



HAL
open science

Looking for the most suitable method for the study of enthesal changes: Application to upper limb's fibrocartilaginous entheses in a human medieval sample

Uxue Perez-arzak, Sébastien Villotte, Alvaro Arrizabalaga, Gonzalo J Trancho

► To cite this version:

Uxue Perez-arzak, Sébastien Villotte, Alvaro Arrizabalaga, Gonzalo J Trancho. Looking for the most suitable method for the study of enthesal changes: Application to upper limb's fibrocartilaginous entheses in a human medieval sample. *International Journal of Osteoarchaeology*, 2022, 32 (3), pp.595 - 606. 10.1002/oa.3085 . hal-03859425

HAL Id: hal-03859425

<https://hal.science/hal-03859425v1>

Submitted on 24 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Looking for the most suitable method for the study of enthesal changes: Application to upper limb's fibrocartilaginous entheses in a human medieval sample

Uxue Perez-Arzak¹  | Sébastien Villotte² | Alvaro Arrizabalaga¹ | Gonzalo J. Trancho³

¹Department of Geography, Prehistory and Archaeology, University of Basque Country (UPV-EHU), Gasteiz, Spain

²PACEA-UMR 5199, Université de Bordeaux, Pessac Cedex, France

³Department of Biodiversity, Ecology and Evolution, University Complutense of Madrid (UCM), Madrid, Spain

Correspondence

Uxue Perez-Arzak, Department of Geography, Prehistory and Archaeology, University of Basque Country (UPV-EHU) - "Micaela Portilla" Research Center. Vélaz de Elorriaga, no number - 01006 Gasteiz, Spain.
Email: uxue.perez@ehu.eus

Funding information

Prehistory Research Group, Grant/Award Number: IT-1223-19; Ministerio de Economía, Industria y Competitividad, Gobierno de España (MINECO), Grant/Award Number: HAR2017-82483-C3-1-P

Abstract

The study of enthesal changes (ECs), that is, osseous changes that occur at muscle attachment sites, has increased in recent years because they have been considered one of the most promising markers of activity in the human skeleton. Their aetiology is highly questioned because several factors unrelated to mechanical stress may be involved in their occurrence. Here, we test the validity of some methods created on different bases for the study of ECs to try to make an approach towards the most beneficial aspects of each method. Seven upper limbs' fibrocartilaginous entheses from a sample of 60 adult medieval individuals have been analysed employing three methods: Villotte (2006), Santana (2011) and the so-called New Coimbra method (Henderson et al., 2016; Henderson, Mariotti, et al., 2013). These three methods have been applied on the same skeletons. The results show that the three methods seem to indicate the same general trends but with different intensity and some nuances. Santana's method always provides the most frequencies of presence and the highest scores of ECs. In turn, the Coimbra method finds the fewest scores and frequencies of ECs probably because its criteria are the most specific of the three methods. Villotte's method seems to be the one that determines the general tendencies of ECs in the clearest way. This is also evidenced by the calculated Kappa coefficients: the highest agreement is detected between Villotte's lowest scores and Santana's highest ones ($k = 0.743$); the moderate agreement ($k = 0.596$; $k = 0.544$) recorded between Villotte's highest scores and Coimbra's lowest indicates that the latest one registers the lowest scores of ECs between the analysed three methods. All this points to a methodological consensus that should be based on current anatomical knowledge, criteria that are intuitive and flexible, and the use of many scores of osseous expression to avoid excessive sample fragmentation.

KEYWORDS

activity-related osseous markers, enthesopathies, methodology, osteoarchaeology, subsistence patterns

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *International Journal of Osteoarchaeology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Enthesal changes (ECs) (i.e., osseous alterations of the attachment sites of muscles, tendons and ligaments on the bone) are morphological conditions often used to reconstruct the way of life of past human individuals or populations. They have been regarded as one of the most promising areas of research on human occupation (Jurmain et al., 2012).

Particularly since the 1980s, numerous studies have considered these changes. Since the beginning, there were problems regarding their aetiology and lack of information on anatomical issues (Dutour, 1986; Hawkey & Merbs, 1995; Jurmain et al., 2012), which led to a gap in methodological consensus.

The first studies assumed that ECs appeared as a direct consequence of muscle activity, reflected in the probably most-widely used term of “musculoskeletal stress markers” (Hawkey & Merbs, 1995). However, interest in understanding the nature and aetiology of ECs has increased in recent years, and this has led to the emergence of more biologically appropriate terms such as “enthesopathy” (Mariotti et al., 2004; Villotte, 2006) or “ECs” (Jurmain & Villotte, 2010; Villotte et al., 2016).

These new research in ECs have brought to light that many conditions are involved in their origin. They can be summarised in two main groups: biomechanical factors and physiological ones, such as sex, age, genetic legacy and hormonal levels, among others (Alves Cardoso & Henderson, 2010; Havelková et al., 2013; Henderson et al., 2017; Henderson, Mariotti, et al., 2013; Jurmain et al., 2012; Lieverse et al., 2009; Mariotti et al., 2004, 2007; Milella et al., 2015; Robb, 1998; Villotte & Knüsel, 2013, 2014). There are also internal anatomical conditions that strongly influence the possible association between ECs and physical activity (Alves Cardoso & Henderson, 2010; Henderson et al., 2017; Henderson, Mariotti, et al., 2013; Villotte, 2006; Villotte et al., 2016, 2010). Thus, two different types of ECs are known: the fibrocartilaginous and fibrous entheses. This is important since mechanical stresses affect each one differently (Benjamin et al., 2002, 2006; Villotte, 2006).

All these issues have led to the development of a wide variety of methods (e.g., Galtés et al., 2006; Hawkey & Merbs, 1995; Henderson et al., 2016; Henderson, Mariotti, et al., 2013; Mariotti et al., 2004, 2007; Myszka & Piontek, 2012; Robb, 1998; Santana, 2011; Stirland, 1998; Villotte, 2006), making it difficult to compare the results of different studies. Here, we attempt to provide an approach to the most appropriate method for the analysis of ECs, thus highlighting the most interesting aspects of each method in order to use them in a discussion towards consensus that has not yet been reached in this area.

For this purpose, we compare three methods (including traditional as well as recent ones) in the case of fibrocartilaginous entheses of the human upper limb from a medieval sample to avoid influences coming from anatomical differences. This comparison could clarify the path to the most suitable method to study ECs as it is important to use a methodology that is both easy to understand and apply and incorporates current medical and anatomical knowledge.

2 | MATERIALS AND METHODS

We studied a sample from a skeletal assemblage from the site of Buradón (Salinillas de Buradón, Araba) that has been excavated in several seasons since 1993 (Rodríguez Fernández, 2013). The large size of the site and its location between the steep side of Buradón Hill and the banks of the River Ebro have given rise to a quite complex sequence of occupations, from the Late Bronze Age-Early Iron Age to a full Middle Age (Cepeda-Ocampo, 2019; Martínez Salcedo & Cepeda-Ocampo, 1994; Quirós Castillo et al., 2009).

The skeletal remains analysed here belong to two important burial times: one group composed mostly of graves with wooden coffins from the fourth to seventh centuries, and a second group from Early Medieval stone slab graves, dated between the ninth and tenth centuries. One of the latest studies on the site has pointed out (Cepeda-Ocampo, 2019) that there are few details available about the archaeological context of the sample, which hampers the exact chronological attribution of the individuals studied here. However, given the methodological character of this work, the selection of this sample was made regarding the good state of conservation.

A sample of 60 adults out of over 250 individuals buried in this necropolis (Cepeda-Ocampo, 2019) has been analysed (Table 1). They were all older than 20 years of age, to avoid bone alterations related to the growth and development of childhood and adolescence. Previous studies have shown that by the age of 20 years, the structural properties of entheses are already well defined (Henderson et al., 2017; Milella et al., 2015; Villotte et al., 2010; Villotte & Knüsel, 2013). Similarly, it appears that ECs are reliable indicators of activity before the age of 50 (Milella et al., 2012; Villotte, 2009). Although some individuals older than 50 have been included, this factor will be considered when looking for information related to that subgroup.

For the estimation of sex, morphological and morphometrical methods were employed for the *os coxae* (Brůžek, 2002; Brůžek et al., 2017; Murail et al., 2005), and the morphology of the skull (Buikstra & Ubelaker, 1994) was also used.

Age at death was based on the changes in the surface of the pubic symphysis (Brooks & Suchey, 1990) and on the appearance of the sacroiliac surface of the *os coxae* (Schmitt, 2005). We also employed the state of fusion of later secondary ossification centres, following Owings Webb and Suchey (1985) and Albert and Maples (1995). Based on these methodologies and the results they provided, we decided to divide the individuals into the following age groups: young adult (between 20 and 40 years old), middle adult (between 40 and 60 years old) and senile adult (those over 60 years old). We added the group “general adult” to refer to those individuals whose age could not be assigned more precisely but who could be placed between 20 and 60 years of age.

This work has focussed on fibrocartilaginous entheses of the upper limb to avoid anatomical influences in the results. In this way, as shown in Table 2, five entheses of the humerus, one of the radius and one of the ulna have been studied. These entheses were analysed on a total sample of 119 humeri, 118 radii and 118 ulnae. The osseous

TABLE 1 Demographic representation of the studied reference sample

	Age groups				Total
	Young adult (20–40)	Middle adult (40–60)	Old adult (>60)	General adult (20–60)	
Female	8	5	0	10	23
Male	13	11	2	5	31
Indeterminate	1	1	0	4	6
Total	22	17	2	19	60

TABLE 2 The selected upper limb entheses. Most of them are insertions. In the case of the origin, it is specified in the name of the muscle

Bone	Muscle	Location of the entheses	Code
Humerus	Subscapularis	Supero-medial part of the lesser tubercle	HSB
Humerus	Supraspinatus	Superior facet of the greater tubercle	HSP
Humerus	Infraspinatus	Middle facet of the greater tubercle	HNF
Humerus	Common origin of extensors	Anterior and lateral part of the lateral epicondyle	HXT
Humerus	Common origin of flexors	Anterior and medial part of the medial epicondyle	HFL
Radius	Biceps brachii	Medial part of the bicipital tuberosity	RBB
Ulna	Triceps brachii	Posterior part of the olecranon	UTB

changes in all these entheses have been well documented in previous studies (Benjamin et al., 2006; Villotte, 2006).

Three different methods have been applied by a single observer (U.P-A.). The method created by S. Villotte (2006) was based mainly on studies of osseous changes and anatomy at the level of enthesis (Benjamin et al., 2002; Benjamin & McGonagle, 2001). This methodology divides the entheses into four groups depending on their biological nature. A three-stage scale (A, B and C) is established depending on the type of bone change observed.

Second, the methodology based on criteria developed by J. A. Santana (2011) was used. It establishes 5° of osseous changes: 0, lack of expression; 1, incipient expression; 2, moderate expression; 3, important expression; and 4, pathological expression. It is based on the best known methods in the field of ECs (Galtés et al., 2006; Galtés & Malgosa, 2007; Hawkey & Merbs, 1995; Mariotti et al., 2007; Robb, 1998), but it analyses the entheses without considering their different anatomies.

Finally, one of the most modern methods was included in the comparison: the new Coimbra method (Henderson et al., 2016; Henderson, Mariotti, et al., 2013). It combines the clinical definitions in the descriptions in Villotte's (2006) method with the different formations that can be observed at enthesis level (bone formation [BF], erosion, textural change, fine porosity, macroporosity and cavitation) following the methodology of Mariotti et al. (2004, 2007). It was not intended to discuss past activities but to record the type of bone change and to see the influence of different factors on the occurrence of ECs (Henderson, Mariotti, et al., 2013).

One of the main limitations for the reproduction and comparison of results from an analysis of ECs is that each research group/researcher uses a different method, as noted in previous studies

(Henderson, Mariotti, et al., 2013; Mariotti et al., 2004; Villotte, 2006). Therefore, the variables that are considered in each method also vary (Figure 1). The Coimbra method does not offer a single score for an enthesis but eight, corresponding to each type of morphological change. For this reason, we only considered the results recorded for BF in both zones in the present comparisons because the character is most often documented in the present study. Moreover, the previous studies showed (Henderson et al., 2017) that the variable obtains higher scores and the most varied ones. This is advantageous to observe the methodological differences more clearly.

Bearing in mind the main objective of comparing methodologies and avoiding problems due to sample size, a basic but illustrative descriptive analysis was made, taking into account means and frequencies of presence/absence of ECs. In order to do so in the most similar way, we made some adaptations: the absence of an EC attributed by each method corresponds to Grade A or 0 in Villotte's method, Grade 0 in the Coimbra method, and Grade 0 and 1 in Santana's.

Asymmetry was also determined using for that only the entheses that could be analysed in both sides within an individual; so, the sample size was reduced because if one of the sides could not be recorded, the asymmetry was taken as a non-recordable (NR). The frequencies of asymmetry were calculated taking into account whether the two sides showed the same score of osseous changes (not-biased), whether the right side showed more EC score (right-biased) or the left side was the most developed (left-biased).

In order to obtain some evidence of the agreement between the different methodologies, the unweighted Cohen's Kappa was applied using IBM SPSS version 24 software. To get over the obstacle of

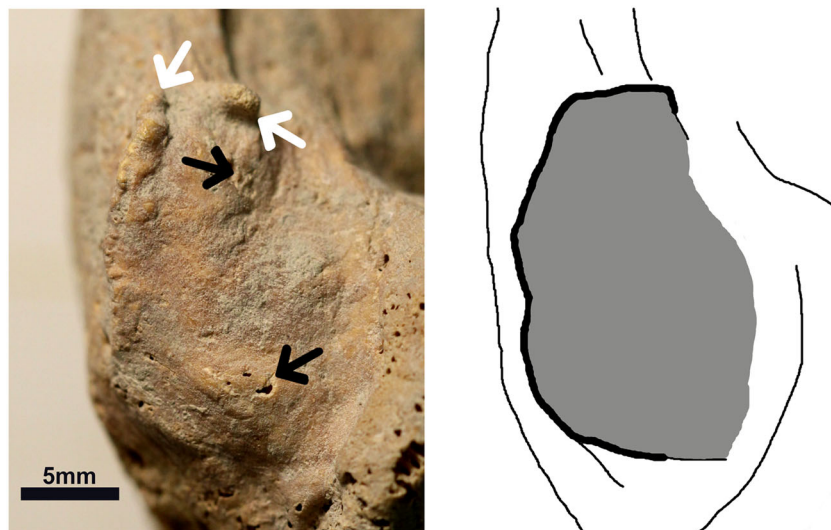


FIGURE 1 Example of an EC at the attachment point of the common extensor (HXT) of a right humerus of a male adult individual. On the right, the thick line indicates the margin of the enthesis (zone 1) and the grey colour shows the total surface (zone 2). On the left, the appearance of the enthesis assessed with different scores allotted with each method: According to Villotte (2006), it is an intermediate stage (regular surface –value 0– and margin with presence of enthesophytes –value 2–); according to Santana (2011) it is the highest stage (grade 4 due to the presence of osteogenic process at the margin); according to Coimbra (Henderson et al., 2016; Henderson, Mariotti, et al., 2013), for BF Z1 it is an intermediate stage (bone formation larger than 1 mm but less than 50% of the margin's extension affected) and for BF Z2 a lower stage (absence of any bone formation). The white arrows indicate enthesophytes on the margin; black arrows point taphonomic alterations [Colour figure can be viewed at wileyonlinelibrary.com]

comparing methodologies based on different numbers and types of scores, the proposal of Natarajan et al. (2007) was followed. In this way, the three ratings were dichotomised at each cutpoints on the ordinal scale: four cutpoints for Santana (0 versus ≥ 1 ; ≤ 1 versus $2 \geq$; ≤ 2 versus $3 \geq$; and ≤ 3 versus 4) and two for Villotte (0 versus ≥ 1 and ≤ 1 versus 2) and Coimbra (0 versus ≥ 1 and ≤ 1 versus 2). Then the respective Kappa coefficients were examined, taking into account the scale of strength of agreement proposed by Landis and Koch (1977).

3 | RESULTS

Starting with the frequencies of the presence of ECs (Table 3), they show the highest ones for Santana's method and, in contrast, the lowest clearly for the Coimbra method. The frequencies obtained by Villotte's method seem to be in an intermediate position. This trend is observed in the whole sample and also if it is divided depending on sex, laterality or age. It is noteworthy that when the sample is classified by those factors, the three methods display the same trends but differ in the intensity.

According to sex, the most frequencies of ECs are seen in the male series according to the three methods. Nevertheless, Villotte's method shows this most clearly, since its frequencies of occurrence of ECs are higher for males in the case of all entheses and the differences between females and males are greater than with the other two methods.

When the sample is classified by laterality, the same pattern is repeated. The predominance of the right arm is clear, but Villotte's

method shows this for most of the ECs with clear differences in the frequencies recorded for the right and left sides, whereas Santana's method exhibits some exceptions (HSB and HSP) and the Coimbra method does not indicate any clear side dominance.

Despite the small size of the sample to make comparisons related to age, the frequencies documented in all the ECs with the three methods indicate a clear influence of this factor in their appearance.

Frequencies of the presence of ECs obtained for individuals displaying different scores on each side, that is, bilateral asymmetry (Figure 2 and Table S1), show that the three methods indicate a clear predominance of the right limb. It should be noted that the analysis of asymmetry only takes into account the assessments of the entheses that have been recorded on both sides within one same individual; so, it should not be surprising that the highest frequencies of the three methods indicate individuals with the same scores of osseous alterations on each side, and therefore an absence of asymmetry (Table S2). In the case of individuals displaying asymmetry, it is significant that Santana's method shows right-biased predominance most noticeably for all the entheses, unlike in the case of laterality. The exceptions of HXT with the Villotte's method and HSB, HSP, HNF and RBB with Coimbra method are clear. Although Santana's method shows right-biased asymmetry, it agrees in recording less evident differences for RBB; again, it is worth noting the preliminary nature of these results due to the sample size.

In the second place, methodological differences were observed in relation not only to the frequencies of presence but also to different scores of osseous changes. For that, the means of the ECs' scores were obtained according different conditions.

TABLE 3 Frequencies (%) of presence of ECs for the humeri, radii and ulnae samples according to the three methods in the whole sample, and according to sex (F = female; M = male; Indet = indeterminate), laterality (R = right, L = left) and age-at-death (YA = young adult, MA = middle adult, OA = old adult, AG = general adult). The highest results are shown in **bold**, in the case of sex, between F and M, and in the case of age, between YA and MA. BF = bone formation; Z1 = zone 1; Z2 = zone 2; n = the total number of analysed entheses, without taking into account the non-recordable (NR) ones

Method	Enthesis	Total sample	Sex			Side		Age			
			F	M	Indet	R	L	YA	MA	OA	AG
			n/%	n/%	n/%	n/%	n/%	n/%	n/%	n/%	n/%
Villotte, 2006	HSB	84/73.8%	36/61.1%	40/85%	8/75%	46/73.9%	38/73.7%	31/51.6%	24/95.8%	3/100%	26/76.9%
	HSP	75/25.3%	34/26.5%	35/28.6%	6/0%	39/30.8%	36/19.4%	33/9.1%	19/26.3%	3/33.3%	20/50%
	HNF	67/47.8%	33/45.5%	30/53.3%	4/25%	37/56.8%	30/36.7%	29/34.5%	18/50%	1/100%	19/63.2%
	HXT	82/57.3%	29/48.3%	43/69.8%	10/30%	44/50%	38/65.8%	34/41.2%	24/75%	2/100%	22/59.1%
	HFL	73/42.5%	26/38.5%	35/54.3%	12/16.7%	39/43.6%	34/41.2%	29/24.1%	18/61.1%	1/100%	25/48%
	RBB	99/69.7%	39/61.5%	51/80.4%	9/44.4%	50/78%	49/61.2%	36/58.3%	26/92.3%	4/100%	33/60.6%
	UTB	87/27.6%	35/22.9%	40/37.5%	12/8.3%	44/36.4%	43/18.6%	33/9.1%	22/59.1%	4/50%	28/21.4%
	UTB	84/92.9%	36/88.9%	40/95%	8/100%	46/91.3%	38/94.7%	31/93.5%	24/95.8%	3/100%	26/88.5%
Santana, 2011	HSB	75/49.3%	34/35.3%	35/57.1%	6/83.3%	39/48.7%	36/50%	33/36.4%	19/52.6%	3/66.7%	20/65%
	HSP	67/76.1%	33/63.6%	30/90%	4/75%	37/81.1%	30/70%	29/69%	18/94.4%	1/100%	19/68.4%
	HNF	82/70.7%	29/69%	43/81.4%	10/30%	44/70.5%	38/71.1%	34/73.5%	24/70.8%	2/100%	22/63.6%
	HXT	73/94.5%	26/92.3%	35/97.1%	12/91.7%	39/94.9%	34/94.1%	29/89.7%	18/94.4%	1/100%	25/100%
	HFL	99/62.6%	39/53.8%	51/62.7%	9/100%	50/66%	49/59.2%	36/36.1%	26/69.2%	4/100%	33/81.8%
	RBB	87/78.2%	35/80%	40/77.5%	12/75%	44/86.4%	43/69.8%	33/66.7%	22/81.8%	4/100%	28/85.7%
	UTB	84/21.4%	36/19.4%	40/20%	8/37.5%	46/15.2%	38/28.9%	31/6.5%	24/29.2%	3/66.7%	26/26.9%
	UTB	75/1.3%	34/0%	35/2.9%	6/0%	39/0%	36/2.8%	33/3%	19/0%	3/0%	20/0%
Coimbra method (Z1 BF)	HNF	67/6%	33/9.1%	30/3.3%	4/0%	37/8.1%	30/3.3%	29/0%	18/0%	1/100%	19/15.8%
	HXT	82/40.2%	29/34.5%	43/48.8%	10/20%	44/43.2%	38/36.8%	34/26.5%	24/45.8%	2/100%	22/50%
	HFL	73/13.7%	26/11.5%	35/20%	12/0%	39/20.5%	34/5.9%	29/17.2%	18/5.6%	1/100%	25/12%
	RBB	99/29.3%	39/28.2%	51/31.4%	9/22.2%	50/28%	49/30.6%	36/13.9%	26/46.2%	4/50%	33/30.3%
	UTB	87/21.8%	35/17.1%	40/32.5%	12/0%	44/27.3%	43/16.3%	33/3%	22/50%	4/25%	28/21.4%
	HSB	84/34.5%	36/25%	40/42.5%	8/37.5%	46/39.1%	38/28.9%	31/12.9%	24/45.8%	3/66.7%	26/46.2%
	HSP	75/5.3%	34/2.9%	35/8.6%	6/0%	39/5.1%	36/5.6%	33/0%	19/10.5%	3/0%	20/10%
	HNF	67/9%	33/9.1%	30/13.3%	4/0%	37/8.1%	30/10%	29/0%	18/5.5%	1/0%	19/26.3%
Coimbra method (Z2 BF)	HXT	82/17.1%	29/17.2%	43/18.6%	10/10%	44/15.9%	38/18.4%	34/8.8%	24/25%	2/0%	22/22.7%
	HFL	73/9.6%	26/3.8%	35/14.3%	12/0%	39/15.4%	34/0%	29/3.4%	18/11.1%	1/0%	25/16%
	RBB	99/13.1%	39/10.3%	51/17.6%	9/0%	50/14%	49/12.2%	36/2.8%	26/23.1%	4/25%	33/15.2%
	UTB	87/2.3%	35/2.9%	40/2.5%	12/0%	44/4.5%	43/0%	33/3%	22/0%	4/0%	28/3.6%

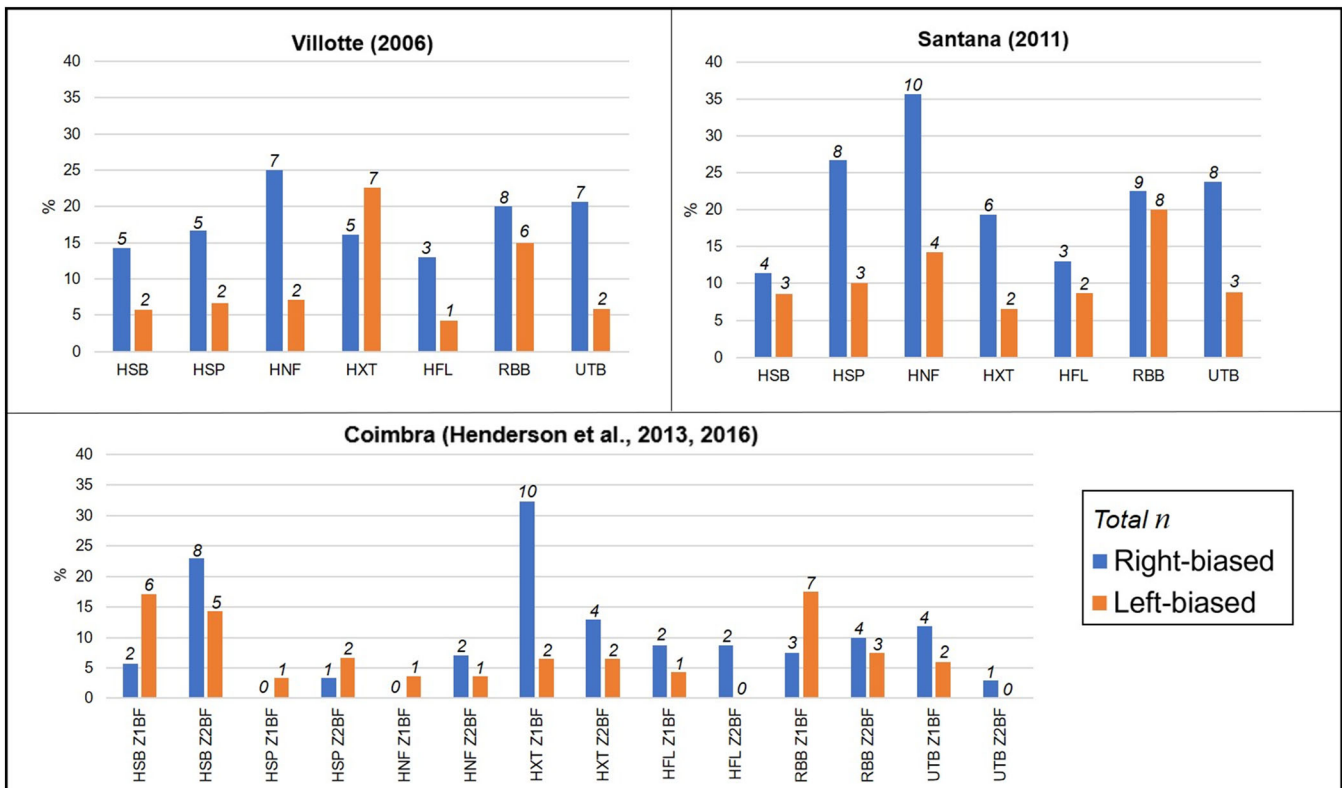


FIGURE 2 Comparison of frequencies (%) of asymmetry according ECs recorded by each method. Above each column the total number (n) of analysed entheses that present some kind of bias related to asymmetry (in this case, only the entheses that could be studied on both sides within the same individual were used) [Colour figure can be viewed at wileyonlinelibrary.com]

The means according to laterality (Figure 3 and Table S2) show that the three methods indicate a predominance of the use of the right limb in the case of most entheses. The differences between the left and the right sides are greater in the results obtained with Villotte's one, with the exception of HXT. Although Santana's method shows right-side predominance for all the entheses, the graph shows that these differences are not as obvious as in the previous case. Coimbra is the method that shows more exceptions (i.e., more left side dominance). When the sex factor is added, the same pattern is seen (Figure 5 and Table S4).

According to sex (Figure 4 and Table S3), the three methods generally coincide in recording the highest scores in ECs for the male sample. The differences between the values recorded for the female and male sample appear to be clearest with Villotte's method, followed by Santana's one. The scores registered by the Coimbra method for BF do not show such a clear trend. If laterality is also considered within sex (Figure 5 and Table S4), the predominance of higher scores in the male sample is maintained. This is clear in the case of both sides in the results of Santana's method. However, in the left side sample, the results obtained with Villotte's method display fewer differences for some ECs and even in some cases, as the HSP, there is a predominance of higher scores for the female sample. These exceptions can also be observed for some variables of the Coimbra method, but in this case, for both sides.

Regarding only the different scores of ECs, it is noteworthy that in general Santana's method documents scores higher than its intermediate degree (2). With Villotte's method, the means only exceed its intermediate degree (B or 1) in the case of the subscapularis (HSB) and biceps brachii (RBB) entheses, and only in the male sample. None of the means in the Coimbra method reaches its intermediate degree (1).

This is a trend confirmed by Kappa coefficients (Table 4). The highest agreements are detected between Santana's and Villotte's results. It is interesting the high agreement between the cutpoint 0 versus ≥ 1 , indicating their similar results in recording the absence of ECs. The agreement detected for Villotte's cutpoint 0 versus ≥ 1 and Santana's ≤ 2 versus $3 \geq$ (or even ≤ 3 versus 4) may confirmed that the last method tends to identify higher scores than the first one. Between Santana's and Coimbra's results happens the similar, but lower Kappa coefficients are identified, indicating less agreement between Santana's and Coimbra methods than between Santana's and Villotte's. Finally, it is important that the moderate Kappa coefficient calculated for the Coimbra's 0 versus ≥ 1 cutpoint and Villotte's ≤ 1 versus 2. This tendency would be in accordance with the observations done according to the ECs' means, presenting Coimbra as the method that indicates the lowest scores of ECs; we must not forget that here is only being evaluated the BF in the case of Coimbra method and, therefore, this trend could be modified if other condition is taken into account.

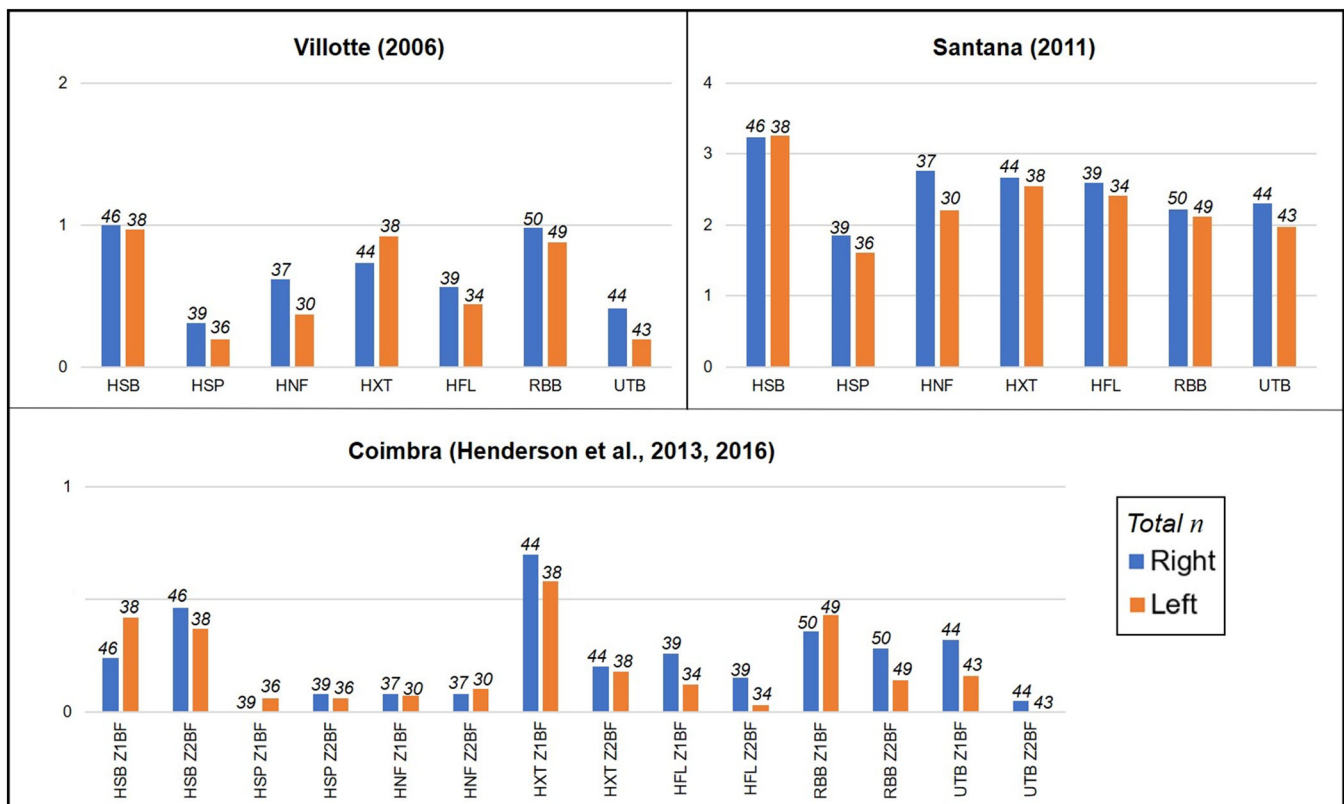


FIGURE 3 Comparison of the means of ECs' scores obtained by each method according to laterality. The vertical axis shows the scores used by each methodology. Above each column the total number (n) of analysed entheses, without taking into account the non-recordable (NR) ones [Colour figure can be viewed at wileyonlinelibrary.com]

If these coefficients are calculated in subgroups (Data S1–S4), the same main trends are observed. Looking, for instance, at the coefficients for each entheses (Data S4), it is highlighted that HSB, HSP and HNF show very low coefficients in general in all of the cases (so a practical absence of agreement); in the other side, HXT always appears to be the entheses where the methods show the most agreement in the way of analysing it (obtaining high coefficients for the most of cutpoints comparing the three methods). So, there may also be differences related to the nature of each entheses.

4 | DISCUSSION

There is no current consensus on the most accurate methodology for the analysis of ECs, which makes it difficult the comparison between the results obtained by different researchers. Here, we have tried to solve this problem by analysing the results obtained by three methods. These were selected mainly because they propose different ways of recording ECs.

The results have shown clear differences between the chosen methods, thus underscoring the need to establish a methodological consensus and the basis with which this can be sought.

The method of Santana (2011) obtains the highest frequencies of presence and the greatest scores of ECs, also confirmed by Kappa coefficients. This is in line with it being the methodology that includes more scores than the other analysed methods, which least emphasises biological and anatomical aspects of the entheses. Therefore, these high scores and frequencies of presence can be understood by regarding that the descriptions in this technique consider osseous changes in relation to the form, volume or the presence of roughness. The point is that all these conditions may be part of a normal variation in the appearance of many entheses. Consequently, this type of methodology attributes intermediate or even high scores to some entheses whose appearance could really be associated with a normal variation, that is, a weak score or even a total absence of an osseous change (Alves Cardoso & Henderson, 2010; Jurmain et al., 2012). Furthermore, the main issue with this type of methods is that they assume a continuous relationship between the score of the EC and the observed osseous change or, ultimately, the physical effort carried out by the individual (Henderson, Mariotti, et al., 2013; Jurmain et al., 2012).

In addition, most of the studies on ECs that use methods with many categories or scores finally reduce their number to enable statistical analyses and comparisons with other studies (Alves Cardoso & Henderson, 2010; Henderson, Craps, et al., 2013; Mariotti

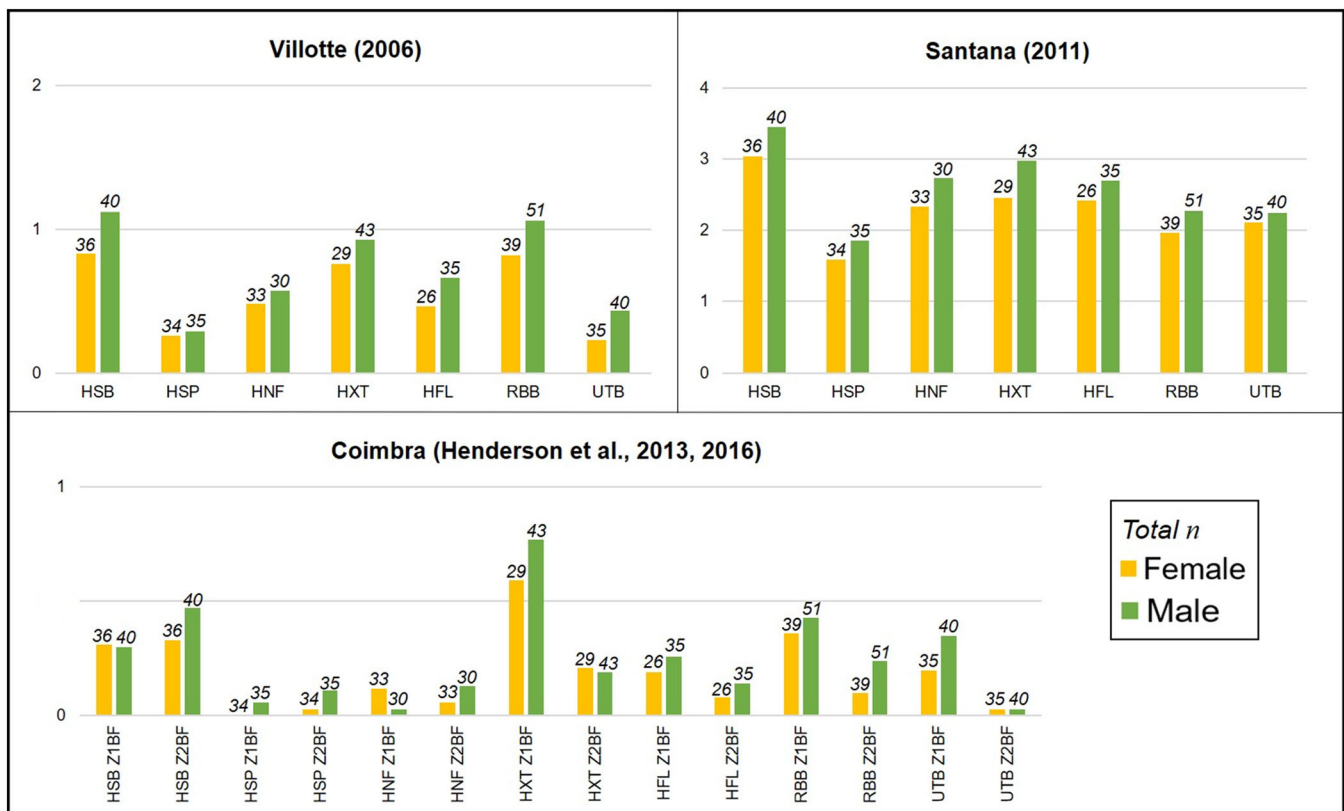


FIGURE 4 Comparison of the means of ECs' scores obtained by each method according to sex. The vertical axis shows the scores used by each methodology. Above each column the total number (n) of analysed entheses, without taking into account the non-recordable (NR) ones [Colour figure can be viewed at wileyonlinelibrary.com]

et al., 2007; Milella et al., 2015; Mysza & Piontek, 2012). A high number of scores increases the risk of segmenting more the sample; therefore, this makes it more difficult to detect general patterns (Mariotti et al., 2007).

The results obtained here support this since when the sample is subdivided based on sex or laterality, Santana's methodology (2011) still maintains the general trends but not in such an evident way as Villotte's method (2006), for instance. Similarly, and always bearing in mind the limitations of small sample size, this explains the asymmetry detected using Santana's method compared with those seen by the Villotte's or Coimbra method (Henderson et al., 2016; Henderson, Mariotti, et al., 2013); if there are more scores, it is more likely that an individual will obtain a different score for each side.

Another important difference is that the Coimbra method has proved to be the one that obtained fewest frequencies of presence of ECs regarding BF, also confirmed with the Kappa coefficients. This can be explained by the fact that this method was designed with more demanding and specific criteria where personal experience and ability to interpret the descriptions are more important than in other methods (Wilczak et al., 2017), as the authors themselves admit and is seen in the low reproducibility recorded at the beginning (Henderson et al., 2016; Henderson, Mariotti, et al., 2013). This is avoided with methods whose principles are more flexible, that is, do not include too specific criteria in their descriptions and thus facilitate the ascription

of a score to the observed bone alteration, as is the case with the Santana's or Villotte's method, compared with the Coimbra's criteria. However, it should not be forgotten that the latter was created with the intention of analysing the influence of different factors on each type of ECs (Henderson, Mariotti, et al., 2013), which may explain the difference in the rigour of its criteria.

Similar is true for the different scores of bone alterations. The present study has verified that the specific criteria—particularly of the Coimbra method but also to a lesser extent of Villotte's method—depend much on the experience of the analyst (Wilczak et al., 2017); this reduces the presence of the highest scores, also seen in other studies (Henderson et al., 2017).

The frequencies of presence of ECs, the mean of the scores and the Kappa coefficients regarding Villotte's method, situate this approach in an intermediate position between Santana's and Coimbra methods. The combination of few scores to be assigned, which has been seen to excessively fragment the sample (Mariotti et al., 2007), and easy descriptions to follow, is reflected in the results recorded with Villotte's method. This method appears to explain the general trends in the clearest way according to the basic analyses done in the present study, that is, where there was a difference between male versus female samples or right versus left samples; Villotte's results were the ones that show greater differences between the two groups. Additionally, its fine resolution in the establishment of correlations

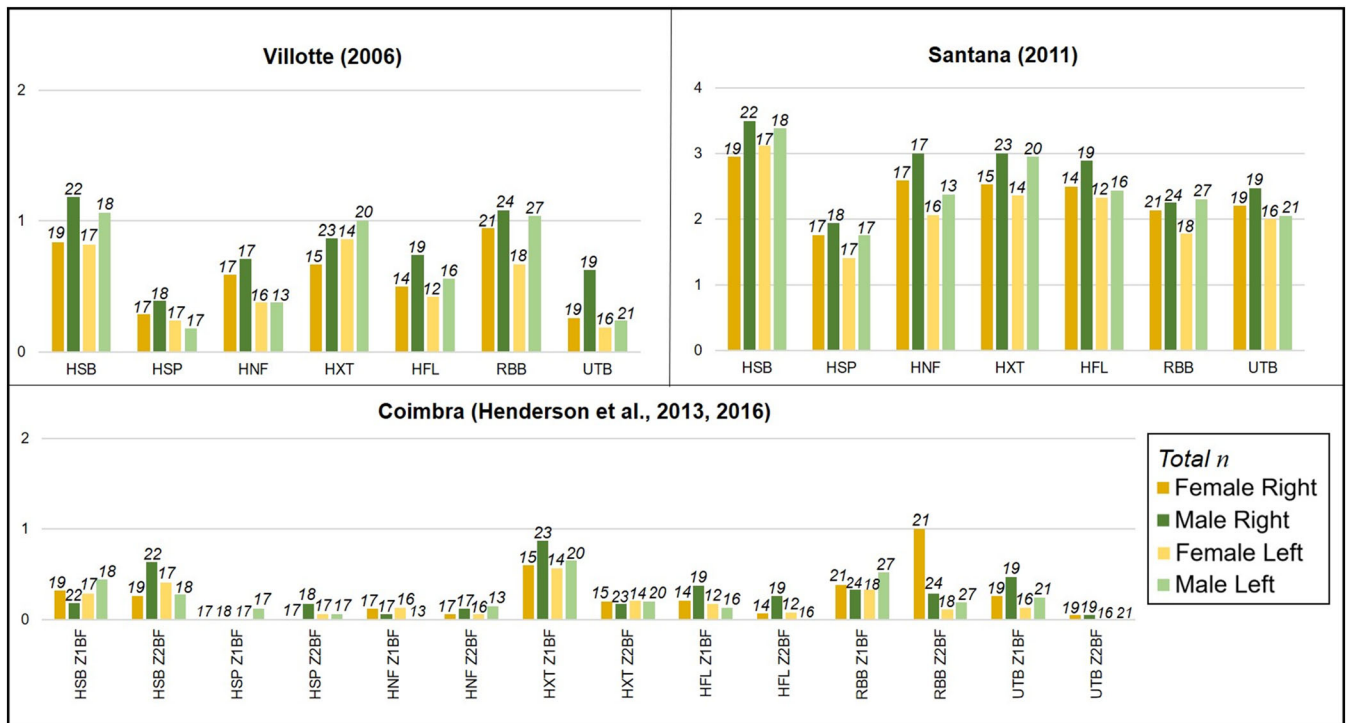


FIGURE 5 Comparison of the means of ECs' scores obtained by each method according to sex and laterality. The vertical axis shows the scores used by each methodology. Above each column the total number (*n*) of analysed entheses, without taking into account the non-recordable (NR) ones [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Kappa coefficients for the whole sample at all cutpoints for all entheses combined, comparing in pairs the results obtained by the three methods applied in this study. In brackets, *p* values; in **bold**, coefficients indicating moderate to almost-perfect (0.41–1) agreement, following Landis and Koch (1977). *n* = 567 entheses, without taking into account the non-recordable (NR) ones; BF = bone formation; Z1 = zone 1; Z2 = zone 2

		Santana			
<i>Santana versus Villotte</i>		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.74 (<i>p</i> < 0.001)	0.319 (<i>p</i> < 0.001)	0.743 (<i>p</i> < 0.001)	0.609 (<i>p</i> < 0.001)
	≤ 1 versus 2	0.013 (<i>p</i> = 0.058)	0.092 (<i>p</i> < 0.001)	0.334 (<i>p</i> < 0.001)	0.405 (<i>p</i> < 0.001)
		Santana			
<i>Santana versus Coimbra</i>		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.021 (<i>p</i> = 0.014)	0.147 (<i>p</i> < 0.001)	0.459 (<i>p</i> < 0.001)	0.52 (<i>p</i> < 0.001)
	≤ 1 versus 2	0.008 (<i>p</i> = 0.141)	0.06 (<i>p</i> < 0.001)	0.238 (<i>p</i> < 0.001)	0.337 (<i>p</i> < 0.001)
Coimbra BF Z2	0 versus ≥1	0.013 (<i>p</i> = 0.056)	0.099 (<i>p</i> < 0.001)	0.355 (<i>p</i> < 0.001)	0.411 (<i>p</i> < 0.001)
	≤ 1 versus 2	0.002 (<i>p</i> = 0.472)	0.015 (<i>p</i> = 0.041)	0.063 (<i>p</i> < 0.001)	0.081 (<i>p</i> < 0.001)
		Coimbra BF Z1		Coimbra BF Z2	
<i>Coimbra versus Villotte</i>		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.401 (<i>p</i> < 0.001)	0.165 (<i>p</i> < 0.001)	0.264 (<i>p</i> < 0.001)	0.042 (<i>p</i> < 0.001)
	≤ 1 versus 2	0.596 (<i>p</i> < 0.001)	0.494 (<i>p</i> < 0.001)	0.544 (<i>p</i> < 0.001)	0.228 (<i>p</i> < 0.001)

between ECs and physical effort has been proven on several occasions (Havelková et al., 2013; Henderson, Craps, et al., 2013; Henderson, Mariotti, et al., 2013; Villotte et al., 2010; Villotte & Knüsel, 2014).

Finally, despite the small number of individuals older than 60 years of age-at-death in the sample, it can be confirmed that the three methods similarly demonstrate the influence of age in the appearance of ECs, as many previous studies have found (Alves

Cardoso & Henderson, 2010; Henderson et al., 2017; Mariotti et al., 2004, 2007; Milella et al., 2020, 2012; Robb, 1998; Villotte et al., 2010).

It is also important to note the biases that may have arisen in these results due to the fact that in this case the analyses were carried out by an individual scorer (Wilczak et al., 2017). In addition to the scorer's own general experience, the experience with each method also comes into play; this is also related to the problems of basing analyses only on published descriptions (Davis et al., 2013). As has been shown, prior training, ideally with the originators of the methods themselves, greatly changes the degree of agreement of the results (Wilczak et al., 2017).

The limitation of this study in terms of sample size also needs to be taken into account. This has conditioned that the analyses carried out here were limited to mainly descriptive questions. Future analyses with a larger sample size are necessary to observe whether the trends evidenced in this study are fulfilled; it would also be interesting to perform the same analyses also taking into account fibrous entheses or entheses of the lower limbs.

5 | CONCLUSIONS

Throughout this work, we have compared the results obtained for fibrocartilaginous entheses of the human upper limb using three different methods. Since the beginning of the use of this type of analysis as a way of reconstructing the daily life of past populations, the lack of terminological and methodological consensus has been evident, which has made comparisons of different studies exceedingly difficult.

The fact that the method with more scores (Santana, 2011) has documented higher frequencies of presence and higher scores of ECs warns about the risk of sample segmentation. Villotte's (2006) method has been shown to point to general trends in a more evident way. At the same time, the Coimbra method (Henderson et al., 2016; Henderson, Mariotti, et al., 2013) has the lowest ECs' frequencies of presence and, together with Villotte's (2006) method, the lowest scores of osseous alterations. Furthermore, the latter two methods agree that they are based on current anatomical data, so they are biologically more appropriate. As for Santana's and Villotte's methods, they both provide detailed and well-illustrated descriptions, without being too specific, which makes the recording process simpler and quicker.

All these remarks highlight the positive features seen in the three methods; this gives an insight into the basis for a methodological consensus on the analysis of ECs. The main thing is not to consider many scores of osseous changes, as a way to avoid segmenting the sample too much, also points to the importance of flexible grading criteria with clear descriptions, so that they do not depend much on the researcher's own experience. Furthermore, new anatomical knowledge on ECs should be included. Ultimately, it is clear that the methodological consensus in ECs should aim at obtaining general patterns that show differences as clearly as possible.

ACKNOWLEDGMENTS

This study was partially funded by the Prehistory Research Group (IT-1223-19) at the University of the Basque Country (UPV/EHU) and the Ministerio de Economía, Industria y Competitividad, Gobierno de España (MINECO) project HAR2017-82483-C3-1-P. We would especially like to thank the Arabako Arkeologia Museoa, and above all Jaione Agirre García and Raúl Sánchez Rincón for facilitating access to the anthropological collection and for all their help. We would also like to thank the editors and the anonymous reviewers whose suggestions undoubtedly helped to improve this paper.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Uxue Perez-Arzak  <https://orcid.org/0000-0002-5209-9968>

REFERENCES

- Albert, A. M., & Maples, W. R. (1995). Stages of epiphyseal union for thoracic and lumbar vertebral centra as a method of age determination for teenage and young adult skeletons. *Journal of Forensic Sciences*, *40*, 623–633. <https://doi.org/10.1520/JFS13838J>
- Alves Cardoso, F., & Henderson, C. Y. (2010). Enthesopathy formation in the humerus: Data from known age-at-death and known occupation skeletal collections. *American Journal of Physical Anthropology*, *141*, 550–560. <https://doi.org/10.1002/ajpa.21171>
- Benjamin, M., Kumai, T., Milz, S., Boszczyk, B. M., Boszczyk, A. A., & Ralphs, J. R. (2002). The skeletal attachment of tendons-tendon "entheses". *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *133*, 931–945. [https://doi.org/10.1016/S1095-6433\(02\)00138-1](https://doi.org/10.1016/S1095-6433(02)00138-1)
- Benjamin, M., & McGonagle, D. (2001). The anatomical basis for disease localisation in seronegative spondyloarthopathy at entheses and related sites. *Journal of Anatomy*, *199*, 503–526. <https://doi.org/10.1046/j.1469-7580.2001.19950503.x>
- Benjamin, M., Toumi, H., Ralphs, J. R., Bydder, G., Best, T. M., & Milz, S. (2006). Where tendons and ligaments meet bone: Attachment sites ('entheses') in relation to exercise and/or mechanical load. *Journal of Anatomy*, *208*, 471–490. <https://doi.org/10.1111/j.1469-7580.2006.00540.x>
- Brooks, S., & Suchey, J. M. (1990). Skeletal age determination based on the os pubis: A comparison of Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution*, *5*, 227–238. <https://doi.org/10.1007/BF02437238>
- Brůžek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*, 157–168. <https://doi.org/10.1002/ajpa.10012>
- Brůžek, J., Santos, F., Dutailly, B., Murail, P., & Cunha, E. (2017). Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology. *American Journal of Physical Anthropology*, *164*, 440–449. <https://doi.org/10.1002/ajpa.23282>
- Buikstra, J. E., & Ubelaker, D. H. (1994). Standards for data collection from human skeletal remains. *Arkansas Archeological Survey Research Series: Arkansas*. 44.
- Cepeda-Ocampo, J. J. (2019). Dos iglesias tardoantiguas en el norte de Hispania: San Martín de Finaga (Bizkaia) y Buradón (Álava). In J. López

- Vilar (Ed.), *Actes 4 Congrès Internacional d'Arqueologia i Món Antic. VII Reunió d'Arqueologia Cristiana Hispànica. El cristianisme en l'Antiguitat Tardana, noves perspectives* (pp. 243–250). Universitat Rovira i Virgili - Institut d'Estudis Catalans.
- Davis, C. B., Shuler, K. A., Danforth, M. E., & Herndon, K. E. (2013). Patterns of interobserver error in the scoring of enthesal changes. *International Journal of Osteoarchaeology*, 23(2), 147–151. <https://doi.org/10.1002/oa.2277>
- Dutour, O. (1986). Enthesopathies (lesions of muscular insertions) as indicators of the activities of Neolithic Saharan population. *American Journal of Physical Anthropology*, 71, 221–224. <https://doi.org/10.1002/ajpa.1330710209>
- Galtés, I., & Malgosa, A. (2007). Atlas metodològic para el estudio de marcadores musculoesqueléticos de actividad en el radio. *Paléo*, 3, 3–33.
- Galtés, I., Rodríguez-Baeza, A., & Malgosa, A. (2006). Mechanical morphogenesis: A concept applied to the surface of the radius. *The Anatomical Record Part A*, 288A, 794–805. <https://doi.org/10.1002/ar.a.20337>
- Havelková, P., Hladík, M., & Velemínský, P. (2013). Enthesal changes: Do they reflect socioeconomic status in the Early Medieval Central European population? (Mikulčice-Klášteřísko, Great Moravian Empire, 9th–10th century). *International Journal of Osteoarchaeology*, 23, 237–251. <https://doi.org/10.1002/oa.2294>
- Hawkey, D. E., & Merbs, C. F. (1995). Activity-induced musculoskeletal stress markers (MSM) and subsistence strategy changes among Hudson Bay eskimos. *International Journal of Osteoarchaeology*, 5, 324–338. <https://doi.org/10.1002/oa.1390050403>
- Henderson, C. Y., Craps, D. D., Caffell, A. C., Millard, A. R., & Gowland, R. (2013). Occupational mobility in 19th century rural England: The interpretation of enthesal changes. *International Journal of Osteoarchaeology*, 23, 197–210. <https://doi.org/10.1002/oa.2286>
- Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Villotte, S., & Wilczak, C. (2013). Recording specific enthesal changes of fibrocartilaginous entheses: Initial tests using the Coimbra method. *International Journal of Osteoarchaeology*, 23, 152–162. <https://doi.org/10.1002/oa.2287>
- Henderson, C. Y., Mariotti, V., Pany-Kucera, D., Villotte, S., & Wilczak, C. (2016). The new 'Coimbra method': A biologically appropriate method for recording specific features of fibrocartilaginous enthesal changes. *International Journal of Osteoarchaeology*, 26, 925–932. <https://doi.org/10.1002/oa.2477>
- Henderson, C. Y., Mariotti, V., Santos, F., Villotte, S., & Wilczak, C. A. (2017). The new Coimbra method for recording enthesal changes and the effect of age-at-death. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, 29, 140–149. <https://doi.org/10.1007/s13219-017-0185-x>
- Jurmain, R., Alves Cardoso, F., Henderson, C. Y., & Villotte, S. (2012). Bioarchaeology's holy grail: The reconstruction of activity. In A. L. Grauer (Ed.), *Companion to paleopathology* (pp. 531–552). Wiley-Blackwell. 10.1002/9781444345940.ch29
- Jurmain, R., & Villotte, S. (2010). Terminology. Entheses in medical literature and physical anthropology: a brief review [Online]. Document published online in 4th February following the Workshop in Musculoskeletal Stress Markers (MSM): Limitations and achievements in the reconstruction of past activity patterns, University of Coimbra, July 2–3, 2009. Coimbra, CIAS - Centro de Investigação em Antropologia e Saúde. [Consulted in 8th January 2020]. Available from: http://www.uc.pt/en/cia/msm/MSM_terminology3
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–117. <https://doi.org/10.2307/2529310>
- Lieverse, A. R., Bazaliiskii, V. I., Goriunova, O. I., & Weber, A. W. (2009). Upper limb musculoskeletal stress markers among Middle Holocene foragers of Siberia's Cis-Baikal region. *American Journal of Physical Anthropology*, 138, 458–472. <https://doi.org/10.1002/ajpa.20964>
- Mariotti, V., Facchini, F., & Belcastro, M. G. (2004). Enthesopathies—Proposal of standardized scoring method and applications. *Collegium Anthropologicum*, 28(1), 145–159.
- Mariotti, V., Facchini, F., & Belcastro, M. G. (2007). The study of entheses: Proposal of a standardised scoring method for twenty-three entheses of the postcranial skeleton. *Collegium Anthropologicum*, 31(1), 291–313.
- Martínez Salcedo, A., & Cepeda-Ocampo, J. J. (1994). Castro de Buradón. In *Arqueología de urgencia en Álava. 1989–1993*, Baldeón A, García E (coords.) (pp. 43–60). Diputación Foral de Álava, Museo de Arqueología de Álava.
- Milella, M., Alves Cardoso, F., Assis, S., Perréard Lopreno, G., & Speith, N. (2015). Exploring the relationship between enthesal changes and physical activity: A multivariate study. *American Journal of Physical Anthropology*, 156, 215–223. <https://doi.org/10.1002/ajpa.22640>
- Milella, M., Belcastro, M. G., Mariotti, V., & Nikita, E. (2020). Estimation of adult age-at-death from enthesal robusticity: A test using an identified Italian skeletal collection. *American Journal of Physical Anthropology*, 173(1), 190–199. <https://doi.org/10.1002/ajpa.24083>
- Milella, M., Belcastro, M. G., Zollikofer, C. P. E., & Mariotti, V. (2012). The effect of age, sex, and physical activity on enthesal morphology in a contemporary Italian skeletal collection. *American Journal of Physical Anthropology*, 148(3), 379–388. <https://doi.org/10.1002/ajpa.22060>
- Murail, P., Brůžek, J., Houët, F., & Cunha, E. (2005). DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, 17, 167–176. <https://doi.org/10.4000/bmsap.1157>
- Myszka, A., & Piontek, J. (2012). Variation of musculoskeletal stress markers in the Medieval population from Cedynia (Poland)—Proposal of standardized scoring method application. *Collegium Anthropologicum*, 36(3), 1009–1017.
- Natarajan, S., McHenry, M. B., Lipsitz, S., Klar, N., & Lipschutz, S. (2007). Agreement between two ratings with different ordinal scales. In J. L. Auget, N. Balakrishnan, M. Mesbah, & G. Molenberghs (Eds.), *Advances in statistical methods for the health sciences. Application to cancer and AIDS studies, genome sequence analysis, and survival analysis* (pp. 139–148). Birkhäuser.
- Owings Webb, P. A., & Suchey, J. M. (1985). Epiphyseal union of the anterior iliac crest and medial clavicle in a modern multiracial sample of American males and females. *American Journal of Anthropology*, 68, 457–466. <https://doi.org/10.1002/ajpa.1330680402>
- Quirós Castillo, J. A., Azkarate Garai-Olaun, A., Bohigas Roldán, R., García Camino, I., Palomino Lázaro, A. L., & Tejado Sebastián, J. M. (2009). Arqueología de la Alta Edad Media en el Cantábrico oriental. In *Actas del Congreso Internacional "Medio siglo de arqueología en el Cantábrico Oriental y su entorno"*, Llanos Ortiz de Landaluze A (coord.) (pp. 462–500). Instituto Alavés de Arqueología.
- Robb, J. E. (1998). The interpretation of skeletal muscle sites: A statistical approach. *International Journal of Osteoarchaeology*, 8, 363–377. [https://doi.org/10.1002/\(SICI\)1099-1212\(199809\)8:5<363::AID-OA438>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-1212(199809)8:5<363::AID-OA438>3.0.CO;2-K)
- Rodríguez Fernández, J. (2013). Conjunto arqueológico de Buradón, en Salinillas de Buradón. *Arkeoikuska*, 13, 87–92.
- Santana, J. A. (2011). El trabajo fosilizado: Patrón cotidiano de actividad física y organización social del trabajo en la Gran Canaria Prehispanica. PhD Dissertation: Universidad de las Palmas de Gran Canaria.
- Schmitt, A. (2005). Une nouvelle méthode pour estimer l'âge au décès des adultes à partir de la surface sacro-pelvienne iliaque. *Bulletins et Mémoires de la Société d'Anthropologie de Paris*, 17(1–2), 89–101. <https://doi.org/10.4000/bmsap.943>
- Stirling, A. J. (1998). Musculoskeletal evidence for activity: Problems of evaluation. *International Journal of Osteoarchaeology*, 8, 354–362. [https://doi.org/10.1002/\(SICI\)1099-1212\(199809\)8:5<354::AID-OA432>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1099-1212(199809)8:5<354::AID-OA432>3.0.CO;2-3)
- Villotte, S. (2006). Connaissances médicales actuelles, cotation des enthésopathies: Nouvelle méthode. *Bulletins et Mémoires de la Société*

- d'Anthropologie de Paris*, 18(1-2), 65–85. <https://doi.org/10.4000/bmsap.1325>
- Villotte, S. (2009). *Enthésopathies et activités des hommes préhistoriques. Recherche méthodologique et application aux fossiles européens du Paléolithique supérieur et du Mésolithique*. Oxford. 10.30861/9781407305264
- Villotte, S., Assis, S., Alves Cardoso, F., Henderson, C. Y., Mariotti, V., Milella, M., Pany-Kucera, D., Speith, N., Wilczak, C. A., & Jurmain, R. (2016). In search of consensus: Terminology for enthesal changes (EC). *International Journal of Paleopathology*, 13, 49–55. <https://doi.org/10.1016/j.ijpp.2016.01.003>
- Villotte, S., Castex, D., Couallier, V., Dutour, O., Knüsel, C. J., & Henry-Gambier, D. (2010). Enthesopathies as occupational stress markers: Evidence from the upper limb. *American Journal of Physical Anthropology*, 142, 224–234. <https://doi.org/10.1002/ajpa.21217>
- Villotte, S., & Knüsel, C. J. (2013). Understanding enthesal changes: Definition and life course changes. *International Journal of Osteoarchaeology*, 23, 135–146. <https://doi.org/10.1002/oa.2289>
- Villotte, S., & Knüsel, C. J. (2014). “I sign of arms and of a man ...”: Medial epicondylitis and the sexual division of labour in prehistoric Europe. *Journal of Archaeological Science*, 43, 168–174. <https://doi.org/10.1016/j.jas.2013.12.009>
- Wilczak, C. A., Mariotti, V., Pany-Kucera, D., Villotte, S., & Henderson, C. Y. (2017). Training and interobserver reliability in qualitative scoring of skeletal samples. *Journal of Archaeological Science: Reports*, 11, 69–79. <https://doi.org/10.1016/j.jasrep.2016.11.033>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Perez-Arzak, U., Villotte, S., Arrizabalaga, A., & Tranco, G. J. (2022). Looking for the most suitable method for the study of enthesal changes: Application to upper limb's fibrocartilaginous entheses in a human medieval sample. *International Journal of Osteoarchaeology*, 1–12. <https://doi.org/10.1002/oa.3085>

Supporting information

Table S1. Frequencies (%) of asymmetry according ECs recorded by each method. Data used to create Figure 3. *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones. Only the entheses that could be analysed on both sides within the same individual were used. “Not-biased” corresponds to the percentage of the ECs that showed the same scores of bone alteration in the case of both sides. BF = Bone formation; Z1= Zone 1; Z2=Zone 2.

Method	Type of bias		HSB	HSP	HNF	HXT	HFL	RBB	UTB
Villette (2006)	Not-biased	<i>n</i>	28	23	19	19	19	26	25
		%	80	76.7	67.9	61.3	82.6	65	73.5
	Right-biased	<i>n</i>	5	5	7	5	3	8	7
		%	14.3	16.7	25	16.1	13	20	20.6
	Left-biased	<i>n</i>	2	2	2	7	1	6	2
		%	5.7	6.7	7.1	22.6	4.3	15	5.9
Santana (2011)	Not-biased	<i>n</i>	28	19	14	23	18	23	23
		%	80	63.3	50	74.2	78.3	57.5	67.6
	Right-biased	<i>n</i>	4	8	10	6	3	9	8
		%	11.4	26.7	35.7	19.4	13	22.5	23.5
	Left-biased	<i>n</i>	3	3	4	2	2	8	3
		%	8.6	10	14.3	6.5	8.7	20	8.8
BF Z1 by Coimbra Method (Henderson et al., 2013, 2016)	Not-biased	<i>n</i>	27	29	27	19	20	30	28
		%	77.1	96.7	96.4	61.3	87	75	82.4
	Right-biased	<i>n</i>	2	0	0	10	2	3	4
		%	5.7	0	0	32.3	8.7	7.5	11.8
	Left-biased	<i>n</i>	6	1	1	2	1	7	2
		%	17.1	3.3	3.6	6.5	4.3	17.5	5.9
BF Z2 by Coimbra Method (Henderson et al., 2013, 2016)	Not-biased	<i>n</i>	22	27	25	25	21	33	33
		%	62.9	90	89.3	80.6	91.3	82.5	97.1
	Right-biased	<i>n</i>	8	1	2	4	2	4	1
		%	22.9	3.3	7.1	12.9	8.7	10	2.9
	Left-biased	<i>n</i>	5	2	1	2	0	3	0
		%	14.3	6.7	3.6	6.5	0	7.5	0

Table S2. Means of ECs' scores obtained by each method according to laterality. Data used to create Figure 2. *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones; BF = Bone Formation; Z1= Zone 1; Z2=Zone 2.

Method	Side		HSB	HSP	HNF	HXT	HFL	RBB	UTB
Villotte (2006)	Right	<i>n</i>	46	39	37	44	39	50	44
		Mean	1	0.31	0.62	0.73	0.56	0.98	0.41
	Left	<i>n</i>	38	36	30	38	34	49	43
		Mean	0.97	0.19	0.37	0.92	0.44	0.88	0.19
Santana (2011)	Right	<i>n</i>	46	39	37	44	39	50	44
		Mean	3.24	1.85	2.76	2.66	2.59	2.22	2.3
	Left	<i>n</i>	38	36	30	38	34	49	43
		Mean	3.26	1.61	2.2	2.55	2.41	2.12	1.98
BF Z1 by Coimbra Method (Henderson et al., 2013, 2016)	Right	<i>n</i>	46	39	37	44	39	50	44
		Mean	0.24	0	0.08	0.7	0.26	0.36	0.32
	Left	<i>n</i>	38	36	30	38	34	49	43
		Mean	0.42	0.06	0.07	0.58	0.12	0.43	0.16
BF Z2 by Coimbra Method (Henderson et al., 2013, 2016)	Right	<i>n</i>	46	39	37	44	39	50	44
		Mean	0.46	0.08	0.08	0.2	0.15	0.28	0.05
	Left	<i>n</i>	38	36	30	38	34	49	43
		Mean	0.37	0.06	0.1	0.18	0.03	0.14	0

Table S3. Means of ECs' scores obtained by each method according to sex. Data used to create Figure 4. *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones; BF = Bone formation; Z1= Zone 1; Z2=Zone 2.

Method	Sex		HSB	HSP	HNF	HXT	HFL	RBB	UTB
Villotte (2006)	Female	<i>n</i>	36	34	33	29	26	39	35
		Mean	0.83	0.26	0.48	0.76	0.46	0.82	0.23
	Male	<i>n</i>	40	35	30	43	35	51	40
		Mean	1.12	0.29	0.57	0.93	0.66	1.06	0.43
Santana (2011)	Female	<i>n</i>	36	34	33	29	26	39	35
		Mean	3.03	1.59	2.33	2.45	2.42	1.97	2.11
	Male	<i>n</i>	40	35	30	43	35	51	40
		Mean	3.45	1.86	2.73	2.98	2.69	2.27	2.25
BF Z1 by Coimbra Method (Henderson et al., 2013, 2016)	Female	<i>n</i>	36	34	33	29	26	39	35
		Mean	0.31	0	0.12	0.59	0.19	0.36	0.2
	Male	<i>n</i>	40	35	30	43	35	51	40
		Mean	0.3	0.06	0.03	0.77	0.26	0.43	0.35
BF Z2 by Coimbra Method (Henderson et al., 2013, 2016)	Female	<i>n</i>	36	34	33	29	26	39	35
		Mean	0.33	0.03	0.06	0.21	0.08	0.1	0.03
	Male	<i>n</i>	40	35	30	43	35	51	03
		Mean	0.47	0.11	0.13	0.19	0.14	0.24	0.03

Table S4. Means of ECs' scores obtained by each method according to sex and laterality. Data used to create Figure 5. *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones; BF = Bone formation; Z1= Zone 1; Z2=Zone 2.

Method	Sex & side		HSB	HSP	HNF	HXT	HFL	RBB	UTB
Villotte (2006)	Female right	<i>n</i>	19	17	17	15	14	21	19
		Mean	0.84	0.29	0.59	0.67	0.5	0.95	0.26
	Female left	<i>n</i>	17	17	16	14	12	18	16
		Mean	0.82	0.24	0.38	0.86	0.42	0.67	0.19
	Male right	<i>n</i>	22	18	17	23	19	24	19
		Mean	1.18	0.39	0.71	0.87	0.74	1.08	0.63
Male left	<i>n</i>	18	17	13	20	18	27	21	
	Mean	1.06	0.18	0.38	1	0.56	1.04	0.24	
Santana (2011)	Female right	<i>n</i>	19	17	17	15	14	21	19
		Mean	2.95	1.76	2.59	2.53	2.5	2.14	2.21
	Female left	<i>n</i>	17	17	16	14	12	18	16
		Mean	3.12	1.41	2.06	2.36	2.33	1.78	2
	Male right	<i>n</i>	22	18	17	23	19	24	19
		Mean	3.5	1.94	3	3	2.89	2.25	2.47
Male left	<i>n</i>	18	17	13	20	16	27	21	
	Mean	3.39	1.76	2.38	2.95	2.44	2.3	2.05	
BF Z1 by Coimbra Method (Henderson et al., 2013, 2016)	Female right	<i>n</i>	19	17	17	15	14	21	19
		Mean	0.32	0	0.12	0.6	0.21	0.38	0.26
	Female left	<i>n</i>	17	17	16	14	12	18	16
		Mean	0.29	0	0.13	0.57	0.17	0.33	0.13
	Male right	<i>n</i>	22	18	17	23	19	24	19
		Mean	0.18	0	0.06	0.87	0.37	0.33	0.47
Male left	<i>n</i>	18	17	13	20	16	27	21	
	Mean	0.44	0.12	0	0.65	0.13	0.52	0.24	
BF Z2 by Coimbra Method (Henderson et al., 2013, 2016)	Female right	<i>n</i>	19	17	17	15	14	21	19
		Mean	0.26	0	0.06	0.2	0.07	1	0.05
	Female left	<i>n</i>	17	17	16	14	12	18	16
		Mean	0.41	0.06	0.06	0.21	0.08	0.11	0
	Male right	<i>n</i>	22	18	17	23	19	24	19
		Mean	0.64	0.17	0.12	0.17	0.26	0.29	0.05
Male left	<i>n</i>	18	17	13	20	16	27	21	
	Mean	0.28	0.06	0.15	0.2	0	0.19	0	

Data S1. Kappa coefficients at all cutpoints for all entheses combined according laterality, comparing in pairs the results obtained by the three methods applied in this study. In brackets, *p*-values; in **bold**, coefficients indicating moderate (0.41-0.6), substantial (0.61-0.8) and almost-perfect (0.81-1) agreement, following Landis and Koch (1977). *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones; BF = Bone formation; Z1= Zone 1; Z2=Zone 2.

RIGHT <i>n</i> = 299		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.04 (<i>p</i> =0.034)	0.337 (<i>p</i> <0.001)	0.742 (<i>p</i> <0.001)	0.608 (<i>p</i> <0.001)
	≤ 1 versus 2	0.008 (<i>p</i> =0.286)	0.091 (<i>p</i> <0.001)	0.328 (<i>p</i> <0.001)	0.405 (<i>p</i> <0.001)

RIGHT <i>n</i> = 299		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.013 (<i>p</i> =0.167)	0.138 (<i>p</i> <0.001)	0.428 (<i>p</i> <0.001)	0.506 (<i>p</i> <0.001)
	≤ 1 versus 2	0.004 (<i>p</i> =0.429)	0.051 (<i>p</i> =0.005)	0.204 (<i>p</i> <0.001)	0.292 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	0.008 (<i>p</i> =0.26)	0.101 (<i>p</i> <0.001)	0.36 (<i>p</i> <0.001)	0.377 (<i>p</i> <0.001)
	≤ 1 versus 2	0.001 (<i>p</i> =0.657)	0.016 (<i>p</i> =0.116)	0.069 (<i>p</i> =0.001)	0.083 (<i>p</i> =0.001)

RIGHT <i>n</i> = 299		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.372 (<i>p</i> <0.001)	0.139 (<i>p</i> <0.001)	0.264 (<i>p</i> <0.001)	0.046 (<i>p</i> =0.008)
	≤ 1 versus 2	0.562 (<i>p</i> <0.001)	0.504 (<i>p</i> <0.001)	0.566 (<i>p</i> <0.001)	0.252 (<i>p</i> <0.001)

LEFT <i>n</i> = 268		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.102 (<i>p</i> <0.001)	0.294 (<i>p</i> <0.001)	0.74 (<i>p</i> <0.001)	0.605 (<i>p</i> <0.001)
	≤ 1 versus 2	0.018 (<i>p</i> =0.122)	0.092 (<i>p</i> <0.001)	0.34 (<i>p</i> <0.001)	0.403 (<i>p</i> <0.001)

LEFT <i>n</i> = 268		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.029 (<i>p</i> =0.046)	0.156 (<i>p</i> <0.001)	0.496 (<i>p</i> <0.001)	0.535 (<i>p</i> <0.001)
	≤ 1 versus 2	0.012 (<i>p</i> =0.206)	0.07 (<i>p</i> =0.002)	0.284 (<i>p</i> <0.001)	0.395 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	0.016 (<i>p</i> =0.143)	0.093 (<i>p</i> <0.001)	0.345 (<i>p</i> <0.001)	0.451 (<i>p</i> <0.001)
	≤ 1 versus 2	0.002 (<i>p</i> =0.612)	0.012 (<i>p</i> =0.209)	0.052 (<i>p</i> =0.007)	0.076 (<i>p</i> =0.001)

LEFT <i>n</i> = 268		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.434 (<i>p</i> <0.001)	0.199 (<i>p</i> <0.001)	0.259 (<i>p</i> <0.001)	0.035 (<i>p</i> =0.029)
	≤ 1 versus 2	0.636 (<i>p</i> <0.001)	0.483 (<i>p</i> <0.001)	0.514 (<i>p</i> <0.001)	0.195 (<i>p</i> <0.001)

Data S2. Kappa coefficients at all cutpoints for all entheses combined according sex, comparing in pairs the results obtained by the three methods applied in this study. In brackets, *p*-values; in **bold**, coefficients indicating moderate (0.41-0.6), substantial (0.61-0.8) and almost-perfect (0.81-1) agreement, following Landis and Koch (1977). *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones; BF = Bone formation; Z1= Zone 1; Z2=Zone 2.

FEMALE <i>n</i> = 232		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.094 (<i>p</i> =0.002)	0.323 (<i>p</i> <0.001)	0.804 (<i>p</i> <0.001)	0.682 (<i>p</i> <0.001)
	≤ 1 versus 2	0.019 (<i>p</i> =0.132)	0.116 (<i>p</i> <0.001)	0.355 (<i>p</i> <0.001)	0.413 (<i>p</i> <0.001)

FEMALE <i>n</i> = 232		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.03 (<i>p</i> =0.058)	0.164 (<i>p</i> <0.001)	0.495 (<i>p</i> <0.001)	0.535 (<i>p</i> <0.001)
	≤ 1 versus 2	0.012 (<i>p</i> =0.229)	0.076 (<i>p</i> =0.003)	0.258 (<i>p</i> <0.001)	0.343 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	0.017 (<i>p</i> =0.159)	0.103 (<i>p</i> <0.001)	0.338 (<i>p</i> <0.001)	0.443 (<i>p</i> <0.001)
	≤ 1 versus 2	0.003 (<i>p</i> =0.583)	0.016 (<i>p</i> =0.167)	0.06 (<i>p</i> =0.007)	0.083 (<i>p</i> =0.002)

FEMALE <i>n</i> = 232		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.42 (<i>p</i> <0.001)	0.194 (<i>p</i> <0.001)	0.256 (<i>p</i> <0.001)	0.044 (<i>p</i> =0.023)
	≤ 1 versus 2	0.705 (<i>p</i> <0.001)	0.583 (<i>p</i> <0.001)	0.67 (<i>p</i> <0.001)	0.235 (<i>p</i> <0.001)

MALE <i>n</i> = 274		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.076 (<i>p</i> <0.001)	0.37 (<i>p</i> <0.001)	0.682 (<i>p</i> <0.001)	0.538 (<i>p</i> <0.001)
	≤ 1 versus 2	0.009 (<i>p</i> =0.261)	0.088 (<i>p</i> <0.001)	0.292 (<i>p</i> <0.001)	0.371 (<i>p</i> <0.001)

MALE <i>n</i> = 274		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.017 (p=0.127)	0.157 (p<0.001)	0.418 (p<0.001)	0.496 (p<0.001)
	≤ 1 versus 2	0.005 (p=0.385)	0.054 (p=0.006)	0.205 (p<0.001)	0.317 (p<0.001)
Coimbra BF Z2	0 versus ≥1	0.011 (p=0.223)	0.103 (p<0.001)	0.339 (p<0.001)	0.363 (p<0.001)
	≤ 1 versus 2	0.001 (p=0.688)	0.012 (p=0.204)	0.048 (p=0.01)	0.057 (p=0.014)

MALE <i>n</i> = 274		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.352 (p<0.001)	0.13 (p<0.001)	0.241 (p<0.001)	0.029 (p=0.044)
	≤ 1 versus 2	0.477 (p<0.001)	0.409 (p<0.001)	0.459 (p<0.001)	0.181 (p<0.001)

Data S3. Kappa coefficients at all cutpoints for all entheses combined according sex and laterality, comparing in pairs the results obtained by the three methods applied in this study. In brackets, *p*-values; in **bold**, coefficients indicating moderate (0.41-0.6), substantial (0.61-0.8) and almost-perfect (0.81-1) agreement, following Landis and Koch (1977). *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones; BF = Bone formation; Z1= Zone 1; Z2=Zone 2.

FEMALE RIGHT <i>n</i> = 122		Santana			
		0 versus ≥ 1	≤ 1 versus $2 \geq$	≤ 2 versus $3 \geq$	≤ 3 versus 4
Villotte	0 versus ≥ 1	0.045 (<i>p</i> =0.195)	0.337 (<i>p</i> <0.001)	0.785 (<i>p</i> <0.001)	0.668 (<i>p</i> <0.001)
	≤ 1 versus 2	0.011 (<i>p</i> =0.411)	0.103 (<i>p</i> =0.01)	0.294 (<i>p</i> <0.001)	0.342 (<i>p</i> <0.001)

FEMALE RIGHT <i>n</i> = 122		Santana			
		0 versus ≥ 1	≤ 1 versus $2 \geq$	≤ 2 versus $3 \geq$	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥ 1	0.021 (<i>p</i> =0.258)	0.16 (<i>p</i> =0.002)	0.473 (<i>p</i> <0.001)	0.517 (<i>p</i> <0.001)
	≤ 1 versus 2	0.007 (<i>p</i> =0.519)	0.065 (<i>p</i> =0.044)	0.219 (<i>p</i> <0.001)	0.29 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥ 1	0.008 (<i>p</i> =0.472)	0.08 (<i>p</i> =0.024)	0.265 (<i>p</i> <0.001)	0.348 (<i>p</i> <0.001)
	≤ 1 versus 2	0.001 (<i>p</i> =0.768)	0.014 (<i>p</i> =0.356)	0.05 (<i>p</i> =0.077)	0.069 (<i>p</i> =0.038)

FEMALE RIGHT <i>n</i> = 122		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥ 1	≤ 1 versus 2	0 versus ≥ 1	≤ 1 versus 2
Villotte	0 versus ≥ 1	0.415 (<i>p</i> <0.001)	0.157 (<i>p</i> =0.001)	0.191 (<i>p</i> <0.001)	0.035 (<i>p</i> =0.141)
	≤ 1 versus 2	0.692 (<i>p</i> <0.001)	0.666 (<i>p</i> <0.001)	0.689 (<i>p</i> <0.001)	0.228 (<i>p</i> <0.001)

FEMALE LEFT <i>n</i> = 110		Santana			
		0 versus ≥ 1	≤ 1 versus $2 \geq$	≤ 2 versus $3 \geq$	≤ 3 versus 4
Villotte	0 versus ≥ 1	0.133 (<i>p</i> =0.005)	0.302 (<i>p</i> <0.001)	0.823 (<i>p</i> <0.001)	0.695 (<i>p</i> <0.001)
	≤ 1 versus 2	0.029 (<i>p</i> =0.201)	0.132 (<i>p</i> =0.005)	0.432 (<i>p</i> <0.001)	0.506 (<i>p</i> <0.001)

FEMALE LEFT <i>n</i> = 110		Santana			
		0 versus ≥ 1	≤ 1 versus $2 \geq$	≤ 2 versus $3 \geq$	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥ 1	0.037 (<i>p</i> =0.149)	0.165 (<i>p</i> =0.002)	0.518 (<i>p</i> <0.001)	0.554 (<i>p</i> <0.001)
	≤ 1 versus 2	0.02 (<i>p</i> =0.297)	0.09 (<i>p</i> =0.023)	0.31 (<i>p</i> <0.001)	0.414 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥ 1	0.029 (<i>p</i> =0.201)	0.132 (<i>p</i> =0.005)	0.432 (<i>p</i> <0.001)	0.564 (<i>p</i> <0.001)
	≤ 1 versus 2	0.004 (<i>p</i> =0.634)	0.019 (<i>p</i> =0.3)	0.073 (<i>p</i> =0.041)	0.103 (<i>p</i> =0.015)

FEMALE LEFT <i>n</i> = 110		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥ 1	≤ 1 versus 2	0 versus ≥ 1	≤ 1 versus 2
Villotte	0 versus ≥ 1	0.419 (<i>p</i> <0.001)	0.244 (<i>p</i> <0.001)	0.345 (<i>p</i> <0.001)	0.056 (<i>p</i> =0.075)
	≤ 1 versus 2	0.722 (<i>p</i> <0.001)	0.497 (<i>p</i> <0.001)	0.651 (<i>p</i> <0.001)	0.243 (<i>p</i> <0.001)

MALE RIGHT <i>n</i> = 142		Santana			
		0 versus ≥ 1	≤ 1 versus $2 \geq$	≤ 2 versus $3 \geq$	≤ 3 versus 4
Villotte	0 versus ≥ 1	0.053 (<i>p</i> =0.05)	0.366 (<i>p</i> <0.001)	0.656 (<i>p</i> <0.001)	0.511 (<i>p</i> <0.001)
	≤ 1 versus 2	0.006 (<i>p</i> =0.521)	0.083 (<i>p</i> =0.013)	0.308 (<i>p</i> <0.001)	0.411 (<i>p</i> <0.001)

MALE RIGHT <i>n</i> = 142		Santana			
		0 versus ≥ 1	≤ 1 versus $2 \geq$	≤ 2 versus $3 \geq$	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥ 1	0.01 (<i>p</i> =0.407)	0.135 (<i>p</i> =0.001)	0.356 (<i>p</i> <0.001)	0.465 (<i>p</i> <0.001)
	≤ 1 versus 2	0.003 (<i>p</i> =0.651)	0.042 (<i>p</i> =0.08)	0.165 (<i>p</i> <0.001)	0.268 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥ 1	0.008 (<i>p</i> =0.452)	0.113 (<i>p</i> =0.004)	0.372 (<i>p</i> <0.001)	0.345 (<i>p</i> <0.001)
	≤ 1 versus 2	0.001 (<i>p</i> =0.786)	0.015 (<i>p</i> =0.293)	0.063 (<i>p</i> =0.031)	0.071 (<i>p</i> =0.059)

MALE RIGHT <i>n</i> = 142		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥ 1	≤ 1 versus 2	0 versus ≥ 1	≤ 1 versus 2
Villotte	0 versus ≥ 1	0.304 (<i>p</i> <0.001)	0.101 (<i>p</i> =0.006)	0.257 (<i>p</i> <0.001)	0.038 (<i>p</i> =0.098)
	≤ 1 versus 2	0.415 (<i>p</i> <0.001)	0.356 (<i>p</i> <0.001)	0.483 (<i>p</i> <0.001)	0.231 (<i>p</i> <0.001)

MALE LEFT <i>n</i> = 132		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.09 (<i>p</i> =0.013)	0.364 (<i>p</i> <0.001)	0.701 (<i>p</i> <0.001)	0.56 (<i>p</i> <0.001)
	≤ 1 versus 2	0.012 (<i>p</i> =0.381)	0.09 (<i>p</i> =0.012)	0.266 (<i>p</i> <0.001)	0.315 (<i>p</i> <0.001)

MALE LEFT <i>n</i> = 132		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.024 (<i>p</i> =0.207)	0.179 (<i>p</i> <0.001)	0.489 (<i>p</i> <0.001)	0.531 (<i>p</i> <0.001)
	≤ 1 versus 2	0.009 (<i>p</i> =0.451)	0.068 (<i>p</i> =0.032)	0.258 (<i>p</i> <0.001)	0.379 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	0.011 (<i>p</i> =0.397)	0.085 (<i>p</i> =0.016)	0.281 (<i>p</i> <0.001)	0.373 (<i>p</i> <0.001)
	≤ 1 versus 2	0.001 (<i>p</i> =0.842)	0.005 (<i>p</i> =0.57)	0.021 (<i>p</i> =0.242)	0.032 (<i>p</i> =0.142)

MALE LEFT <i>n</i> = 132		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.407 (<i>p</i> <0.001)	0.167 (<i>p</i> =0.001)	0.206 (<i>p</i> <0.001)	0.013 (<i>p</i> =0.359)
	≤ 1 versus 2	0.55 (<i>p</i> <0.001)	0.475 (<i>p</i> <0.001)	0.411 (<i>p</i> <0.001)	0.098 (<i>p</i> =0.009)

Data S4. Kappa coefficients at all cutpoints for each enthesis, comparing in pairs the results obtained by the three methods applied in this study. In brackets, *p*-values; in **bold**, coefficients indicating moderate (0.41-0.6), substantial (0.61-0.8) and almost-perfect (0.81-1) agreement, following Landis and Koch (1977). * = constant trends; *n* = the total number of analysed entheses, without taking into account the non-recordable (NR) ones; BF = Bone formation; Z1= Zone 1; Z2=Zone 2.

HSB <i>n</i> = 84		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	*	0.356 (<i>p</i> <0.001)	0.912 (<i>p</i> <0.001)	0.731 (<i>p</i> <0.001)
	≤ 1 versus 2	*	0.05 (<i>p</i> =0.142)	0.248 (<i>p</i> <0.001)	0.298 (<i>p</i> <0.001)

* = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for HSB

HSB <i>n</i> = 84		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	*	0.41 (<i>p</i> =0.184)	0.169 (<i>p</i> =0.011)	0.162 (<i>p</i> =0.035)
	≤ 1 versus 2	*	0.18 (<i>p</i> =0.379)	0.097 (<i>p</i> =0.039)	0.138 (<i>p</i> =0.013)
Coimbra BF Z2	0 versus ≥1	*	0.078 (<i>p</i> =0.065)	0.365 (<i>p</i> <0.001)	0.446 (<i>p</i> <0.001)
	≤ 1 versus 2	*	0.12 (<i>p</i> =0.481)	0.063 (<i>p</i> =0.098)	0.09 (<i>p</i> =0.046)

* = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for HSB

HSB <i>n</i> = 84		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.176 (<i>p</i> =0.004)	0.082 (<i>p</i> =0.059)	0.315 (<i>p</i> <0.001)	0.053 (<i>p</i> =0.13)
	≤ 1 versus 2	0.7 (<i>p</i> <0.001)	0.529 (<i>p</i> <0.001)	0.549 (<i>p</i> <0.001)	0.375 (<i>p</i> <0.001)

HSP <i>n</i> = 75		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.11 (<i>p</i> =0.036)	0.463 (<i>p</i> <0.001)	0.888 (<i>p</i> <0.001)	0.764 (<i>p</i> <0.001)
	≤ 1 versus 2	*	*	*	*

* = ≤ 1 versus 2 according Villotte (2006) is constant, there is no score 2 for HSP

HSP <i>n</i> = 75		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.005 (p=0.676)	0.027 (p=0.308)	0.095 (p=0.053)	0.121 (p=0.028)
	≤ 1 versus 2	0.005 (p=0.676)	0.027 (p=0.308)	0.095 (p=0.053)	0.121 (p=0.028)
Coimbra BF Z2	0 versus ≥1	0.19 (p=0.394)	0.109 (p=0.037)	0.344 (p<0.001)	0.295 (p=0.002)
	≤ 1 versus 2	0.005 (p=0.676)	0.027 (p=0.308)	0.095 (p=0.053)	(-)0.025 (p=0.645)

HSP <i>n</i> = 75		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.077 (p=0.084)	0.077 (p=0.084)	0.285 (p<0.001)	0.077 (p=0.084)
	≤ 1 versus 2	*	*	*	*

* = ≤ 1 versus 2 according Villotte (2006) is constant, there is no score 2 for HSP

HNF <i>n</i> = 67		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.027 (p=0.335)	0.329 (p=0.001)	0.759 (p<0.001)	0.728 (p<0.001)
	≤ 1 versus 2	0.001 (p=0.86)	0.019 (p=0.421)	0.082 (p=0.09)	0.111 (p=0.047)

HNF <i>n</i> = 67		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.002 (p=0.8)	0.039 (p=0.248)	0.162 (p=0.015)	0.217 (p=0.004)
	≤ 1 versus 2	0.000 (p=0.901)	0.009 (p=0.573)	0.041 (p=0.234)	0.056 (p=0.163)
Coimbra BF Z2	0 versus ≥1	0.003 (p=0.752)	0.06 (p=0.15)	0.241 (p=0.002)	0.236 (p=0.008)
	≤ 1 versus 2	*	*	*	*

* = ≤ 1 versus 2 according Coimbra Method (Henderson et al., 2013, 2016) for BF Z2 is constant, there is no score 2 for HNF

HNF <i>n</i> = 67		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.13 (p=0.031)	0.033 (p=0.292)	0.194 (p=0.007)	*
	≤ 1 versus 2	0.653 (p<0.001)	(-)0.02 (p=0.86)	0.215 (p=0.039)	*

* = ≤ 1 versus 2 according Coimbra Method (Henderson et al., 2013, 2016) for BF Z2 is constant, there is no score 2 for HNF

HXT <i>n</i> = 82		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	*	0.403 (<i>p</i> <0.001)	0.783 (<i>p</i> <0.001)	0.736 (<i>p</i> <0.001)
	≤ 1 versus 2	*	0.195 (<i>p</i> =0.006)	0.493 (<i>p</i> <0.001)	0.532 (<i>p</i> <0.001)

* = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for HXT

HXT <i>n</i> = 82		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	*	0.391 (<i>p</i> <0.001)	0.827 (<i>p</i> <0.001)	0.875 (<i>p</i> <0.001)
	≤ 1 versus 2	*	0.236 (<i>p</i> <0.001)	0.544 (<i>p</i> <0.001)	0.584 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	*	0.157 (<i>p</i> =0.008)	0.385 (<i>p</i> <0.001)	0.417 (<i>p</i> <0.001)
	≤ 1 versus 2	*	0.02 (<i>p</i> =0.357)	0.056 (<i>p</i> =0.123)	0.062 (<i>p</i> =0.106)

* = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for HXT

HXT <i>n</i> = 82		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.668 (<i>p</i> <0.001)	0.387 (<i>p</i> <0.001)	0.266 (<i>p</i> <0.001)	0.037 (<i>p</i> =0.217)
	≤ 1 versus 2	0.648 (<i>p</i> <0.001)	0.669 (<i>p</i> <0.001)	0.558 (<i>p</i> <0.001)	0.144 (<i>p</i> =0.012)

HFL <i>n</i> = 73		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	*	0.082 (<i>p</i> =0.077)	0.688 (<i>p</i> <0.001)	0.393 (<i>p</i> <0.001)
	≤ 1 versus 2	*	0.01 (<i>p</i> =0.538)	0.252 (<i>p</i> =0.001)	0.585 (<i>p</i> <0.001)

* = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for HFL

HFL <i>n</i> = 73		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	* ₁	0.018 (<i>p</i> =0.412)	0.407 (<i>p</i> <0.001)	0.846 (<i>p</i> <0.001)
	≤ 1 versus 2	* ₁	0.007 (<i>p</i> =0.62)	0.17 (<i>p</i> =0.009)	0.422 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	* ₁	0.012 (<i>p</i> =0.503)	0.291 (<i>p</i> <0.001)	0.429 (<i>p</i> <0.001)
	≤ 1 versus 2	* _{1,2}	* ₂	* ₂	* ₂

*₁ = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for HFL
*₂ = ≤ 1 versus 2 according Coimbra Method (Henderson et al., 2013, 2016) for BF Z2 is constant, there is no score 2 for HFL

HFL <i>n</i> = 73		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.354 (<i>p</i> <0.001)	0.146 (<i>p</i> =0.017)	0.251 (<i>p</i> =0.001)	*
	≤ 1 versus 2	0.582 (<i>p</i> <0.001)	0.358 (<i>p</i> =0.002)	0.747 (<i>p</i> <0.001)	*

* = ≤ 1 versus 2 according Coimbra Method (Henderson et al., 2013, 2016) for BF Z2 is constant, there is no score 2 for HFL

RBB <i>n</i> = 99		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	0.388 (<i>p</i> <0.001)	0.349 (<i>p</i> <0.001)	0.412 (<i>p</i> <0.001)	0.293 (<i>p</i> <0.001)
	≤ 1 versus 2	0.073 (<i>p</i> =0.053)	0.306 (<i>p</i> <0.001)	0.533 (<i>p</i> <0.001)	0.447 (<i>p</i> <0.001)

RBB <i>n</i> = 99		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	0.098 (<i>p</i> =0.024)	0.36 (<i>p</i> <0.001)	0.594 (<i>p</i> <0.001)	0.581 (<i>p</i> <0.001)
	≤ 1 versus 2	0.028 (<i>p</i> =0.238)	0.126 (<i>p</i> =0.01)	0.317 (<i>p</i> <0.001)	0.443 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	0.37 (<i>p</i> =0.171)	0.165 (<i>p</i> =0.003)	0.305 (<i>p</i> <0.001)	0.316 (<i>p</i> <0.001)
	≤ 1 versus 2	0.008 (<i>p</i> =0.534)	0.037 (<i>p</i> =0.174)	0.1 (<i>p</i> =0.023)	0.147 (<i>p</i> =0.005)

RBB <i>n</i> = 99		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.305 (<i>p</i> <0.001)	0.093 (<i>p</i> =0.028)	0.123 (<i>p</i> =0.011)	0.027 (<i>p</i> =0.246)
	≤ 1 versus 2	0.585 (<i>p</i> <0.001)	0.259 (<i>p</i> =0.004)	0.533 (<i>p</i> <0.001)	0.187 (<i>p</i> =0.001)

UTB <i>n</i> = 87		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Villotte	0 versus ≥1	*	0.192 (<i>p</i> =0.002)	0.791 (<i>p</i> <0.001)	0.508 (<i>p</i> <0.001)
	≤ 1 versus 2	*	0.013 (<i>p</i> =0.449)	0.138 (<i>p</i> =0.011)	0.307 (<i>p</i> <0.001)

* = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for UTB

UTB <i>n</i> = 87		Santana			
		0 versus ≥1	≤ 1 versus 2 ≥	≤ 2 versus 3 ≥	≤ 3 versus 4
Coimbra BF Z1	0 versus ≥1	* ₁	0.145 (<i>p</i> =0.009)	0.676 (<i>p</i> <0.001)	0.635 (<i>p</i> <0.001)
	≤ 1 versus 2	* ₁	0.013 (<i>p</i> =0.449)	0.138 (<i>p</i> =0.011)	0.307 (<i>p</i> <0.001)
Coimbra BF Z2	0 versus ≥1	* ₁	0.013 (<i>p</i> =0.449)	0.138 (<i>p</i> =0.011)	0.133 (<i>p</i> =0.084)
	≤ 1 versus 2	* _{1,2}	* ₂	* ₂	* ₂
* ₁ = 0 versus ≥1 according Santana (2011) is constant, there is no score 0 for UTB					
* ₂ = ≤ 1 versus 2 according Coimbra Method (Henderson et al., 2013, 2016) for BF Z2 is constant, there is no score 2 for UTB					

UTB <i>n</i> = 87		Coimbra BF Z1		Coimbra BF Z2	
		0 versus ≥1	≤ 1 versus 2	0 versus ≥1	≤ 1 versus 2
Villotte	0 versus ≥1	0.846 (<i>p</i> <0.001)	0.116 (<i>p</i> =0.02)	0.116 (<i>p</i> =0.02)	*
	≤ 1 versus 2	0.155 (<i>p</i> =0.007)	0.488 (<i>p</i> <0.001)	(-)0.024 (<i>p</i> =0.826)	*
* = ≤ 1 versus 2 according Coimbra Method (Henderson et al., 2013, 2016) for BF Z2 is constant, there is no score 2 for UTB					