Henderson, CY. Mariotti, V. Pany-Kucera, D. Villotte, S. and Wilczak, C. 2013. Recording specific features of fibrocartilaginous entheses: Initial tests using the Coimbra method. International Journal of Osteoarchaeology 23(2): 152-162. DOI: 10.1002/oa.2287.

Short title:
Recording entheseal changes

CY. HENDERSON[1,2], V. MARIOTTI[3,4], D. PANY-KUCERA[5, 6], S.VILLOTTE[7] and C. WILCZAK[8]

1. CIAS - Centro de Investigação em Antropologia e Saúde, Universidade de Coimbra, Portugal
2. Department of Archaeology, Durham University, UK.
3. Department of Experimental Evolutionary Biology, University of Bologna, Italy.
4. Unité Mixte de Recherche 6578 – Unité d’Anthropologie Bioculturelle, Centre National de la Recherche Scientifique/Université de la Méditerranée/Etablissement Français du Sang, Marseille, France
5. Natural History Museum, Department of Anthropology, Vienna, Austria.
6. University of Vienna, Department of Anthropology, Vienna, Austria
7. Department of Archaeological, Geographical and Environmental Sciences, University of Bradford, UK.
8. Department of Anthropology, San Francisco State University, USA.

ABSTRACT

A working group was established in 2009 during a workshop in Coimbra, Portugal to review the various methodologies used to record entheseal changes (EC), and develop a standardized system to facilitate comparisons across studies. This paper presents the first results of the "Coimbra method"; a new qualitative method for recording fibrocartilaginous entheses based on the types of changes observed.

Materials and Methods: The new method divides the enthesis into a margin (only the area opposite the acute angle of fibre attachment) and surface (which also includes the remaining margin). Five features are recorded: bone formation, erosion, fine porosity, macroporosity, and cavitation. A total of 67 male skeletons from the identified SIMON collection, Geneva, Switzerland all of whom were manual workers and aged between 20 and 79 years, were used for this study. Six skeletons were used by the authors as exemplars to determine standard criteria for recording each change. Thirty male skeletons were selected to test intra- and inter-observer error of the new method. An additional 31 skeletons were used for a preliminary test of the relationship between EC and age, using exploratory statistics and ordinal regression.

Results: Intra- and inter-observer error had a similar percentage agreement of around 70%. The exploratory statistics indicated a general trend for increased scores of each feature with age, but ordinal regression demonstrated that this was not statistically significant ($p<0.05$) for all features.

Discussion: The recording method is repeatable for some entheses. The effect of the ageing process is dependent on enthesis and EC feature. Unlike most methods, the "Coimbra method" records EC features in detail; this has the advantage of allowing studies of the relationship between different EC and age, as well as sex and occupation. Further studies on larger identified skeletal collections are needed to test the effect of age, sex and occupation.

Keywords: musculoskeletal stress markers (MSM); ageing; methodology; enthesis; enthesopathy; SIMON Collection, Geneva

Introduction

The bioarchaeological study of activity-related bone changes has been an alluring prospect for over a century (*e.g.* Testut, 1889; Wood Jones, 1910; Angel, 1966; Wells, 1962, 1963). However, this allure has led to imaginative interpretations of skeletal changes from rape victims (Hawkes and Wells, 1975: 119) to, more recently, acrobats (Oates *et al.*, 2008), despite debate of the limitations in methodology and interpretation (*e.g.* Peterson and Hawkey, 1998; Jurmain,1999; Jurmain *et al.*, 2012). To address this challenge, a workshop focussing on the progress and limitations of research on musculoskeletal stress was arranged in 2009 in Coimbra, Portugal (Santos *et al.,* 2011). The discussion at this workshop raised three primary issues and working groups were set up to deal with these, specifically: terminology (Jurmain and Villotte, 2010), recording methods (Henderson *et al.*, 2010; 2012) and classifying occupation (see Perréard Lopreno *et al.*, in this volume). The first recommendation from the terminology group was to replace the term musculoskeletal stress markers with entheseal changes (EC) to better reflect the complex aetiology of lytic and hyperostotic reactions at entheses. In this paper the first results from the working group on recording methods are presented. To date several visual recording methods have been proposed (for a review see Jurmain, *et al.,* 2012). Amongst these the most commonly used was developed in 1988 (Hawkey, 1988; Hawkey and Merbs, 1995). However, more recently methods have been tested on identified skeletal collections (Mariotti *et al.*, 2004, 2007; Cardoso, 2008; Alves Cardoso and Henderson 2010; Villotte, 2006, 2009) and two of these (Villotte, 2006, 2009; and Mariotti *et al.* 2004, 2007) have formed the basis for the construction of the “Coimbra method”.

The most widely used recording method is that developed by Hawkey (1988) and published eight years later (Hawkey and Merbs, 1995). This method divides EC into 3 distinct types of bone changes: robusticity markers, stress lesions and ossification exostoses. Robusticity markers, in their most extreme expression are seen as sharp ridges or crests of bone (Hawkey and Merbs, 1995: 328). Stress lesions are those EC which resemble lytic lesions and cover all forms of pitting or furrowing whereas ossification exostoses involve bone production forming an exostosis or spur (Hawkey and Merbs, 1995: 329). Each of these changes is recorded on a scale representing increasing expression. The initial reports of the method reported low intra- and inter-observer error, but a recent study found very poor repeatability (Davis *et al.*, 2012). The primary problem (Jurmain *et al.*, 2012) with the method is the assumption, without clinical or histological justification, that ossification exostoses are due to macrotrauma and robusticity markers and stress lesions are caused by microtrauma, and consequently considered as a continuum in a scale of increasing development (Hawkey and Merbs, 1995: 329). Moreover, the method was neither tested on an identified skeletal collection nor were differential diagnoses (*e.g.* the possibility that the changes were pathological or caused by age) discussed. Subsequent research using this method has demonstrated that the scores are highly correlated with age and that the ageing process, rather than occupation, is most closely associated with these changes (Cardoso, 2008).

The method developed by Mariotti (1998) and later published in two parts (Mariotti *et al.*, 2004, 2007) was conceived to describe morphological variability at entheses. The method retained the subdivision by Hawkey into 3 types of entheseal changes, but re-named the last two in a more descriptive way, as enthesophytic and osteolytic formations, respectively. These two features were considered pathological, *i.e.* abnormal morphological variations (*cf.* International Dictionary of Medicine and Biology, 1986) not manifestations of specific diseases. and, therefore, collectively referred to as "enthesopathies" (Mariotti *et al.*, 2009). A 3-degree scale is provided for the robusticity of 23 entheses, while the standard for enthesopathies (which can be absent, or present in 3 degrees) can be applied to any enthesis. The main objections to this method concern the high intra- and inter-observer error for robusticity (about 20%) and the fact that the entheses studied and the scoring method were chosen without reference to the medical literature on entheses, their anatomy or the aetiology of changes (Villotte 2009; Jurmain *et al.,* 2012).

The method developed by Villotte (2006) is based on the anatomy and histology of entheses (fibrous or fibrocartilaginous), taking into account the clinical literature. The entheses studied are subdivided into 4 groups ('scoring systems') on the basis of their anatomical characteristics. Three stages (A, B, C) are described for each scoring system. In general, for scoring groups 1 and 2, stage A represents the normal enthesis, while stages B and C represent pathological conditions of increasing severity. For groups 3 and 4 the borderline between normal and pathological cannot be clearly established on the basis of medical and anatomical research. This method reports a low inter- and intra-observer error. Later, this method was simplified into presence (stages B and C pooled together) and absence (stage A) for a subset of fibrocartilaginous entheses (Villotte *et al.*, 2010). The main objection to this method is that different entheseal features (new bone formation, vascularisation, and lytic lesions) are included in stages B and C, while it is quite clear from other studies that some of these features behave differently in relation to age and sex (Mariotti *et al.*, 2004; 2007; Milella *et al.*, 2012).

The aim for this working group was to use the most current knowledge of entheses and utilise the differences in skill sets, samples seen and approaches to develop a new consensus based recording method. It was decided that the anatomical understanding of entheses, *i.e.* the differentiation between fibrocartilaginous and fibrous entheses (described in this issue by Villotte and Knüsel), should form the basis for the method. Due to the currently limited clinical data on fibrous entheses, it was decided to develop the method for the clinically better understood fibrocartilaginous entheses, following Villotte's approach, but developing separate recording criteria and scores for the various kinds of features that can be observed at entheses, following Mariotti's system, thereby retaining the strengths of both methods. The aim of this paper is to present the first output of the work of this working group: the new 'Coimbra method' for recording EC at fibrocartilaginous entheses, the intra- and inter-observer error results, alongside preliminary results testing the effect of age on EC.

Materials and Methods

Materials

Male skeletons (n=67) with physically demanding occupations (Perréard Lopreno, 2007) from the identified SIMON collection in Geneva were used. The individuals in this collection died between 1900 and 1969 and were buried in 27 cemeteries from *Canton de Vaud*, an area lying predominantly on the northern side of Lake Geneva. Only males were used, due to time-constraints. The sample was chosen to have a normal age distribution, ranging from 20-79 years. The sample of occupations chosen predominantly consists of agricultural workers, but includes gardeners, glass workers, masons, winemakers and workers. This sample was randomly divided into three sub-samples: A. initial discussion (n=6), B. intra- and inter-observer error (n=30) and C. effect of age (n=31). These samples were not combined to test the effect of age, as the scores of samples A and B were discussed by the whole group and may have biased the results of C. The working group chose five of the most commonly recorded entheses (all insertions unless otherwise specified) to focus on: *m. subscapularis*, common extensor origin, *m. biceps brachii*, *m. iliopsoas*, and the *m. triceps surae*.

Methods

Online Discussion

The first discussion, development of the method and recording form, took place online and included photographs of entheses (all fibrocartilaginous). These were examples from the authors' own samples. It was decided that the method should focus on recording different types of EC rather than a single score per enthesis in the initial stage to enable testing of age and asymmetry effects on individual features. The types of changes considered important to record, based on the consensus experience were: bone formation (BF), erosion (ER), fine porosity (FPO), macroporosity (MPO) and cavitation (CA) (Table 1, Figs 1 and 2). The anatomy of the fibrocartilaginous enthesis, noting in particular the fibrous nature of the margin (Benjamin *et al.*, 1986; Villotte, 2006), was also considered important to include. Consequently, the enthesis was divided into two portions: 'zone 1' the contour opposite the acute angle at which the fibres attach (*i.e.* the more fibrous margin) and 'zone 2' the remaining surface and margin (Figs. 1 and 2). Only bone formation and erosions were scored in zone 1 while all five features were scored in zone 2. Zones 1 and 2 were scored separately, therefore, each zone was considered missing (score of 99) if more than 50% was damaged. The scores for each feature are considered differences in expression and not severity. The relationship between features such as Hawkey and Merbs (1995) proposed between their "ruggedness and stress lesions", is not endorsed. In this method each feature is scored and interpreted separately.

Initial Discussion in Geneva

The group met in Geneva in 2010 to finalise the method on skeletal remains. Sub-sample A was recorded by the five authors and G. Perréard Lopreno (n=6) using the recording form agreed upon during the online discussion. The results were then compared and each enthesis was discussed. It became apparent that slight differences existed in the interpretation of the features, leading to refinement of the definitions, *e.g.* the size of pores and area of coverage. A new list of definitions was then created for day two's recording (Table 1). On the second day, one observer photographed the entheses recorded during day one and two. These photographs were discussed at the end of that day and each photograph given a score agreed by all observers for publication of the method and training purposes (Figs.1 and 2).

Intra- and Inter-observer Error

Thirty skeletons (although due to missing data n=28 for intra-observer error and n=27 for inter-observer error) with a mean age of 53 (range 20-79) were recorded by five observers. Inter-observer error was calculated in terms of the percentage agreement between scores. This was calculated by counting the number of exact agreements in scores (*e.g.,* 99 versus 99; 2 versus 2) between two observers for each feature and enthesis. An overall percentage agreement score was calculated along with calculations between all observers, by observer, by enthesis and by feature. Percentage agreement was also recalculated to test the effect of reducing the number of degrees for each feature, *i.e.* by pooling degrees 2 and 3 for BF, 1 and 2 of ER, FPO and MPO and secondly by removing the subdivision into zones for BF and ER.

Four days after the initial recording and subsequent to the discussion of the photographs, observer 1 re-recorded these thirty skeletons. Percentage agreement by enthesis and by feature between the first and second observation were calculated.

Asymmetry and the Effect of Age

Bilateral asymmetry of scores was also tested to understand the relationship between the EC features and handedness. Asymmetry was calculated for each enthesis and feature by subtracting the left side score from the right side score for each individual. The following were then counted: equal scores, right side higher scores (positive integers) and left side higher scores (negative integers).

The skeletons (n=31) selected to test the effect of age had a normal age distribution (range 23-73, mean=45.9, standard deviation 14.4, Z=0.524, p=0.946). All of these males were from a similar occupational background (manual workers): farmers, gardeners, winemakers and those listed as workers. These skeletons were only recorded by observer 1 on day two of the meeting in Geneva, prior to the discussion of the photographs.

Scores were plotted against mean age for each enthesis and side to explore the effect of age. Ordinal regression was performed on left and right sides for each enthesis in SPSS 19.0 with the score for each feature as the dependent variable and age as the covariate. Cumulative probabilities were transformed using the negative log-log function [f(x) = -log (- log(x))] because the data were negatively skewed, with lower scores more probable than higher ones (Norusis, 2010). In cases of poorly fitting models, fit was not significantly improved by including scale parameters.

Results

Intra- and Inter-observer Error

For the intra-observer error, the percentage agreement of all features was lowest for the *m. triceps surae* (66.8%) and *m. subscapularis* insertions (68.9%) and highest for the common extensor origin (82.7%). All features were scored lower on the second observation (Fig. 3).

For the inter-observer error (the first set of data collected by observer 1 was used), the overall percentage agreement is 71.8%. The greatest variability in agreement (range 52.5 - 92.1%) is seen for the test by feature (Table 2). Interestingly, just as with the intra-observer error, the common extensor origin has the highest percentage agreement (75.4%). While for the intra-observer error the *m. triceps surae* insertion scored lowest, this had the second highest percentage agreement for the inter-observer error (74.8%), and the *m. subscapularis* insertion had the lowest percentage agreement (67.4%). Figure 4 demonstrates the mean scores for each feature by observer. Observer 1 scored consistently higher across most features, whereas observer 4 scores consistently lower. Percentage agreement between observer 1 and 4 was 69.1% with the lowest percentage agreement between observers being between observer 1 and 2 (65.3%). The highest percentage agreement was between observer 4 and 5 (78.1%). To test whether the number of stages was causing the problem the scores were combined and analysed by enthesis and feature (see Appendix 1). The lowest agreement is seen for BF, but it does improve reproducibility for some features, most notably ER (both zones) and MPO and improves reproducibility for some entheses. The overall agreement (mean of all the mean scores) is 71.7%.

Asymmetry and the Effect of Age

No cavitations were recorded in this sample and this feature was, therefore, excluded from analysis.

In general, the left and right side scores were equal, but where there are differences these tend to favour higher scores on the right side (Table 3). The opposite only occurs for ER(Z1) on both the *m. triceps surae* (2 scores out of 22 are higher on the left) and the *m. biceps brachii* (2 scores out of 29 are higher on the left) insertions and FPO on the common extensor origin (2 scores out of 20 are higher on the left).

The effect of age is not the same for each feature, enthesis or even side. Figure 5 presents the plots of mean age and enthesis by feature. For many entheses there is a trend for an increased score with age, but sample size, causing limited variability in the range of scores, is a severe limitation. Ordinal regression demonstrated that there was only a statistically significant association between bone formation and age for some entheses, and it was not consistent for left and right sides (Table 4).

Discussion

Intra-observer error scores were consistently lower on the second observation. This may be due to tiredness, different lighting conditions or a psychological effect with the observer considering the need for more caution on the second attempt. Inter-observer error was highly variable between observers. Overall both intra- and inter-observer error could be improved, but these results have a higher reproducibility than found by an independent test of other methods (Davis *et al.,* 2012). However, the lower than expected reproducibility of this new method may be caused by the high degree of experience of the observers, who have all studied and developed recording methods. This may have led individuals to perceive changes differently leading to different scores. This is illustrated by the systematic discrepancy between observers 1 (consistently higher) and observer 4 (consistently lower). Cavitations and their recording was a particular problem this may in part be due to their definition, but also the difficulty of observing the internal shape of these holes, which by definition are larger on the inside than at their opening. However, it is important to note that in the discussion of photographs at the end of the scoring session, an agreement between all observers was found on initially different scores including the cavitations. Interestingly the reproducibility is fairly good for some entheses, especially the common extensor origin.

Therefore the method may be easier to apply to or better defined for certain types of enthesis, perhaps due to less variability in entheseal changes. The reproducibility of the method needs to be improved. Photographs defining each feature and score need to be produced for each enthesis. It is possible that observation time should be standardised to a minimum and maximum duration.

Data on the percentage of agreement by adding up some degrees of development and zones demonstrates that the errors are more related to the number of developmental degrees than to the subdivision of entheses in zone 1 and 2 (see Appendix 1). The fact that in many cases more than 80% agreement was reached by some pairs of observers demonstrates that the many factors influencing reproducibility act in random, unpredictable ways. A scale with fewer degrees could improve reproducibility, but would capture less of the variation. A future workshop and preceding online discussion will aim to resolve the reproducibility by improving the definition of the method. This will also include a discussion of how many scores of expression are required.

Increased scores with age were not unexpected as this is the general trend for all methods tested on identified skeletal collections (*e.g.* Cunha and Umbelino 1995; Mariotti *et al.,* 2004: 2007; Cardoso, 2008; Villotte, 2009; Alves Cardoso and Henderson, 2010; Villotte *et al.*, 2010; Milella *et al.*, 2012), but no previous methods have tested these individual features. Few correlations are significant, probably due to the sample size and, for some entheses, the limited variability in scores for features. However, it is clear that the effect of increased age is dependent on enthesis and on EC feature. The effect of age seems more important for bone formation compared to porosity for instance. The differential influence of age on types of EC has previously been demonstrated (Mariotti *et al.*, 2004; 2007; Milella *et al.*, 2012). A previous study (Milella *et al.*, 2012) also demonstrated that the frequency of bone formation is more strongly correlated with age than "osteolytic enthesopathies". These results illustrate the necessity to record the changes separately for a better understanding of the effect of age on entheseal aspect.

For almost all the entheses and the types of changes, the most common result is a similar score for both sides. The results may indicate that intrinsic factors play an important role in the occurrence of ECs and this requires testing on a larger sample and by taking into account symmetry caused by no EC. Bilateral symmetry was also found commonly by Alves Cardoso and Henderson (2010). However, when upper and lower limbs are considered separately (Table 3) a pattern emerged. For the upper limb, asymmetries are more frequent and clearly more common on the right side for bone production (zones 1 and 2) and to a lesser extent for MPO. Such a trend is not seen for the lower limb (excepting BF in zone 2). This result is consistent with previous studies of laterality and asymmetry (Villotte 2009) and with many studies reporting more EC on the right side for upper limbs (*e.g.* Villotte *et al.*, 2010; Millela *et al.*, 2012). Given that the majority of people are right handed, it is consistent with mechanical factors as these stimulate bone formation. Bone formation seems more sensitive to mechanical stress and better correlated with age. Both can be related, as it has been suggested that the effect of age is associated with the accumulation of micro-trauma through life (Robb 1998; Milella *et al.*, 2012). However, this theory requires further testing with a larger sample size.

Conclusions

The development of a new recording method was considered necessary based on discussion after a workshop on EC held in 2009 (Santos *et al.,* 2011).The aim of this paper was to present this new method, the "Coimbra method", for recording EC based on anatomical knowledge while accounting for the variability in types of changes.

The goal of developing a new method is partially achieved. Currently this method cannot be considered as fully applicable, mainly because of the limited reproducibility. However, it is good at distinguishing types of changes, which need to be studied. Currently the method should be considered as a tool to try to understand, in more subtle ways, the effect of factors (especially age) on the entheses. The working group will continue to develop the method, taking into consideration also the problems that emerged during a workshop held at the European Meeting of the Paleopathology Association in Vienna (Pany-Kucera *et al.,* 2010). This workshop not only highlighted people's interest in recording EC, but also the many problems associated with identifying features and whether they belonged to the enthesis. Further group training on using the method is planned to discuss the problems highlighted during this workshop and to improve inter-observer error rates. The method will also soon be tested on a larger sample of identified skeletons to assess the effect of age and sex on the features.

Acknowledgements:

The authors would like to thank Dr Perréard-Lopreno for access to the SIMON collection and laboratory facilities. The contribution of C. Henderson to this research was supported by the FCT – Fundação para a Ciência e a Tecnologia, reference: SFRH/BPD82559/2011. S. Villotte's participation in the research was funded by the Fyssen Foundation.

Authors' contribution

All the authors contributed equally to the preparation of this paper.

References

Alves Cardoso F. and Henderson CY. 2010. Enthesopathy formation in the humerus: Data from known age-at-death and known occupation skeletal collections. *American Journal of Physical Anthropology* 141:550-560.

Angel JL. 1966. Early skeletons from Tranquility, California. *Smithsonian Contributions to Anthropology* 2:1-19.

Benjamin M. Copp L. and Evans EJ. 1986. The histology of tendon attachments to bone in man. *Journal of Anatomy* 149:89-100.

Cardoso FA. 2008. *A portrait of gender in two 19th and 20th century Portuguese populations: a palaeopathological perspective*. PhD thesis. Durham University, UK.

Cunha E. and Umbelino C. 1995. What can bones tell about labour and occupation: the analysis of skeletal markers of occupational stress in the Identified Skeletal Collection of the Anthropological Museum of the University of Coimbra (preliminary results). *Antropologia Portuguesa* 13:49-68.

Davis CB. Shuler K. Danforth M. and Herndon K. 2012. Patterns of interobserver error in the scoring of entheseal changes. *International Journal of Osteoarchaeology. doi: 10.1002/oa.2277.*

Hawkes SC. and Wells C. 1975. Crime and punishment in an Anglo-Saxon cemetery. *Antiquity* 49:118-122.

Hawkey DE. 1988. *Use of upper extremity enthesopathies to indicate habitual activity patterns*. MA thesis. Arizona State University.

Hawkey DE. and Merbs CF. 1995. Activity-induced musculoskeletal stress markers (MSM) and subsistence strategy changes among ancient Hudson Bay Eskimos. *International Journal of Osteoarchaeology*. 5: 324-338.

Henderson CY. 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*. 18th European Meeting of the Paleopathology Association, Vienna, Austria. 23rd -26th of August. URL: https://www.uc.pt/en/cia/msm/Vienna2010.pdf. Accessed 8th of March, 2012.

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

International Dictionary of Medicine and Biology. 1986. Wiley and Sons: New York, Chichester, Brisbane, Toronto, Singapore.

Jurmain RD. 1999. *Stories from the Skeleton. Behavioral Reconstruction in Human Osteology*. Gordon and Breach: Amsterdam.

Jurmain RD. and Villotte S. 2010. Terminology – entheses in medical literature: a brief review. In *Workshop in Musculoskeletal Stress markers (MSM): Limitations and Achievements in the Reconstruction of Past Activity Patterns*. URL: http://www.uc.pt/en/cia/msm/MSM_terminology3.pdf. Accessed 9th of March, 2012.

Jurmain RD. Alves Cardoso F. Henderson CH. and Villotte S. 2012. Bioarchaeology's Holy Grail: the reconstruction of activity. In *Companion to Paleopathology*, Grauer A. (ed.) Wiley-Blackwell: Chichester, UK; 531-552.

Mariotti V. 1998. *Ricerche sugli indicatori scheletrici morfologici di attività*. PhD Dissertation, University of Bologna.

Mariotti V. Facchini F. and Belcastro MG. 2004. Enthesopathies - Proposal of a Standardized Scoring Method and Applications. *Collegium Anthropologicum* 28:145-159.

Mariotti V. Facchini F. and Belcastro MG. 2007. The Study of Entheses: Proposal of a Standardised Scoring Method for Twenty-three Entheses of the Postcranial Skeleton. *Collegium Antropologicum* 31:291-313.

Mariotti V. Milella M. and Belcastro MG. 2009. Musculoskeletal stress markers (MSM): methodological reflections. In *Workshop in Musculoskeletal Stress markers (MSM): Limitations and Achievements in the Reconstruction of Past Activity Patterns*. Coimbra, Portugal, July 2-3, 2009. URL: http://www.uc.pt/en/cia/msm/MSM_podium. Accessed 26th of June, 2012.

Milella M. Belcastro MG. Zollikofer CPE. 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* 148:379-388. DOI 10.1002/ajpa.22060.

Oates J. Molleson T. and Soltysiak A. 2008. Equids and an acrobat: closure rituals at Tel Brak. *Antiquity* 82:390-400.

Pany-Kucera D. Viola TB. Henderson C. Mariotti V. and Villotte S. 2010. *Scoring entheseal changes: hands-on workshop in search of shared solutions for recording fibrocartilaginous entheses.* 18th European Meeting of the Paleopathology Association, Vienna, Austria. 23rd -26th of August.

Perréard Lopreno G. 2007. Adaptation structurelle des os du membre supérieur et de la clavicule à l'activité: analyse de l'asymétrie des propriétés géométriques de sections transverse et de mesures linéaires dans une population indentifiée (Collection SIMON). PhD thesis. University of Geneva.

Peterson J. and Hawkey DE. 1998. Preface. *International Journal of Osteoarchaeology* 8:303-304.

Robb JE. 1998. The interpretation of skeletal muscle sites: a statistical approach. *International Journal of Osteoarchaeology* 8:363-377.

Santos AL. Alves Cardoso F. Assis S. and Villotte S. 2011. The Coimbra Workshop in

Musculoskeletal Stress Markers (MSM): an annotated review. *Antropologia Portuguesa* 28:135-161.

Testut L. 1889. *Recherches Anthropologiques sur le squelette quaternaire de Chancelade (Dordogne)*. Pitrat Aîné: Lyon.

Villotte S. 2006. Connaissances Médicales Actuelles, Cotation des Enthésopathies: Nouvelle Méthode. *Bulletins et Mémoires de la Société d'Anthropologie de Paris* 18:1-2.

Villotte S. 2009. *Enthésopathies et activités des Hommes préhistoriques - Recherche méthodologique et application aux fossiles européens du Paléolithique supérieur et du Mésolithique*. Archaeopress: Oxford.

Villotte S. Castex D. Couallier V. Dutour O. Knüsel CJ. and Henry-Gambier D. 2010. Enthesopathies as occupational stress markers: evidence from the upper limb. *American Journal of Physical Anthropology* 142: 224-234.

Villotte, S. and Knusel, C. 2012. Understanding Entheseal Changes: Definition and Life Course Changes. *International Journal of Osteoarchaeology*

Wells C. 1962. Joint pathology in ancient Anglo-Saxons. *Journal of Bone and Joint Surgery*. 44B:948-949.

Wells C. 1963. Hip disease in ancient man. Report of three cases. *Journal of Bone and Joint Surgery*. 45B:790-791.

Wood Jones F. 1910. Fractured bones and dislocations. In *The Archaeological Survey of Nubia, Volume 2*. Elliot-Smith G. and Wood Jones F. (eds.) National Printing Department: Cairo.

Table 1. Features and their definitions along with degrees of expression recorded.

Zone	Feature	Abbreviation	Definition	Degrees of expression
Zone 1 (margin opposite acute angle of fibre attachment)	Bone Formation	BF (Z1)		1= small, nodular or slightly raised margin <1mm 2= distinctive sharp crests or other enthesophytes ≥1mm but <50% of margin 3= distinctive sharp crests or other enthesophytes ≥1mm but ≥50% of margin
	Erosion	ER (Z1)	Depressions or excavations of any shape and involving discontinuity of the cortex generally grater in width than depth with irregular margins. Only erosions >2mm were recorded.	1=<25% of margin 2= 25-50% of margin 3= >50% of margin
Zone 2 (remaining margin and surface)	Bone Formation	BF		1= “roughness”/rugosity, change is diffuse not a distinct structure 2= distinct structure measuring >1mm, affecting <50% of surface 3= distinct structure measuring >1mm, affecting ≥50% of surface
	Erosion	ER	Depressions or excavations of any shape and involving discontinuity of the cortex generally grater in width than depth with irregular margins. Only erosions >2mm were recorded.	1=<25% of surface 2= 25-50% of surface 3= >50% of surface
	Fine Porosity	FPO	Small, round to oval perforations with smooth, rounded margins ≤1mm	1= <50% of surface 2= ≥50% of surface
	Macro-porosity	MPO	Small, round to oval perforations with smooth, rounded margins >1mm	1= one or two pores 2= >2 pores
	Cavitation	CA	Subcortical cavity with an external opening smaller than the maximum diameter of the cavity	1= 1 cavitation 2= >1 cavitation

Table 2. Inter-observer error percentage agreement

	Test	% of Agreement
Total	Total	71.8%
By trait	BF (Z1)	61.5%
	ER(Z1)	82.4%
	BF	52.5%
	FPO	61.9%
	MPO	79.0%
	ER	73.3%
	CA	92.1%
By enthesis	*Triceps surae*	74.8%
	Common extensor	75.4%
	Biceps brachii	72.9%
	Iliopsoas	68.5%
	Subscapularis	67.4%
By comparison between obs.	Obs 1 vs Obs 2	65.3%
	Obs 1 vs Obs 3	68.8%
	Obs 1 vs Obs 4	69.1%
	Obs 1 vs Obs 5	69.6%
	Obs 2 vs Obs 3	74.6%
	Obs 2 vs Obs 4	74.9%
	Obs 2 vs Obs 5	73.7%
	Obs 3 vs Obs 4	71.4%
	Obs 3 vs Obs 5	72.5%
	Obs 4 vs Obs 5	78.1%
By obs.	Obs 1	68.2%
	Obs 2	72.1%
	Obs 3	71.8%
	Obs 4	73.4%
	Obs 5	73.5%

Table 3 Bilateral asymmetry calculated as right side minus left side. N=total n for each enthesis. Equal scores = number with equal scores. Left side higher score = number with a higher score on the left side. Right side higher score = number with a higher score on the right side.

		BF (Z1)	ER (Z1)	BF	FPO	MPO	ER	CA
Subscapularis	n	25	25	25	25	25	25	25
	equal scores	11	24	12	13	15	18	25
	left side higher score	1	0	3	5	2	3	0
	right side higher score	13	1	10	7	8	4	0
Common extensor origin	n	20	20	20	20	20	20	20
	equal scores	11	20	13	18	20	17	20
	left side higher score	2	0	3	2	0	1	0
	right side higher score	7	0	4	0	0	2	0
Biceps brachii	n	29	29	29	29	29	29	29
	equal scores	9	26	15	17	25	25	29
	left side higher score	2	2	5	4	1	3	0
	right side higher score	18	1	9	8	3	1	0
Iliopsoas	n	20	20	20	20	20	20	20
	equal scores	15	19	8	16	19	14	20
	left side higher score	4	1	3	2	0	2	0
	right side higher score	1	0	9	2	1	4	0
Triceps surae	n	22	22	22	22	22	22	22
	equal scores	11	19	11	14	22	20	22
	left side higher score	2	2	4	4	0	0	0
	right side higher score	9	1	7	4	0	2	0
Upper limb	n	74	74	74	74	74	74	74
	equal scores	31	70	40	48	60	60	74
	left side higher score	5	2	11	11	3	7	0
	right side higher score	38	2	23	15	11	7	0
Lower limb	n	42	42	42	42	42	42	42
	equal scores	26	38	19	30	41	34	42
	left side higher score	6	3	7	6	0	2	0

	right side higher score	10	1	16	6	1	6	0
all entheses	n	145	145	145	145	145	145	145
	equal scores	72	136	76	101	130	121	145
	left side higher score	18	5	24	19	3	11	0
	right side higher score	55	4	45	25	12	13	0

Table 4. Ordinal regression: N/A means that there is not enough variation for estimates. Statistical significance which does not violate assumptions is highlighted in bold.

Feature	**Enthesis**	**Side**	**Model significance**	**Nagelkerke pseudo R-square**	**Age coefficient**	**Significance (Wald)**
Marginal bone formation (BF1)	Achilles	Right	0.406	0.029	0.014	0.414
		Left	0.052	0.156	0.040	0.054
	Biceps	Right	**0.038**	**0.150**	**0.035**	**0.054**
		Left	n/a			
	Common extensor origin	Right	**0.010**	**0.273**	0.056	0.024
		Left	**0.001**	**0.414**	**0.080**	**0.005**
	Iliopsoas	Right	n/a			
		Left	**0.021**	**0.271**	**0.057**	**0.032**
	Subscapularis	Right	**0.020**	**0.207**	**0.041**	**0.021**
		Left	**0.015**	**0.218**	**0.047**	**0.027**
Marginal erosion (ER1)	All entheses	Right and left	Low variation not estimated using ordinal regression			
Surface bone formation (BF2)	Achilles	Right	0.843	0.002	-0.004	0.841
		Left	**0.020**	**0.272**	**0.058**	**0.026**
	Biceps	Right	0.124	0.038	0.025	0.166
		Left	0.110	0.093	0.032	0.118
	Common extensor origin	Right	0.068	0.171	0.054	0.091
		Left	**0.013**	**0.309**	**0.086**	**0.036**
	Iliopsoas	Right	0.412	0.035	0.014	0.411
		Left	0.060	0.157	0.030	0.091
	Subscapularis	Right	**0.040**	**0.156**	**0.036**	**0.051**
		Left	0.350	0.035	0.015	0.369
Surface erosion (ER2)	Achilles	Right	n/a			
	Biceps	Right	n/a			
	Common extensor origin	Right	0.029	0.359	0.188	0.149

	Iliopsoas	Right	n/a			
	Subscapularis	Right	0.154	0.091	0.024	0.191
	All entheses	Left	Low variation not estimated using ordinal regression			
Fine porosity (FPO)	Achilles	Right	0.505	0.021	-0.016	0.519
		Left	0.238	0.058	-0.310	0.271
	Biceps	Right	0.165	0.077	0.031	0.186
		Left	0.121	0.104	0.046	0.108
	Common extensor origin	Right and left	n/a			
	Iliopsoas	Right	0.516	0.037	-0.030	0.512
		Left	n/a			
	Subscapularis	Right	0.187	0.075	-0.028	0.176
		Left	0.450	0.024	-0.018	0.424
Macro-porosity (MPO)	Subscapularis	Right	0.047	0.155	0.043	0.064
		Left	0.366	0.042	0.029	0.393
	All entheses excluding subscapularis	Right and left	n/a			

Figure 1. *M. subscapularis*. A. Margin (black outline) and surface (grey area) B. Scores for each feature.

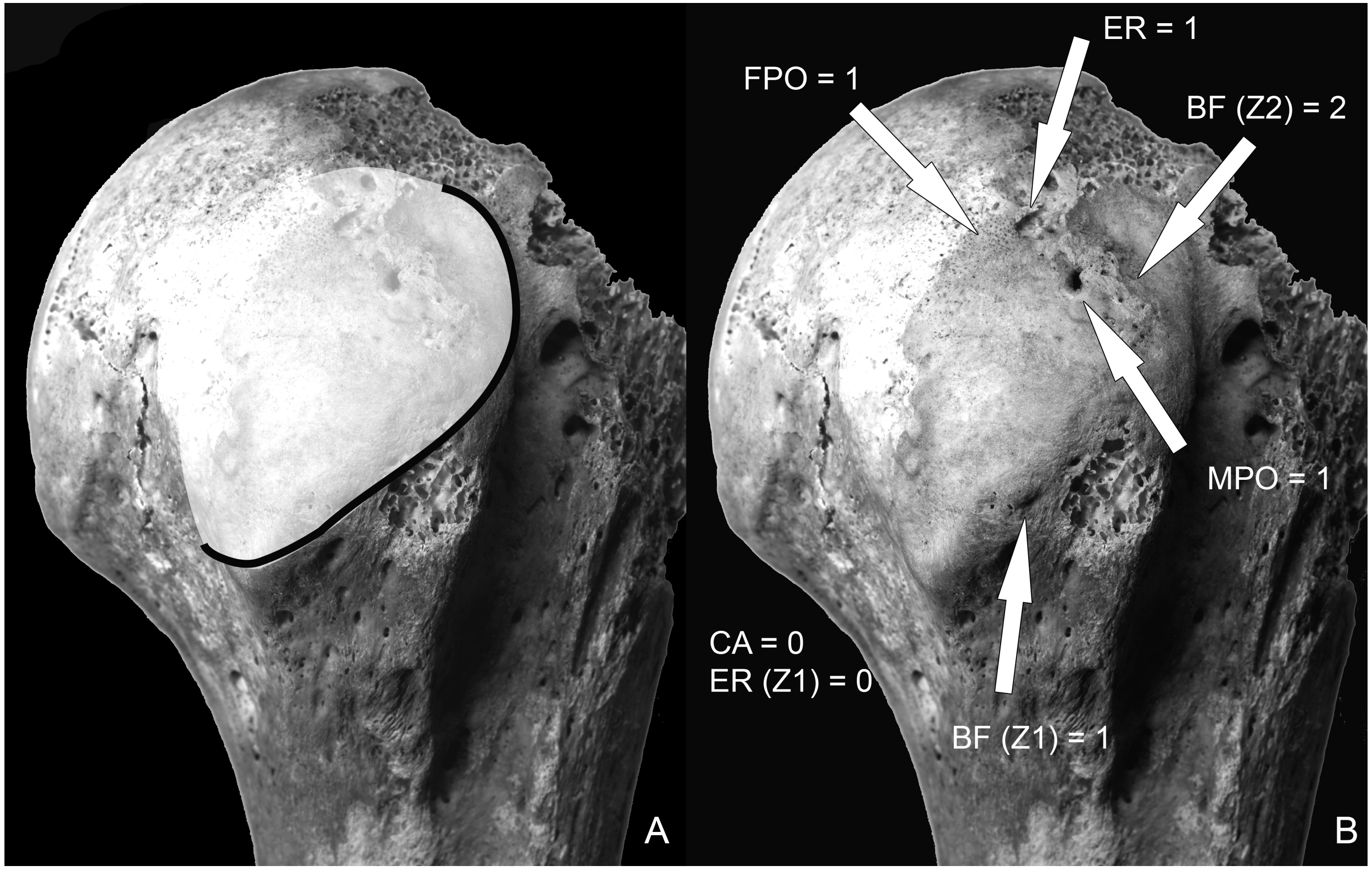


Figure 2. *M. biceps brachii*. A. Margin (black outline) and surface (grey area) B. Scores for each feature.

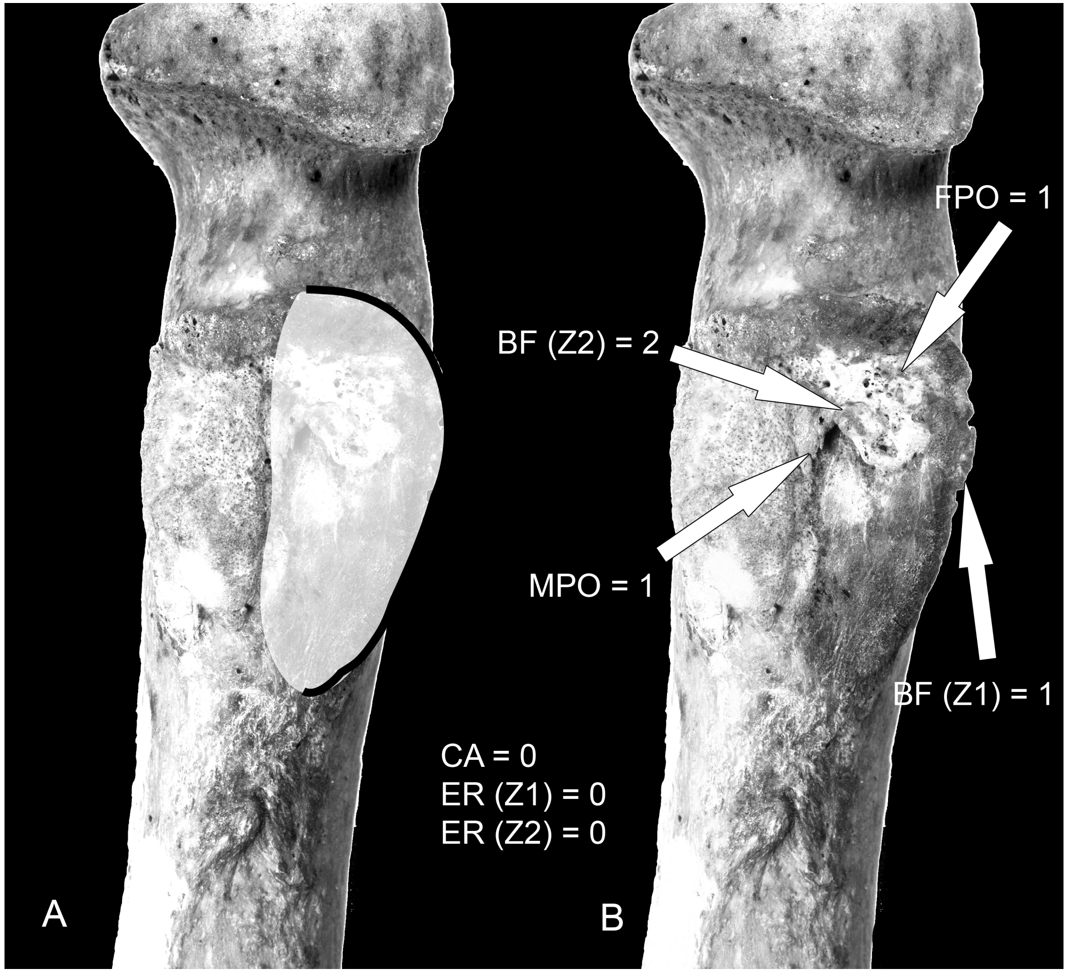


Figure 3. Intra-observer error: mean scores by feature for first and second observation.

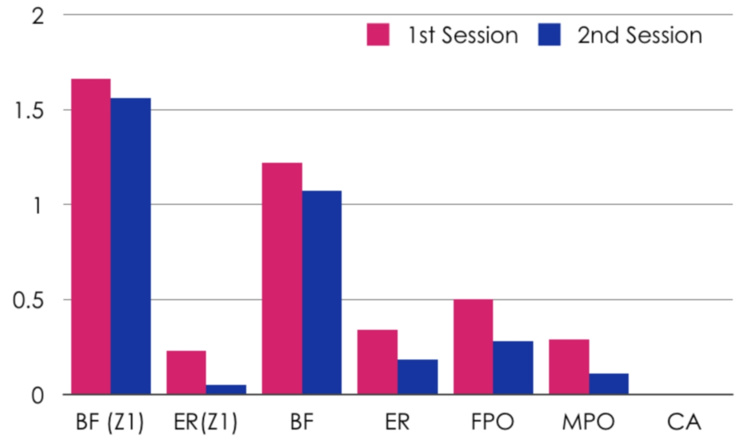


Figure 4. Inter-observer error: mean scores by feature by observer.

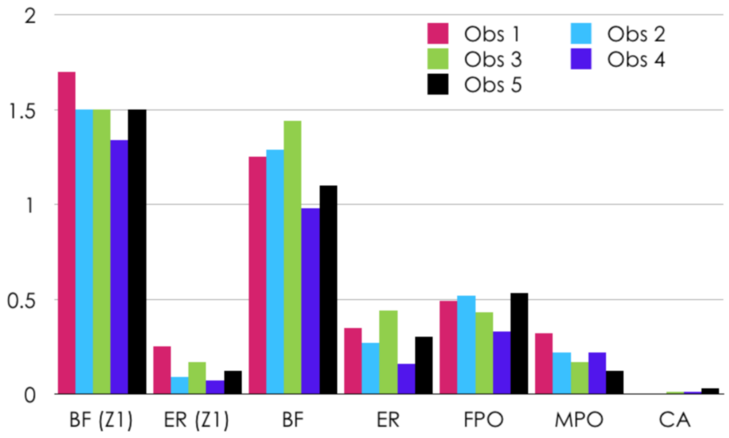
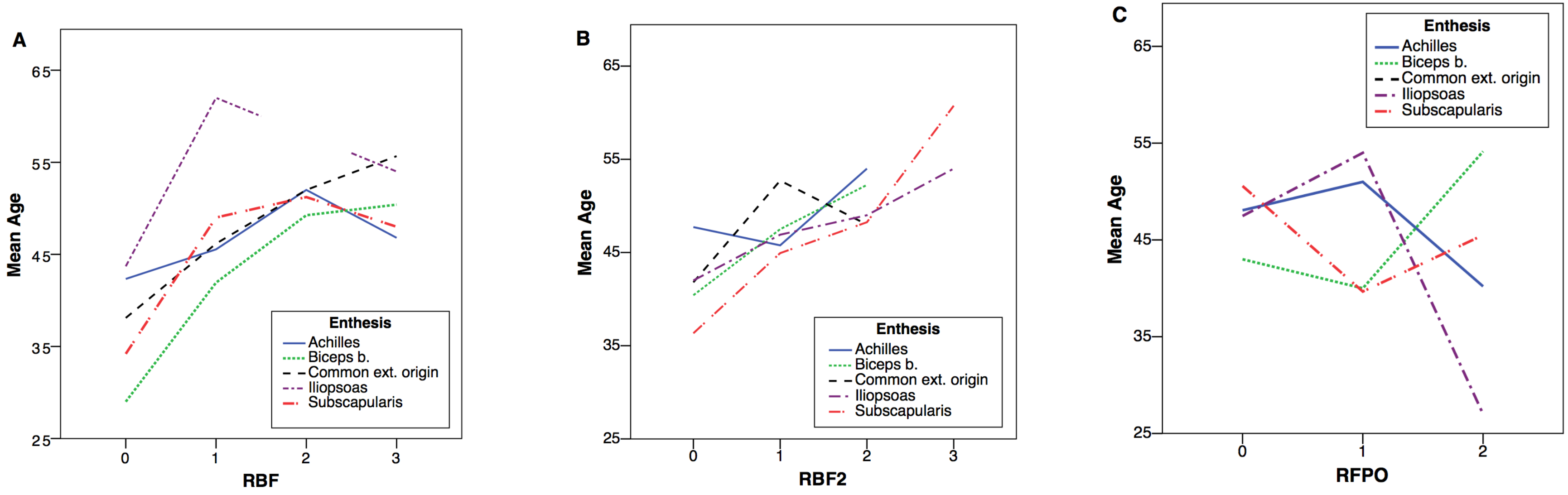


Figure 5. Plots of mean age by enthesis for three features (all right side). A. Bone formation zone 1. B. Bone formation zone 2. C. Fine porosity.



Appendix 1.
Inter-observer error percentage agreement following the reduction in the number of scores for each feature.

Appendix 1. Percentage agreement between pairs of observers by polling degrees 2 and 3 for BF, 1 and 2 for ER, FPO and MPO. In red the maximum agreement, in blue the minimum agreement.

Feature	Enthesis	1 vs 2	1 vs 3	1 vs 4	1 vs 5	2 vs 3	2 vs 4	2 vs 5	3 vs 4	3 vs 5	4 vs 5	mean
BF (Z1)	*Triceps surae*	51.9	63.0	77.8	77.8	70.4	66.7	70.4	81.5	70.4	81.5	71.1
	Common extensor	59.3	77.8	63.0	70.4	74.1	74.1	74.1	63.0	74.1	77.8	70.7
	Biceps brachii	70.4	66.7	63.0	59.3	70.4	59.3	63.0	81.5	70.4	74.1	67.8
	Iliopsoas	48.1	55.6	48.1	63.0	55.6	51.9	59.3	37.0	55.6	70.4	54.4
	Subscapularis	77.8	81.5	63.0	70.4	77.8	81.5	81.5	59.3	74.1	77.8	74.4
ER (Z1)	*Triceps surae*	77.8	81.5	81.5	77.8	92.6	92.6	96.3	92.6	96.3	88.9	87.8
	Common extensor	85.2	85.2	81.5	77.8	100.0	96.3	85.2	96.3	85.2	88.9	88.1
	Biceps brachii	81.5	88.9	77.8	85.2	85.2	81.5	88.9	88.9	92.6	88.9	85.9
	Iliopsoas	81.5	66.7	77.8	81.5	85.2	81.5	85.2	74.1	77.8	88.9	80.0
	Subscapularis	59.3	66.7	63.0	59.3	70.4	96.3	85.2	74.1	77.8	81.5	73.3
BF (Z2)	*Triceps surae*	18.5	48.1	59.3	51.9	44.4	48.1	48.1	37.0	70.4	55.6	48.1
	Common extensor	51.9	44.4	59.3	63.0	66.7	70.4	59.3	48.1	66.7	63.0	59.3
	Biceps brachii	55.6	59.3	51.9	44.4	81.5	55.6	48.1	51.9	44.4	70.4	56.3
	Iliopsoas	44.4	59.3	48.1	63.0	63.0	59.3	66.7	51.9	70.4	66.7	59.3
	Subscapularis	59.3	66.7	63.0	40.7	66.7	77.8	63.0	59.3	40.7	77.8	61.5
ER (Z2)	*Triceps surae*	81.5	85.2	88.9	81.5	96.3	92.6	88.9	96.3	85.2	81.5	87.8
	Common extensor	66.7	55.6	66.7	66.7	70.4	92.6	66.7	77.8	55.6	66.7	68.5
	Biceps brachii	74.1	63.0	88.9	88.9	74.1	85.2	77.8	66.7	66.7	92.6	77.8
	Iliopsoas	66.7	66.7	70.4	63.0	88.9	74.1	70.4	77.8	77.8	77.8	73.3
	Subscapularis	77.8	88.9	59.3	59.3	81.5	66.7	66.7	63.0	63.0	85.2	71.1
FPO	*Triceps surae*	40.7	63.0	66.7	70.4	63.0	59.3	44.4	66.7	55.6	59.3	58.9
	Common extensor	66.7	74.1	81.5	81.5	70.4	74.1	70.4	81.5	77.8	85.2	76.3
	Biceps brachii	66.7	51.9	81.5	77.8	51.9	63.0	66.7	40.7	51.9	74.1	62.6
	Iliopsoas	70.4	70.4	70.4	63.0	77.8	66.7	81.5	74.1	74.1	63.0	71.1
	Subscapularis	63.0	77.8	55.6	59.3	63.0	66.7	63.0	44.4	48.1	88.9	63.0
MPO	*Triceps surae*	85.2	88.9	88.9	85.2	96.3	88.9	100.0	92.6	96.3	88.9	91.1
	Common extensor	74.1	81.5	81.5	77.8	92.6	92.6	81.5	100.0	88.9	88.9	85.9
	Biceps brachii	74.1	81.5	85.2	85.2	85.2	81.5	88.9	88.9	96.3	92.6	85.9
	Iliopsoas	63.0	55.6	55.6	59.3	85.2	77.8	77.8	74.1	85.2	85.2	71.9
	Subscapularis	59.3	74.1	70.4	70.4	51.9	66.7	66.7	85.2	66.7	70.4	68.1
Overall agreement:												71.7