Assessing individual and population variability in degenerative joint disease prevalence using generalized linear mixed models

Abbreviated title: On the use of GLMM to analyze DJD.

Alonso-Llamazares, Carmen[1]; Blanco Márquez, Beatriz[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

Keywords:

Degenerative joint disease, generalized linear mixed models, random effects, archaeological population, Spain.

Corresponding authors:

Antonio F. Pardiñas
Telephone number: +44 (0)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/ Catedratico Valentín Andrés Álvarez s/n, Asturias 33006, Spain.
E-mail address: lopezbelen@uniovi.es

ABSTRACT

Objectives: In this paper we introduce the use of generalized linear mixed models (GLMM) as a better alternative to traditional statistical methods for studying factors associated to the prevalence of degenerative joint disease (DJD) in bioarchaeological contexts.

Materials and Methods: DJD prevalence was assessed for the appendicular joints and the spine of a Spanish population dated from the 15th to the 18th century. Data was analyzed using contingency tables, logistic regression models, and logistic GLMM.

Results: In general, results from GLMMs find agreement in other methods. However, by being able to analyze the data at the level of individual bones instead of aggregated joints or limbs, GLMMs are capable of revealing associations that are not evident in other frameworks.

Discussion: Currently widely available in statistical analysis software, GLMMs can accommodate a wide array of data distributions, account for hierarchical correlations and return estimates of DJD prevalence within individuals and skeletal locations that are unbiased by the effect of covariates. This gives clear advantages for the analysis of bioarchaeological datasets which can lead to more robust and comparable analyses across populations.

Degenerative joint disease (DJD) is, along with trauma and infection, the most frequently observed skeletal lesion in paleopathology, with osteoarthritis being the most common form of joint disease (Ortner, 2003; Plomp & Boylston, 2016). Even though the term DJD is commonly used as synonym of osteoarthritis (Jurmain & Kilgore, 1995; Lieverse, Weber, Bazaliiskiy, Goriunova, & Savel'ev, 2007; Weiss, 2018), the latter is defined as a disease of synovial joints while DJD encompasses other degenerative changes such as those manifested at the vertebral bodies (Jurmain, 2000; Waldron, 2019). These are chronic and slowly progressive alterations of the joint articulation that can cause severe pain, as well as physical disability in its advanced stages (Eng, 2016; Ortner, 2003). Their main cause is the breakdown of protective articular cartilage, affecting the subchondral bone (Lieverse, Mack, Bazaliiskii, & Weber, 2016; Shin et al., 2016; Suzuki et al., 2016). DJD is still a common ailment in present times and its rates are expected to increase in coming years due to population ageing and rising levels of obesity in high-income countries (Gustafsson, Kvist, Eriksson, Dahlberg, & Rolfson, 2020). However, since its prevalence varies among populations, as well as among articulations, accurate global frequencies are difficult to establish. Usually, the joints with higher DJD rates are hip, knees and hands (Barbour et al., 2018; Gupta & Gupta, 2018; Peña Ayala & Fernández-López, 2007; Turkiewicz et al., 2014). In Spain, for instance, a recent clinical study of adults above 20 years of age shows frequencies of 13.9% for the knee and 15.5% for lumbar segment of the spine (Seoane-Mato, Sánchez-Piedra, Díaz-González, & Bustabad, 2018). It is important to note that these current prevalence estimates cannot be directly extrapolated to osteological contexts, as they are based on radiographic evidence and clinical symptoms reported by living patients, such as the sensation of joint pain, for which the underlying causes might not be visible in dry bones.

DJD has a multifactorial etiology with mechanical factors playing a major role. It is known that is affected by age, sex, ancestry, nutrition, body mass, genetics, trauma, mechanical stress and physical activity (Knüsel, Göggel, & Lucy, 1997; Lieverse et al., 2016; Shimoda et al., 2012), with age being the factor with a stronger relationship to DJD. Clinical studies in living populations show that more than 50% of individuals older than 60 years have radiological signs of the pathology (Guilak, 2011; Waldron, 2019). Moreover, autopsy reports indicate that almost all people over 65 years of age present evidence of degenerative changes in the articular cartilage (Waldron, 2019), although these diagnoses cannot be made in dry bone. Despite these other factors involved in the development of this disorder, physical activity and mechanical stress

associated to movement seem to be essential conditions; if a joint does not move, it does not develop this disease (Becker, 2020; Waldron, 2009).

For that reason, the distribution of joint disease within ancient populations can be used to make inferences about the intensity of physical demand. This can then be used to reconstruct lifestyles, such as the existence of division of labor or diachronic variations in activity patterns (Cheverko & Bartelink, 2017; Eng, 2016). While the DJD patterns of an individual are not informative enough to determine specific physical activities, they can, with the proper archaeological contextualization, reveal data about the relative intensities of functional stress within a population and through time (Cheverko & Bartelink, 2017). Additionally, due to its debilitating nature, the study of DJD can bring insights about the quality of life, considering that severe states may have a major impact on the mobility of individuals, preventing them to perform their daily activities and contribute to their communities (Domett, Evans, Chang, Tayles, & Newton, 2017).

When working with archaeological samples, it is important to consider the unavoidable bias that occurs in data collection due to the fragmentary, and often commingled, state in which human remains are generally recovered at archaeological sites (Knüsel & Outram, 2004). Despite this, it is widely agreed that skeletal remains can and do provide valuable information for understanding ancient diseases (Bertsatos & Chovalopoulou, 2019; Klaus, 2014), but their limitations have to be addressed at the level of data recording methods and through the introduction of robust approaches to data analysis (Baker & Pearson, 2006; Bertsatos & Chovalopoulou, 2019).

Statistical methodologies are thus a core part of the assessment of DJD in past populations. Traditionally, a plethora of tests have been applied to study the distribution of this and other pathologies in the bioarchaeological literature (Cheverko & Hubbe, 2017), with the most common being contingency table tests (Chi-square and Fisher's exact) and analyses of variance (ANOVA and ANCOVA). These tests are often used to answer broad questions, such as the existence of differences in DJD prevalence between sexes and age categories (Shimoda et al., 2012; Suzuki et al., 2016; Woo & Pak, 2013), and for this purpose their performance is equivalent (Cheverko & Hubbe, 2017). However, given that analyses of variance are derived from the generalized linear model (GLM) of regression (Gelman, 2005; Ziglari, 2017), the lack of adoption of the standard form of the latter in routine paleopathological analyses seems noteworthy. Previous research has shown that, despite a concern that GLM analyses might be more complex to conduct or harder to interpret (Kesleman, Othman, & Wilcox, 2016; Mai & Zhang, 2017), their results converge with those of traditional tests in common bioarchaeological study designs (Nikita, 2014; Nikita, Mattingly, & Lahr, 2013). Also, their flexibility for dealing simultaneously with multiple predictors allows for a better assessment of the inter-individual variability of ancient populations, which is essential for an accurate comparison of results between studies (Jurmain, Cardoso, Henderson, & Villotte, 2012).

In this paper, we study the frequencies of DJD in a medieval and modern population from Burgos (Spain), using the traditional and GLM approaches summarized above. Aiming to consider as much of the available information as possible, we also showcase the use of a flexible statistical framework called the “generalized linear mixed model” (GLMM) or “multilevel regression” (McCulloch & Neuhaus, 2005). A GLMM allows the analysis of correlated data, avoiding the inflation of test statistics caused by not meeting the independence assumption of the GLM (Forstmeier, Wagenmakers, & Parker, 2017). Paleopathological studies might not seem to fit the traditional “repeated measures” design where this issue is well-known, but as each single bone is usually considered a sample (Stodder, 2012), at least two levels of non-independence are possible: One from the individuals, as genetic and other risk factors might make some people prone or refractory to skeletal disease (Sandell, 2012); and another from the anatomical locations, as the presence of disease markers might also vary between joints or limbs (Turkiewicz et al., 2014). Ignoring or inappropriately accounting for this hierarchical structure can result in confounding

("pseudoreplication"; Arnqvist, 2020), which many studies avoid by collapsing data into single measures per individual or anatomical location (e.g. Henderson & Nikita, 2016). However, alternatively recognizing and integrating these sources of variability into a GLMM leads to increased statistical power and a more complete assessment of the dataset (Harrison et al., 2018). Our comparison of GLMMs with other approaches acknowledges that they are rarely in the statistical toolkit of the bioarchaeologist, but the advantages they bring have been well-established through their use in other fields such as social, ecological, clinical and genetic research (Bell & Jones, 2015; Bolker et al., 2009; White & Barnett, 2019; Ziyatdinov et al., 2018).

MATERIAL AND METHODS

Skeletal sample

The population used in this study comes from the church of San Nicolás de Bari (SNB), a medieval and 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 extant building housing the necropolis was built in 1408, replacing another Romanesque temple that already appears in records from the 12th century (Florez, 1771-72).

The skeletal remains were recovered in 3 different excavation phases, between 2007 and 2008, and correspond to different periods between the 15th and 18th centuries. A total of 78 adult SNB individuals were analyzed, 49 females and 29 males. A cut-off point of 45 years was used to establish relatively balanced age groups within each sex (Table 1), though for 12 adult individuals age determination was not possible.

Historical context

The Middle Ages was a period of intense urbanization both in Spain and in the rest of Europe. The city of Burgos was founded in the 9th century and by the 15th century it had already reached 10,000 inhabitants (Sebastián Moreno, 2017). It was one of the wealthiest cities in the Kingdom of Castile, with a profitable market and extensive production of artisan goods, and it was located in the commercial axis of the kingdom. As such, Burgos was a diverse city with a high number of members of the nobility, clergy, and urban elites (Goicolea Julián, 2019; Sebastián Moreno, 2017; Sebastián Moreno & Guerrero Navarrete, 2018).

Dietary habits in Burgos were based mainly on cereals such as barley and wheat, grown on land close to the city (Sebastián Moreno & Guerrero Navarrete, 2018). Meat, fish, and wine were also easily accessible by almost all citizens (Sebastián Moreno, 2017). Commonly traded products included cloth, leather, wood, clay, coal, stone and metal; as required by thriving guilds of specialized artisans and builders (Sebastián Moreno, 2017). These manufacturers were generally found near the city centre forming working communities, and their work was in high demand due to the intensity of the economic activity and the fact that the city was usually under construction and renovation (Sebastián Moreno & Guerrero Navarrete, 2018). Throughout the Middle Ages, it is considered that 40% of the population of Burgos was broadly dedicated to craftsmanship, with a similarly large proportion engaged in the local and regional trade (Goicolea Julián, 2019; Sebastián Moreno, 2017).

Osteological methodology

For the determination of sex and age, methods commonly used in physical anthropology were applied. Sex was established primarily through the morphological features of the skull and pelvis (Buikstra & Ubelaker, 1994). When these bone elements were not in good condition, discriminant functions generated from Spanish populations were used (López-Bueis, Robledo, Roselló, &

Trancho, 1996; Trancho, Robledo, López-Bueis, & Sánchez, 1997; Trancho, Robledo, & Sánchez, 2012). For determining age-at-death, the pubic symphysis was assessed when possible. When this was not possible, other methods were applied based on the morphology of the auricular surface, the closure of cranial sutures, dental wear or the ossification degree of the thyroid cartilage (Buikstra & Ubelaker, 1994; Krenzer, 2006). Non-adult individuals, as well as those with pathological features that could bias the DJD analysis were excluded from the study.

Throughout this paper we use the term DJD for all the changes we evaluated, since some of them may not qualify as osteoarthritis. Different criteria were used to evaluate DJD at the spine and at the appendicular joints. For the diagnosis of DJD on the spine, the method used was based on Shimoda et al. (2012), for both the vertebral body and the apophyseal joints, modified to include a more objective value to porosity: Grade 0 is the normal condition, without evidence of degenerative changes; Grade 1 are slight degenerative changes present, with small, horizontal growth osteophytes and porosity absent or covering less than 10% of the joint surface; Grade 2 are moderate lesions with osteophytes that can grow around the rim and porosity covering 10-50% of the joint surface; Grade 3 are severe lesions with significant growth of the osteophytes in a vertical direction and porosity covering more than 50% of the joint surface; and Grade 4 implies bridging between adjacent vertebrae or joint fusion (Figure 1).

To evaluate the degenerative changes on the appendicular joints, the method used was based on Steckel, Larsen, Sciulli, and Walker (2005): Grade 0 is the typical condition with absence of degenerative changes; Grade 1 is a slight marginal lipping with osteophytes less than 3mm and slight degenerative changes on the articular surface; Grade 2 is a severe marginal lipping with osteophytes greater than 3mm and severe degenerative changes on the articular surface, with possible eburnated areas; Grade 3 is a complete or nearly complete destruction of the articular surface; finally, Grade 4 implies joint fusion (Figure 2).

Table 2 shows the articular surfaces where the presence of degenerative changes was evaluated, as well as the joint and limb complexes of which they are part. Data were collected on joints that preserved at least 50% of the articular surface. The presence of any bone with an analyzable joint surface of a compound joint was sufficient to evaluate DJD, thus enabling our analysis of joints without all the elements present in incomplete skeletons. The severity of the changes reported in each joint system corresponds to the maximum severity of the joint surfaces that form it (Supplementary Table 1). As the analysis of the discrete categories of severity scores is limited by the resolution and reliability of macroscopic evaluations (Plomp, Roberts, & Strand Viðarsdóttir, 2015), the work presented here solely concerns the prevalence of DJD at each skeletal location, assessed as the observation of a severity score greater than zero.

Statistical methodology

Fisher's exact test for contingency tables was applied to our data as recommended by Cheverko & Hubbe (2017). For GLM regression, which includes ANOVA/ANCOVA as particular cases, a mathematical description and rationale for its use in osteological assessments can be found in Henderson & Nikita (2016). Briefly, a GLM extends ordinary linear regression to the analysis of outcome variables conforming to distributions other than the normal (McCullagh & Nelder, 1989). This requires specifying a relationship between the predictors and the expectation of the response which will usually be a member of the exponential family of distributions (Hedeker, 2005; Rabe-Hesketh, Skrondal, & Pickles, 2002). Common choices include the Poisson distribution for event counts and the Binomial distribution for presence/absence data, commonly called logistic regression (Imrey, 2000). We chose the latter to model DJD prevalence in our data.

The effects of predictors in traditional tests and GLM models are estimated as constant across all the samples in the analysis, which grants them the term "fixed". GLMMs introduce "random"

predictors as those with effects that differ across groups of samples, though a number of definitions exist (Gelman, 2005). A standard GLMM then takes the form:

$$y = X\beta + Z\gamma + e$$

Where y is the outcome variable; X is a matrix of predictor variables with a corresponding matrix β of fixed effects; Z is a matrix of grouping variables with a corresponding matrix γ of random effects; and e is an error ("residual") term of unexplained variability (UCLA: Statistical Consulting Group, 2012). Fixed effects are the regression coefficients of each predictor conditional on the random effects, which are estimates of the mean outcome within each group. The usual calculation of random effects, called "partial pooling", considers the variance of the entire dataset in order to reduce the influence of outlier observations (Gelman & Hill, 2007). Thus, particularly in small datasets with a large relative number of groups, random effect estimates are less prone to bias than a series of fixed-effect predictors (Clark & Linzer, 2015).

Statistical testing

We used R v3.53 (R Core Team, 2019) to perform all analyses. Fisher's exact and GLM logistic regression were carried out with functions implemented in the base *stats* package, and the packages *lme4* and *lmerTest* were used to perform GLMM logistic analyses (Bates, Maechler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2017). Our main research question was to study the relationship between DJD prevalence and sex, age group, and bone laterality across joints and limbs. In all analyses these three variables were included as fixed effect predictors. The outcome prevalence variable was dichotomized from the DJD severity metric, by assigning a value of one to every bone graded 1-4 and zero otherwise. To preserve sample independence for Fisher's exact and GLM tests, data was aggregated into joints and limbs by considering that DJD was present when it was observed in at least one of their constituting bones. Equivalent GLMMs were run directly on the prevalence data from the bones, with a random effect group including every individual in the SNB population. We also ran GLMMs for the entire SNB sample, which included an additional random effect group for joints.

While effect sizes and confidence intervals should be the gold standard for evaluating association test results (Smith, 2018), our study is built on the analysis of multiple anatomical locations in the same individuals. As it is well known, serial testing of null hypotheses ("multiplicity") leads to the inflation of the false positive rate, though the magnitude of this problem is lessened if tests results are correlated (Westfall & Young, 1989). As this could be argued for skeletal markers which share common aetiologies (Jurmain et al., 2012), we corrected our p-values for multiplicity using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995), which can accommodate partially correlated test results. FDR correction was applied across all results from the Fisher's exact test and within covariates in the regression models, though we also report all uncorrected p-values following recent methodological recommendations (Smith, 2018; Streiner, 2015).

Including more fixed or random variables in a statistical model does not necessarily imply that it will explain more accurately the observed data. Coefficient of determination statistics (R^2) allow to estimate the variance explained by a regression model, and equivalent definitions exist for GLMs and GLMMs (Nakagawa & Schielzeth, 2013). To quantitatively assess the performance of our models, we generated R^2 values for our GLMs using the method by Nagelkerke (1991). For GLMMs, we generated marginal R^2 values to assess the variance explained by the fixed effects, and conditional intra-class correlation coefficients (ICCs) to assess the variance explained by the random effects (Nakagawa, Johnson, & Schielzeth, 2017).

RESULTS

Table 3 shows the empirical prevalence of DJD in every analyzed joint. Average prevalence of DJD in the SNB population was 82.3%, ranging from 68.0% in the wrist to 93.0% in the thoracic vertebrae. For more detailed information, Table 4 shows the prevalence at each joint stratified by sexes and age groups.

We have compared the distribution of DJD prevalence by sex, age group and laterality with different statistical tests. Table 5 shows the odds ratios (ORs) for sex and age group obtained with Fisher's exact test, GLM logistic regression and GLMM logistic regression for each joint. A longer version of this table with laterality results and FDR corrected p-values is shown in Supplementary Table 2. All joints that display nominally significant differences between age groups with the Fisher' exact test show consistent and stronger results with GLM regression analysis. However, the GLMM analysis indicates differences between age groups existing in other skeletal locations but not for the elbow, and highlights differences between sexes that are not apparent with the other methods. Most of these differences remain significant after FDR correction. Laterality analyses showed no significant results with any method (Supplementary Table 2).

Another important piece of information the GLMM analysis provides is a measure of the variability in the response variable that is due to the random factors (Winter, 2013), which in our case are the individuals. This variance is shown in Table 6. Note that, for most skeletal locations, close to 50% of the variability in our data can only be explained as inter-individual even after accounting for the fixed effects. Though the GLMs were based on data aggregated within joints and limbs, and thus their absolute baseline variance is different, we have also noted fixed effects explained similar amounts of variability in both GLM and GLMM analyses. This shows no obvious advantage to aggregating data in our study, at least in terms of model fitting performance, and supports our preference of GLMMs as a more comprehensive analytic framework.

Finally, we fitted a GLMM for our entire population using all the available bone data, sex and age as fixed effects, and random effects for the individuals and skeletal locations (joints). Fixed effects explained 9.9% of the variance, with the random effects explaining 51.5%. Thus, in our data, DJD frequencies are determined more by their skeletal location and the idiosyncrasies of each individual (the activities it performed in life and other unobserved variables), than by sex and age, though the influence of the latter is not negligible. From this model, we extracted the random effect coefficients, sometimes called "best linear unbiased predictors" (BLUPs; Houslay & Wilson, 2017). These account for sex, age and missing data, and yield a probability value for the presence of DJD in each skeletal location and individual (Figure 3). This is a more nuanced reflection of the data than crude frequency estimates (Table 3), and allows for other populations to be straightforwardly compared with our sample via the use of similar modelling approaches.

DISCUSSION

Frequency and patterns of disease markers in San Nicolas de Bari

The frequencies of DJD observed in the SNB population are quite high when compared with other archaeological populations that have been assessed using the same scoring system as this study. For the appendicular skeleton, Austin (2017) analyzed osteoarthritis on a laborer population from Deir el-Medina (Egypt), obtaining an overall prevalence of 34% for the hip, 28% for the knee and 22% for the ankle. Eng (2016) studied pastoral populations of several sites and periods from Northern China and Mongolia, obtaining higher DJD frequencies for female knees (6.7%) and male vertebrae (11.1%) during the Bronze Age and male elbows (13.0%) and female vertebrae

(18.8%) during the Iron Age. Regarding the spine, observations of DJD were also more frequent than in other medieval Spanish populations, both from rural contexts, which reported rates between 30%-13% in males and 19%-7% in females (Jiménez-Brobeil, Roca-Rodríguez, Al Oumaoui, & Du Souich, 2012). In this regard, the SNB population resembles the study from a Korean Joseon necropolis (Kim, Kim, Kim, Oh, & Shin, 2012), in which rates of DJDs were 82.1% for the cervical vertebrae, 94.7% for the thoracic vertebrae, and 90.7% for the lumbar vertebrae. This study also showed increased DJD prevalence for older male individuals, which according to the authors was likely related to Joseon labourers using traditional instruments which imposed heavy loads on the spine (Kim et al., 2012).

The shoulders of the SNB population not only show the highest rate of DJD presence of the appendicular skeleton, but also a significant amount of severe DJD (10.3% of grade 3, Supplementary Table 1). This is consistent with a population of artisans undertaking construction and craftwork since these occupations imply the use of the upper limb with repetitive and physically demanding tasks. Degenerative changes of that degree must have caused pain and discomfort for most individuals, who probably experienced reduced mobility. More severe is the degeneration observed in the spine, with a range of 15.8%-27.5% of grade 3 changes and 10.5%-12.5% of fused vertebrae, which would have caused complete mobility impairment.

The significant differences found with the GLMM analysis evidence the relationship between DJD and age for the upper limb, a well-known association shown in many populations worldwide (Cheverko & Bartelink, 2017; Eng, 2016; Shin et al., 2016), and expected given the cumulative effects of age on degenerative diseases. However, the only association found in the lower limb is between DJD and sex, with pathology being more frequent in males. Austin (2017) also found this relationship and ascribed it to commuting, with males working further from home than females. A similar situation could have occurred in this population with men walking more frequently, as raw materials and industry would have been outside the city. However, we currently do not have enough information on the lifestyle of our population to corroborate this assertion. Lastly, both predictors in the spine, sex and age, seemed to have an impact on DJD, which has also been suggested by other studies (Kim et al., 2012; Shimoda et al., 2012).

Statistical assessment of DJD prevalence

Applying the different statistical analyses on our data gave broadly convergent results, with some differences. Variation in effect sizes was expected, as ORs from the Fisher's exact test are a direct expression of the raw ratio of DJD prevalence between the two categories within each predictor, while ORs from GLMs and GLMMs are conditional on all other terms within each model. This influences their interpretation but, in our analyses, most of the results of each method were within the 95% CI of other methods (Table 5), supporting previous studies which discussed the consistency of contingency table and regression approaches in bioarchaeology (Cheverko & Hubbe, 2017; Nikita et al., 2013). However, we also observed that GLMMs resulted in larger numbers of statistically significant associations, even after multiple testing correction. These highlighted a general influence of older age and male sex in DJD development in our population across different joints and limbs, particularly noteworthy in the spine and vertebrae.

These results are likely a consequence of the greater sample size of GLMMs, by the fact that these models integrate the information of each bone into a single analysis. We point out the importance of this capability by graphically representing the data assessed by the GLM and GLMM analyses (Figure 4). This shows clearly that the individuals we analyzed do not have the same preservation status, i.e., they have different amounts of missing bones, a recurrent limitation of archaeological populations. The common approach of aggregating DJD prevalence into joints across a sample of individuals fills-in this missingness, altering the population data distribution, which can be simply assessed by comparing the amount of dark grey between the raw and aggregated panels in Figure

4. Under certain scenarios, data aggregation can also lead to model fitting concerns, such as quasi-complete separation (Table 5), which in logistic regression occurs when the variability in the outcome is so low that it can be perfectly predicted by its covariates (Albert & Anderson, 1984). In naturally complex biological data, such unexpected accuracy often indicates sampling or statistical issues, rather than true results. While in large and moderately complete samples the aforementioned problems can be ameliorated by adopting location scoring schemes which account for absent bones (e.g. Jurmain, 1991), GLMMs can be used without aggregating the data at all. For this reason, in our view they provide a tried-and-tested way forward to building more complete statistical models from paleopathological data, which has been argued before as a requirement for the field to transition from a purely observational to an epidemiological/causal framework (Jurmain et al., 2012).

It must be noted that GLMMs are not the only statistical approach that can easily accommodate correlated data. Some researchers have previously proposed using generalized estimating equations (GEEs) to address the problem of non-independence between bioarchaeological samples (Nikita, 2014; Villotte et al., 2010). GEEs are based on specifying the relationship and correlation structure between predictors, and can also handle non-normal distributions and estimate the effect of several covariates simultaneously (Halekoh, Højsgaard, & Yan, 2006; Pekár & Brabec, 2018). However, their "population average" effects do not allow to infer individual-specific parameter estimates as GLMMs do (Figure 3), and are not equivalent to random effects in the case of logistic models (Pekár & Brabec, 2018). While GEEs can be a suitable method when only fixed effects are of relevance, computational developments in statistical software over the last decade have made the use of GLMMs straightforward and user-friendly, which makes them preferred for arbitrarily complex studies (Harrison et al., 2018; Muff, Held, & Keller, 2016).

Finally, we acknowledge that modelling the prevalence of DJD is a simplified scenario that only allows us to make broad inferences about the influence of risk factors in the appearance of these osteopathological changes. This is independent of the statistical approach used, and reflects limitations in the experimental design and material available for most studies, which bounds the questions that can be meaningfully answered (Zampetti, Mariotti, Radi, & Belcastro, 2016). A fuller evaluation would require the use of severity scores, given that some risk factors have been previously related to degenerative disease onset and not progression or vice-versa (Vina & Kwoh, 2018). However, such an analysis is complex, ideally requiring specific modelling approaches for discrete data (Milanzi, Alonso, & Molenberghs, 2012) and ways of accounting for the uncertainties in the estimation of these scores (Hillson, 2005; Plomp & Boylston, 2016; Plomp et al., 2015). While those issues are outside of the scope of this manuscript, they are also promising research avenues for the field.

CONCLUSION

Robust statistical methods are needed to properly estimate the effect of the different variables that impact osteopathological processes. While identifying these can be complex in ancient populations, DJD is known to be affected by factors that can usually be inferred in an anthropological study, such as sex and age. These characteristics do not act independently from each other, or sporadically throughout the skeleton. While different methods have been proposed to analyze their relationship to the prevalence patterns of DJD, regression approaches are the most flexible. Among them, GLMMs are noteworthy for being able to analyze large amounts of paleopathological assessments without the need of aggregating data. This study shows that such a feature can lead to revealing potentially interesting associations that are missed by other methods. Also, the use of complementary approaches such as variance-partitioning or the estimation of random effect coefficients can provide new ways of quantitatively comparing the

results of paleopathological studies. For this reason, we believe they are a novel and powerful incorporation to the statistical framework used in physical anthropology, with the potential of superseding currently standard testing procedures that do not accurately reflect the complexity and hierarchical structure of bioarcheological data.

DATA AVAILABILITY STATEMENT

Full syntaxis of each model and code for reproducing the analyses is available at [Github URL; available in the final published version]. Raw DJD prevalence data from the SNB population is also available at [Figshare URL; available in the final published version].

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). Antonio F. Pardiñas acknowledges support from a Medical Research Council Project Grant (MC_PC_17212).

LITERATURE CITED

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika, 71*(1), 1-10.

Arnqvist, G. (2020). Mixed Models Offer No Freedom from Degrees of Freedom. *Trends in Ecology & Evolution, 35*(4), 329-335. doi:https://doi.org/10.1016/j.tree.2019.12.004

Austin, A. E. (2017). The Cost of a Commute: A Multidisciplinary Approach to Osteoarthritis in New Kingdom Egypt. *International Journal of Osteoarchaeology, 27*(4), 537-550. doi:10.1002/oa.2575

Baker, J., & Pearson, O. M. (2006). Statistical methods for bioarchaeology: applications of age-adjustment and logistic regression to comparisons of skeletal populations with differing age-structures. *Journal of Archaeological Science, 33*(2), 218-226. doi:https://doi.org/10.1016/j.jas.2005.07.019

Barbour, K. E., Moss, S., Croft, J. B., Helmick, C. G., Theis, K. A., Brady, T. J., . . . Lu, H. (2018). Geographic variations in arthritis prevalence, health-related characteristics, and management—United States, 2015. *MMWR Surveillance Summaries, 67*(4), 1.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1)*, 1-48. doi:doi:10.18637/jss.v067.i01

Becker, S. K. (2020). Osteoarthritis, entheses, and long bone cross-sectional geometry in the Andes: Usage, history, and future directions. *International Journal of Paleopathology, 29*, 45-53. doi:https://doi.org/10.1016/j.ijpp.2019.08.005

Bell, A., & Jones, K. (2015). Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods, 3*(1), 133-153. doi:10.1017/psrm.2014.7

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Bertsatos, A., & Chovalopoulou, M.-E. (2019). Validation study of osteometric techniques for sorting commingled human skeletal remains in archaeological samples. *International Journal of Osteoarchaeology, 29*(2), 253-259. doi:10.1002/oa.2733

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*(3), 127-135. doi:https://doi.org/10.1016/j.tree.2008.10.008

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.

Cheverko, C. M., & Bartelink, E. J. (2017). Resource intensification and osteoarthritis patterns: changes in activity in the prehistoric Sacramento-San Joaquin Delta region. *American journal of physical anthropology*.

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

Clark, T. S., & Linzer, D. A. (2015). Should I use fixed or random effects? *Political Science Research and Methods, 3*(2), 399-408.

Domett, K., Evans, C., Chang, N., Tayles, N., & Newton, J. (2017). Interpreting osteoarthritis in bioarchaeology: Highlighting the importance of a clinical approach through case studies from prehistoric Thailand. *Journal of Archaeological Science: Reports, 11*, 762-773.

Eng, J. T. (2016). A bioarchaeological study of osteoarthritis among populations of northern China and Mongolia during the Bronze Age to Iron Age transition to nomadic pastoralism. *Quaternary International, 405*, 172-185.

Florez, P. E. (1771-72). España Sagrada. In: Ed. facsímil de Burgos: Ayuntmiento, 1983.

Forstmeier, W., Wagenmakers, E.-J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings – a practical guide. *Biological Reviews, 92*(4), 1941-1968. doi:10.1111/brv.12315

Gelman, A. (2005). Analysis of variance--why it is more important than ever. *Ann. Statist., 33*(1), 1-53. doi:10.1214/009053604000001048

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge university press.

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.

Guilak, F. (2011). Biomechanical factors in osteoarthritis. *Best Practice & Research Clinical Rheumatology, 25*(6), 815-823. doi:https://doi.org/10.1016/j.berh.2011.11.013

Gupta, B., & Gupta, R. (2018). Knee Osteoarthritis: A Scientometric Assessment of Global Publications Output during 2008-17. *EC Orthopaedics, 9*, 275-284.

Gustafsson, K., Kvist, J., Eriksson, M., Dahlberg, L. E., & Rolfson, O. (2020). Socioeconomic status of patients in a Swedish national self-management program for osteoarthritis compared with the general population—a descriptive observational study. *BMC Musculoskeletal Disorders, 21*(1), 10. doi:10.1186/s12891-019-3016-z

Halekoh, U., Højsgaard, S., & Yan, J. (2006). The R package geepack for generalized estimating equations. *Journal of Statistical Software, 15*(2), 1-11.

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

Hedeker, D. (2005). Generalized Linear Mixed Models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Vol. 2, pp. 729–738). Chichester: John Wiley & Sons, Ltd.

Henderson, C. Y., & Nikita, E. (2016). Accounting for multiple effects and the problem of small sample sizes in osteology: a case study focussing on entheseal changes. *Archaeological and Anthropological Sciences, 8*(4), 805-817. doi:10.1007/s12520-015-0256-1

Hillson, S. (2005). *Teeth*: Cambridge University Press.

Houslay, T. M., & Wilson, A. J. (2017). Avoiding the misuse of BLUP in behavioural ecology. *Behavioral Ecology, 28*(4), 948-952.

Imrey, P. B. (2000). Poisson Regression, Logistic Regression, and Loglinear Models for Random Counts. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 391-437). San Diego: Academic Press.

Jiménez-Brobeil, S., Roca-Rodríguez, M., Al Oumaoui, I., & Du Souich, P. (2012). Vertebral pathologies and related activity patterns in two mediaeval populations from Spain. *Collegium antropologicum, 36*(3), 1019-1025.

Jurmain, R. D. (1991). Degenerative changes in peripheral joints as indicators of mechanical stress: Opportunities and limitations. *International Journal of Osteoarchaeology, 1*(3-4), 247-252. doi:10.1002/oa.1390010319

Jurmain, R. D. (2000). Degenerative joint disease in African great apes: an evolutionary perspective. *Journal of Human Evolution, 39*(2), 185-203.

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., & Kilgore, L. (1995). Skeletal evidence of osteoarthritis: a palaeopathological perspective. *Annals of the Rheumatic Diseases, 54*(6), 443.

Kesleman, H. J., Othman, A. R., & Wilcox, R. R. (2016). Generalized linear model analyses for treatment group equality when data are non-normal. *Journal of Modern Applied Statistical Methods, 15*(1), 4.

Kim, D. K., Kim, M. J., Kim, Y.-S., Oh, C. S., & Shin, D. H. (2012). Vertebral osteophyte of pre-modern Korean skeletons from Joseon tombs. *Anatomy & cell biology, 45*(4), 274-281.

Klaus, H. D. (2014). Frontiers in the bioarchaeology of stress and disease: Cross-disciplinary perspectives from pathophysiology, human biology, and epidemiology. *American journal of physical anthropology, 155*(2), 294-308. doi:10.1002/ajpa.22574

Knüsel, C. J., Göggel, S., & Lucy, D. (1997). Comparative degenerative joint disease of the vertebral column in the medieval monastic cemetery of the Gilbertine Priory of St. Andrew, Fishergate, York, England. *American Journal of Physical Anthropology, 103*, 481-495.

Knüsel, C. J., & Outram, A. K. (2004). Fragmentation: The Zonation Method Applied to Fragmented Human Remains from Archaeological and Forensic Contexts. *Environmental Archaeology, 9*(1), 85-98. doi:10.1179/env.2004.9.1.85

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *2017, 82*(13), 26. doi:10.18637/jss.v082.i13

Lieverse, A. R., Mack, B., Bazaliiskii, V. I., & Weber, A. W. (2016). Revisiting osteoarthritis in the Cis-Baikal: Understanding behavioral variability and adaptation among middle Holocene foragers. *Quaternary International, 405*, 160-171.

Lieverse, A. R., Weber, A. W., Bazaliiskiy, V. I., Goriunova, O. I., & Savel'ev, N. A. (2007). Osteoarthritis in Siberia's Cis-Baikal: Skeletal indicators of hunter-gatherer adaptation and cultural change. *American Journal of Physical Anthropology, 132*(1), 1-16. doi:10.1002/ajpa.20479

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

Mai, Y., & Zhang, Z. (2017). Statistical Power Analysis for Comparing Means with Binary or Count Data Based on Analogous ANOVA. In (pp. 381-393). Asheville, North Carolina, USA: Springer International Publishing.

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

McCulloch, C. E., & Neuhaus, J. M. (2005). Generalized Linear Mixed Models. In P. Armitage & T. Colton (Eds.), *Encyclopedia of Biostatistics* (Vol. 2). Chichester: Wiley-Blackwell.

Milanzi, E., Alonso, A., & Molenberghs, G. (2012). Ignoring overdispersion in hierarchical loglinear models: Possible problems and solutions. *Statistics in Medicine, 31*(14), 1475-1482. doi:10.1002/sim.4482

Muff, S., Held, L., & Keller, L. F. (2016). Marginal or conditional regression models for correlated non-normal data? *Methods in ecology and evolution, 7*(12), 1514-1524.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691-692.

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface, 14*(134), 20170213. doi:doi:10.1098/rsif.2017.0213

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution, 4*(2), 133-142.

Nikita, E. (2014). The use of generalized linear models and generalized estimating equations in bioarchaeological studies. *American journal of physical anthropology, 153*(3), 473-483. doi:10.1002/ajpa.22448

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.

Ortner, D. J. (2003). Osteoarthritis and diffuse idiopathic skeletal hyperostosis. In *Identification of pathological conditions in human skeletal remains* (pp. 545-560): Academic Press.

Pekár, S., & Brabec, M. (2018). Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. *Ethology, 124*(2), 86-93. doi:10.1111/eth.12713

Peña Ayala, A. H., & Fernández-López, J. C. (2007). Prevalencia y factores de riesgo de la osteoartritis. *Reumatología Clínica, 3*, 6-12.

Plomp, K. A., & Boylston, A. (2016). Frequency and patterns of costovertebral osteoarthritis in two Medieval English populations. *International Journal of Paleopathology, 14*, 64-68.

Plomp, K. A., Roberts, C. A., & Strand Viðarsdóttir, U. (2015). Morphological Characteristics of Healthy and Osteoarthritic Joint Surfaces in Archaeological Skeletons. *International Journal of Osteoarchaeology, 25*(4), 515-527. doi:10.1002/oa.2319

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/.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal, 2*(1), 1-21.

Sandell, L. J. (2012). Etiology of osteoarthritis: genetics and synovial joint development. *Nature Reviews Rheumatology, 8*(2), 77-89. doi:10.1038/nrrheum.2011.199

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.

Seoane-Mato, D., Sánchez-Piedra, C., Díaz-González, F., & Bustabad, S. (2018). THU0684 Prevalence of rheumatic diseases in adult population in spain. episer 2016 study. *Annals of the Rheumatic Diseases, 77*(Suppl 2), 535. doi:10.1136/annrheumdis-2018-eular.6463

Shimoda, Y., Nagaoka, T., Moromizato, K., Sunagawa, M., Hanihara, T., Yoneda, M., . . . Fukumine, T. (2012). Degenerative changes of the spine in people from prehistoric Okhotsk culture and two ancient human groups from Kanto and Okinawa, Japan. *Anthropological Science, 120*(1), 1-21.

Shin, D. H., Jung, G.-U., Oh, C. S., Kim, M. J., Shin, E.-K., & Kim, Y.-S. (2016). Paleopathological Patterns of Degenerative Arthropathy: Prevalence of Limb-Joint Osteoarthritis in Joseon People Skeletons. *Anthropologist, 24*(3), 702-710.

Smith, R. J. (2018). The continuing misuse of null hypothesis significance testing in biological anthropology. *American journal of physical anthropology, 166*(1), 236-245. doi:10.1002/ajpa.23399

Steckel, R., Larsen, C., Sciulli, P., & Walker, P. (2005). The Global History of Health Project: data collection codebook. Ohio State University, Columbus, OH. In.

Stodder, A. L. W. (2012). Data and Data Analysis Issues in Paleopathology. In A. L. Grauer (Ed.), *A companion to Paleopathology* (pp. 339-356). Chichester, UK: Wiley-Blackwell.

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

Suzuki, S., Sunagawa, M., Shindo, M., Kimura, R., Yamaguchi, K., Sato, T., . . . Wakebe, T. (2016). Degenerative changes in the appendicular joints of ancient human populations from the Japan Islands. *Quaternary International, 405*, 147-159.

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.

Turkiewicz, A., Petersson, I. F., Björk, J., Hawker, G., Dahlberg, L. E., Lohmander, L. S., & Englund, M. (2014). Current and future impact of osteoarthritis on health care: a population-based study with projections to year 2032. *Osteoarthritis and Cartilage, 22*(11), 1826-1832. doi:https://doi.org/10.1016/j.joca.2014.07.015

UCLA: Statistical Consulting Group. (2012). Introduction to Generalized Linear Mixed Models. Retrieved from https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/ [accessed 15/06/20]

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.

Vina, E. R., & Kwoh, C. K. (2018). Epidemiology of osteoarthritis: literature update. *Current Opinion in Rheumatology, 30*(2), 160-167. doi:10.1097/BOR.0000000000000479

Waldron, T. (2009). Diseases of Joints, Part 1. In *Palaeopathology* (pp. 24-45): Cambridge University Press.

Waldron, T. (2019). Chapter 20 - Joint Disease. In J. E. Buikstra (Ed.), *Ortner's Identification of Pathological Conditions in Human Skeletal Remains (Third Edition)* (pp. 719-748). San Diego: Academic Press.

Weiss, E. (2018). Degenerative Joint Disease. In *The Encyclopedia of Archaeological Sciences* (pp. 1-3).

Westfall, P. H., & Young, S. S. (1989). p Value Adjustments for Multiple Tests in Multivariate Binomial Models. *Journal of the American Statistical Association, 84*(407), 780-786. doi:10.2307/2289666

White, N. M., & Barnett, A. G. (2019). Analysis of multisite intervention studies using generalized linear mixed models. *Infection Control & Hospital Epidemiology, 40*(8), 910-917.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]*.

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.

Zampetti, S., Mariotti, V., Radi, N., & Belcastro, M. G. (2016). Variation of skeletal degenerative joint disease features in an identified Italian modern skeletal collection. *American Journal of Physical Anthropology, 160*(4), 683-693. doi:10.1002/ajpa.22998

Ziglari, L. (2017). Interpreting multiple regression results: β weights and structure coefficients. *General Linear Model Journal, 43*(2), 13-22.

Ziyatdinov, A., Vázquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., & Soria, J. M. (2018). lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics, 19*(1), 68. doi:10.1186/s12859-018-2057-x

Figure 1. Criteria used to evaluate degenerative changes on the spine. a) Grade 0: normal condition. b) Grade 1: small, horizontal growth osteophytes and porosity absent or covering less than 10% of the joint surface. c) Grade 2: osteophytes that can grow around the rim and porosity covering 10-50% of the joint surface. d) Grade 3: significant growth of the osteophytes in a vertical direction and porosity covering more than 50% of the joint surface. e) Grade 4: joint fusion.

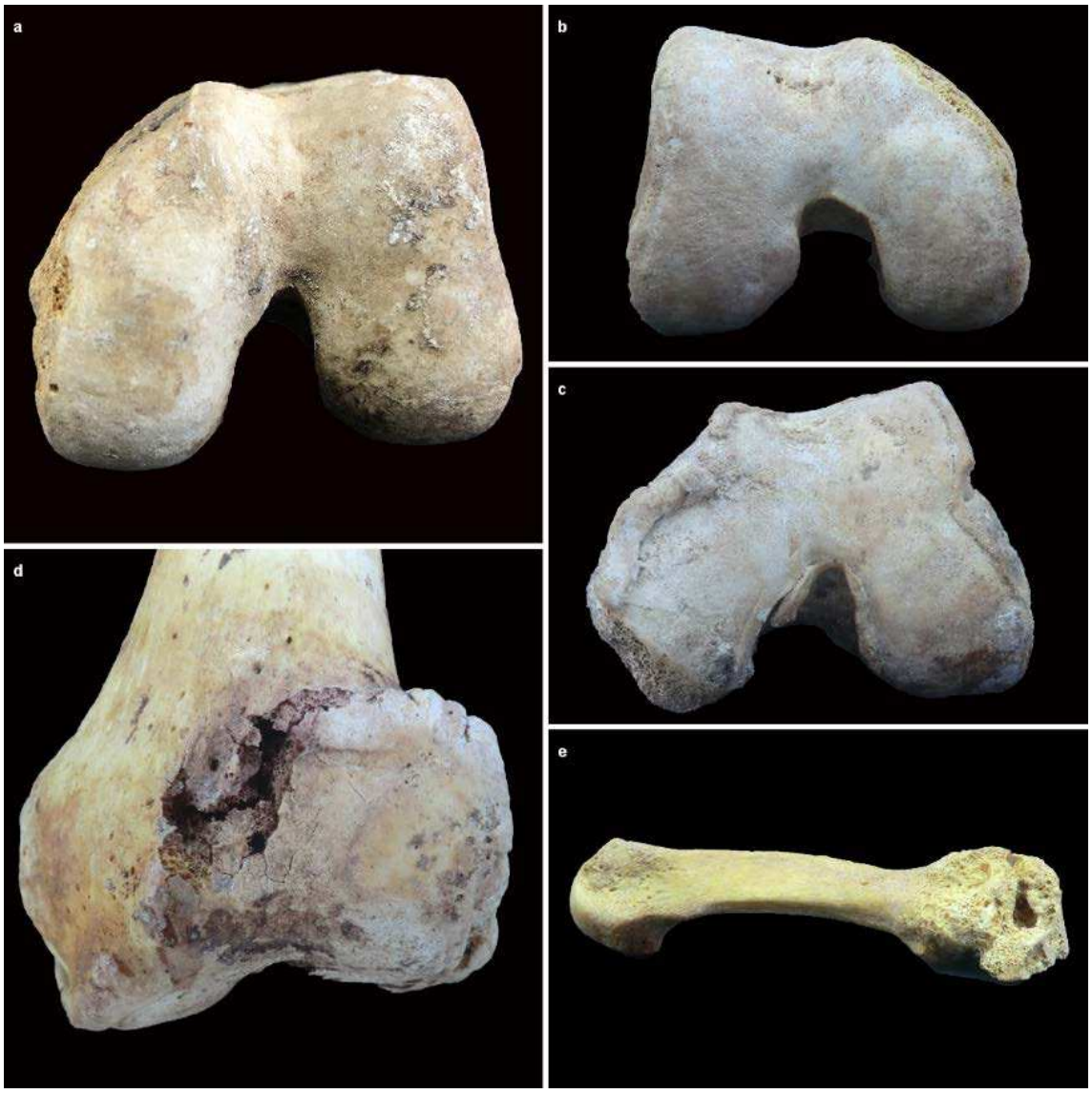

Figure 2. Criteria used to evaluate degenerative changes on the appendicular joints. a) Grade 0: normal condition. b) Grade 1: slight marginal lipping with osteophytes less than 3 mm and slight degenerative changes on the articular surface. c) Grade 2: severe marginal lipping with osteophytes greater than 3 mm and severe degenerative changes on the articular surface. d) Grade 3: complete or nearly complete destruction of the articular surface. e) Grade 4: joint fusion.

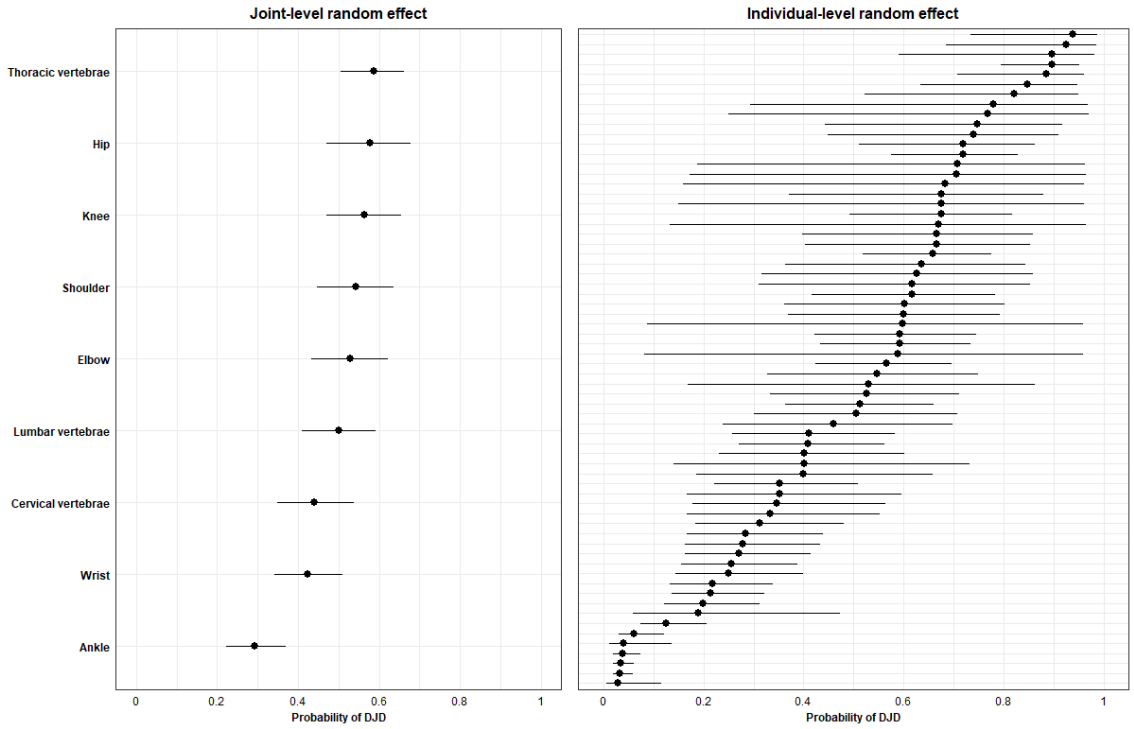

Figure 3. Random effect coefficients from a GLMM model including joints (left) and SNB individuals (right), with their corresponding 95% confidence interval. The scale has been converted from the raw log-odds into the corresponding DJD probability by applying an inverse logit transformation.

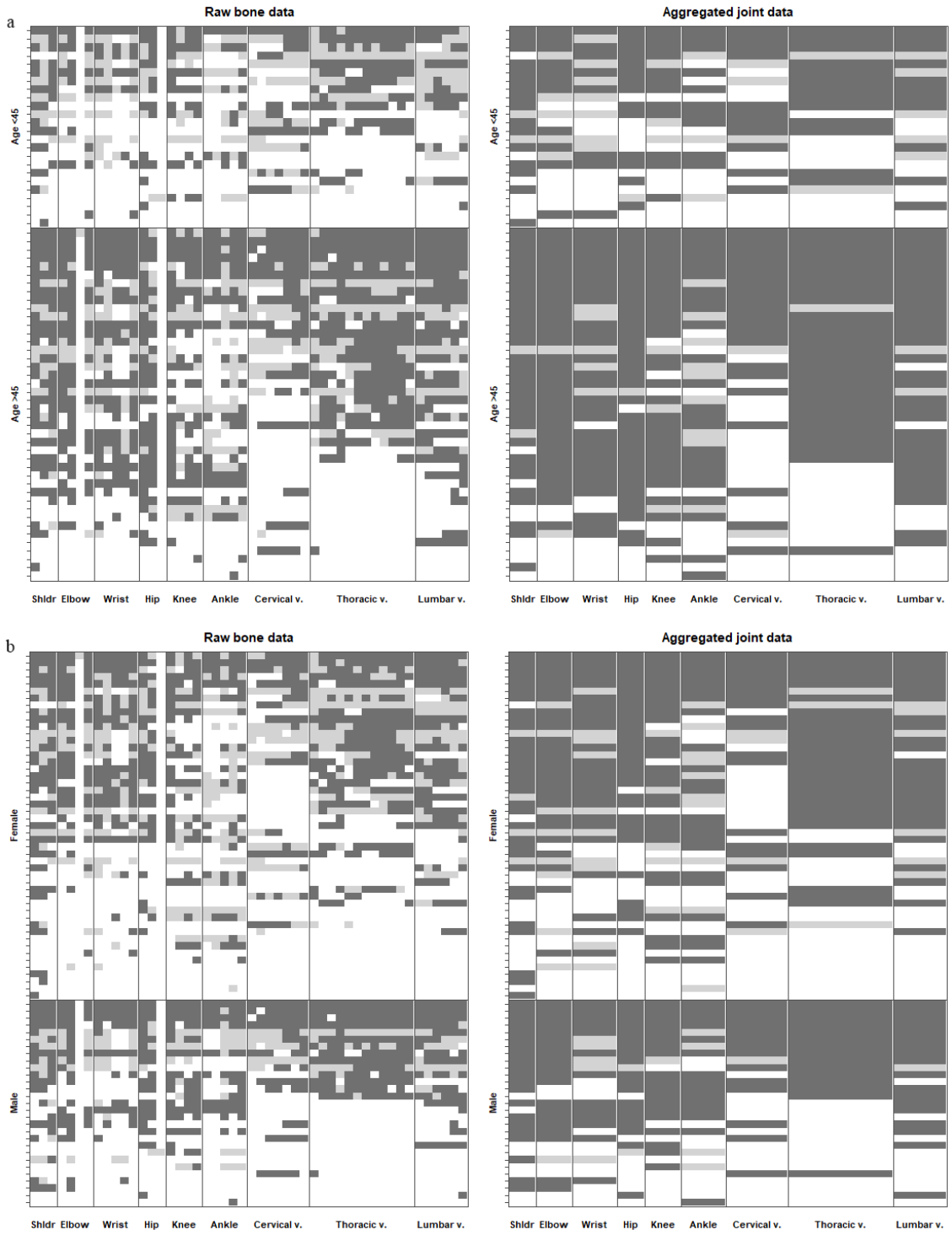

Figure 4. Plots of the DJD prevalence data. Each cell is an individual bone (left), as assessed by GLMMs; or joint (right), as assessed by GLMs. Colors: Dark grey if DJD is present, light gray if absent, and white for missing data. a) Prevalence sorted by age; only individuals with estimated age are shown, and their order is the same in both panels. b) Prevalence sorted by sex; only individuals with estimated sex are shown, and their order is the same in both panels.

Table 1. Frequencies of SNB individuals stratified by sexes and age groups.

AGE GROUP	MALE	FEMALE	TOTAL
20 - 45 years	6	18	24
> 45 years	18	24	42
Adult (>20 years)	5	7	12
TOTAL	29	49	78

Table 2. Summary of joint systems and analyzed articular surfaces.

Articular surfaces	**Joint system**	**Limb**
Acromial end of the clavicle	Shoulder	Arm
Scapular glenoid fossa		
Proximal humerus		
Distal humerus	Elbow	
Proximal ulna		
Proximal radius		
Distal ulna	Wrist	
Distal radius		
Carpal bones		
Acetabulum	Hip	Leg
Proximal femur		
Distal femur	Knee	
Patella		
Proximal tibia		
Proximal fibula		
Distal tibia	Ankle	
Distal fibula		
Tarsal bones		
C1	Cervical vertebrae	Spinal column
C2		
C3		
C4		
C5		
C6		
C7		
T1	Thoracic vertebrae	
T2		
T3		
T4		
T5		
T6		
T7		
T8		
T9		
T10		
T11		
T12		
L1	Lumbar vertebrae	
L2		
L3		
L4		
L5		
Superior articular facets of the sacrum		

Table 3. Prevalence of DJDs in the SNB population.

Joint		Absence of DJD	Presence of DJD	Total
Shoulder	N	13	84	97
	%	13.4	86.6	100.0
Elbow	N	15	79	94
	%	16.0	84.0	100.0
Wrist	N	32	68	100
	%	32.0	68.0	100.0
Hip	N	10	82	92
	%	10.9	89.1	100.0
Knee	N	16	78	94
	%	17.0	83.0	100.0
Ankle	N	24	68	92
	%	26.1	73.9	100.0
Cervical vertebrae	N	6	34	40
	%	15.0	85.0	100.0
Thoracic vertebrae	N	4	53	57
	%	7.0	93.0	100.0
Lumbar vertebrae	N	7	46	53
	%	13.2	86.8	100.0
Total	N	127	592	719
	%	17.7	82.3	100.0

Table 4. Stratified prevalence of DJD in the SNB population.

Joint		Age group		Sex		Side	
		20-45	>45	♂	♀	Right	Left
Shoulder	**A**	5	7	4	9	9	4
	A %	15.2	12.5	11.8	14.3	18.0	8.5
	P	28	49	30	54	41	43
	P %	84.8	87.5	88.2	85.7	82.0	91.5
Elbow	**A**	8	5	3	12	11	4
	A %	29.6	8.1	8.6	20.3	22.9	8.7
	P	19	57	32	47	37	42
	P %	70.4	91.9	91.4	79.7	77.1	91.3
Wrist	**A**	15	15	9	23	14	18
	A %	51.7	22.1	27.3	34.3	29.2	34.6
	P	14	53	24	44	34	34
	P %	48.3	77.9	72.7	65.7	70.8	65.4
Hip	**A**	2	8	4	6	5	5
	A %	7.1	12.5	11.4	10.5	11.1	10.6
	P	26	56	31	51	40	42
	P %	92.9	87.5	88.6	89.5	88.9	89.4
Knee	**A**	7	8	4	12	8	8
	A %	25.9	13.8	11.4	20.3	17.4	16.7
	P	20	50	31	47	38	40
	P %	74.1	86.2	88.6	79.7	82.6	83.3
Ankle	**A**	6	15	8	16	10	14
	A %	26.1	25.4	23.5	27.6	21.7	30.4
	P	17	44	26	42	36	32
	P %	73.9	74.6	76.5	72.4	78.3	69.6
Cervical vertebrae	**A**	3	2	1	5	n/a	n/a
	A %	20.0	8.3	6.7	20.0	n/a	n/a
	P	12	22	14	20	n/a	n/a
	P %	80.0	91.7	93.3	80.0	n/a	n/a
Thoracic vertebrae	**A**	3	1	0	4	n/a	n/a
	A %	15.8	2.9	0.0	10.3	n/a	n/a
	P	16	33	18	35	n/a	n/a
	P %	84.2	97.1	100.0	89.7	n/a	n/a
Lumbar vertebrae	**A**	4	2	1	6	n/a	n/a
	A %	25.0	5.9	5.0	18.2	n/a	n/a
	P	12	32	19	27	n/a	n/a
	P %	75.0	94.1	95.0	81.8	n/a	n/a

A (absence): Individuals without DJD in the joint. P (presence): Individuals with DJD in the joint.

Table 5. Analysis of DJD prevalence in the SNB population using three statistical methods.

	Fisher's Exact				GLM logistic regression				GLMM logistic regression			
	Sex (Male)		Age Group (>45)		Sex (Male)		Age Group (>45)		Sex (Male)		Age Group (>45)	
Location	**OR [95% CI]**	**P-value**	**OR [95% CI]**	**P-value**	**OR [95% CI]**	**P-value**	**OR [95% CI]**	**P-value**	**OR [95% CI]**	**P-value**	**OR [95% CI]**	**P-value**
Arm	2.92 [0.59-28.49]	0.217	4.25 [0.84-27.90]	0.062	3.1 [0.5-59.95]	0.956	3.58 [0.86-18.16]	0.089	2.14 [0.57-8.10]	0.261	**3.6 [1.01-12.8]**	**0.047**
Shoulder	1.25 [0.31-6.02]	1	1.25 [0.28-5.07]	0.755	1.53 [0.4-7.51]	0.558	1.11 [0.29-3.94]	0.869	0.81 [0.16-4.04]	0.794	2.71 [0.54-13.53]	0.223
Elbow	2.7 [0.65-16.07]	0.157	**4.7 [1.19-20.64]**	**0.018**	2.97 [0.67-21.04]	0.194	**4.34 [1.22-16.88]**	**0.026**	1.97 [0.3-12.71]	0.477	6.91 [0.98-48.61]	0.052
Wrist	1.45 [0.54-4.13]	0.501	**3.73 [1.35-10.56]**	**0.007**	1.2 [0.45-3.38]	0.722	**3.61 [1.42-9.42]**	**0.007***	3.03 [0.36-25.36]	0.307	**17.05 [1.97-147.27]**	**0.010***
Leg	1.55 [0.24-16.97]	0.709	2.36 [0.3-18.66]	0.369	1.12 [0.2-8.44]	0.903	2.28 [0.4-13.09]	0.332	**3.46 [1.28-9.38]**	**0.015***	2.24 [0.81-6.24]	0.122
Hip	0.94 [0.2-4.9]	1	0.55 [0.05-3.04]	0.718	0.93 [0.24-3.88]	0.912	0.54 [0.08-2.35]	0.456	2.05 [0.6-7.03]	0.251	2.04 [0.58-7.15]	0.263
Knee	1.97 [0.53-9.13]	0.396	2.17 [0.58-7.9]	0.223	1.62 [0.49-6.39]	0.451	2.09 [0.65-6.65]	0.209	9.36 [0.83-105.6]	0.070	3.69 [0.37-37.29]	0.268
Ankle	1.24 [0.42-3.83]	0.807	1.03 [0.28-3.44]	1	1.65 [0.57-5.25]	0.366	0.95 [0.29-2.82]	0.927	**5.1 [1.25-20.84]**	**0.023***	1.95 [0.45-8.39]	0.368
Spinal column	Inf	0.535	Inf	0.119	Inf †	0.997 †	Inf †	0.997 †	**4.07 [1.28-12.96]**	**0.018***	**4.44 [1.5-13.12]**	**0.007***
Cervical vertebrae	3.41 [0.33-177]	0.381	2.68 [0.27-36.25]	0.354	2.19 [0.25-47.47]	0.521	2.25 [0.31-20.21]	0.426	7.77 [0.73-82.8]	0.09	3.74 [0.44-32.12]	0.229
Thoracic vertebrae	Inf	0.294	6.35 [0.44-355.88]	0.114	Inf	0.995	4.75 [0.54-102.39]	0.199	**5.5 [1.05-28.69]**	**0.043**	2.3 [0.51-10.36]	0.276
Lumbar vertebrae	4.13 [0.44-204.04]	0.233	5.13 [0.64-63.73]	0.074	2.68 [0.35-55.41]	0.398	4.66 [0.78-37.58]	0.103	1.63 [0.33-8.01]	0.548	**17.55 [3.09-99.75]**	**0.001***

Significant p-values in **bold** (p-value < 0.05). Asterisks (*) indicate results significant at an FDR threshold of 10%.

Infinite ORs are reported when any contingency table cell values are zero or when the upper bound of the confidence interval was infinity.

†: Model diagnostics reported quasi-complete separation.

Table 6. Variance explained in the regression analyses of DJD prevalence in the SNB population.

Location	GLM R^2 (%)	GLMM marginal R^2 (%)	GLMM conditional ICC (%)
Arm	11.34%	6.94%	52.05%
Shoulder	3.87%	2.86%	55.96%
Elbow	21.27%	9.56%	57.96%
Wrist	11.29%	16.55%	58.10%
Leg	4.39%	8.84%	40.84%
Hip	1.38%	4.63%	32.99%
Knee	4.58%	12.41%	64.14%
Ankle	2.91%	11.42%	44.66%
Spinal column	33.70% †	15.54%	40.17%
Cervical vertebrae	7.25%	14.88%	58.15%
Thoracic vertebrae	22.94%	10.22%	53.10%
Lumbar vertebrae	15.88%	20.12%	46.12%

†: Model diagnostics reported quasi-complete separation.

Supplementary Table 1. Prevalence and severity of DJDs in the SNB population .

Joint		Severity Grade					Subtotal 1-4 (prevalence)	Total
		0	1	2	3	4		
Shoulder	N	13	45	29	10	0	84	97
	%	13.4	46.4	29.9	10.3	0.0	86.6	100.0
Elbow	N	15	55	23	1	0	79	94
	%	16.0	58.5	24.5	1.1	0.0	84.0	100.0
Wrist	N	32	52	13	1	2	68	100
	%	32.0	52.0	13.0	1.0	2.0	68.0	100.0
Hip	N	10	52	30	0	0	82	92
	%	10.9	56.5	32.6	0.0	0.0	89.1	100.0
Knee	N	16	47	27	4	0	78	94
	%	17.0	50.0	28.7	4.3	0.0	83.0	100.0
Ankle	N	24	59	5	1	3	68	92
	%	26.1	64.1	5.4	1.1	3.3	73.9	100.0
Cervical vertebrae	N	6	15	3	11	5	34	40
	%	15.0	37.5	7.5	27.5	12.5	85.0	100.0
Thoracic vertebrae	N	4	12	26	9	6	53	57
	%	7.0	21.1	45.6	15.8	10.5	93.0	100.0
Lumbar vertebrae	N	7	13	20	13	0	46	53
	%	13.2	24.5	37.7	24.5	0.0	86.8	100.0
Total	N	127	350	176	50	16	592	719
	%	17.7	48.7	24.5	7.0	2.2	82.3	100.0

Supplementary Table 2. Analysis of DJD prevalence in the SNB population using three statistical methods.

Location	Location Type	Test	Covariate	OR	95% CI lower	95% CI upper	P-value	P-value (FDR-corrected)
Arm	Limb	Fisher's Exact	Sex (Male)	2.92	0.59	28.49	0.217	0.629
Arm	Limb	Fisher's Exact	Age Group (>45)	4.25	0.84	27.90	0.062	0.544
Arm	Limb	Fisher's Exact	Laterality (Right)	1.25	0.34	4.84	0.774	0.953
Shoulder	Joint	Fisher's Exact	Sex (Male)	1.25	0.31	6.02	1.000	1.000
Shoulder	Joint	Fisher's Exact	Age Group (>45)	1.25	0.28	5.07	0.755	0.953
Shoulder	Joint	Fisher's Exact	Laterality (Right)	0.43	0.09	1.68	0.236	0.629
Elbow	Joint	Fisher's Exact	Sex (Male)	2.70	0.65	16.07	0.157	0.627
Elbow	Joint	Fisher's Exact	Age Group (>45)	**4.70**	**1.19**	**20.64**	**0.018**	**0.292**
Elbow	Joint	Fisher's Exact	Laterality (Right)	0.32	0.07	1.18	0.089	0.544
Wrist	Joint	Fisher's Exact	Sex (Male)	1.45	0.54	4.13	0.501	0.778
Wrist	Joint	Fisher's Exact	Age Group (>45)	**3.73**	**1.35**	**10.56**	**0.007**	**0.238**
Wrist	Joint	Fisher's Exact	Laterality (Right)	1.31	0.53	3.34	0.532	0.778
Leg	Limb	Fisher's Exact	Sex (Male)	1.55	0.24	16.97	0.709	0.953
Leg	Limb	Fisher's Exact	Age Group (>45)	2.36	0.30	18.66	0.369	0.745
Leg	Limb	Fisher's Exact	Laterality (Right)	2.48	0.39	27.15	0.439	0.778
Hip	Joint	Fisher's Exact	Sex (Male)	0.94	0.20	4.90	1.000	1.000
Hip	Joint	Fisher's Exact	Age Group (>45)	0.55	0.05	3.04	0.718	0.953
Hip	Joint	Fisher's Exact	Laterality (Right)	0.95	0.20	4.48	1.000	1.000
Knee	Joint	Fisher's Exact	Sex (Male)	1.97	0.53	9.13	0.396	0.745
Knee	Joint	Fisher's Exact	Age Group (>45)	2.17	0.58	7.90	0.223	0.629
Knee	Joint	Fisher's Exact	Laterality (Right)	0.95	0.28	3.23	1.000	1.000
Ankle	Joint	Fisher's Exact	Sex (Male)	1.24	0.42	3.83	0.807	0.956
Ankle	Joint	Fisher's Exact	Age Group (>45)	1.03	0.28	3.44	1.000	1.000
Ankle	Joint	Fisher's Exact	Laterality (Right)	1.57	0.56	4.55	0.477	0.778
Spinal column	Limb	Fisher's Exact	Sex (Male)	Inf	0.10	Inf	0.535	0.778

Spinal column	Limb	Fisher's Exact	Age Group (>45)	Inf	0.35	Inf	0.119	0.544
Cervical vertebrae	Joint	Fisher's Exact	Sex (Male)	3.41	0.33	177.00	0.381	0.745
Cervical vertebrae	Joint	Fisher's Exact	Age Group (>45)	2.68	0.27	36.25	0.354	0.745
Thoracic vertebrae	Joint	Fisher's Exact	Sex (Male)	Inf	0.31	Inf	0.294	0.723
Thoracic vertebrae	Joint	Fisher's Exact	Age Group (>45)	6.35	0.47	355.88	0.114	0.544
Lumbar vertebrae	Joint	Fisher's Exact	Sex (Male)	4.13	0.44	204.04	0.233	0.629
Lumbar vertebrae	Joint	Fisher's Exact	Age Group (>45)	5.13	0.64	63.73	0.074	0.544
Arm	Limb	GLM logistic regression	Sex (Male)	3.10	0.50	59.95	0.304	0.956
Arm	Limb	GLM logistic regression	Age Group (>45)	3.58	0.86	18.16	0.089	0.308
Arm	Limb	GLM logistic regression	Laterality (Right)	0.84	0.19	3.46	0.803	0.962
Shoulder	Joint	GLM logistic regression	Sex (Male)	1.53	0.40	7.51	0.558	0.956
Shoulder	Joint	GLM logistic regression	Age Group (>45)	1.11	0.29	3.94	0.869	0.997
Shoulder	Joint	GLM logistic regression	Laterality (Right)	0.47	0.12	1.64	0.254	0.877
Elbow	Joint	GLM logistic regression	Sex (Male)	2.97	0.67	21.04	0.194	0.956
Elbow	Joint	GLM logistic regression	Age Group (>45)	**4.34**	**1.22**	**16.88**	**0.026**	**0.154**
Elbow	Joint	GLM logistic regression	Laterality (Right)	0.31	0.07	1.13	0.089	0.712
Wrist	Joint	GLM logistic regression	Sex (Male)	1.20	0.45	3.38	0.722	0.997
Wrist	Joint	GLM logistic regression	Age Group (>45)	**3.61**	**1.42**	**9.42**	**0.007***	**0.089***
Wrist	Joint	GLM logistic regression	Laterality (Right)	1.14	0.46	2.86	0.770	0.962
Leg	Limb	GLM logistic regression	Sex (Male)	1.12	0.20	8.44	0.903	0.997
Leg	Limb	GLM logistic regression	Age Group (>45)	2.28	0.40	13.09	0.332	0.569
Leg	Limb	GLM logistic regression	Laterality (Right)	2.00	0.37	14.99	0.439	0.877
Hip	Joint	GLM logistic regression	Sex (Male)	0.93	0.24	3.88	0.912	0.997
Hip	Joint	GLM logistic regression	Age Group (>45)	0.54	0.08	2.35	0.456	0.608
Hip	Joint	GLM logistic regression	Laterality (Right)	0.97	0.25	3.75	0.962	0.962
Knee	Joint	GLM logistic regression	Sex (Male)	1.62	0.49	6.39	0.451	0.956
Knee	Joint	GLM logistic regression	Age Group (>45)	2.09	0.65	6.65	0.209	0.419
Knee	Joint	GLM logistic regression	Laterality (Right)	1.10	0.35	3.52	0.866	0.962
Ankle	Joint	GLM logistic regression	Sex (Male)	1.65	0.57	5.25	0.366	0.956
Ankle	Joint	GLM logistic regression	Age Group (>45)	0.95	0.29	2.82	0.927	0.997

Ankle	Joint	GLM logistic regression	Laterality (Right)	1.59	0.59	4.47	0.363	0.877
Spinal column	Limb	GLM logistic regression	Sex (Male)	4.49×10^{7} †	0.00 †	Inf †	0.997 †	0.997 †
Spinal column	Limb	GLM logistic regression	Age Group (>45)	1.53×10^{8} †	4.82×10^{-308} †	Inf †	0.997 †	0.997 †
Cervical vertebrae	Joint	GLM logistic regression	Sex (Male)	2.19	0.25	47.47	0.521	0.956
Cervical vertebrae	Joint	GLM logistic regression	Age Group (>45)	2.25	0.31	20.21	0.426	0.608
Thoracic vertebrae	Joint	GLM logistic regression	Sex (Male)	2.57×10^{7}	8.44×10^{-105}	Inf	0.995	0.997
Thoracic vertebrae	Joint	GLM logistic regression	Age Group (>45)	4.75	0.54	102.39	0.199	0.419
Lumbar vertebrae	Joint	GLM logistic regression	Sex (Male)	2.68	0.35	55.41	0.398	0.956
Lumbar vertebrae	Joint	GLM logistic regression	Age Group (>45)	4.66	0.78	37.58	0.103	0.308
Arm	Limb	GLMM logistic regression	Sex (Male)	2.14	0.57	8.10	0.261	0.392
Arm	Limb	GLMM logistic regression	**Age Group (>45)**	**3.60**	**1.01**	**12.80**	**0.047**	**0.125**
Arm	Limb	GLMM logistic regression	Laterality (Right)	1.27	0.84	1.93	0.252	0.753
Shoulder	Joint	GLMM logistic regression	Sex (Male)	0.81	0.16	4.04	0.794	0.794
Shoulder	Joint	GLMM logistic regression	Age Group (>45)	2.71	0.54	13.53	0.223	0.302
Shoulder	Joint	GLMM logistic regression	Laterality (Right)	1.29	0.58	2.88	0.528	0.753
Elbow	Joint	GLMM logistic regression	Sex (Male)	1.97	0.30	12.71	0.477	0.572
Elbow	Joint	GLMM logistic regression	Age Group (>45)	6.91	0.98	48.61	0.052	0.125
Elbow	Joint	GLMM logistic regression	Laterality (Right)	1.17	0.51	2.70	0.708	0.753
Wrist	Joint	GLMM logistic regression	Sex (Male)	3.03	0.36	25.36	0.307	0.410
Wrist	Joint	GLMM logistic regression	**Age Group (>45)**	**17.05**	**1.97**	**147.27**	**0.010***	**0.040***
Wrist	Joint	GLMM logistic regression	Laterality (Right)	1.23	0.64	2.39	0.535	0.753
Leg	Limb	GLMM logistic regression	**Sex (Male)**	**3.46**	**1.28**	**9.38**	**0.015***	**0.094***
Leg	Limb	GLMM logistic regression	Age Group (>45)	2.24	0.81	6.24	0.122	0.243
Leg	Limb	GLMM logistic regression	Laterality (Right)	1.20	0.82	1.78	0.350	0.753
Hip	Joint	GLMM logistic regression	Sex (Male)	2.05	0.60	7.03	0.251	0.392
Hip	Joint	GLMM logistic regression	Age Group (>45)	2.04	0.58	7.15	0.263	0.302
Hip	Joint	GLMM logistic regression	Laterality (Right)	1.18	0.51	2.74	0.699	0.753
Knee	Joint	GLMM logistic regression	Sex (Male)	9.36	0.83	105.60	0.070	0.169
Knee	Joint	GLMM logistic regression	Age Group (>45)	3.69	0.37	37.29	0.268	0.302

Knee	Joint	GLMM logistic regression	Laterality (Right)	1.14	0.51	2.56	0.753	0.753
Ankle	Joint	GLMM logistic regression	Sex (Male)	**5.10**	**1.25**	**20.84**	**0.023***	**0.094***
Ankle	Joint	GLMM logistic regression	Age Group (>45)	1.95	0.45	8.39	0.368	0.368
Ankle	Joint	GLMM logistic regression	Laterality (Right)	1.36	0.73	2.53	0.336	0.753
Spinal column	Limb	GLMM logistic regression	Sex (Male)	**4.07**	**1.28**	**12.96**	**0.018***	**0.094***
Spinal column	Limb	GLMM logistic regression	Age Group (>45)	**4.44**	**1.50**	**13.12**	**0.007***	**0.040***
Cervical vertebrae	Joint	GLMM logistic regression	Sex (Male)	7.77	0.73	82.80	0.090	0.179
Cervical vertebrae	Joint	GLMM logistic regression	Age Group (>45)	3.74	0.44	32.12	0.229	0.302
Thoracic vertebrae	Joint	GLMM logistic regression	Sex (Male)	**5.50**	**1.05**	**28.69**	**0.043**	**0.130**
Thoracic vertebrae	Joint	GLMM logistic regression	Age Group (>45)	2.30	0.51	10.36	0.276	0.302
Lumbar vertebrae	Joint	GLMM logistic regression	Sex (Male)	1.63	0.33	8.01	0.548	0.597
Lumbar vertebrae	Joint	GLMM logistic regression	Age Group (>45)	**17.55**	**3.09**	**99.75**	**0.001***	**0.015***

Significant p-values in **bold** (p-value < 0.05). Asterisks (*) indicate results significant at an FDR threshold of 10%.
Infinite ORs are reported when any contingency table cell values are zero or when the upper bound of the confidence interval was infinity.
†: Model diagnostics reported quasi-complete separation.