Citation for final published version:

Alonso-Llamazares, Carmen, Lopez, Belen and Pardinas, Antonio ORCID: https://orcid.org/0000-0001-6845-7590 2022. Sex differences in the distribution of entheseal changes Meta-analysis of published evidence and its use in Bayesian paleopathological modelling. American Journal of Physical Anthropology 177 (2) , pp. 249-265. 10.1002/ajpa.24425 file

Publishers page: https://doi.org/10.1002/ajpa.24425 <https://doi.org/10.1002/ajpa.24425>

Sex differences in the distribution of entheseal changes: Meta-analysis of published evidence and its use in Bayesian paleopathological modelling

Abbreviated title: Sex differences in the distribution of entheseal changes.

Alonso-Llamazares, Carmen[1]; Lopez, Belen[1]; Pardiñas, Antonio F[2].

[1] Department of Biology of Organisms and Systems, University of Oviedo, Asturias 33006, Spain.

[2] MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, United Kingdom

10 text pages, plus 6 bibliography pages, 4 figures, 4 tables.

Keywords:

Entheseal changes, multivariate generalized linear mixed models, Bayesian inference, Meta-analysis, Spain.

Corresponding author:

Antonio F. Pardiñas
Telephone number: +44 29 2068 8407
Address: MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, United Kingdom.
E-mail address: pardinasa@cardiff.ac.uk

Belen Lopez
Telephone number: +34 985104768
Address: Department of Biology of Organisms and Systems, University of Oviedo. C/ Catedrático Valentín Andrés Álvarez s/n, Asturias 33006, Spain.
E-mail address: lopezbelen@uniovi.es

ABSTRACT

Objectives: We studied the sex differences in the distribution of entheseal changes (EC) in an archaeological population through a Bayesian approach that allows incorporating existing knowledge while controlling for confounder factors that may affect EC development.

Materials and Methods: We performed a meta-analysis of past research on sex differences in EC frequencies from archaeological populations. Also, EC were assessed for fibrocartilaginous entheses following the 'New Coimbra Method' in a Spanish population that dates from the 15th to the 18th century. Data were analyzed with multivariate generalized linear mixed models (MGLMM).

Results: Meta-analysis showed a consistent but small effect of males usually manifesting higher EC frequencies. Similarly, our MGLMM analysis showed that bone formation and erosion is unequally distributed in the archeological population we studied, with bone formation more present in male lower limbs and erosion more frequent in male upper limbs.

Discussion: Bayesian inference makes it possible to assess more complex models than traditional frequentist methods, and can be informed by meta-analysis to reflect the current state of knowledge on any given topic. MGLMM are an appropriate technique for the study of EC as they can accommodate several response variables in a single model, controlling for well-known confounders of EC formation to establish sex differences that could be attributed to daily behavior.

Inferring lifestyle is one of the most important research interests of studying ancient populations, particularly how it relates to differences in habitual activity in relation to sex and gender. In bioarchaeology, entheseal changes (EC) are commonly analyzed, which can be linked to physical activity patterns. EC are defined as modifications occurring at the entheses, i.e., the origin and insertion attachment points of tendons, ligaments, and joint capsules (Benjamin et al., 2002; Claudia Mercedes Rojas-Sepúlveda & Dutour, 2014). Despite several terminologies having been proposed to designate and categorize these modifications, the term 'entheseal changes' does not necessarily imply pathological conditions, nor assume any specific cause or nature for the changes (Jurmain & Villotte, 2010).

EC have been used as indicators of the habitual activity, which allows undertaking certain inferences about past populations (al-Oumaoui, Jiménez-Brobeil, & Du Souich, 2004; Giannotti, 2020; Lieverse, Bazaliiskii, Goriunova, & Weber, 2013; Mazza, 2020; Steen & Lane, 1998). The biomechanical principle behind the use of EC rests upon the number of small blood vessels in the periosteum increasing when an enthesis is repeatedly under stress, stimulating bone remodeling and osseous hypertrophy to change the enthesis morphology, accommodating a larger muscle mass (He & de Almeida Prado, 2020; Nikita, Xanthopoulou, Bertsatos, Chovalopoulou, & Hafez, 2019; Claudia Mercedes Rojas-Sepúlveda & Dutour, 2014). As visually assessing EC is complex and traditional scoring systems can be subjective (Henderson & Alves Cardoso, 2013), recent methods have been proposed to take advantage of a better anatomical understanding of entheses to anticipate plausible observable alterations, resulting in increased reproducibility and less interobserver error (Wilczak, Mariotti, Pany-Kucera, Villotte, & Henderson, 2017). In the recent literature, studies based on EC have attempted to understand many different aspects of the daily life and cultural milieu of ancient populations, including sexual division of labour, subsistence strategies, frequency of certain activities, social stratification or even physical disabilities (Hawkey, 1998; Lieverse, Bazaliiskii, Goriunova, & Weber, 2009; Lieverse, Stock, Katzenberg, & Haverkort, 2011; Molnar, 2010; Palmer, 2012; Peterson, 2010; C. M. Rojas-Sepúlveda, Rivera-

Sandoval, & Martín-Rincón, 2011; Villotte, Castex, et al., 2010; Villotte, Churchill, Dutour, & Henry-Gambier, 2010).

All of these sources of mechanical stress can lead to differences in the frequency and distribution of EC, but societal customs and behaviors only partly explain EC development. Sex, age, trauma, preexisting conditions, body mass, nutrition, metabolism, and genetic predisposition are some of the factors that have been related to the multifactorial aetiology of EC (Nikita et al., 2019; Woo & Pak, 2013). Sex is of particular interest because of the apparent differences in body size between males and females (Weiss, Corona, & Schultz, 2012), which lead to larger muscular insertion sites in males. Some methods to score EC already take this into account (Weiss, 2015). In addition, hormones such as estrogens and androgens can influence bone deposition, being contributors to sexual dimorphism in EC development and expression (Santana-Cabrera, Velasco-Vázquez, & Rodríguez-Rodríguez, 2015). Despite these biological influences, it is important to remark that the sexual division of labor is implicated in the expression of EC in human skeletal remains (al-Oumaoui et al., 2004; Eshed, Gopher, Galili, & Hershkovitz, 2004; Lieverse et al., 2009, 2013; Stefanović & Porčić, 2013).

A limitation for archaeological studies on factors related to EC, including sex, is that those can be partial or difficult to obtain in ancient studies, and their effects are rarely independent of each other, entangling the task of establishing a link between a marker of activity and a confounding factor (Jurmain, Cardoso, Henderson, & Villotte, 2012). Some authors have attempted to control for these issues through the sampling process by using only males or a specific age range, in a procedure akin to the "matched" or "stratified" experimental designs of the clinical literature (Godde & Taylor, 2013; Niinimäki & Baiges Sotos, 2013; Nikita et al., 2019; Wilczak, 1998). However, these procedures can incur the loss of a substantial part of the individuals available for study, and thus the use of statistical approaches based on the linear regression model (including classic ANOVA designs) has been preferred. The use of many of these methods for bioarchaeological research has been reviewed and discussed elsewhere (Alonso-Llamazares, Blanco Márquez, Lopez, & Pardiñas, 2021; Cheverko & Hubbe, 2017; Nikita, Mattingly, & Lahr, 2013). Nevertheless, a prevailing drawback of the usual framework for these analyses, based on "frequentist" null hypothesis significance testing, is that each new study carried out disregards that a large body of evidence might already exist on the questions being explored (van de Schoot et al., 2014). The alternative Bayesian paradigm allows for the incorporation of existing ("prior") information in the analysis, can be used for testing null hypotheses if desired, and allows for building statistical models which can account for highly complex correlations and dependencies between observations and variables (Depaoli, 2014; Kruschke, 2013; van de Schoot & Depaoli, 2014; Zyphur & Oswald, 2015). A more in-depth introduction to Bayesian statistics and its advantages and differences in respect to traditional frequentist testing can be found in van de Schoot et al. (2021). While there is a long-standing debate about the suitability of each statistical framework (frequentist or Bayesian) to different fields of research, there is agreement that none should be applied without a thorough understanding of the data, including its potential sources of errors, and the question(s) to be addressed (Silva, 2018). In anthropology and archaeology, extensive methodological developments have been applied to the implementation of Bayesian models in forensic settings and age-at-death estimation (Konigsberg & Frankenberg, 2013; Otárola-Castillo & Torquato, 2018). However, these methods remain underutilized in other topics, including paleopathology. Specifically, their potential advantages for the analyses of EC, which often yield complex and multivariate data, have not been explored.

In the present study we analyze the distribution of EC in a population of medieval and early modern times from the north of Spain, focusing on sex differences as an indicator of social organization in the daily behavior in this community. From an experimental point of view, similar research has been carried out before in other populations and timeframes. Thus, we first conduct a systematic review of the relevant literature to extract the prior information needed to exploit this similarity in a Bayesian framework. Then, taking advantage of recent proposals to include different sources of uncertainty in the statistical analysis of skeletal markers (Alonso-Llamazares et al., 2021; Beier, Anthes, Wahl, & Harvati, 2021), we define a multivariate mixed-effect model which integrates eight different features of EC and their variation across individuals, sexes, ages and skeletal locations. In addition, our analysis illustrates a reproducible statistical analysis pipeline that could be adapted to other study designs in physical anthropology, contributing to enhancing the methodological diversity within our field (Martin, 2019).

MATERIAL AND METHODS

Anthropological sample and historical context

The population used in this study comes from the church of San Nicolás de Bari, a medieval and early modern necropolis from Burgos (Spain). This church is located behind the Cathedral of Burgos and used to be part of the urban route of the Way of St. James (López Sobrino, 2000). The church of San Nicolás de Bari, as it is today, was built in 1408, replacing another Romanesque temple that already appears in the texts of the existing churches in 1163 (Florez, 1771-72).

The Middle Ages was a period of intense urbanization both in Spain and in the rest of Europe. The urban districts of Burgos arose as a result of the expansion of the Christian kingdoms in the Iberian Peninsula during the Reconquista War (8th-15th centuries). The Cathedral of Burgos underwent construction in 1221 and became the activity centre of the city, undergoing reforms and extensions for several centuries. This was one of the wealthiest cities in the kingdom, with a profitable market and extensive production of artisan goods, and a high number of members of the nobility, clergy, and urban elites (Goicolea Julián, 2019; Sebastián Moreno, 2017; Sebastián Moreno & Guerrero Navarrete, 2018).

The skeletal remains studied here were recovered in three different excavation phases, between 2007 and 2008, and correspond to the period between the 15th and the beginning of the 18th century. A total of 60 adult individuals were analyzed, of whom 40 were female and 20 were male.

Osteological methodology

Sex and age-at-death estimation

Sex was established primarily through the morphological features of the skull and pelvis (Buikstra & Ubelaker, 1994). If these skeletal elements were not well-preserved, we used discriminant functions generated from Spanish populations from the humerus, femur, and tibia (López-Bueis, Robledo, Roselló, & Trancho, 1996; Trancho, Robledo, López-Bueis, & Sánchez, 1997; Trancho, Robledo, & Sánchez, 2012). Since this research focuses on sex differences, only individuals with a reliable sex estimation were included, though we acknowledge this estimation might not be completely accurate for every individual. For age-at-death, the pubic symphysis was used when possible (Buikstra & Ubelaker, 1994; Meindl, Lovejoy, Mensforth, & Walker, 1985; Todd, 1920, 1921). When this was not possible, we used other methods such as the morphology of the auricular surface, the closure of cranial sutures, or the ossification degree of the thyroid cartilage (Buikstra & Ubelaker, 1994; Krenzer, 2006; Lovejoy, Meindl, Pryzbeck, & Mensforth, 1985; Meindl & Lovejoy, 1985). The individuals were not stratified into age groups. Instead, they were given a single value based on the results of the age estimation method(s) used on them, either an average point estimate or the central value of an age range. These values were paired with a standard deviation for the age-at-death estimate, again obtained from the specific method(s) used on each individual (**Supplementary Table 1**). Non-adult individuals, as well as those with pathological features that could affect the EC analysis were excluded from the study.

Methodology for the analysis of entheseal changes

There are two types of entheses that are defined according to the nature of the tissue connecting muscle and bone: fibrous and fibrocartilaginous (Benjamin et al., 2002; Villotte & Knüsel, 2013). In fibrous entheses, soft tissue can attach directly into the bone or through a mediating layer of periosteum, being known as bony fibrous entheses and periosteal fibrous entheses, respectively. These can be found on the spine and on long bone diaphysis, where they have large insertion sites and are related with large and strong muscles such as the deltoid or the ones attached to the linea aspera (Benjamin et al., 2002; Villotte & Knüsel, 2013). They leave a wide mark on the bone, with poorly defined ends, being common osseous irregularities on those attachments (Benjamin et al., 2002; Villotte & Knüsel, 2013), which makes the definition of a healthy fibrous enthesis more complex.

In fibrocartilaginous entheses, on the other hand, the union is produced through fibrocartilage and four zones can be distinguished: (1) tendon or ligament, (2) uncalcified fibrocartilage, (3) calcified fibrocartilage, and (4) subchondral bone (Benjamin et al., 2002; Villotte & Knüsel, 2013). These entheses have a small attachment area and they are located on the epiphysis of long bones next to articular surfaces, on short bones, and also on some parts of the spine (Nikita, 2016; Villotte & Knüsel, 2013). The separation between calcified and uncalcified fibrocartilage is produced by a well-defined calcification front called the tidemark, and it is also where soft tissue and bone are separated, exposing a clear area of avascular calcified fibrocartilage which is observed in dry bones (Benjamin et al., 2002; Villotte, 2013). Therefore, a healthy fibrocartilaginous enthesis is a smooth and delimited bone area without vascular foramina, although there are some exceptions such as the *brachialis* insertion on the ulna (Villotte & Knüsel, 2013).

Due to those variations, the changes occurring on each type of entheses should not be evaluated in the same way. Also, some studies suggest that fibrous entheses are less appropriate to analyze activity patterns since the stress is more widespread across bone surface dissipating the effect (Nikita et al., 2019; Villotte, Castex, et al., 2010), so they have been excluded from this study. For the analysis of the EC we used the 'New Coimbra Method' published by Henderson, Mariotti, Pany-Kucera, Villotte, and Wilczak (2016), which was developed to record the specific changes that occur at fibrocartilaginous entheses.

To evaluate the changes, each enthesis should be divided into two zones: *zone 1* usually corresponds to the outer margin of the enthesis, i.e., the margin at the opposite side of the acute angle formed by the tendon-bone attachment; *zone 2* corresponds with the rest of the enthesis, which normally is closer to the articular surface (Henderson et al., 2016). In *zone 1*, two features are scored with two degrees of expression each: bone formation, defined as a sharp and demarcated bone formation on the margin; and erosion, seen as depressions or excavations of any shape greater than 1 mm (Henderson et al., 2016). In *zone 2*, six different features are evaluated: textural change, which is the only one with just one degree of expression and is defined as a non-smooth, granular surface; bone formation, which in this zone includes any shape of bone formation greater than 1 mm; erosion, depressions or excavations of any shape in this case greater than 2 mm; fine porosity, seen as small, round or oval perforations smaller than 1 mm in diameter; macro-porosity, same shape as fine porosity but larger than 1 mm; and cavitations, which are subcortical cavities with a clear floor (Henderson et al., 2016).

The main advantage of this method is the thorough analysis of possible changes that may occur at an insertion site through the analysis of all the mentioned features in two adjacent but independent zones. A total of 22 fibrocartilaginous entheses were analyzed, 12 from the upper limb and 10 from the lower limb (**Tables 1** and **2; Supplementary Note**). To avoid inter-observer error, EC were scored by one of the authors (CAL). Intra-observer variability was assessed by a correlation analysis between the scores of 5 individuals studied twice several weeks apart (Spearman's correlation: Spearman's $\rho > 0.84$ and p-values < 0.05 for all features).

Statistical methodology

Meta-analysis

A meta-analysis is a statistical approach, often performed as part of a review of past research, employed to synthesize the evidence from empirical studies (Marín Martínez, Sánchez Meca, & López López, 2009). In this paper, we review the literature on sex-based differences on EC, aiming to capture, in each study, the estimated effect that the sex of an individual has on the presence of EC. For this we followed the four-phase flow diagram of the PRISMA statement for standardizing reporting of meta-analyses and systematic reviews (Liberati et al., 2009), although we did not refer to all elements of its checklist as several of them (particularly those referring to study biases) were designed to accommodate medical research and health interventions in contemporary populations.

To summarize the process, we first defined a search query and inclusion criteria for studies. We centred our review on archaeological studies that report EC frequencies for both sexes, population sizes, and the results of a statistical analysis comparing the distribution of EC between sexes. The search was conducted on Google Scholar between December 2020 and January 2021, and, after screening for duplicate citations, 104 papers were found on which presence of EC on ancient populations was analyzed. Of those, 28 matched our criteria, allowing us to extract a total of 532 estimated effect sizes (see **Supplementary Note** for further detail). There were 51 articles excluded from the analysis after a full-text screen and other 25 publications were excluded because they do not report their results in a way that allows them to be converted into a standardized effect size metric (e.g., insufficient description of the statistical procedure used, report restricted to p-values with no indication of either magnitude or sign of the effects, etc.). Such exclusions were performed a posteriori after a detailed revision of each publication, and are a standard procedure that ensures that the included studies have a sufficient level of methodological detail and consistency (Haidich, 2010; Meline, 2006). A graphical representation of the full process is shown in **Figure 1**.

We used R v4.0.2 (R Core Team, 2019) to carry out all of our statistical analyses. Before meta-analysis, all the effect size estimates reported in the 28 studies we obtained through the literature review were transformed into a common metric that integrates both variance and sample size (Hedges'g, the "standardized median difference"; SMD). Transformations were carried out with functions included in the packages *esc* (Lüdecke, 2019) and *effectsize* (Ben-Shachar, Lüdecke, & Makowski, 2020) for the results of Student's t-tests, Pearson correlation coefficients, Chi-squared tests and 2-by-2 contingency tables. Additionally, we used the equations provided by Cumming, Churilov, and Sena (2015) to transform the results of Mann-Whitney U tests, and Bonett and Wright (2000) for Spearman rank correlation coefficients. Following recommendations to account for between-study heterogeneity (Debray et al., 2017), we conducted a random-effect meta-analysis using the restricted maximum likelihood procedure (REML; Viechtbauer, 2005) with the Hartung and Knapp adjustment (Hartung & Knapp, 2001a, 2001b) as implemented in the package *meta* (Schwarzer, 2010). For interpretability and consistency with much of the published literature, and since our definition of sex is that of a binary variable, we report the meta-analytic effect sizes as odds-ratios (ORs). SMD values are also provided in **Supplementary Table 2.**

Bayesian model definition and prior information

In our anthropological sample, we aimed to evaluate the effect of sex on EC frequencies while controlling for the effect of other variables that could also influence EC, like age or skeletal location. As in a previous analysis of degenerative joint disease markers on this population

(Alonso-Llamazares et al., 2021), we determined that a robust approach for this would involve considering every observation on every bone as an independent sample, and controlling for intra- and inter-individual variability. This can be achieved through the use of a regression model with both fixed and random effects, also called a generalized linear mixed model (GLMM; Harrison et al., 2018). However, our dataset has the added complexity of including eight different features which have been examined on each enthesis. Serial independent analyses of these features would ignore that they could arise as the result of related processes (e.g., bone formation and bone destruction), which suggests that a procedure in which all the entheses are modelled at the same time, accounting for their correlation, should be preferred. For this reason, we carried out our regression analysis in a Bayesian framework using the *brms* package (Bürkner, 2017b), which allows for defining and resolving statistical models with arbitrary levels of hierarchical complexity. Specifically, the model we evaluated belongs to the class of multivariate generalized linear mixed models (MGLMMs; Williams, Martin, Liu, & Rast, 2020), and includes the eight entheseal features evaluated with the 'New Coimbra Method' as independent, but potentially correlated, response variables (Bürkner, 2020). All these response variables were analysed with a Poisson link function except for textural change (TCZ2), for which a Bernoulli function was chosen due to its binary nature. As predictor variables, *sex* (male/female), *age-at-death* (years) and *side* (left/right) were included as fixed effects, while *enthesis* and *individual* were included as random effects. Since we recorded a mean value and standard deviation for age-at-death estimates in each individual, both values were included in the predictor variable via a "measurement error" specification (McElreath, 2018), also known as "errors-in-variables" in frequentist regression models (Al-Sharadqah, 2017).

The Bayesian framework enables the incorporation of background knowledge into statistical models through what is called "prior distributions" or simply "priors" (Bürkner, 2015; Correa Morales, 2018; van de Schoot & Depaoli, 2014; van de Schoot et al., 2014). According to the influence priors have on the results of a model, they can be classified as informative, weakly informative and diffuse; although ascribing a prior into one category is subjective and different definitions exist in the literature (Banner, Irvine, & Rodhouse, 2020). Since priors are required for the computation of Bayesian probability estimates, there is a large body of literature exploring approaches to define them in situations when the information available is considered limited or unreliable (Chen, Ibrahim, Shao, & Weiss, 2003; Elfadaly & Garthwaite, 2017; Garthwaite, Al-Awadhi, Elfadaly, & Jenkinson, 2013; Hanea, Nane, Bedford, & French, 2021; Morris, Oakley, & Crowe, 2014; O'Hagan et al., 2006; Stefan, Evans, & Wagenmakers, 2020; van de Schoot et al., 2021; van de Schoot et al., 2018). As an example, the default prior for regression coefficients in *brms* accommodates this scenario through the use of a uniform distribution over the real numbers, which does not exclude any potential estimate. This can lead to Bayesian parameter estimates that coincide with the maximum likelihood estimator, sometimes interpreted as equivalent to hypothetical frequentist results, though this is not always the case (Zhu & Lu, 2004). While in general uniform priors are considered to add little bias to inferences (Sarma & Kay, 2020), they also lack many of the advantages of more informative alternatives (Gelman, Simpson, & Betancourt, 2017). Thus, we use default uniform priors for all parameters within our model except for effect sizes derived from the *sex* predictor. As these are central to the questions we explore in this research, we sought to define an informative prior for them based on our literature review and meta-analysis.

Our primary choice for an informative prior for the effect of sex on the prevalence of EC was the meta-analytic "prediction interval", which is the distribution containing the range of effect sizes that we would expect to see when conducting a new study (Higgins, Thompson, & Spiegelhalter, 2009). While this interval takes different sources of heterogeneity into account, its interpretation assumes that the studies included in the meta-analysis are representative of their topic and do not exclude any particular results or experimental designs. This is difficult to conclude due to the known biases of the published peer-reviewed literature (Mathur & VanderWeele, 2021), particularly "publication bias" i.e. the fact that an article is more likely to be published if the results are positive or significant. While this is a particular concern in health-related research (Page, Sterne, Higgins, & Egger, 2021), it is also difficult to assess and might not impact the bioarchaeological literature as much. However, this challenge prompted us to consider some alternatives to the prediction interval for defining a prior: First, we followed the insight of Higgins et al. (2009) at pointing out that while the empirical distribution of effect sizes included in a meta-analysis is overdispersed and inadequate for predictive inference, an "impractically wide" estimate could be derived through the application of Chebyshev's inequality, which implies that 95% of all true effects must lie within 4.47 standard deviations from their mean. Second, we followed the approach of van Zwet and Gelman (2020) in transforming all the effect sizes included in the meta-analysis into z-scores by dividing them by their standard deviation, and fitting a generalized Student's-t distribution with mean zero to their range. In contrast to our primary choice, we can consider these two latter priors to be only weakly informative, in particular the second one which should only have minimal influence on the magnitude or direction of the effects we are assessing.

Bayesian analysis and model checking

Our model, as defined above, was fitted in *brms* using the settings described by Bürkner (2017a), which include 4 parallel chains and 10000 interactions (5000 for the warmup phase and 5000 for the sampling phase, leading to a total of 20000 posterior samples). Visual inspection of model performance indices (chain mixing, posterior densities) was performed with the functions implemented in *bayesplot* (Gabry & Mahr, 2021). Following recommendations from Vehtari, Gelman, Simpson, Carpenter, and Bürkner (2021) and Gabry and Modrák (2021), we assessed model convergence by checking the values of the potential scale reduction factor "Rhat" ($\hat{R}$) and the effective sample size (ESS) estimate. A visual inspection of convergence diagnostics was also carried out using empirical cumulative distribution function (ECDF) difference plots, following Säilynoja, Bürkner, and Vehtari (2021).

Despite their advantages in accommodating complex models and experimental designs, Bayesian analysis methods require resampling and permutation approaches (such as the Hamiltonian Monte Carlo algorithm employed by *brms*) to explore and characterize the posterior probability distributions that are then used for inference. There are several circumstances under which these approaches can fail, rendering inaccurate results (Betancourt, 2020). While many of these can be detected in standard assessments after model fitting, we also carried out a pipeline of additional model checks, implemented in the WAMBS protocol, to ensure that our results were unaffected by other sources of errors (Depaoli & Van de Schoot, 2017; Schoot, Veen, Smeets, Winter, & Depaoli, 2020). We provide a completed WAMBS checklist, which also includes all model performance and convergence checks described above, in the **Supplementary Note**.

Finally, as a secondary analysis, we split our dataset into observations from the upper and lower limbs and fitted the same model to each split. With this procedure, we sought to acquire deeper insights into possible sex differences in behavior and daily activities of the SNB population. In this analysis, we kept the prior distribution for the effects of sex as our primary choice above, as well as the *brms* settings and the use of model performance and convergence criteria.

RESULTS

Meta-analysis

Results from the meta-analysis of the 28 papers extracted after the literature review (**Figure 1**, **Supplementary Note**) are shown in **Figure 2** and **Table 3**. The meta-analytic OR of all the data retrieved is 1.24 [95%CI=1.20-1.29], indicating that male individuals usually have an excess of EC. This is generally consistent with individual publications, as 22/28 show within-study ORs greater than one, although the confidence intervals of 15 of those include values below one. As advocated by IntHout, Ioannidis, Rovers, and Goeman (2016), we also compute and show the meta-analytic prediction interval, which ranges between 0.73 and 2.11, and several heterogeneity statistics. These, in our case, are measures of how similar the effect sizes of different entheses are, within and between every publication. Within-study heterogeneity was, in general, low or moderate, only high in 3/28 studies, though the heterogeneity between studies was large. Full meta-analysis results are available in **Supplementary Table 2**.

Prior definitions

Our prior for the effect size of sex, used in our multivariate Bayesian model, was derived from the meta-analytic prediction interval after transformation to the log-OR scale, and follows a normal distribution with mean 0.39 and standard deviation 0.488. For simplicity, we note this as N(0.39,0.488) elsewhere. From the data included in the meta-analysis, we also defined alternative distributions based on Chebyshev's interval of the effect sizes, N(0.43, 1.231), and a Generalized Student's T distribution of the z-scores, T(7, 0, 1.375). As required by the WAMBS checklist, we used these alternative priors for a sensitivity analysis of our regression estimates (**Supplementary Note**), together with the *brms* default uniform prior, $U(-\infty,\infty)$. All prior distributions are graphically represented in **Figure 3**.

Bayesian statistical modelling

The main results from the MGLMM are shown in **Table 4**, and completion of the WAMBS checklist did not highlight any computational problems in estimating this model (**Supplementary Note**). All estimates for the effect of sex were positive, indicating the presence of EC was more likely in male skeletons. However, only for bone formation in both zones and erosion in zone 1 the 95% credible interval did not contain zero. For prioritizing results of interest, we focus on those intervals not containing zero as this is a common approximation to the usual frequentist inference based on $p<0.05$ (Greenland, 2019). Regression coefficients from all fixed-effect covariates are shown in **Supplementary Table 3**.

The secondary examination of upper and lower limbs showed these results could be driven by erosion on the upper limb and bone formation in the lower limb (**Table 4**, **Figure 4**). The frequencies of the observed changes for each feature and enthesis, separated by sex, are presented in **Table 5**.

DISCUSSION

Meta-analysis of published EC studies

A meta-analysis allows for the efficient handling of large amounts of quantitative information, ideally leading to the synthesis of accurate, objective, and verifiable knowledge (Marín Martínez et al., 2009). In our case, this procedure confirmed that, generally, males do show larger EC scores than females, validating the role of sex as a contributing factor in the distribution of these osseous modifications. However, perhaps due to the heterogeneity between studies and assessed populations, the meta-analytic prediction interval highlighted that the opposite direction of effect (females having larger EC scores) could not be discarded. This likely reflects the diversity of populations and archaeological timeframes and contexts included in our review (Supplementary Table 2). Similarly, the average effect across all studies was small, corresponding to less than a

30% increase in EC in males compared to females. Thus, it is conceivable that biological characteristics (sexual dimorphism and body size variation) favouring finding more EC in male skeletons could be overcome by the environmental or cultural context of a population, and thus future studies should be open to this possibility. It is also worth noting that even the studies in our review with the stronger effect sizes favouring males, such as Weiss (2003, 2004), explicitly cautioned that differentiating between biological and non-biological sources of variation was not possible with their data.

Bayesian statistical analysis

The 'New Coimbra Method' for the study of EC is a standardized system that provides a thorough description of all the occurring changes on an enthesis through eight different features. However, it is not entirely clear how these features should be statistically analyzed and indeed different methods have been applied in empirical scenarios (Michopoulou, Nikita, & Henderson, 2017). Following well-known precedents in medical research, we can consider that a series of independent tests would imply a statistical multiplicity burden (Streiner, 2015), while deriving composite metrics from all or groups of features might lead to information loss (Gorter, Fox, & Twisk, 2015). The MGLMM approach we chose for analyzing EC frequencies avoids both of these drawbacks by fitting all response and predictor variables into a single model, although more flexible parameterizations selecting specific predictors for specific responses can also be used (Williams et al., 2020). Such a model would have been likely unsolvable on a frequentist maximum-likelihood setting, but it is valid in a Bayesian framework even with limited sample sizes (Carvalho, Gonçalves, Grosgeorge, & Paes, 2017; van de Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & van Loey, 2015). Due to their use of a hierarchical random-effect structure, MGLMMs also allow us to account for different sources of variability at the level of individuals and skeletal locations (entheses in this case), which as we have shown previously are relevant for the analysis of paleopathological data (Alonso-Llamazares et al., 2021).

Often, a controversial stage in Bayesian inference is the decision of what constitutes prior information or even if this can be considered to be available. There are several ways to obtain a prior distribution, such as consulting experts opinion, also known as elicitation, performing a pilot study or extracting data from a meta-analysis (Correa Morales, 2018; van de Schoot et al., 2014). For our prior on the effect size of sex, we have chosen the latter option. Importantly, we made our prior distribution compatible with the meta-analytic prediction interval, since this range should be inclusive of all effect size values plausible for our study. Such an approach is common in practice (Sarma & Kay, 2020), although we also benchmarked our prior choice via comparison to weakly informative and diffuse alternatives, all leading to largely overlapping posterior distributions. This leads us to consider that our dataset, by itself, contains enough information to determine a plausible estimate for the effect of sex without requiring a highly informative prior (Etz, 2018; Gelman et al., 2017). Additionally, it validates that our choice of prior distribution might be applicable to future research on the same topic, given that it systematically captures information from a wide array of previous studies and populations. However, further work might be needed to assess whether it can also be of use for samples from understudied regions not captured by our literature review, which might display different adaptations to activity patterns.

Entheseal changes and sex differences

Bone tissue has two possible answers to stress: bone formation and bone resorption. EC will take one of these two forms, with different sizes and shapes (Jurmain et al., 2012). The thin cortical of fibrocartilaginous entheses undergoes those changes due to repetitive biomechanical loading occurring during muscle contraction (Karakostis, Hotz, Scherf, Wahl, & Harvati, 2018). The detail in which the 'New Coimbra Method' record EC represent an advantage for study the relationship between these changes and other factors such as sex. Besides, this method has proven to be less affected by body size than others, partially because fibrocartilaginous entheses have a more compact attachment site (Michopoulou et al., 2017; Weiss, 2015), leading to a more favorable position to evaluate sex differences due to daily behavior.

Results of our analysis show a remarkable consistency in the estimation of sex-based differences in EC between previous studies, which extends to our current sample. This contrasts with the concern that such inter-population inferences would be "doomed" by confounding effects, and supports the feasibility of the "epidemiological" perspective advocated for moving forward the paradigms of reconstructing past activities through skeletal markers (Jurmain & Villotte, 2010). Particularly, our MGLMM analysis showed that erosion in zone 1 and bone formation in both zones have greater development in males. Our separate analysis at each limb validates this observation and suggests that erosion only has a higher effect in males for the upper limb while bone formation is greater in lower limbs (**Figure 4**). Erosion is an osteolytic type of lesion and it has been argued that it could be the result of insufficient remodeling and healing after daily trauma, whereas bone formation can be the response to mild trauma with a low repetition rate (Barbe et al., 2018). That is to say, erosion is a more probable response of the bone when the physical activity performed is constant and prolonged in time, while bone formation is produced by efforts of lower intensity, not necessarily daily.

Although the analysis of EC does not permit establishing specific activities as responsible for their appearance, the level and distribution of the observed changes are compatible with what we know of this medieval and early modern population. Burgos was a city of artisans with intense economic activity, during a time in which constructions and building renovations were nearly continuous (Sebastián Moreno & Guerrero Navarrete, 2018). As expected, high frequencies of bone formation in the upper limb suggest physically demanding work in both sexes, maybe more consistently in males since they also present more erosion changes. Among the most demanding jobs during that time were woodworkers, roofers, glassmakers, and other construction-related professionals, besides the manufacture of clothing, footwear, and work tools (Sebastián Moreno, 2017). Although those jobs are usually considered male jobs, urban women were known to work with their fathers and husbands in a wide variety of tasks, besides the domestic responsibilities, like as mason and bricklayer assistants (Bovey, 2015; Herrero, 2006; López Beltrán, 2010). In general, women had many occupations and worked as hard as men, even though their names rarely appear in contracts (Bennett & Karras, 2013). Specifically in Burgos, it is known that women usually carried the water and help with mixes in state constructions (López Beltrán, 2010), but they probably performed more strenuous activities in those constructions as it happened in other cities of the time (Hatipkarasulu & Roff, 2011; Herrero, 2006). All that information is compatible with the EC frequencies observed in SNB.

Finally, even though we have controlled for most known confounding variables to isolate the effect of sex on EC, there are hormonal factors and sex differences in how bone adapts that make it difficult to determine the extent of EC that can be safely attributed to physical activity (Santana-Cabrera et al., 2015). Assuming that they are at least in part due to daily activities, the differences in bone formation in the lower limb are consistent with previous results from a degenerative joint disease study of this population in which differences at the level of the lower limbs were attributed to men having to walk more frequently outside the city to pick raw materials for their workshops or for construction work (Alonso-Llamazares et al., 2021).

CONCLUSIONS

A meta-analysis is a useful tool for quantitative evidence reviews, which has given us an overview of the existing literature on the effect of sex on EC frequencies and an expectation for future studies. In the Bayesian framework we used, it also allowed us to calibrate a regression model applied to a novel archaeological sample.

Attempting to model the wide range of variables that can be informative on the development of EC, we chose to estimate a Bayesian MGLMM. Despite methodological complexities, in relation to traditional methods, its performance in our sample suggests it could be a better alternative to estimate sex differences in EC distributions since it allows controlling for confounding factors and the hierarchical nature of paleopathological data. Our analysis illustrates that the EC observed in individuals from the SNB population exhibit greater effects in males for bone formation and erosion, that are consistent with the known activities (crafts and construction work) carried out by this urban population for economic sustenance.

DATA AVAILABILITY STATEMENT

Syntaxis to generate the prior distributions and of the main model for reproducing the analyses is available at https://github.com/karmenhbc/SNB_ECdata_analysis. Raw EC data from the SNB population, as well as the Hedges'g data used in the meta-analysis is also available at https://doi.org/10.6084/m9.figshare.14610492.

ACKNOWLEDGMENTS

This research has received support in part from the 'Ministerio de Educación, Cultura y Deporte' of the Junta de Castilla y León, Spain (Project CN-08-045).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

LITERATURE CITED

Al-Sharadqah, A. (2017). A new perspective in functional EIV linear model: Part I. *Communications in Statistics - Theory and Methods, 46*(14), 7039-7062. doi:10.1080/03610926.2016.1143009

al-Oumaoui, I., Jiménez-Brobeil, S., & Du Souich, P. (2004). Markers of activity patterns in some populations of the Iberian Peninsula. *International Journal of Osteoarchaeology, 14*(5), 343-359.

Alonso-Llamazares, C., Blanco Márquez, B., Lopez, B., & Pardiñas, A. F. (2021). Assessing individual and population variability in degenerative joint disease prevalence using generalized linear mixed models. *American Journal of Physical Anthropology, [in press].* doi:https://doi.org/10.1002/ajpa.24195

Banner, K. M., Irvine, K. M., & Rodhouse, T. J. (2020). The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in ecology and evolution, 11*(8), 882-889. doi:https://doi.org/10.1111/2041-210X.13407

Barbe, M. F., Massicotte, V. S., Assari, S., Monroy, M. A., Frara, N., Harris, M. Y., . . . Popoff, S. N. (2018). Prolonged high force high repetition pulling induces osteocyte apoptosis and trabecular bone loss in distal radius, while low force high repetition pulling induces bone anabolism. *Bone, 110*, 267-283. doi:https://doi.org/10.1016/j.bone.2018.02.014

Beier, J., Anthes, N., Wahl, J., & Harvati, K. (2021). Prevalence of cranial trauma in Eurasian Upper Paleolithic humans. *American Journal of Physical Anthropology, 174*(2), 268-284. doi:https://doi.org/10.1002/ajpa.24163

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software, 5*(56), 2815.

Benjamin, M., Kumai, T., Milz, S., Boszczyk, B., Boszczyk, A., & Ralphs, J. (2002). The skeletal attachment of tendons—tendon 'entheses'. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 133*(4), 931-945.

Bennett, J. M., & Karras, R. M. (2013). Women, gender and medieval historians. In J. M. Bennett & R. M. Karras (Eds.), *The Oxford Handbook of Women and Gender in Medieval Europe* (pp. 1-17).

Betancourt, M. (2020). Identity Crisis. Retrieved from https://github.com/betanalpha/knitr_case_studies/tree/master/identifiability,

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika, 65*(1), 23-28. doi:10.1007/BF02294183

Bovey, A. (2015). Women in medieval society. *British Library.* Retrieved from https://www.bl.uk/the-middle-ages/articles/women-in-medieval-society#

Buikstra, J. E., & Ubelaker, D. H. (1994). *Standards for data collection from human skeletal remains. Proceedings of a seminar at the Field Museum of Natural History.*: Arkansas Archaeological Survey Research Series No. 44. Fayetteville, AK: Arkansas Archaeological Survey.

Bürkner, P.-C. (2015). brms: An R Package for Bayesian Generalized Linear Mixed Models using Stan. *URL: https://cran. r-project. org/web/packages/brms/vignettes/brms. pdf.*

Bürkner, P.-C. (2017a). Advanced Bayesian multilevel modeling with the R package brms. *arXiv preprint arXiv:.11123*.

Bürkner, P.-C. (2017b). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80*(1). doi:10.18637/jss.v080.i01

Bürkner, P.-C. (2020). Estimating Multivariate Models with brms. *https://cran.r-project.org/web/packages/brms/vignettes/brms_multivariate.html*.

Carvalho, H. M., Gonçalves, C. E., Grosgeorge, B., & Paes, R. R. (2017). Validity and usefulness of the Line Drill test for adolescent basketball players: a Bayesian multilevel analysis. *Research in Sports Medicine, 25*(3), 333-344. doi:10.1080/15438627.2017.1314296

Chen, M.-H., Ibrahim, J. G., Shao, Q.-M., & Weiss, R. E. (2003). Prior elicitation for model selection and estimation in generalized linear mixed models. *Journal of Statistical Planning and Inference, 111*(1), 57-76. doi:https://doi.org/10.1016/S0378-3758(02)00285-9

Cheverko, C. M., & Hubbe, M. (2017). Comparisons of statistical techniques to assess age-related skeletal markers in bioarchaeology. *American Journal of Physical Anthropology, 163*(2), 407-416. doi:10.1002/ajpa.23206

Correa Morales, J. C. (2018). *Introducción a la Estadística Bayesiana con R*. Medeññín: Universidad Nacional de Colombia.

Cumming, T. B., Churilov, L., & Sena, E. (2015). The missing medians: exclusion of ordinal data from meta-analyses. *PLoS ONE, 10*(12), e0145580.

Debray, T. P. A., Damen, J. A. A. G., Snell, K. I. E., Ensor, J., Hooft, L., Reitsma, J. B., . . . Moons, K. G. M. (2017). A guide to systematic review and meta-analysis of prediction model performance. *BMJ, 356*, i6460. doi:10.1136/bmj.i6460

Depaoli, S. (2014). The Impact of Inaccurate "Informative" Priors for Growth Parameters in Bayesian Growth Mixture Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(2), 239-252. doi:10.1080/10705511.2014.882686

Depaoli, S., & Van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods, 22*(2), 240.

Elfadaly, F. G., & Garthwaite, P. H. (2017). Eliciting Dirichlet and Gaussian copula prior distributions for multinomial models. *Statistics Computing, 27*(2), 449-467.

Eshed, V., Gopher, A., Galili, E., & Hershkovitz, I. (2004). Musculoskeletal stress markers in Natufian hunter-gatherers and Neolithic farmers in the Levant: The upper limb. *American Journal of Physical Anthropology, 123*(4), 303-315.

Etz, A. (2018). Introduction to the Concept of Likelihood and Its Applications. *Advances in Methods and Practices in Psychological Science, 1*(1), 60-69. doi:10.1177/2515245917744314

Florez, P. E. (1771-72). España Sagrada. In (Vol. XXVI y XXVII).

Gabry, J., & Mahr, T. (2021). bayesplot: Plotting for Bayesian Models. *URL: https://mc-stan.org/bayesplot/, R package version 1.8.0.*

Gabry, J., & Modrák, M. (2021). Visual MCMC diagnostics using the bayesplot package. *URL: https://cran.r-project.org/web/packages/bayesplot/vignettes/visual-mcmc-diagnostics.html#rhat-potential-scale-reduction-statistic*.

Garthwaite, P. H., Al-Awadhi, S. A., Elfadaly, F. G., & Jenkinson, D. J. (2013). Prior distribution elicitation for generalized linear and piecewise-linear models. *Journal of Applied Statistics, 40*(1), 59-75.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy, 19*(10), 555.

Giannotti, P. S. (2020). Resultados preliminares de evaluación de cambios entesiales mediante método Coimbra en poblaciones coloniales españolas de América meridional: ciudad de Mendoza, Argentina (siglos XVII-XIX). *Revista Argentina de Antropología Biológica, 22*(1), 011-011.

Godde, K., & Taylor, R. W. (2013). Distinguishing body mass and activity level from the lower limb: Can entheses diagnose obesity? *Forensic Science International, 226*(1), 303.e301-303.e307. doi:https://doi.org/10.1016/j.forsciint.2013.01.027

Goicolea Julián, F. J. (2019). J. Sebastián Moreno. La ciudad medieval como capital regional. Burgos en el siglo XV. *Studia Historica. Historia Medieval, 37*(2), 212-214.

Gorter, R., Fox, J.-P., & Twisk, J. W. R. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Medical Research Methodology, 15*(1), 55. doi:10.1186/s12874-015-0050-x

Greenland, S. (2019). Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution With S-Values. *The American Statistician, 73*(sup1), 106-114. doi:10.1080/00031305.2018.1529625

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia, 14*(Suppl 1), 29-37.

Hanea, A. M., Nane, G. F., Bedford, T., & French, S. (2021). *Expert Judgement in Risk and Decision Analysis* (Vol. 293): Springer Nature.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., . . . Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ, 6*, e4794-e4794. doi:10.7717/peerj.4794

Hartung, J., & Knapp, G. (2001a). On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine, 20*(12), 1771-1782. doi:https://doi.org/10.1002/sim.791

Hartung, J., & Knapp, G. (2001b). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine, 20*(24), 3875-3889. doi:https://doi.org/10.1002/sim.1009

Hatipkarasulu, Y., & Roff, E. (2011). *Women in construction: An early historical perspective.* Paper presented at the 47th ASC Annual International Conference Proceedings of Associated Schools of Construction.

Hawkey, D. E. (1998). Disability, compassion and the skeletal record: using musculoskeletal stress markers (MSM) to construct an osteobiography from early New Mexico. *International Journal of Osteoarchaeology, 8*(5), 326-340.

He, L. R., & de Almeida Prado, P. S. (2020). An evaluation of the relationship between the degree of entheseal changes and the severity of osteodegenerative processes at fibrocartilaginous entheses. *The Anatomical Record, n/a*(n/a). doi:10.1002/ar.24541

Henderson, C. Y., & Alves Cardoso, F. (2013). Special issue entheseal changes and occupation: technical and theoretical advances and their applications. *International Journal of Osteoarchaeology, 23*(2), 127-134.

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 entheseal changes. *International Journal of Osteoarchaeology, 26*(5), 925-932.

Herrero, M. d. C. G. (2006). Actividades laborales femeninas en la Baja Edad Media turolense. *Aragón en la Edad Media*(19), 181-200.

Higgins, J. P., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A, 172*(1), 137-159.

IntHout, J., Ioannidis, J. P., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7).

Jurmain, R. D., Cardoso, F. A., Henderson, C., & Villotte, S. (2012). Bioarchaeology's Holy Grail: The Reconstruction of Activity. In A. L. Grauer (Ed.), *A Companion to Paleopathology* (pp. 531-552). Chichester, UK: Wiley-Blackwell.

Jurmain, R. D., & Villotte, S. (2010). Terminology. Entheses in medical literature and physical anthropology: a brief review. *http://www.uc.pt/en/cia/msm/MSM_terminology3*.

Karakostis, F. A., Hotz, G., Scherf, H., Wahl, J., & Harvati, K. (2018). A repeatable geometric morphometric approach to the analysis of hand entheseal three-dimensional form. *American Journal of Physical Anthropology, 166*(1), 246-260. doi:https://doi.org/10.1002/ajpa.23421

Konigsberg, L. W., & Frankenberg, S. R. (2013). Bayes in biological anthropology. *American Journal of Physical Anthropology, 152*(S57), 153-184. doi:https://doi.org/10.1002/ajpa.22397

Krenzer, U. (2006). Compendio de Métodos Antropológicos Forenses para la Reconstrucción del Perfil Osteo-biológico. *Estimación de la edad osteológica en adultos, III*.

Kruschke, J. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142*(2), 573.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ, 339*, b2700. doi:10.1136/bmj.b2700

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*(4), 458-472.

Lieverse, A. R., Bazaliiskii, V. I., Goriunova, O. I., & Weber, A. W. (2013). Lower limb activity in the Cis-Baikal: Entheseal changes among middle holocene siberian foragers. *American Journal of Physical Anthropology, 150*(3), 421-432.

Lieverse, A. R., Stock, J. T., Katzenberg, M. A., & Haverkort, C. M. (2011). The bioarchaeology of habitual activity and dietary change in the Siberian Middle Holocene. In R. Pinhasi & J. T. Stock (Eds.), *Human bioarchaeology of the transition to agriculture* (pp. 263-291): John Wiley and Sons: Chichester.

López-Bueis, I., Robledo, B., Roselló, J., & Trancho, G. J. (1996). *Funciones discriminantes para la determinación sexual de la tibia en una serie española de sexo y edad conocidos*. Paper presented at the IX Congreso de Antropología Biológica.

López Beltrán, M. T. (2010). El trabajo de las mujeres en el mundo urbano medieval. *Mélanges de la Casa de Velázquez. Nouvelle série*(40-2), 39-57.

López Sobrino, J. (2000). *La iglesia de San Nicolás de Bari, Burgos* (2 ed.): Amabar SL.

Lovejoy, C. O., Meindl, R. S., Pryzbeck, T. R., & Mensforth, R. P. (1985). Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death. *American journal of physical anthropology, 68*(1), 15-28.

Lüdecke, D. (2019). Effect Size Computation for Meta Analysis (Version 0.5.1). *https://CRAN.R-project.org/package=esc*. doi:10.5281/zenodo.1249218

Marín Martínez, F., Sánchez Meca, J., & López López, J. (2009). El metaanálisis en el ámbito de las Ciencias de la Salud: una metodología imprescindible para la eficiente acumulación del conocimiento. *Fisioterapia, 31*(3), 107-114.

Martin, M. A. (2019). Biological Anthropology in 2018: Grounded in Theory, Questioning Contexts, Embracing Innovation. *American Anthropologist, 121*(2), 417-430. doi:https://doi.org/10.1111/aman.13233

Mathur, M. B., & VanderWeele, T. J. (2021). Estimating publication bias in meta-analyses of peer-reviewed studies: A meta-meta-analysis across disciplines and journal tiers. *Research Synthesis Methods, 12*(2), 176-191. doi:https://doi.org/10.1002/jrsm.1464

Mazza, B. P. (2020). Cambios entésicos en guaraníes pre-hispánicos del Delta inferior del Paraná. *Revista Argentina de Antropología Biológica, 22*(1), 010-010.

McElreath, R. (2018). Missing Data and Other Opportunities. In *Statistical rethinking: A Bayesian course with examples in R and Stan* (pp. 423-439). London, UK: Chapman and Hall/CRC.

Meindl, R. S., & Lovejoy, C. O. (1985). Ectocranial suture closure: A revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American journal of physical anthropology, 68*(1), 57-66.

Meindl, R. S., Lovejoy, C. O., Mensforth, R. P., & Walker, R. A. (1985). A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. *American journal of physical anthropology, 68*(1), 29-45.

Meline, T. (2006). Selecting Studies for Systemic Review: Inclusion and Exclusion Criteria. *CONTEMPORARY ISSUES IN COMMUNICATION SCIENCE AND DISORDERS, 33*(Spring), 21-27. doi:doi:10.1044/cicsd_33_S_21

Michopoulou, E., Nikita, E., & Henderson, C. (2017). A Test of the Effectiveness of the Coimbra Method in Capturing Activity-induced Entheseal Changes. *International Journal of Osteoarchaeology, 27*(3), 409-417.

Molnar, P. (2010). Patterns of physical activity and material culture on Gotland, Sweden, during the Middle Neolithic. *International Journal of Osteoarchaeology, 20*(1), 1-14. doi:10.1002/oa.1000

Morris, D. E., Oakley, J. E., & Crowe, J. A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software, 52*, 1-4. doi:https://doi.org/10.1016/j.envsoft.2013.10.010

Niinimäki, S., & Baiges Sotos, L. (2013). The relationship between intensity of physical activity and entheseal changes on the lower limb. *International Journal of Osteoarchaeology, 23*(2), 221-228.

Nikita, E. (2016). Activity patterns. In *Osteoarchaeology: A guide to the macroscopic study of human skeletal remains* (pp. 269-300): Academic Press.

Nikita, E., Mattingly, D., & Lahr, M. M. (2013). Methodological considerations in the statistical analysis of degenerative joint and disc disease. *International Journal of Paleopathology, 3*(2), 105-112.

Nikita, E., Xanthopoulou, P., Bertsatos, A., Chovalopoulou, M.-E., & Hafez, I. (2019). A three-dimensional digital microscopic investigation of entheseal changes as skeletal activity markers. *International Journal of Osteoarchaeology, 169*(4), 704-713. doi:10.1002/ajpa.23850

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., . . . Rakow, T. (2006). *Uncertain judgements: eliciting experts' probabilities*: John Wiley & Sons.

Otárola-Castillo, E., & Torquato, M. G. (2018). Bayesian Statistics in Archaeology. *Annual Review of Anthropology, 47*(1), 435-453. doi:10.1146/annurev-anthro-102317-045834

Page, M. J., Sterne, J. A., Higgins, J. P., & Egger, M. (2021). Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. *Research Synthesis Methods, 12*(2), 248-259.

Palmer, J. (2012). *Busy Bones Osteoarthritis and musculoskeletal markers as evidence of physical activity and social differentiation in post-medieval the Netherlands.* (Master Thesis), University of Leiden, Leiden.

Peterson, J. (2010). Domesticating gender: Neolithic patterns from the southern Levant. *Journal of Anthropological Archaeology, 29*(3), 249-264. doi:https://doi.org/10.1016/j.jaa.2010.03.002

PRISMA Flow Diagram Generator. *(2021). http://prisma.thetacollaborative.ca/.*

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL https://www.R-project.org/.

Rojas-Sepúlveda, C. M., & Dutour, O. (2014). Enfermedad articular degenerativa y cambios entesiales en seis colecciones óseas prehispánicas del noroccidente de América del Sur. *Chungará (Arica), 46*(1), 153-169.

Rojas-Sepúlveda, C. M., Rivera-Sandoval, J., & Martín-Rincón, J. G. (2011). Paleoepidemiology of pre-Columbian and Colonial Panamá Viejo: a preliminary study. *Bulletins et Mémoires de la Société d'anthropologie de Paris, 23*(1), 70-82. doi:10.1007/s13219-011-0033-3

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225-237. doi:10.3758/PBR.16.2.225

Säilynoja, T., Bürkner, P.-C., & Vehtari, A. (2021). Graphical Test for Discrete Uniformity and its Applications in Goodness of Fit Evaluation and Multiple Sample Comparison. *arXiv preprint arXiv:.10522*.

Santana-Cabrera, J., Velasco-Vázquez, J., & Rodríguez-Rodríguez, A. (2015). Entheseal changes and sexual division of labor in a North-African population: The case of the pre-Hispanic period of the Gran Canaria Island (11th–15th c. CE). *HOMO - Journal of Comparative Human Biology, 66*(2), 118-138. doi:https://doi.org/10.1016/j.jchb.2014.10.005

Sarma, A., & Kay, M. (2020). *Prior Setting in Practice: Strategies and Rationales Used in Choosing Prior Distributions for Bayesian Analysis*. Paper presented at the Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376377

Schoot, R. v. d., Veen, D., Smeets, L., Winter, S. D., & Depaoli, S. (2020). A tutorial on using the WAMBS checklist to avoid the misuse of Bayesian statistics. In R. an de Schoot & M. Miocevic (Eds.), *Small Sample Size Solutions: A Guide for Applied Researchers Practitioners* (pp. 30-49): Routledge. Taylor & Francis Group.

Schwarzer, G. (2010). meta: Meta-Analysis with R. R package version 1.6-1. *http://CRAN. R-project. org/package= meta*.

Sebastián Moreno, J. (2017). *La ciudad medieval como capital regional, Burgos, siglo XV*. Universidad Autónoma de Madrid, Madrid.

Sebastián Moreno, J., & Guerrero Navarrete, Y. (2018). Todos los caminos confluyen en Burgos. Centralidad y jerarquización urbanas en la Castilla bajomedieval. *Anuario de Estudios Medievales, 48*(1), 181-211.

Silva, I. R. (2018). On the correspondence between frequentist and Bayesian tests. *Communications in Statistics - Theory and Methods, 47*(14), 3477-3487. doi:10.1080/03610926.2017.1359296

Steen, S. L., & Lane, R. W. (1998). Evaluation of habitual activities among two Alaskan Eskimo populations based on musculoskeletal stress markers. *International Journal of Osteoarchaeology, 8*(5), 341-353.

Stefan, A. M., Evans, N. J., & Wagenmakers, E.-J. (2020). Practical challenges and methodological flexibility in prior elicitation. *Psychological Methods*.

Stefanović, S., & Porčić, M. (2013). Between-group differences in the patterning of musculo-skeletal stress markers: avoiding confounding factors by focusing on qualitative aspects of physical activity. *International Journal of Osteoarchaeology, 23*(1), 94-105.

Streiner, D. L. (2015). Best (but oft-forgotten) practices: the multiple problems of multiplicity—whether and how to correct for many statistical tests. *The American journal of clinical nutrition, 102*(4), 721-728. doi:10.3945/ajcn.115.113548

Todd, T. W. (1920). Age changes in the pubic bone. I. The male white pubis. *American journal of physical anthropology, 3*(3), 285-334.

Todd, T. W. (1921). Age changes in the pubic bone. *American journal of physical anthropology, 4*(1), 1-70.

Trancho, G. J., Robledo, B., López-Bueis, I., & Sánchez, J. A. (1997). Sexual determination of the femur using discriminant functions. Analysis of a Spanish population of known sex and age. *Journal of Forensic Science, 42*(2), 181-185.

Trancho, G. J., Robledo, B., & Sánchez, J. A. (2012). Dimorfismo sexual del húmero en una población española de sexo y edad conocidos. *Biodiversidad Humana y Evolución*, 364-369.

van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology, 6*(1), 25216. doi:10.3402/ejpt.v6.25216

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *The European Health Psychologist, 16*(2), 75-84.

van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., . . . Yau, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers, 1*(1), 1. doi:10.1038/s43586-020-00001-2

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development, 85*(3), 842-860. doi:https://doi.org/10.1111/cdev.12169

van de Schoot, R., Sijbrandij, M., Depaoli, S., Winter, S. D., Olff, M., & Van Loey, N. E. (2018). Bayesian PTSD-trajectory analysis with informed priors based on a systematic literature search and expert elicitation. *Multivariate Behavioral Research, 53*(2), 267-291.

van Zwet, E., & Gelman, A. (2020). A proposal for informative default priors scaled by the standard error of estimates. *arXiv preprint arXiv:.15037.*

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *arXiv:1903.08008v3, 1*(1), 1-28.

Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. *Journal of Educational and Behavioral Statistics, 30*(3), 261-293. doi:10.3102/10769986030003261

Villotte, S. (2013). Practical protocol for scoring the appearance of some fibrocartilaginous entheses on the human skeleton. *Villotte. Online. Available: http://www. academia. edu/1427191/Practical_protocol_for_scoring_the_appearance_of_some_fibrocartilaginous_entheses_on_the_human_skeleton.*

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*(2), 224-234.

Villotte, S., Churchill, S. E., Dutour, O. J., & Henry-Gambier, D. (2010). Subsistence activities and the sexual division of labor in the European Upper Paleolithic and Mesolithic: Evidence from upper limb enthesopathies. *Journal of Human Evolution, 59*(1), 35-43.

Villotte, S., & Knüsel, C. J. (2013). Understanding entheseal changes: definition and life course changes. *International Journal of Osteoarchaeology, 23*(2), 135-146.

Weiss, E. (2003). Understanding muscle markers: aggregation and construct validity. *American Journal of Physical Anthropology, 121*(3), 230-240.

Weiss, E. (2004). Understanding muscle markers: lower limbs. *American Journal of Physical Anthropology, 125*(3), 232-238.

Weiss, E. (2015). Examining activity patterns and biological confounding factors: differences between fibrocartilaginous and fibrous musculoskeletal stress markers. *International Journal of Osteoarchaeology, 25*(3), 281-288.

Weiss, E., Corona, L., & Schultz, B. (2012). Sex differences in musculoskeletal stress markers: problems with activity pattern reconstructions. *International Journal of Osteoarchaeology, 22*(1), 70-80.

Wilczak, C. A. (1998). Consideration of sexual dimorphism, age, and asymmetry in quantitative measurements of muscle insertion sites. *International Journal of Osteoarchaeology, 8*(5), 311-325.

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. doi:https://doi.org/10.1016/j.jasrep.2016.11.033

Williams, D. R., Martin, S. R., Liu, S., & Rast, P. (2020). Bayesian Multivariate Mixed-Effects Location Scale Modeling of Longitudinal Relations among Affective Traits, States, and Physical Activity. *European Journal of Psychological Assessment, 36*, 981-997. doi:https://doi.org/10.1027/1015-5759/a000624

Woo, E. J., & Pak, S. (2013). Degenerative joint diseases and enthesopathies in a Joseon Dynasty population from Korea. *HOMO - Journal of Comparative Human Biology, 64*(2), 104-119.

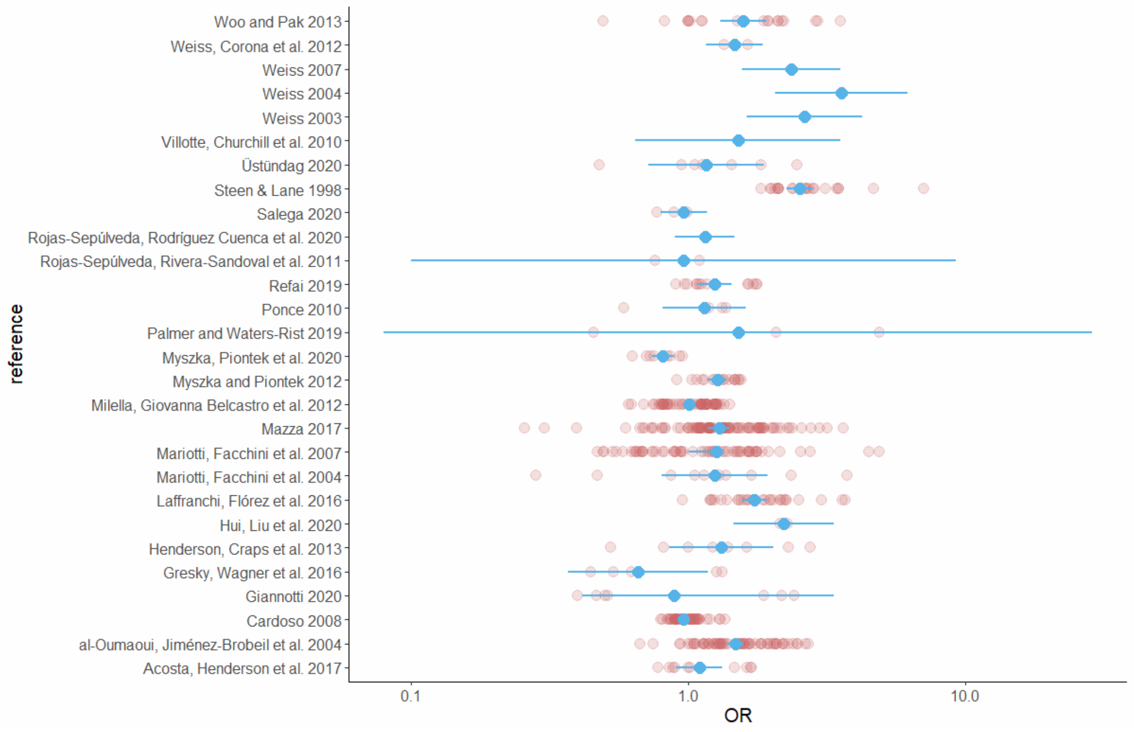
Zhu, M., & Lu, A. Y. (2004). The Counter-intuitive Non-informative Prior for the Bernoulli Family. *Journal of Statistics Education, 12*(2), null-null. doi:10.1080/10691898.2004.11910734

Zyphur, M. J., & Oswald, F. L. (2015). Bayesian Estimation and Inference: A User's Guide. *Journal of Management, 41*(2), 390-420. doi:10.1177/0149206313501200

Figure 1

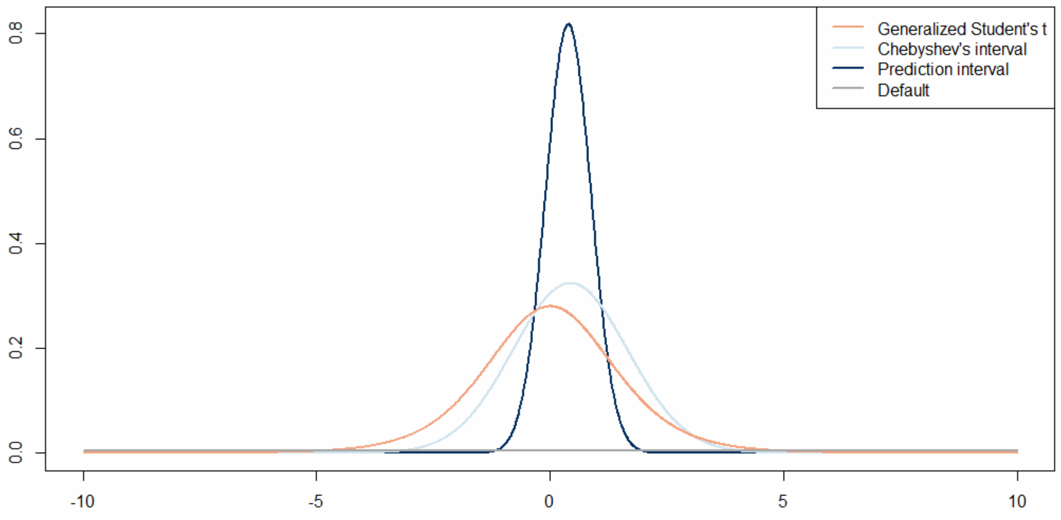
PRISMA flow diagram of the systematic literature review conducted before meta-analysis. Generated with "PRISMA Flow Diagram Generator" (http://prisma.thetacollaborative.ca/).

Figure 2



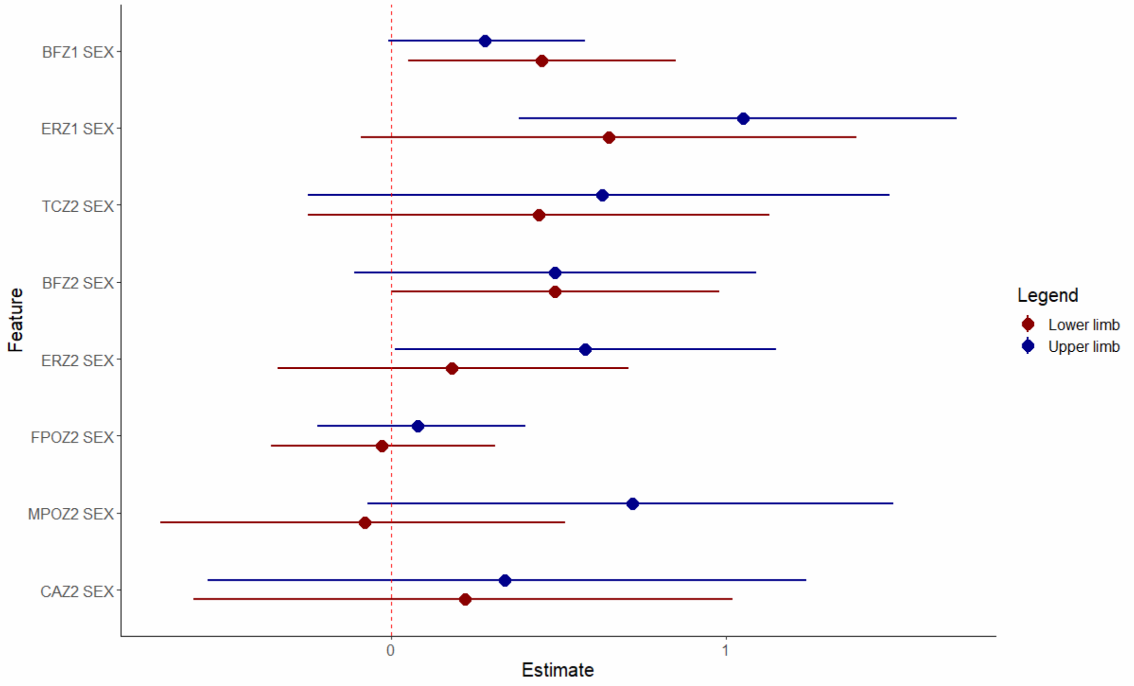
Results from the meta-analysis of 28 articles that assessed sex-based differences on EC prevalence. Odds ratios (OR) derived from each estimate reported in each publication are shown in red. For illustration purposes, within publication summary ORs and 95% confidence intervals, estimated via fixed-effect meta-analyses, are shown in blue.

Figure 3



Prior distributions for the effect of sex in EC frequency within the entheses used in our main Bayesian analysis and additional model checking tests. Note the scale of the distributions refers to the regression beta coefficient (“estimate” in our result tables), which in a logistic model would be the natural logarithm of the odds ratio. Note that all our proposed priors give negligible probabilities to absolute effect sizes larger than 6, as those would be so drastic as to be appreciable without experimentation (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

Figure 4



Graphical representation of the estimates and credible intervals for the effect of sex on each feature for the upper limb (blue) and the lower limb (red) separately. Note that credible intervals for ERZ1/2 in the lower limb and BFZ1 in the upper limb do not contain zero and indicate greater EC frequency in males. [BFZ1: bone formation zone 1; ERZ1: erosion zone 1; TCZ2: textural change zone 2; BFZ2: bone formation zone 2; ERZ2: erosion zone 2; FPOZ2: fine porosity zone 2; MPOZ2: macroporosity zone 2; CAZ2: cavitations zone 2].

Table 1

Entheses analyzed from the upper limb, attachment site and main movement involved.

Entheses	Attachment site	Movement
Origin of the long head of the *biceps brachii*	Supraglenoid tubercle of the scapula	Forearm flexion
Origin of the long head of the *triceps brachii*	Infraglenoid tubercle of the scapula	Forearm extension
Insertion of the *supraspinatus*	Superior facet on the greater tuberosity of the humerus	Shoulder rotation
Insertion of the *infraspinatus*	Middle facet on the greater tuberosity of the humerus	Shoulder rotation
Insertion of the *teres minor*	Inferior facet on the greater tuberosity of the humerus	Shoulder rotation
Insertion of the *subscapularis*	Lesser tuberosity of the humerus	Shoulder rotation
Common origin of the extensors	Lateral epicondyle of the humerus	Wrist and hand extension
Common origin of the flexors	Medial epicondyle of the humerus	Wrist and hand flexion
Insertion of the *triceps brachii*	Proximal end of the olecranon process of the ulna	Forearm extension
Insertion of the *brachialis*	Ulna tuberosity	Forearm flexion
Insertion of the *biceps brachii*	Radius tuberosity	Forearm flexion
Insertion of the *brachioradialis*	Lateral surface of distal end of the radius	Forearm flexion

Table 2

Entheses analyzed from the lower limb, attachment site and main movement involved.

Entheses	Attachment site	Movement
Common origin of *biceps femoris*, *semitendinosus* and *semimembranosus*	Ischial tuberosity	Leg extension
Insertion of the *gluteus medius*	Lateral and superior surfaces of the greater trochanter of the femur	Leg abduction
Insertion of the *gluteus minimus*	Anterior surface of the greater trochanter of the femur	Leg abduction
Insertion of the *iliopsoas*	Lesser trochanter of the femur	Hip and torso flexion
Insertion of the *obturator externus*	Posteromedial surface of the greater trochanter of the femur	Hip abduction
Origin of the medial and lateral heads of the *gastrocnemius*	Posterior nonarticular surface of medial femoral condyle and the lateral surface of lateral femoral condyle	Sole of foot flexion
Insertion of *quadriceps femoris*	Anterior-superior surface of the patella	Knee extension
Origin of the patellar tendon	Anterior-inferior surface of the patella	Knee extension
Insertion of the patellar tendon	Tibial tuberosity	Knee extension
Insertion of the *triceps surae* (Achilles' tendon)	Middle third of the posterior calcaneal surface	Sole of foot flexion

Table 3

Main results from the meta-analysis of the reviewed EC literature. Statistics shown for each reference were derived, for illustration purposes, from a fixed-effect meta-analysis of all effect sizes described within the study. Final total effect size, prediction interval and between-study heterogeneity were derived from the main random-effect meta-analysis described in the main text.

Reference	**OR**	**95%-CI**	**Q**	**I^2**	**τ^2**
Woo and Pak 2013	1.58	1.305 – 1.922	47.92	58.3%	0.098
Weiss, Corona et al. 2012	1.47	1.161 – 1.869	0.15	0%	0
Weiss 2007	2.36	1.568 – 3.565	0.00	-	-
Weiss 2004	3.58	2.070 – 6.199	0.00	-	-
Weiss 2003	2.64	1.629 – 4.268	0.00	-	-
Villotte, Churchill et al. 2010	1.52	0.646 – 3.565	0.00	-	-
Üstündag 2020	1.16	0.721 – 1.879	9.41	36.2%	0.097
Steen & Lane 1998	2.54	2.275 – 2.835	9.54	0%	0
Salega 2020	0.97	0.796 – 1.170	0.25	0%	0
Rojas-Sepúlveda, Rodríguez Cuenca et al. 2020	1.15	0.898 – 1.476	0.00	-	-
Rojas-Sepúlveda, Rivera-Sandoval et al. 2011	0.96	0.100 – 9.293	0.10	0%	0
Refai 2019	1.25	1.084 – 1.438	26.65	51.2%	0.030
Ponce 2010	1.14	0.811 – 1.614	11.63	65.6%	0.035
Palmer and Waters-Rist 2019	1.51	0.080 – 28.760	15.95	87.5%	1.172
Myszka, Piontek et al. 2020	0.81	0.739 – 0.890	4.28	0%	0
Myszka and Piontek 2012	1.28	1.184 – 1.383	14.08	0%	0
Milella, Giovanna Belcastro et al. 2012	1.01	0.960 – 1.066	97.03	38.2%	0.016
Mazza 2017	1.30	1.184 – 1.416	66.47	0%	0
Mariotti, Facchini et al. 2007	1.27	1.005 – 1.263	91.87	31.4%	0.052
Mariotti, Facchini et al. 2004	1.25	0.803 – 1.940	14.55	38.1%	0.064
Laffranchi, Flórez et al. 2016	1.73	1.564 – 1.916	29.07	0%	<0.0001
Hui, Liu et al. 2020	2.21	1.456 – 3.360	0.03	0%	0
Henderson, Craps et al. 2013	1.31	0.8523 – 2.025	2.60	0%	0
Gresky, Wagner et al. 2016	0.66	0.368 – 1.181	4.51	11.2%	0.019
Giannotti 2020	0.89	0.416 – 1.886	34.55	82.6%	0.550
Cardoso 2008	0.96	0.922 – 1.004	7.38	0%	0
al-Oumaoui, Jiménez-Brobeil et al. 2004	1.48	1.382 – 1.591	64.48	0%	0.002
Acosta, Henderson et al. 2017	1.10	0.906 – 1.324	118.32	91.5%	0.065
Total	1.24	1.20 – 1.29	1132.72	53%	0.072
Prediction interval					
0.73 – 2.11					
Heterogeneity among publications					
Q: 695.87 (p-value <0.0001)					

OR: Odds Ratio; CI: Confidence Interval; Q: Cochran's Q; I^2: Higgin's & Thompson's I^2; τ^2: Tau-squared.

* Within-study estimates from a fixed-effect meta-analysis.

Table 4

Sex effects obtained in the MGLMM analysis. Results where the 95% credible intervals not containing zero in **bold**.

	Feature	Estimate	Est. error	l-95% CI	u-95% CI
Both limbs	BFZ1 SEX	**0.37**	0.13	**0.12**	**0.63**
	ERZ1 SEX	**0.90**	0.28	**0.34**	**1.46**
	TCZ2 SEX	0.55	0.37	-0.18	1.26
	BFZ2 SEX	**0.56**	0.23	**0.11**	**1.02**
	ERZ2 SEX	0.35	0.23	-0.09	0.79
	FPOZ2 SEX	0.06	0.14	-0.21	0.34
	MPOZ2 SEX	0.23	0.29	-0.34	0.80
	CAZ2 SEX	0.17	0.39	-0.61	0.94
Upper limb	BFZ1 SEX	0.28	0.15	-0.01	0.58
	ERZ1 SEX	**1.05**	0.34	**0.38**	**1.69**
	TCZ2 SEX	0.63	0.44	-0.25	1.49
	BFZ2 SEX	0.49	0.30	-0.11	1.09
	ERZ2 SEX	**0.58**	0.29	**0.01**	**1.15**
	FPOZ2 SEX	0.08	0.16	-0.22	0.40
	MPOZ2 SEX	0.72	0.40	-0.07	1.50
	CAZ2 SEX	0.34	0.46	-0.55	1.24
Lower limb	BFZ1 SEX	**0.45**	0.20	**0.05**	**0.85**
	ERZ1 SEX	0.65	0.37	-0.09	1.39
	TCZ2 SEX	0.44	0.35	-0.25	1.13
	BFZ2 SEX	0.49	0.25	0.00	0.98
	ERZ2 SEX	0.18	0.27	-0.34	0.71
	FPOZ2 SEX	-0.03	0.17	-0.36	0.31
	MPOZ2 SEX	-0.08	0.31	-0.69	0.52
	CAZ2 SEX	0.22	0.41	-0.59	1.02

BFZ1: bone formation zone 1; ERZ1: erosion zone 1; TCZ2: textural change zone 2; BFZ2: bone formation zone 2; ERZ2: erosion zone 2; FPOZ2: fine porosity zone 2; MPOZ2: macroporosity zone 2; CAZ2: cavitations zone 2; Est. error: error of the effect size estimate; l-95% CI: lower 95% credible interval; u-95% CI: upper 95% credible interval.

Table 5

EC frequencies for each observed feature in every analyzed enthesis, separated by sex (higher value among sexes in bold).

Enthesis	BFZ1 Female %	BFZ1 Male %	ERZ1 Female %	ERZ1 Male %	TCZ2 Female %	TCZ2 Male %	BFZ2 Female %	BFZ2 Male %	ERZ2 Female %	ERZ2 Male %	FPOZ2 Female %	FPOZ2 Male %	MPOZ2 Female %	MPOZ2 Male %	CAZ2 Female %	CAZ2 Male %
Origin long head *biceps brachii*	**26.0**	17.4	0.0	**4.3**	0.0	0.0	**4.0**	0.0	0.0	**8.7**	14.0	**21.7**	0.0	0.0	0.0	0.0
Origin long head *triceps brachii*	32.6	**50.0**	10.9	**27.3**	0.0	**13.6**	19.6	**27.3**	10.9	**22.7**	**84.8**	77.3	4.3	**4.5**	**2.2**	0.0
Insertion *supraspinatus*	**20.9**	15.4	0.0	**7.7**	0.0	**3.8**	**9.3**	7.7	**20.9**	11.5	**23.3**	19.2	0.0	**3.8**	2.3	**3.8**
Insertion *infraspinatus*	**24.4**	12.0	2.2	**12.0**	0.0	**4.0**	**8.9**	4.0	26.7	**32.0**	20.0	**24.0**	6.7	**8.0**	2.2	**4.0**
Insertion *subscapularis*	64.4	**77.8**	6.7	**14.8**	0.0	**3.7**	15.6	**33.3**	13.3	**18.5**	33.3	**37.0**	2.2	**3.7**	2.2	**7.4**
Insertion *teres minor*	25.6	**29.2**	4.7	**8.3**	4.7	**29.2**	**7.0**	4.2	2.3	**12.5**	**44.2**	41.7	0.0	**4.2**	2.3	**4.2**
Insertion *triceps brachii*	53.3	**60.0**	0.0	**8.0**	8.9	**28.0**	0.0	0.0	11.1	**20.0**	**48.9**	24.0	0.0	**4.0**	0.0	0.0
Insertion *braquialis*	74.4	**77.3**	0.0	0.0	**2.3**	0.0	25.6	**45.5**	11.6	**50.0**	46.5	**59.1**	0.0	**9.1**	0.0	0.0
Insertion *biceps braquii*	68.8	**80.0**	2.1	**25.0**	12.5	**20.0**	10.4	**15.0**	25.0	**40.0**	**37.5**	20.0	0.0	**10.0**	0.0	0.0
Insertion *brachioradialis*	63.4	**76.2**	0.0	**9.5**	**7.3**	4.8	**4.9**	0.0	2.4	**14.3**	58.5	**90.5**	0.0	0.0	0.0	0.0
Origin common extensors	68.3	**80.0**	0.0	**4.0**	0.0	**4.0**	**19.5**	12.0	0.0	**20.0**	**31.7**	28.0	0.0	0.0	0.0	**4.0**
Origin common flexors	57.8	**59.3**	0.0	**7.4**	0.0	0.0	4.4	**11.1**	4.4	**18.5**	28.9	**29.6**	0.0	0.0	0.0	0.0
Ischial tuberosity	56.0	**66.7**	16.0	**22.2**	**26.0**	14.8	40.0	**55.6**	50.0	**59.3**	**36.0**	33.3	**8.0**	3.7	**14.0**	7.4
Insertion *gluteus medius*	55.6	**69.0**	**8.9**	3.4	**8.9**	6.9	**13.3**	6.9	**24.4**	3.4	24.4	**27.6**	2.2	**3.4**	4.4	**6.9**
Insertion *gluteus minimus*	50.0	**67.7**	10.4	**22.6**	18.8	**41.9**	20.8	**38.7**	22.9	**41.9**	**33.3**	32.3	**22.9**	6.5	**10.4**	3.2
Insertion *iliopsoas*	58.3	**81.5**	**8.3**	3.7	**12.5**	7.4	25.0	**48.1**	18.8	**37.0**	47.9	**74.1**	0.0	0.0	**10.4**	3.7
Insertion *obturator externus*	**47.8**	43.3	0.0	0.0	2.2	**10.0**	32.6	**56.7**	6.5	**6.7**	**43.5**	30.0	**8.7**	0.0	**4.3**	3.3
Insertion *cuadriceps femoris* (patella superior)	56.1	**85.2**	4.9	**11.1**	4.9	**22.2**	9.8	**11.1**	14.6	**22.2**	**43.9**	29.6	0.0	0.0	**4.9**	0.0
Insertion patellar tendon (tibial tuberosity)	18.9	**54.2**	0.0	0.0	5.4	**16.7**	0.0	0.0	0.0	**12.5**	**43.2**	33.3	0.0	0.0	0.0	0.0
Origin patellar tendon (patella inferior)	36.6	**60.0**	0.0	0.0	14.6	**28.0**	0.0	**8.0**	**7.3**	4.0	**19.5**	16.0	0.0	0.0	**4.9**	4.0
Origin *gastrocnemius*	**59.6**	54.2	10.6	**16.7**	12.8	**25.0**	17.0	**20.8**	**40.4**	25.0	17.0	**50.0**	10.6	**29.2**	12.8	**25.0**
Insertion *triceps surae*	82.5	**84.0**	2.5	**8.0**	**15.0**	12.0	7.5	**8.0**	**22.5**	12.0	**65.0**	60.0	2.5	**4.0**	**7.5**	4.0
Total	50.1	**59.2**	8.4	**9.7**	11.5	**13.7**	13.8	**19.6**	15.7	**22.3**	38.1	**38.3**	3.3	**4.1**	**4.0**	3.8

BFZ1: bone formation zone 1; ERZ1: erosion zone 1; TCZ2: textural change zone 2; BFZ2: bone formation zone 2; ERZ2: erosion zone 2; FPOZ2: fine porosity zone 2; MPOZ2: macroporosity zone 2; CAZ2: cavitations zone 2.