of the method were done to assess and improve method usability and repeatability. Only after these tests was the final method created (section 9).

4.7 Results

4.7.1 Trait expression per enthesis

To assess EC trait expression, left and right entheses were scored for all individuals. Thus, a maximum of 58 entheses could be scored per specific muscle attachment site, given our sample collection of 29 individuals. The number of individuals and entheses observed varies due to completeness and preservation, so "all individuals" in the entheses subsections refers to all individuals for whom the left, right, or both entheses could be observed. Exact numbers of observable entheses are given per subsection. For full detailed data, see appendix B2.

The long head of the M. Triceps Brachii could always be delineated (n=19) with a total of 36 observable entheses. The enthesis was delineated based on the contour of the bone and its distinct location on the lateral edge of the scapular blade just below the glenoid. Porosity is common (53%), as is surface irregularity (69%). Tubercle formation (8%), depression (6%), and ridge formation (6%) were rarer.

The costoclavicular ligament attachment on the clavicle could be delineated in all but one individual, with a total of 39 scorable entheses (n=22). Porosity was present in 64% of the 39 observed entheses, surface irregularity in 44% and ridge formation in 18%. Depressions are relatively common (51%), often with sharp edges (31% of cases of depression). Tubercle formation was observed in one instance in a 5-year old boy (2.6%). Enthesophyte formation and surface flattening were not observed.

The conoid could always be delineated (n=19) with a total of 33 analyzable entheses. Thirty percent of entheses showed porosity. Surface irregularity is very common (79%) and tubercle formation is relatively common (27%). Ridge formation (6%), surface flattening (3%) and depressions (9%) were much less common.

The trapezoid could always be delineated (n=20) with a total of 36 analyzable entheses. Porosity is present in one third (33%) of observed entheses. Surface irregularity was observed in 44% of cases, and ridge formation in 14%. Depression is common (72%), usually in combination with porosity and/or surface irregularity. No enthesophyte formation or surface flattening was observed.

The M. Pectoralis Major could always be delineated (n=22), with a total of 42 analyzable entheses. Fifty-seven percent of entheses showed porosity. Surface irregularity (57%)

and ridge formation (50%) are common, whereas surface flattening and depression were both only observed in 9% of cases. No tubercle formation of enthesophyte formation was observed.

The M. Latissimus Dorsii/Teres Major could be delineated in all individuals (n=22) with a total of 38 analyzable entheses. In one individual, a 10-year old female, the enthesis could be delineated on the right but not on the left. Porosity was observed in 53% of entheses. Surface irregularity was less common (16%), as was ridge formation (8%). Surface flattening (32%) and depression formation (53%) are relatively more common at this enthesis.

The M. Deltoideus could be delineated in all analyzed individuals (n=22) with a total of 37 analyzable entheses. Porosity (46%) and surface irregularity (49%) were common, with some instances of ridge formation (30%) but only two instances of depression (5%).

The M. pronator teres could be delineated in all analyzed individuals (n=21), but only on one side in two cases (94% delineable). A total of 34 entheses could be analyzed. Surface flattening (74%) and porosity (50%) are common, and some individuals show depressions (26%). No ridge, tubercle, or enthesophyte formation was observed.

The M. pronator quadratus could be delineated in all analyzed individuals (n=18), with a total of 30 analyzable entheses. Entheseal change was subtle, with porosity and surface irregularity occurring most commonly.

The M. brachialis could be delineated in all analyzed individuals (n=20) with a total of 37 analyzable entheses. One third exhibit porosity (30%). Surface irregularity is less common (14%) than ridge formation (27%). Some instances of surface flattening were observed (14%) but depressions were more prevalent (54%). Enthesophyte formation was rare (5%) and no tubercle formation was observed.

The M. gluteus maximus could be delineated in all analyzed individuals (n=26) with a total of 38 analyzable entheses. Porosity was relatively uncommon (21%), while surface irregularity (39%) and ridge formation (61%) were observed more frequently. Surface flattening (8%) was observed less often than depression, which was observed relatively frequently (34%), sometimes with sharp edges (31% of depressions). Tubercle and enthesophyte formation were not observed.

Delineation of the M. vastus medialis was possible in 14 of 26 analyzable individuals (53%), although in 3 of these 14 nonadults, delineation was only possible on one side. At the 44 analyzable entheses, the only observed traits were porosity (5%) and depression (18%).

The M. pectineus could be delineated in all but two individuals (n=29; 93%), with a total of 41 analyzable entheses. Porosity was commonly observed (80%). In more rare instances, surface irregularity (17%) or ridge formation (2%) were observed. No other osseous characteristics were observed in these individuals.

The medial head of the M. gastrocnemius could be delineated in all individuals (n=26) with a total of 45 analyzable entheses. Porosity was not scored in these individuals as it was always present on the entire attachment area. Surface irregularity was common (78%), often covering more than half of the entheseal area (in 40% of cases). Depression formation was also common, sometimes with sharp edges (in 18% of cases). In some instances (9%), ridge formation was observed along the edge of the depression. Surface flattening, tubercle, and enthesophyte formation were not observed.

The lateral head of the M. gastrocnemius on the femur could be delineated in all analyzed individuals (n=25) with a total of 41 analyzable entheses. As with the medial head of the M. gastrocnemius, porosity was not scored. Surface irregularity was common (41%) sometimes covering more than half of the entheseal surface (24% of cases). Depressions were commonly observed (66%), with a sharp edge in one instance (4%). Ridge formation only occurred in one individual (2%), along the edge of the depression.

The M. Soleus entheses could be delineated on all individuals (n=25) with a total of 50 analyzable entheses. Porosity was the most commonly observed trait (88%), sometimes covering more than half of the attachment area (23% of cases). Depression formation was also common (66%), sometimes with sharp edges (12%). Some instances of surface irregularity (22%) and ridge formation (10%) were observed, but no tubercles or enthesophyte formation.

From these data it is clear that every enthesis has its own unique range of expression. Thus, it is not appropriate to compare results from different entheses to each other. For instance, the M. Pectineus will never show as much variation in new bone formation or lytic activity as the M. Pectoralis Major; therefore, a lower score on the first does not mean it was less developed or used than the latter. For this reason, expression of each entheses is discussed separately.

4.7.2 Entheseal changes, sex, and age

No clear pattern between EC results (in the form of a composite score) and sex appears for any of the 16 analyzed entheses, whether regarding the data or through statistical testing (appendix B3 and table 4). As physical maturation rates differ between boys and girls, establishing if and how EC relates with sex is going to take considerable research. The current preliminary EC data do not show a clear relationship with sex, but this needs more substantiation from future studies.

	Test statistic	p	n
Long head of M. triceps brachii	1.53	0.675	18
Costoclavicular	2.768	0.597	21
Conoid	2.057	0.561	18
Trapezoid	1.515	0.824	20

M. pectoralis major	0.924	0.82	21
M. latissimus dorsii/M. teres major	5.931	0.052	21
M. deltoideus	2.489	0.288	21
M. pronator teres	0.267	0.606	20
M. pronator quadratus	2.137	0.343	17
M. brachialis	5.964	0.113	19
M. gluteus maximus	7.051	0.133	25
M. vastus medialis	0.135	0.935	25
M. pectineus	5.741	0.125	24
M. gastrocnemius medial head	3.634	0.603	25
M. gastrocnemius lateral head	2.625	0.622	24
M. soleus	2.143	0.543	24

Table 4: Chi square tests for relation between sex and EC based on the composite score per enthesis.

In relation to age, for the specific EC osseous characteristics, only tubercle formation at the conoid ligament attachment on the clavicle stands out as appearing more common with increasing age (see appendix B3). Of all entheseal osseous characteristics this tubercle formation is the most likely to be related to the skeletal maturation process, as the formation of bone eminences is, at least in part, regulated by the expression of specific genes involved in growth and development (Zelzer *et al.*, 2014).

Pearson's correlation tests between age and EC did not reveal a consistent pattern of correlation between EC and age in the sample. The Long head of the M. triceps brachii, the conoid, and both gastrocnemius site showed a positive correlation with age, whereas the M. pectoralis major showed a negative correlation with age (table 5).

	Correlation factor	**p**	**n**
Long head of M. triceps brachii	.613	0.005	19
Costoclavicular	-0.183	0.416	22
Conoid	0.492	0.033	19
Trapezoid	0,07	0.762	21
M. pectoralis major	-5.27	0.012	22
M. latissimus dorsii/M. teres major	-0,.93	0.186	22
M. deltoideus	-0.288	0.194	22
M. pronator teres	-0.074	0.749	21
M. pronator quadratus	-0.362	0.14	18
M. brachialis	0.415	0.069	20
M. gluteus maximus	-0.233	0.253	26
M. vastus medialis	0.302	0.133	26
M. pectineus	-0.177	0.398	25
M. gastrocnemius medial head	0.552	0.003	26
M. gastrocnemius lateral head	0.483	0.014	25
M. soleus	0.222	0.286	25

Table 5: Pearson's correlation tests between age and EC based on the composite score per enthesis. Statistically significant correlations ($p<0.05$) highlighted in bold.

The inconsistency of the relation between age and EC suggests that at this point we do not yet comprehend sufficiently how EC interact with age and/or other etiological factors to accurately analyze relations in this way. The relatively small sample size, especially when divided up per enthesis, and the less than straightforward dataset of this exploratory study do not allow us to draw founded conclusions on correlation with age. As EC data are complex, and, in nonadults, not yet thoroughly understood, further research is necessary to understand the interaction between EC and age.

4.7.3 Inter-and intra-observer agreement

4.7.3.1 Intra-observer agreement

For intra-observer agreement assessment, nine individuals were re-analyzed by the same researcher a month after the original data collection. Individuals were selected for re-analysis based on their level of completeness, with a variety of ages. Data are presented as percentages of agreement per trait and per enthesis in table 6 . When an enthesis was deemed not recordable in one assessment yet scored in the other, this difference was not included in the percentage calculation because it represents the assessment of the entire enthesis rather than observation of the specific trait. The percentage of agreement as to whether an enthesis is scorable is calculated separately (row entitled Entheses Observable).

	Long head of triceps brachii	Costaclavicular ligament	Conoid	Trapezoid ligament	M. pectoralis major	M. Latissimus dorsii/teres major	M. Deltoideus	M. pronator teres	M. Pronator quadratus	M. brachialis	M. Gluteus Maximus	Vastus Medialis	Pectineus	M. gastrocnemius, medial head	M. gastrocnemius, lateral head	M. Soleus
Delineation possible	100	94	100	100	100	88	100	100	100	93	100	85	80	100	93	100
Porosity	86	81	71	76	71	75	69	55	70	93	92	100	53	/	/	83
Surface irregularity	43	94	86	76	65	88	75	82	60	71	62	100	93	94	93	83
Ridge formation	79	100	93	88	65	94	88	100	100	100	69	100	93	88	93	100
Surface flattening	100	100	100	94	88	75	100	73	100	71	62	77	100	100	100	94
Depression	100	88	100	71	82	88	100	73	100	64	38	69	100	94	87	83
Tubercle formation	93	100	93	88	100	100	100	100	100	100	100	100	100	100	100	100
Enthesophyte formation	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
n =	14	16	14	17	17	16	16	11	10	14	13	13	15	16	15	18
Entheses observable	94	94	94	100	100	94	94	83	94	94	89	89	94	100	94	100

Table 6: Percentage of intra-observer agreement when scoring the same enthesis with a one-month interval. All values are percentages except the n-values, which represent the number of entheses scored (9 re-analyzed individuals so a maximum n=18 entheses per ligament/muscle)

Intra-observer agreement was generally high, with 89% on average over all traits and entheses. It must be noted that this agreement percentage is inflated by traits that are highly unlikely to occur at particular entheses and therefore consistently scored as zero (e.g.

enthesophyte formation at the M. pectoralis major). This is a factor for the overall agreement percentage of all entheses discussed for intra-observer error. Inconsistencies where intra-observer agreement was lower than 75% are discussed. Therefore, per enthesis, the more divergent scores between the two moments of observation are discussed, i.e., where intra-observer agreement was less than 75% for a specific enthesis on all scored individuals.

For the long head of the M. triceps brachii on the scapula, there was inconsistency both as to the extent (score 1 or 2) and presence (score 0 or 1) of surface irregularity, y*et all* other traits showed minimal difference between the two scoring instances. On average for this enthesis, intra-observer agreement was 88%. On the clavicle, intra-observer agreement is high for the costoclavicular ligament, with average agreement reaching 94%. The conoid ligament attachment site has some discrepancy for porosity scores, mainly as to the severity (score 1 or 2), with an overall average agreement of 93%. For the trapezoid ligament, some discrepancies occurred for porosity, surface irregularity, and depression formation. Here, the average agreement was 88%. Of the three entheses analyzed on the humerus, the M. pectoralis major was the most prone to disagreement, with some observers scoring as surface irregularity what others scored as ridge formation (average agreement 86%). The M. latissimus dorsii/teres major showed higher consistency (average 89%), while the M. deltoideus showed some disagreement as to the presence of porosity (score 1 vs. 0) with an average agreement of 92%. For the M. pronator teres on the radius, porosity was the least consistently scored, with some instances of disagreement between depression formation and surface flattening. Overall agreement was 85%. On the ulna, some disagreement occurred as to the presence of porosity and surface irregularity of the M. pronator quadratus (average agreement 92%), whereas for the M. brachialis surface flattening and depression were interchanged at times, and surface irregularity only found in the second scoring round. Overall agreement for the M. brachialis was 87%. On the femur, the M. gluteus maximus showed the lowest intra-observer agreement (average 79%), with surface irregularity and ridge formation being interchanged in some instances, as well as surface flattening and depression. The M. vastus medialis was scored as having a depression in the first round of observation more often than in the second round (average agreement 91%). For the M. pectineus, agreement was good (average 90%) for all traits except porosity. Both the medial and lateral head of m. gastrocnemius showed high intra-observer agreement (on average 86% and 85%, respectively). Finally, on the tibia, the M. soleus had good overall intra-observer agreement, with an average of 94%.

Overall, this assessment shows that while intra-observer error is, on average, low, some traits are more consistently scored than others. Specifically, when two traits are on a more continuous spectrum, they can be scored under one trait in one instance and the other in the second instance (i.e., depression vs. surface flattening).

4.7.3.2 Inter-observer agreement

For inter-observer error, two osteoarchaeologists applied the method, using only a scoring guide that defined the traits and score thresholds (as in section 5), provided images showing the location of the entheses, and supplied a scoring form. Their results were tested against those obtained by the first author (referred to here as observer one). The testing observers had PhDs in osteoarchaeology, with observer two being experienced with scoring EC in adults, and observer three unfamiliar with EC scoring. The observers scored five individuals from various age groups. They were not trained by the method creators in any other way than to give them the written scoring guide. Applying a method without being trained by the original developers can be challenging (Davis *et al.*, 2013; Wilczak *et al.*, 2017), thus the goal of this un-aided testing was to assess whether the descriptions were detailed and clear enough to be understood by other researchers, and whether the scoring method would be applicable auto-didactically. Results of the inter-observer agreement tests will be discussed here per enthesis.

Scapula

Scapula	*Long head of triceps brachii*			
	1 vs.2	**1 vs.3**	**2 vs.3**	**Av.**
Delineation possible	70	70	70	70
Porosity	40	60	30	43
Surface irregularity	80	50	30	53
Ridge formation	30	90	40	53
Surface flattening	100	80	80	87
Depression	80	50	50	60
Tubercle formation	90	90	100	93
Enthesophyte formation	90	100	90	93
n=	10	10	10	
Entheses observable	100	100	100	100

Table 7: inter-observer agreement percentages for the long head of the M. triceps brachii on the scapula (Av. = average).

When regarding the long head of the M. triceps brachii on the scapula (table 7), observer 3, who was unfamiliar with EC research, was more conservative in whether delineation of an enthesis was possible, whereas observer 2 agreed with the method creators. Porosity has the lowest inter-observer agreement, with no clear pattern of one observer consistently over-or underscoring relative to the others. Observer 2 tended to give higher scores for all traits than observers 1 and 3.

Clavicle

Clavicle	Costaclavicular ligament				Conoid				Trapezoid ligament			
	1 vs. 2	1 vs.3	2 vs. 3	Av.	1 vs. 2	1 vs.3	2 vs. 3	Av.	1 vs. 2	1 vs.3	2 vs. 3	Av.
Delineation possible	100	90	90	93	100	100	100	94	100	100	100	100
Porosity	50	70	40	53	10	30	30	34	20	40	50	37
Surface irregularity	0	60	20	27	50	80	30	32	20	50	20	30
Ridge formation	40	80	60	60	50	80	30	57	70	90	80	80
Surface flattening	100	90	90	93	100	100	100	94	100	60	60	73
Depression	60	80	60	67	80	100	80	69	80	80	60	73
Tubercle formation	100	100	100	100	50	70	20	83	100	80	80	87
Enthesophyte formation	80	90	90	87	60	100	60	79	40	100	40	60
n=	10	10	10		10	10	10		10	10	10	
Entheses observable	100	100	100	100	100	100	100	100	100	100	100	100

Table 8: inter-observer agreement percentages for the costoclavicular, conoid, and trapezoid on the clavicle (Av. = average).

The costoclavicular ligament attachment on the clavicle was relatively easy to delineate for all observers. Here, surface irregularity proved the most debatable trait (table 8), with observer 2 consistently scoring higher than the others. Porosity also proved somewhat problematic. Observer 1 and 3 showed relatively high consistency.

The conoid ligament attachment on the clavicle had a similar average agreement of 77%. It showed very low agreement for porosity between all observers. Observer 2 spotted more instances of surface irregularity and ridge formation, where observer 1 and 2 usually agreed on these traits.

In the third enthesis scored on the clavicle, the trapezoid ligament attachment, average agreement was 76%. Porosity scoring was once again inconsistent. What one observer scored as surface irregularity was interpreted as ridge formation or porosity by the other observers, and vice versa. Overall, we see a good agreement, again, between observer 1 and 3.

On the humerus, interobserver agreement for delineation was high for the M. pectoralis major and the M deltoideus, with the M. latissimus dorsii/teres major being scored less consistently (table 9). The M. pectoralis major was scored similarly by all three observers for most traits. Inconsistencies were found at the porosity, surface irregularity, and ridge formation. This is due to observer 1 scoring as porosity what observer three scored as surface irregularity, whereas observer two would score both. For ridge formation there is disagreement as to severity (i.e. score 1 or 2), and as to whether it is surface irregularity or a ridge. The M. latissimus dorsii/teres major showed similar disagreement as to porosity and surface irregularity, less confusion as to what constitutes ridge formation. For the M deltoideus the main discrepancies are found in the scoring of the presence and extent of porosity and surface irregularity. Overall, inter-observer agreement was highest between observer 1 and 3.

Humerus

Humerus	*M. pectoralis major*				*M. Latissimus dorsii/teres major*				*M. Deltoideus*			
	1 vs. 2	**1 vs.3**	**2 vs. 3**	**Av.**	**1 vs. 2**	**1 vs.3**	**2 vs. 3**	**Av.**	**1 vs. 2**	**1 vs.3**	**2 vs. 3**	**Av.**
Delineation possible	100	100	100	100	89	78	60	76	90	90	100	93
Porosity	20	40	20	27	44	44	50	46	44	33	50	42
Surface irregularity	60	60	60	60	44	67	50	54	33	22	60	38
Ridge formation	40	40	60	47	78	89	70	79	33	78	50	54
Surface flattening	80	90	90	87	89	89	100	93	100	100	100	100
Depression	70	80	70	73	78	100	80	86	89	100	90	93
Tubercle formation	100	100	100	100	100	100	100	100	100	100	100	100
Enthesophyte formation	90	100	90	93	100	100	100	100	100	100	100	100
n=	10	10	10		9	9	10		9	9	10	
Entheses observable	100	100	100	100	90	90	100	93	90	90	100	93

Table 9: inter-observer agreement percentages for the M. pectoralis major, M. latissimus dorsii/teres major, and M deltoideus on the humerus (Av. = average).

Radius

Radius	*M. pronator teres*			
	1 vs. 2	**1 vs.3**	**2 vs. 3**	**Av.**
Delineation possible	60	100	63	57
Porosity	80	0	38	52
Surface irregularity	60	60	75	72
Ridge formation	60	100	75	59
Surface flattening	20	60	38	56
Depression	60	60	100	87
Tubercle formation	100	100	100	100
Enthesophyte formation	100	100	100	53
n=	5	5	8	
Entheses observable	50	60	90	67

Table 10: inter-observer agreement percentages for the M. pronator teres on the radius (Av. = average).

The M. pronator teres entheses on the radius were delineated consistently by observer 1 and 3, while observer 2 could not delineate it in several cases (table 10). There was also discussion as to whether the entheses was well enough preserved taphonomically to be analyzed, with observer 1 being the most conservative assessor. Further interobserver disagreement occurred when one observer scored as surface flattening what the other labelled a depression.

Ulna

Ulna	*M. Pronator quadratus*				*M. brachialis*			
	1 vs. 2	**1 vs.3**	**2 vs. 3**	**Av.**	**1 vs. 2**	**1 vs.3**	**2 vs. 3**	**Av.**
Delineation possible	100	83	75	86	100	100	100	100
Porosity	0	50	50	33	30	67	67	55
Surface irregularity	29	83	13	42	30	89	11	43
Ridge formation	71	33	63	56	60	78	56	65
Surface flattening	100	100	100	100	100	89	89	93
Depression	86	100	75	87	50	78	44	57
Tubercle formation	100	100	100	100	100	100	100	100
Enthesophyte formation	100	100	100	100	100	89	89	93
n=	7	6	8	7	10	9	9	9,3
Entheses observable	70	60	90	73	100	90	100	97

Table 11: inter-observer agreement percentages for the M. pronator quadratus and M. brachialis on the ulna (Av. = average).

The two entheses scored on the ulna, the M. pronator quadratus and M Brachialis, both show inter-observer disagreement for the scoring of porosity and surface irregularity (table 11). M. brachialis was consistently delineated, as expected given the distinct anatomy of this enthesis site. For the pronator quadratus, there is some disagreement, both as to whether preservation was sufficient to allow scoring, and as to whether the enthesis could be delineated. As on the rest of the upper limb, observer 2 gave higher scores on average than observers 1 and 3.

Femur

Of the five muscles analyzed on the femur, agreement on whether delineation was possible was lowest for the pectineus (table 12), although the vastus medialis also showed variation in attributed score for this trait. In the M. Gluteus maximus, given scores varied the most for surface irregularity and porosity, most likely because of confusion of these two traits. The observers also often disagreed on what constituted a depression. EC occurring at the M. vastus medialis is inherently rather slight. This led to differences in scoring for porosity, where one observer saw just enough to score whereas the other found it insufficient. The same pattern is seen for depression at this enthesis. For the pectineus, also a trickier enthesis to delineate, porosity was the main source of inter-observer score variation, with incongruities for surface irregularity being notable as well. The lateral and medial attachment sites of the M. gastrocnemius have a very similar inter-observer agreement pattern. At these entheses, the extent of surface irregularity is the main source of difference, with disagreements as to the presence and extent of depression formation notable as well. For all entheses scored on the femur, observer 2 tended to give higher scores than observers 1 and 3.

Femur	*M. Gluteus Maximus*				*Vastus Medialis*				*Pectineus*				*M. gastrocnemius, medial head*				*M. gastrocnemius, lateral head*			
	1 vs. 2	**1 vs.3**	**2 vs. 3**	**AV.**	**1 vs. 2**	**1 vs.3**	**2 vs. 3**	**Av.**	**1 vs. 2**	**1 .3**	**2 vs. 3**	**v.**	**1 vs. 2**	**1 s.3**	**2 vs. 3**	**Av.**	**1 vs. 2**	**1 .3**	**2 vs. 3**	**AV.**
Delineation possible	100	100	100	100	56	67	89	71	78	56	33	56	100	100	100	100	88	100	88	92
Porosity	13	13	56	27	100	44	44	63	56	33	33	41	100	100	100	100				
Surface irregularity	13	13	22	16	89	100	89	93	33	67	56	52	33	56	56	48	13	63	13	30
Ridge formation	63	75	56	65	100	100	100	100	78	100	78	85	67	78	89	78	88	88	100	92
Surface flattening	75	75	78	76	100	89	89	93	100	78	78	85	100	89	89	93	100	75	75	83
Depression	38	50	78	55	78	22	44	48	100	78	78	85	56	78	56	63	75	50	50	58
Tubercle formation	88	88	78	85	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Enthesophyte formation	75	100	78	84	100	100	100	100	100	100	100	100	89	89	78	85	100	100	100	100
n=	8	8	9		9	9	9		9	9	9		9	9	9		8	8	8	
Entheses observable	90	90	100	93	100	100	100	100	100	100	100	100	100	100	100	100	90	100	90	93

Table 12: inter-observer agreement percentages for the M. gluteus maximus, vastus Medialis, Pectineus, M. gastrocnemius, medial head, and M. gastrocnemius, lateral head on the femur (Av. = average).

Tibia

Tibia	*M. Soleus*			
	1 vs. 2	**1 vs.3**	**2 vs. 3**	**Av.**
Delineation possible	100	100	100	100
Porosity	90	100	90	93
Surface irregularity	30	50	40	40
Ridge formation	60	80	70	70
Surface flattening	100	70	70	80
Depression	70	80	70	73
Tubercle formation	100	100	100	100
Enthesophyte formation	100	100	100	100
n=	10	10	10	
Entheses observable	100	100	100	100

Table 13: inter-observer agreement percentages for the M. soleus on the tibia.

All entheses of the M. soleus could be observed on all five individuals (n=10). Interobserver agreement was lowest for surface irregularity (table 13), potentially because of confusion with ridge formation which was also somewhat inconsistent. Observer 2 tended to score higher than observers 1 and 3.

4.7.3.3 Inter-observer agreement general

The inter-observer testing provided valuable information for future implementation of the scoring method. The most prominent result is that assessment of porosity was the least consistently scored trait between observers. This is not surprising, given that nonadult bone is, on average, more porous than adult bone on the entire cortical surface. This makes scoring porosity and assessing when an enthesis is more porous than the surrounding bone challenging. All three observers show high agreement for traits that are uncommon at an enthesis (i.e. tubercle/enthesophyte formation on entheses where this is rare). These traits were not considered as much when assessing inter-observer agreement, as they would inflate agreement rates unrealistically. Many instances of inter-observer disagreement were categorization issues. While observers tended to agree that a topography change was present or absent at an enthesis, they did at times score it under a different trait. This was the case for surface flattening and depression, or pronounced depression (score 2) and ridge formation. The same issue arose for traits that reflect more textural osseous changes, with one observer scoring as porosity an osseous characteristic that the other scored as surface irregularity.

It must be noted that the highest inter-observer agreement was between observer 1 (method creator) and observer 3 (non-EC specialist), rather than between observer 1 and 2 (EC specialist). Observer 2 tended to attribute higher scores overall, a trend that is not obscured in the composite scores (Appendix B2). This could be because observers 1 and 3 were trained for a large part by the same researchers in the European/UK field, whereas observer 2 was trained in the North American field and on very different collections. Both education (i.e., the subtle nuances in knowledge imparted by different researchers) and experience (i.e., the skeletal collections one is familiar with) will impact the way a researcher looks at human bone, and thus the scores ultimately accorded to the observed traits.

This means that if composite scores are used in which all traits per EC are combined, the inter-observer disagreement will be, at least in part, obscured. Although it could be argued that, in this way, composite scores can to a certain extent correct for inter-observer error, it also implies that composite scores are less useful for further method testing.

For some entheses there was a noteworthy amount of disagreement on whether the entheses could be delineated or not. This appeared to be partially dependent on whether the observer had previous experience with EC research, as observer 1 and 2 agreed in the majority of cases. For the M. pronator teres and M. vastus medialis however, agreement on delineation was low among all observers, likely due to their less prominent cortical markings. In some cases, there was also disagreement about whether cortical bone preservation was sufficient to allow observation. This is a factor that cannot be corrected for in the method, except by suggesting that researchers err on the side of caution when in doubt.

4.7.4 Method amelioration based on testing

The most striking result of both the intra-and interobserver method testing was that porosity was the least consistently scored trait on average. Because of this, it was decided to remove porosity as a trait to score for EC in the current method. Although this does not mean that porosity cannot be a trait of EC, it is excluded from the recording method because it is not useful to include a trait if it cannot be scored with acceptable consistency by all observers.

Additionally, the inter-observer testing highlighted that the proposed recording method presented a challenge to observers by allowing too many scoring options for some traits (i.e. absent-present-extensive change). Therefore, scoring options are now reduced to absent-present (i.e. 0-1) for all traits at all entheses.

With these improvements to the method, inter-observer error is considerably reduced. When the data are reworked to exclude porosity and reduce all scores to absent (i.e. 0) or present (i.e. 1), average interobserver agreement percentages are: 74% for the long head of the M. triceps brachii on the scapula; respectively 79%, 77%, and 76% for the costoclavicular, conoid, and trapezoid ligaments on the clavicle; 87% for the M. Pectoralis major, 84% for the M. Latissimus dorsii/teres major, and 85% for the M. deltoideus on the humerus; 76% for the M. pronator teres entheses on the radius, and respectively 81% and 79% for the M. pronator quadratus and M. Brachialis on the ulna. On the femur, average inter-observer agreement was 73% for the M. gluteus maximus, 86% for the M. vastus medialis, 81% for the M. pectineus, 87% for the medial branch of the M. gastrocnemius, and 82% for its lateral branch, and finally on the tibia average inter-observer agreement was 91% for the M. soleus. Inter-observer agreement percentages thus obtained are similar to those presented by Wilczak *et al.*, (2017) for the most prominent adult EC recording method, i.e. the Coimbra method (Henderson *et al.*, 2010, 2013, 2016).

As well as the abovementioned ameliorations to the scoring process, method testing also highlighted the need to provide a workable way to process the obtained data. Both the primary observer and the additional observers found that the current method produces a large dataset. Although this large dataset is inevitable in order to obtain the high-resolution information on changes occurring at the enthesis, it can be unwieldy for some research questions. To create a manageable dataset, a system of categorization is proposed, which sorts entheses into categories of minimal-mild-moderate-pronounced EC. Simply put, the scores for all changes occurring at an entheses are combined into a composite score by adding them up. This composite score is then sorted into a category (table 14) for each enthesis. In addition to creating a workable dataset, creating these categories further minimizes inter-observer error. This extra step of establishing EC categories will be especially useful for large scale studies where researchers aim to evaluate differences in EC within or between skeletal collections, particularly when data collection is done by different observers. For researchers who wish to research the interaction between one specific EC trait and other factors (e.g. osteophyte formation vs. skeletal maturity, or depression formation vs. surface irregularity at an enthesis), this categorization step is not necessary. Thus, the data can be processed according to the requirements of the research question, and the method retains maximum applicability.

Composite Score	Category
0-1	I
2-4	II
5-7	III
8+	IV

Table 14: Composite score categories.

Finally, it must be noted that the inter-observer method testing results underline that it can be beneficial to learn the method from someone already familiar with it. Especially researchers less familiar with entheses and/or nonadult remains could be aided by this.

4.8 Final method creation

The final proposed recording method is presented here. It is the product of the stages of method creation and testing described in the previous sections. This final method represents a viable approach to recording nonadult entheses in a standardized and structured way. With future research and potential learning of the method from those already familiar with it, interobserver agreement can be further increased and method repeatability further optimized.

Eight EC characteristics are retained, all scored as absent (=0) or present (=1) (table 15) (see section 5 for definitions of the traits).

Scored traits		
Delineation possible	0=no	1=yes
Surface irregularity	0=no	1=yes
Ridge	0=no	1=yes
Flattening	0=no	1=yes
Sulcus	0=no	1=yes
Depression	0=no	1=yes
Tubercle	0=no	1=yes
Enthesophyte	0=no	1=yes

Table 15: Scored traits with attributed score based on absent or present.

The traits are structured into a recording form that allows recording of all traits for each side and per enthesis (figure 1). This form also contains columns for initial data processing, with a column for composite score (i.e. the sum of all trait scores for the enthesis) and a column for the category that composite score falls in (indicated with Roman numerals I-IV). All information necessary for method implementation is bundled in a recording guide which contains outlines of enthesis location on the bone, trait definitions, and the scoring form (appendix B4).

	Left								Right							
	Delineation possible	Surface Irregularity	Ridge Formation	Surface flattening	Depression/Sulcus Formation	Tubercle formation	enthesophyte formation	COMPOSITE SCORE CATEGORY	Delineation possible	Surface Irregularity	Ridge Formation	Surface flattening	Depression/Sulcus Formation	Tubercle formation	enthesophyte formation	COMPOSITE SCORE CATEGORY
Clavicle																
Costaclavicular ligament																
Conoid																
Trapezoid ligament																
Humerus																
M. pectoralis major																
M. lat. dorsii/teres major																
M. Deltoideus																
Radius																
M. pronator teres																
Ulna																
Pronator Quadratus																
M. Brachialis																
Femur																
Gluteus maximus																
Vastus Medialis																
Pectineus																
M. gastrocnemius, med. head																
M. gastrocnemius, lat. head																
Tibia																
M. Soleus																

Figure 1: Standardized recording form for subadult EC.

4.9 Discussion and conclusion

The most prominent result of this study of EC in nonadults is that each enthesis has its own unique spectrum of variation, with some osseous characteristics prone to change and others unlikely to occur. This is not entirely surprising, given the different biomechanical properties of different entheses; i.e., their size both absolute and relative to the muscle/ligament attached, and the angle of pull. However, this enthesis-unique spectrum is of the utmost importance to any study analyzing EC, as it means that the results from one enthesis can never be directly compared to another, and that different studies using averages of different sets of entheses cannot be compared. Results also show high inter-individual variation in EC, and even variation between left and right sides within one individual.

No clear unidirectional pattern of age- or sex-related EC occurred in this explorative study. This indicates that other etiological factors are likely also in play. It cannot at this point be excluded that idiosyncratic individual variation is responsible for some entheseal variation. However, given the wide range of observable trait variation, between and within individuals, other factors such as activity are also likely to have impacted the appearance of the entheses. Extensive research on identified skeletal collections of various socio-historical contexts is necessary to help determine the extent of impact different factors have on EC in nonadults.

The proposed standardized method presented here is developed to aid and stimulate this future research. As in all qualitative research, repeatability is one of the main challenges in nonadult EC recording. Intra-observer agreement was high, but inter-observer error showed how some traits presented challenges for consistent scoring. Porosity, especially, was scored rather inconsistently by the three observers, causing it to be discarded in the final method. Many inter-observer points of confusion occurred on the threshold between different traits. Different topographic traits which form a spectrum were at times confused (e.g. surface flattening and depression formation), as were extensive surface irregularity and slight ridge formation. The same issues occurred with textural traits, such as porosity versus slight surface irregularity. The tests also showed that while previous experience with EC can help researchers accurately delineate the entheses, the skeletal collections one is experienced with likely also impacts scoring.

With the invaluable feedback provided by testing in the method creation process, a final scoring method was created that scores each enthesis for eight traits, all of which are scored as being either present or absent. These traits can then be summed for each enthesis to create a composite score, which represents the amount of change occurring at that particular attachment site. These composite scores can then further be sorted into categories that represent a level of EC. This categorization allows for a more manageable dataset, which can be preferable for some research questions as well as partially limiting inter-observer variation when several observers are involved in data collection. When possible, learning the method from someone already acquainted with it can further improve repeatability. A scoring guide and standardized recording form are provided in the appendices to increase utility (appendix B4 and B5).

This paper has shown the wealth of individual variation occurring at nonadult entheses, with pilot results indicating that age and/or sex are not the only determining factors. The proposed standardized scoring method will allow a better comprehension of nonadult EC, thus providing a new insight into the lives of children in the past.

Acknowledgements

The authors would like to thank Dr. R. Schats and Dr. S. Schrader for testing the method and providing feedback.