- Havelková P. et Villotte S., 2007. Enthesopathies: Test of reproducibility of the new scoring system based on current medical data. *Slovenská antropológia*, 10 (1): 51-57.

ENTHESOPATHIES: TEST OF THE REPRODUCIBILITY OF THE NEW SCORING SYSTEM BASED ON CURRENT MEDICAL DATA

Havelková Petra[1], Villotte Sébastien[2]

[1] Department of Anthropology, National Museum, Václavské náměstí 68, 115 79, Prague, Czech Republic, e-mail: havelkova.petra@gmail.com, petra_havelkova@nm.cz
[2] Laboratoire d'Anthropologie des Populations du Passé, Université Bordeaux 1, avenue des Facultés, 33405 Talence CEDEX, France, e-mail: s.villotte@anthropologie.u-bordeaux1.fr

Abstract: Many anthropological studies have focused on enthesopathies to find the correlation between their occurrence on muscle insertion sites and physical activities. However, these works are rarely aimed at methodology of the scoring system. A new method (Villotte 2006) is based on current anatomical and histological data. The aim of this work was to propose criteria for recording or not a stage with this method and test its reproducibility (inter-observer error). We used a sample of twenty adult male skeletons from the medieval settlement agglomeration – Mikulčice. According to the Kappa coefficient, reproducibility of the method was substantial, and almost perfect for fibrocartilaginous insertions.
Key-words: enthesopathies, reproducibility of method, insertion site

Introduction:
The first research attempting to find more information about musculoskeletal stress markers (term used by Hawkey, Merbs 1995) on the skeleton appeared in 16th century, especially in medical literature focusing on trade and military diseases (Kennedy 1989). However, only in 19th century scientists became aware that some irregularities in the region of enthesis, i.e. the insertion site of ligament, tendon and joint capsule (Benjamin et al. 2002), could be related to life habits (Cunha, Umbelino 1995). Recently, many anthropological works have focused on the study of the relationship between pathological changes of enthesis and occupational and physical stress (Dutour 1986, 1992, Knüsel 1993, Cunha, Umbelino 1995, Hawkey, Merbs 1995, Crubézy et al. 2002, Cappaso et al. 2004). The manifestations of physical activities at enthesis, associated with the development of cortical defects (enthesophytes, lesions, foramina, cysts) are termed enthesopathies (Niepel, Siťaj 1966, Lagier 1991).
A number of methods for evaluation of enthesopathies were proposed in the last twenty years (Crubézy 1988, Hawkey 1988, Mariotti et al. 2004, Villotte 2006). All the mentioned methods are visual and use the scale system of evaluation. In contrast to the other, Villotte's method is based on current anatomical and histological studies of entheses (insertion site, attachment site) (e.g. Resnick, Niwayama 1983, Benjamin, Ralph 1998, Benjamin et al. 2002). In accordance with medical knowledge, it is possible to distinguish two types of insertions by the character of the tissue at the bone-tendon interface (Benjamin et al. 1986, Benjamin, Ralph 1998, Benjamin et al. 2002). The first type of entheses is termed fibrous and it occurs in all regions of the appendicular skeleton with a thick layer of cortical bone (especially mid-shaft of long bones). The second type, the fibrocartilaginous entheses, is associated mainly with the region of the epiphyses and apophyses. These two types of insertions differ in their mechanical properties and appearance and the methods of evaluation of enthesopathies should take into account the character of the entheses. The categorisation of insertion sites according

to the character of entheses is thus the most important advantage of the newly proposed scoring system for evaluating enthesopathies (Villotte 2006).
The aim of this study was to conduct a test of reproducibility based on inter-observer error on this method. Good result of test reproducibility is the main presumption for the next application of this method.

Materials:
The material contains 20 skeletons from the Mikulčice – Valy (Czech Republic) housed in National Museum in Prague. Mikulčice is the medieval (Middle "Hillfort" Period) settlement agglomeration, dating from the 9^{th}-10^{th} century and is considered one of the possible centres of Great Moravia.
Since this work was focused on the evaluation of enthesopathies related to physical activity, only male adult skeletons were selected. The upper age limit was set at 50-60 years. We also excluded individuals with DISH or spondyloarthropathies for the same reason. We evaluated the insertions on all long bones, (except the fibula), the pelvis, the patella, the calcaneus and the vertebrae (C2-S1).

Methods:
This study uses a new scoring system proposed by S. Villotte (Villotte 2006) for the evaluation of enthesopathies which was applied to 36 insertion sites (18 on each side) of appendicular skeleton and 46 insertions (proximal and distal part) of yellow ligament at the spinal column (Table 1).

Table 1. Evaluated insertion sites, their scoring groups (G1-G4) and the type of enthesis (FC – fibrocartilaginous, F – fibrous)

Group	Muscle(s)	type of ins.
G1	m. subscapularis	FC
	m. supraspinatus, m. infraspinatus	FC
	caput commune ulnare	FC
	caput commune radiale	FC
	m. biceps brachii	FC
	m. semimembranosus, m. biceps femoris	FC
	m. gluteus minimus	FC
	m. gluteus medius	FC
	m. iliopsoas	FC
G2	m. triceps brachii	FC
	m. quadriceps femoris	FC
	m. triceps surae	FC
G3	insertions of yellow ligament at the spinal column	FC
G4	m. pectoralis major	F
	m. deltoideus	F
	m. pronator teres	F
	m. gluteus maximus	F
	m. vastus madialis, m. adductor magnus, m. adductor longus	F
	m. soleus	F

Based on a study of differences between normal (non-pathological) and pathological aspects of individual entheses and comparison with medical descriptions, Villotte differentiates four scoring systems (four muscle groups). The first three groups include fibrocartilaginous entheses (including vertebrae), while the fourth group includes fibrous entheses. Enthesopathies are classified using a three-stage scale (A, B, C). In the case of fibrocartilaginous insertions, especially those in the first group, the contour and surface of the insertion must be evaluated separately.

1st group (G1): All types of remodelling (enthesophytes, lesions, foramina or cysts) can be found on the surface of the insertion. The contour is most commonly remodelled in the form of enthesophytes. The contour and surface is evaluated separately by partial stage (0-2). The final stage (A, B, C) classification is based on the sum of these individual evaluations.
2nd group (G2): The most common form of remodelling of the insertion contour is enthesophytes. Stages are defined metrically, by the size of enthesophytes. The surface is rarely affected (erosive zones).
3rd group (G3). The method used for evaluation of enthesopathies in 3rd group was made according to Crubézy (1988). This group includes the most of insertions of yellow ligaments in the medial part of vertebrae (C2-S1). The proximal and distal part of the vertebral arc in the medial section is classified separately. Stages are also defined metrically.
4th group (G4). This is the most problematic group regarding to the methodology. The lack of anatomical and pathological descriptions does not allow define the stage precisely. This group includes fibrous entheses whose remodelling is expressed as an increasing irregularity of the surface or, rarely, as a cortical lacuna.
The second part of this work, beside the scoring system, is focused on the evaluation of taphonomy, i.e. if it is possible to evaluate the insertion due to damage (A, B or C) or not (NR – non-recordable). The evaluation of enthesopathies was realized by two observes. One of these is the author of the methodology. The reproducibility test was focused on two aspects. First, we monitored the reproducibility of the scoring system itself, i.e. the number of disagreements between the two observers determining the stage for a concrete insertion. The second part of the test involved the evaluation of taphonomy, i.e. the number of disagreements when one of the observers considered the insertion to be non-recordable (NR) while the second one assigned it a certain stage (A, B,C).
Only those insertions that were completely missing, not those that were damaged (NR), were taken into consideration as an absent (Abs). The number of missing insertions was identical for both observers. NR were considered when less than 50% of the inner or the outer part of an enthesis were observable for G1 and G2 and when less than 50% of the insertion were observable for G3 and G4. This method is based on a study of V. Mariotti (Mariotti at al. 2004), according to which, it is only possible to evaluate insertions if more than 50% of the surface has been preserved. The disadvantage of Mariotti proposal is that it does not distinguish between the contour and surface of insertion
The concordance rate between the observers was expressed as a percentage value as well as with the aid of the Kappa's coefficient that also assesses the quality of agreement between the observers (Kappa value was interpreted according to Landis, Koch 1977).

Results:
The study of reproducibility of scoring system and taphonomy involved 1210 insertions (430 insertions were absent).
The reproducibility test results for both systems (scoring - A, B, C and taphonomy – NR) are summarized in Table 2 (Part I). The overall agreement between the two observers for all the groups and all the variants of evaluation (A, B, C, NR) was 88% (kappa 0, 81). It is important to note that this result is based on partial results that are very different. For groups G1 and G2, the overall concordance rate when evaluating taphonomy and developmental stages was greater than 92%, while that for groups G3 and G4 was less than 85%.

Table 2. **Results of the reproducibility test for each scoring system (G1-4) and overall result (percentage value of direct agreement among observers (Part I-III), the Kappa's coefficients and the interpretation according to these coefficients, ABC- scoring system, NR – non-recordable).**

Gr.	N - insertions	Part I			Part II			Part III			
		NR+ABC			NR			ABC			
		% of concordance	Kappa		% of concordance	Kappa		N of insertions	% of concordance	Kappa	
G1	342	92,7	0,89	almost perfect	95,0	0,88	almost perfect	228	96,4	0,93	almost perfect
G2	97	92,8	0,90	almost perfect	95,9	0,90	almost perfect	65	95,5	0,93	almost perfect
G3	544	84,9	0,71	substantial	97,4	0,72	good	511	86,7	0,72	substantial
G4	227	83,7	0,76	substantial	96,9	0,87	almost perfect	193	84,5	0,74	substantial
total	1210	87,5	0,81	almost perfect	96,5	0,87	almost perfect	997	89,0	0,80	almost perfect

The results of test for taphonomy (NR) are contained in Table 2 (Part II).When evaluating taphonomy of the insertions we observed less than 5% of disagreements between both observers for all the groups as well as for each group alone. If we focus only on groups G1 and G2, we recorded more disagreements when evaluating taphonomy than when evaluating the developmental stage itself. However, these differences were not statistically significant. If we rule out all situations that include one of the variants of evaluation, NR/NR or NR/developmental stage, we can conduct the statistical test of reproducibility only for the stages of the scoring system (A, B, C; 997 insertions). The frequency of these is detailed in Table 2 (Part III). The consequent concordance rate for the scoring system was 89% (kappa koeficient = 0,8). This overall value is similar to the result achieved in previous work (Villotte 2006, Table 3). Inter-observer agreement in groups G1 and G2 is almost perfect (kappa koeficient = 0, 93 for both groups). Conversely, groups G4 (kappa koeficient = 0, 74) and G3 (kappa koeficient = 0, 72) appear to be the most problematic. This difference is supposed to be due to the different methodology of evaluation of enthesopathies.

Discussion:

The aim of this work was to test the reproducibility of a newly proposed method for the evaluation of enthesopathies in the region of muscle or ligament insertions. Two observers evaluated 20 adult male skeletons from the Early Medieval burial ground Mikulčice – Valy. Percentage value and the Kappa's coefficient were used for expression of the concordance rate between the observers. The Kappa's coefficient is used to assess inter-observer reliability and reproducibility of method when observing or otherwise coding qualitative or categorical variables. Kappa is considered an improvement over using % agreement to evaluate this type of reliability. The Kappa value, indicates how much better the observers are compared to a throw of the dice, and therefore gives the real credit to the agreement, which was found (Svanholm et al. 1989). The Kappa coefficient is statistical method for observers using nominal and ordinal categories (e.g. Svanholm et al. 1989, Sidor et al. 1993, Cummings et al. 1998, Villotte 2006).

The concordance rate achieved for the evaluation of taphonomy and developmental stages was 88%. We may thus consider the test of reproducibility almost perfect. Compared to other works that focused on the evaluation of enthesopathies, this result is by no means significantly different. While the reproducibility of the method presented by E. Crubézy (Crubézy et al. 2002) was lower (the inter-observer error is cited as 20%), V. Mariotti (Mariotti et al. 2004) presents a degree of concordance higher than 95%, as well as D. Hawkey (Hawkey, Merbs 1995).

The lower concordance rate in this case is mainly due to the low partial resultes of inter-observer error for groups G3 (vertebral insertions) and G4 (fibrous insertions). Concordance rate for these two groups was less than 85%. The reason for such different results is given by

the different methodology of evaluation in the individual groups for both taphonomy and the scoring system.
Concerning just a taphonomy evaluation, G1 and G2 are the most problematic groups. Attributing NR depends both on conservation of inner and outer part for these groups. This may probably be the reason for a greater number of disagreements between the observers. Despite this, we observed less than 5% of errors.
On the other hand, if developmental stage only (A, B nebo C) was evaluated, the best results were achieved in these groups, G1 and G2, including fibrocartilaginous insertions. For group G1, we recorded concordance rate 98%, for G2 97% (Table 2, Part III). These results show that the newly proposed methodology, from the aspect of these types of insertions, is reproducible very well. A similarly high degree of agreement was also achieved in the previous study (Villotte 2006, Table 3). The concordance rate for these two groups is better than in other comparative studies (Crubézy et al. 2002, Hawkey 1988, Mariotti et al. 2004).

Table 3. The comparison of reproducibility test results for the scoring system (ABC) with the study of Villotte (2006)

ABC	Villotte, 2006			Havelková, Villotte		
	% of concordance	Kappa		% of concordance	Kappa	
G1	85	0,75	substantial	96,4	0,93	almost perfect
G2	100	1	perfect	95,5	0,93	almost perfect
G3	92	0,82	almost perfect	86,7	0,72	substantial
G4	89	0,62	substantial	84,5	0,74	substantial
total	90	0,83	almost perfect	89,0	0,80	almost perfect

The differences in inter-observer error between these studies may also be due to the different methodology of evaluating the incidence of enthesopathies. Frequently used methodology is method of E. Crubézy (Crubézy et al. 1988, 2002). This methodology focuses mainly on the evaluation of enthesophytes. It does not evaluate lesions. Besides scopic evaluation, this methodology also uses metric assessment in the case of vertebral insertions. The methodology according to D. Hawkey (Hawkey 1998) is more precisely elaborated. It differentiates between the robusticity and two variants of a defect (enthesophytes and cortical lesions). It is specified for concrete muscle insertions. The methodology of V. Mariotti (Mariotti et al. 2004) is based on the previous two. It also evaluates separately robusticity and defects (enthesopathies). Concerning to enthesopathies, similarly as in the case of D. Hawkey, it differentiates between osteophytic (enthesophytes) and osteolytic (lesions, foramina, cysts) formations. It uses metric evaluation for both enthesopyhtes and lesions. None of these methodologies takes into account the anatomical and histological structure of the insertions. When calculating inter-observer error, scopic evaluation is always problematic. This evaluation is subjective to a various degree and impossible to verify. In methodology of S. Villotte, sub-groups were used for group G1 to render the final stages more precise. This form of evaluation may be one of the reasons for the greater degree of agreement between observers for group G1. In the group G2, the individual stages are defined metrically, thus the concordance rate is relatively high too.
This does not hold for the remaining two groups – G3 (vertebral insertions) and G4 (fibrous insertions). The different results for group G3 in this study and that of S. Villotte (2006; Table 3.) are probably due the incorrect understanding of the methodology in the first case. An error also may occur during measuring, when the gauge is placed differently or when enthesophytes are mistaken for osteophytes near the articular facets. The concordance rate for group G4, i.e. fibrous insertions, is the worst in both previously mentioned studies (Table 3.). Although the

methodology of evaluation is detailed for each insertion (especially stage C) evaluation of fibrous insertions appears to be the most difficult and the resulting kappa coefficient is only good.
Based on these conclusions (and the results of inter-observer error in the study Villotte, 2006), this method was applied to another sample of 250 skeletons from Mikulčice (by the first author) and recent, documented human skeletal collections from Greece (Havelková, unpublished report) and from Portugal, Great Britain and Italy (Villotte, in prep.).

Conclusion:
Compared with currently used methodologies for evaluating muscle insertions and enthesopathies, we may conclude that the test of reproducibility of the new method is generally good, and for evaluating fibrocartilaginous insertions, it is excellent. This method may be applied on the archaeological material as well as on the recent bone collections.

Acknowledgements:
The authors thank Prof. Karel Zvára for the assistance in the statistical analysis. This project was financed by grants from the GAČR 206/03/0699, Eco-Net No. 16368PB, the Czech Ministry of Culture VZ PM MK00002327201 and Research project MSM 0021620843.

References:

BENJAMIN M., KUMAI, T., MILZ, S., BOSZCZYK, B.M., BOSZCZYK, A.A., RALPHS, J.R., 2002: The skeletal attachment of tendons-tendon "entheses". Comp. Biochem. Physiol. A, 133:931-945.
BENJAMIN, M., EVANS, E.J., COPP, L., 1986: The histology of tendon attachments to bone in man. J. Anat., 149: 89-100.
BENJAMIN, M., RALPHS, J.R., 1998: Fibrocartilage in tendons and ligaments –an adaptation to compressive load. J. Anat., 193:481-494.
CAPASSO, L., PIERFELICE,L., MICHETTI, E., DI FABRIZIO, A., D'ANASTASIO, R., 2004: Lesions linked to althletic activities in the ancient roman population from herculaneum (Italy, first century AD). Anthropologie, XLII/2:181-187.
CRUBÉZY, E., 1988: Interactions entre facteurs bio-culturels, pathologie et caractères discrets, Exemple d'une population médiévale : Canac (Aveyron). 417 pp. (Thèse de doctorat de Médecine, Université de Montpellier).
CRUBÉZY, E., GOULET, J., BRUZEK, J., JELINEK, J., ROUGÉ, D., LUDES, F., 2002: Épidémiologie de l'arthrose et des enthésopathies dans une population européenne d'il y a 7700 ans. Revue du Rhumatisme (Éd.fr), 69:1217-1225.
CUMMINGS, R.J. , LOVELESS, E.A. , CAMPBELL, J., SAMELSON, S., MAZUR, J.M., 1998: Interobserver Reliability and Intraobserver Reproducibility of the System of King et al. for the Classification of Adolescent Idiopathic Scoliosis, J. Bone Joint Surg. Am., 80:1107-11.
CUNHA, E., 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. Antrop. Port., 13:49-68.
DUTOUR, O., 1986: Enthesopathies (Lesions of Muscular Insertions) as Indicators of the Activities of Neolithic Saharan Populations. Am. J. Phys. Anthropol., 71:221-224.
DUTOUR, O., 1992: Activités physique et squelette humain: Le difficile passage de l'actuel au fassile. Bull. Mém. Soc. Anthropol. Paris, 4:233-241

HAVELKOVÁ, P. (unpublished report) Enthesopathies in the modern, documented human collection from Greece: Application of the new scoring system. Archive of Dpt. Anthrop. of National Museum in Prague.
HAWKEY, D.E., 1988: Use of upper extremety enthesopathies to indicate habitual activity patterns. 86pp. (PhD thesis, Arizona State University).
HAWKEY, D.E., MERBS, C.F., 1995: Activity-induces musculoskeletal stress markers (MSM) and subsistence strategy changes among ancient Hudson Bay eskimos. Int. J. Osteoarch., 5:324-338.
KENNEDY, K.A.R., 1989: Skeletal markers of occupational stress. In: Iscan, M.Y., Kennedy, K.A.R. (eds.): Reconstruction of life from the skeleton. New York, Liss, p.130-160.
KNÜSEL, C.J., 1993: On the Biomechanical and Osteoarthritic Differeces Between Hunter-Gatherers and Agriculturalists. Am. J. Phys. Anthropol., 91:523-527.
LAGIER, R., 1991: Enthèses normales et enthésopathies. In: Simon, L., Hérisson, C., Rodineau, J. (eds.). Pathologie des insertions et enthésopathies. Paris-Milan-Barcelone-Bonn, Masson, p. 1-7.
LANDIS, J.R., KOCH, G.G., 1977: The measurement of observer agreement for categorical data. Biometrics, 33:159-174.
MARIOTTI, V., FACCHINI, F., BELCASTRO, M.G., 2004: Enthesopathies - Proposal of a standardized scoring method and applications. Collegium Anthropologicum, 28(1):145-159.
NIEPEL, G.A., SIT'AJ, S.,1979: Enthesopathy. Clin. Rheum. Dis., 5(3):857-872.
RESNICK, D., NIWAYAMA, G., 1983: Entheses and enthesopathy. Radiology, 146:1-9.
SIDOR, M.L., ZUCKERMAN, J.D., LYON, T., KOVAL, K., CUOMO, F., SCHOENBERG, N., 1993: The Neer classification system for proximal humeral fractures: An assessment of interobserver reliability and intraobserver reproducibility. J. Bone Joint Surg. Am., 75:1745-1750.
SVANHOLM, H., STARKLINT, H., GUNDERSEN, H.J.G., FABRICIUS, J., BARLEBO, H., OLSEN, S.,1989: Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. Acta Pathol. Microbiol. Immunol. Scand., 97(8):689-698.
VILLOTTE, S., 2006: Connaissances médicales actuelles, cotation des enthésopathies: nouvelle méthode. Bull. Mém. Soc. Anthropol. Paris, 18(1-2): 65-85.
VILLOTTE S., (in prep.) L'étude des enthésopathies, une approche pour l'interprétation des activités des populations du passé. Recherche méthodologique et application aux populations du Paléolithique supérieur et du Mésolithique d'Europe. (Thèse de doctorat, Université Bordeaux 1.)