PUBLISHED VERSION

Henderson, CY. and 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. http://rdcu.be/mHq2

C. Y. Henderson, CIAS – Research Centre for Anthropology and Health.
E. Nikita, Fitch Laboratory, British School at Athens

corresponding author: C. Y Henderson, CIAS – Research Centre for Anthropology and Health, Department of Life Sciences, University of Coimbra, Apartado 3046, 3001-401, Coimbra, Portugal. Email: c.y.henderson@uc.pt or c.y.henderson@dunelm.org.uk

Abstract

Osteoarchaeological studies provide valuable information concerning living conditions and life course changes in past societies. However, many skeletal markers, such as entheseal changes, are multifactorial in aetiology, thus their interpretation is not straightforward. Generalised linear models (GLM) are ideal for analysing such phenomena, i.e. those with multiple underlying causative factors, but, to date, their use has been limited. This paper focusses attention on using these models to test hypotheses regarding the aetiology of entheseal changes, widely regarded as indicative of activity-patterns, but which are also affected by ageing and body size. To demonstrate the use and limitations of these models, this paper provides an independent test of a previously developed GLM on an identified skeletal sample comprised of skeletons from four British sites (n=58) which has a typical sample size for archaeological osteological analysis. In addition to this model, GLM were developed to include the factor of body size and expand the models to test individual entheses, as well as joint complexes whereby multiple entheses for muscles which act synergistically have been pooled.

The results indicate that the original model did not compare well with the frequencies of entheseal changes found in the British assemblage under study. The new models found no clear pattern of influence, although both ageing and body size were important for some entheses. Generalised linear models are appropriate for testing the interaction of biological variables, but future studies need to take into account and test their applicability to archaeological sample sizes.

Keywords
activity-related stress markers; musculoskeletal stress markers (MSM); generalised linear models; identified skeletal collections

1.1 Introduction

Entheses are the attachments of the soft and hard musculoskeletal tissues, e.g. bone and tendon (see review in Jurmain et al. 2012). Lytic lesions, new mineralised tissue formation, among other *in vivo* structural alterations are now named “entheseal changes” (ECs) (Jurmain et al., 2012). Entheseal changes (ECs) have been widely used to study the social stratification of labour in past societies because they are perceived to provide direct evidence of repetitive muscular use from individual skeletons (see review in Jurmain et al. 2012). The importance of distinguishing between the anatomy of entheses has been highlighted in recent years, with distinctions being made between fibrous and fibrocartilaginous entheses (Henderson, 2009; Villotte, 2006, 2008). This research has demonstrated that, in skeletal remains, there is currently no way to identify normal fibrous entheses, because the boundary between their normal surface roughness and the presence of EC is unclear (Jurmain et al., 2012). Therefore studies have focussed almost exclusively on fibrocartilaginous entheses (Alves Cardoso and Henderson, 2013; Henderson, 2009; Jurmain et al. 2012; Villotte, 2010).

Fibrocartilaginous entheses are affected by a number of non-activity-related factors. Diseases, such as diffuse idiopathic skeletal hyperostosis, are widely recognised in the palaeopathological literature and, along with many others, are known to cause ECs (Henderson, 2008). Other factors known to affect these entheses include biological sex, body size and genetic factors (Jurmain et al. 2012; Weiss et al. 2012; Wilczak, 1998). Previous studies that have used identified skeletal collections have demonstrated that the ageing process, rather than occupation, is the primary cause of ECs (Alves Cardoso and Henderson, 2010; 2013; Cardoso, 2008; Milella et al. 2012). However, other studies, have supported the link between activity-patterns and ECs (Kuorinka & Forcier, 1995; Niinimaki, 2012; Niinimaki et al. 2013; Shaw & Benjamin, 2007). The most notable support for this link, which tested a presence/absence method of recording ECs using several large identified skeletal collections, found differences between heavy manual labour and those in other occupations (classified as light manual and non-manual) (Villotte et al. 2010). Villotte et al. (2010) utilised generalised linear methods (GLM), a statistical approach which has recently been proposed for research questions involving multifactorial phenomena, such as ECs (Nikita, 2014; Villotte et al., 2010). This approach offers the advantage over traditional methods that it can explore the simultaneous effect of multiple factors, continuous, binary or ordinal, as well as their interactions. It should therefore be the most appropriate method to test the relationship between ECs and occupation, as it is able to take into account the many other factors known to be associated with EC formation. However, to date the model itself has not been used to determine activity-patterns in archaeological remains.

The sample size used in the GLM analysis described above was over 300 skeletons (n=367) (Villotte et al. 2010). However, sample sizes from archaeological sites are normally under 50 individuals often with significantly fewer elements with observable entheses (Henderson, 2013a). A meta-analysis (ibid.) demonstrated that the median number of individuals represented when divided up by site, period and enthesis was 15 with a maximum of 44. This is likely an over-estimate of average sample sizes because sites without an enthesis present were removed from the analysis. For this reason, it is vital that the model developed by Villotte et al. (2010) is tested on a sample of identified skeletons with a typical size (and therefore with inherent biases in demographic profile) for an archaeological site before it is used to identify activity-patterns in past populations. The model is also based on pooling several entheses into a single frequency score for each individual, but this means that the effect of single muscles (e.g. whether flexing or extending the elbow) cannot be identified nor can joint usage be determined. Furthermore the effect of body size was not taken into account in the model, which some authors have found to affect EC presence (Weiss et al., 2012).

The aim of this paper is to test whether the model developed from the exploratory analysis by Villotte et al. (2010) could be applied to a typical archaeological skeletal sample with known occupation. The model successfully developed by Villotte and colleagues modelled their data based on activity-patterns, therefore this model should be more widely applicable to similar samples based on its robustness. An additional aim of the current paper is to determine whether GLM models applied to specific muscles and accounting for age and body size could achieve a similarly good modelling of the data, while providing improved specificity for the types of movement undertaken by the different occupation categories. Finally, this paper aims at determining whether models of joint usage based on EC presence at joint complexes, while accounting for confounding factors, would provide greater specificity than the original model for studying occupation from EC. To achieve this, British identified skeletal collections representing both a diversity of occupations and a typically sized archaeological sample were used.

2.1 Materials and methods

Male skeletons (n=58) from four British postmediaeval sites were selected to provide a range of occupational categories to provide comparable data to Villotte et al. (2010). The burial dates range in time from 1673 to 1895 and are represented by three urban sites from London (Cowie et al. 2008; Miles et al. 2008; Scheuer and Bowman, 1995) and one rural site from North Yorkshire (Caffell and Holst, 2010; Henderson et al. 2013), for details see Table 1. This is a geographically and socially heterogeneous sample, similar to that of the original study which found no differences between the populations used, indicating that pooling samples in this manner should not skew the results (Villotte et al. 2010). The occupational categorisation follows that used in the original paper except in the case of the tailors from Fewston. Documentary evidence from this site indicates that they were likely to have been engaged in farming activities (Henderson et al. 2013) and they have been classified with this group of individuals. The jewellers (for whom no category was found in the original paper), have been classified with the heavy manual workers, because they likely engaged in relatively heavy repetitive tasks similar to other occupations in this category. Data on sex, age and occupation were all collected from the documentary evidence associated with the skeletal remains. To avoid biasing results, the only variable known was the sex of the skeleton. Only male individuals were recorded because of the limited available data on female activities based on documentary evidence

(Alves Cardoso and Henderson, 2013; Henderson et al. 2013).

Entheseal changes were recorded as absent or present (Villotte, 2006; Villotte et al. 2010). The entheses recorded and the joint complexes in which these entheses were pooled are listed in Tables 2 and 5. Note that these are all fibrocartilaginous entheses because there is currently no biologically appropriate method for recording fibrous entheses (Jurmain, et al. 2012). The method used to create joint complexes is described below in section 2.2.2.2.

Long bone measurements were taken to create proxies for body size and mean values per side and activity are given in Table 3. All measurements were taken following Buikstra and Ubelaker (1994) except for the antero-posterior diameter of the radius which was measured immediately distal to the level of the pronator teres insertion, identified by a roughened often darker area on the bone. This redefinition enables measurements at a comparable level in all individuals in relation to their musculature, without incorporating any ECs associated with these entheses. The averaged z-scores of all humeral dimensions were used as a proxy for humeral size, and the corresponding radial values were used for the radius and ulna.

For the shoulder the averaged z-scores included the vertical and transverse humeral head diameter and the maximum humeral length, for the elbow the condylar width (which avoids including the size of the common extensor and flexor origins) and humeral maximum length, and for the hand/wrist the antero-posterior and medio-lateral radial diameter as well as radial maximum length. A single proxy for body size was deemed inappropriate because of local variation in skeleton size which may impact on the biomechanics of the musculoskeletal system (Henderson, 2013b).

Disease presence was taken into account, such that individuals displaying signs of sacro-iliac joint, vertebral body or apophyseal ankylosis were classified as “boneformers” according to previously published criteria (Henderson, 2008). These individuals were excluded from the main analysis, because the generalised changes to the fibrocartilaginous zones of the body indicate a systemic alteration which may have a pathological aetiology (*ibid.*). However, they were included for the study of joint complexes because they are another known compounding factor for EC presence and their effect on identifying occupation categories needs to be tested.

2.2 Statistical analysis

2.2.1 Testing the Villotte et al. (2010) GLM method

Since one of the aims of the paper was to test whether a previously derived equation predicting entheseal change frequency accounting for age at death and occupation category could be applied to this sample, a model was calculated for the age and occupation profile of this sample (Villotte et al. 2010). Ten year age categories (20-29, 30-39, 40-49, 50-59, 60+) were used based on the data present. The model used is:

$$\eta = \exp(-4.941+0.072*age+0.260*side+0.612*occupation)/$$
$$(1+\exp(\ 4.941+0.072*age+0.260*side+0.612*occupation)$$

where side and occupation are binary variables of which both left and nonmanual are zero; and right and manual are one (see below for an explanation of the terms of the GLM model) (Villotte et al. 2010).

This model was plotted and compared to a graph of the frequencies of EC found in the sample. These frequencies were calculated by adding up the number of EC scored as present and dividing by the total number of entheses observable for the entheses: subscapularis, supraspinatus, infraspinatus and biceps brachii insertions, with the common extensor and flexor origins. Boneformers, as described above, were excluded from this analysis.

2.2.2 New models: the GLM method

An additional aim of this paper was to determine whether the existing model could be improved further for the sample under study. To examine this, generalised linear models were created to study the impact of the factors age, body size, and type of activity on the dependent (response) variable EC presence. As discussed above, generalised linear models (GLM) extend traditional linear regression to encompass response variables that may have non-normal distributions (see detailed discussion in Liang and Zeger, 1986; McCullagh and Nelder, 1989; Agresti, 2002; Molenberghs, 2010 and brief summary in Nikita, 2014). As such, the response variable may be binary, ordinal or a scale while both the main effects of each predictor as well as their interactions may be explored.

In particular, GLM is applied when n response values, $y_1, y_2, \ldots, y_n$, are recorded as a function of p explanatory variables, $X_1, X_2, \ldots, X_p$, which can be either continuous or categorical, and the response values come from any exponential family distribution (i.e., normal, binomial, Poisson, gamma, etc.). The mathematical expression of GLM may be written as:

$$\eta = b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p$$

where $\eta = g(\mu)$ and g is any smooth monotonic *link* function of the mean (μ) of the distribution function of the response variable *y*.

There are several options for the distribution function and, therefore, for the nature of the response variable *y*. For example, y may be a scale, an ordinal or a binary response. If *y* is a scale variable following the normal distribution, the link function is the identity function and, therefore, $\eta = \mu$, where μ is the predicted by the model *y* value. In this case, GLM become identical to a General Linear Model (ANCOVA). When the response is a binary variable, the link function may be expressed as $\eta = \ln(P/(1\text{-}P))$, where P is the probability that the binary variable takes the value 1. In SPSS η is defined from $\eta = \text{-}\ln(P/$

(1-*P*)) and therefore in SPSS the mathematical expression of GLM under binary response may be written as:

$$\ln \frac{P}{1-P} = -(b_0 + b_1X_1 + b_2X_2 + ... + b_pX_p)$$

where now *P* is the probability that the binary variable takes the value 0.

Generalised linear models were run in SPSS 19.0 with a binary logistic response. The covariance matrix used was the robust estimator because this is a corrected model-based estimator that provides a consistent estimate of the covariance (Chrisletta and Spini, 2004). Note that in order for GLM to be applied, there must be no quasi-complete separation in the data; otherwise, the maximum likelihood estimates do not exist. This is a serious limitation with small sample sizes when multiple predictors are explored and as a consequence there are categories with no or very few cases. This is also one of the issues addressed in the current paper. Due to the large number of analyses performed, the p-values were recalculated using a Holm-Bonferroni correction for multiple comparisons.

For all analyses descriptive statistics and odds ratios were calculated for the EC data and an effect size, an unbiased version of Cohen's d (Nakagawa and Cuthill, 2007), was used to study the continuous data, e.g. age and bone size. This approach was taken to enable comparison with other studies and to enable comparisons where assumptions of the GLM were violated.

For the new model real age-at-deaths were used rather than the age categories because the method used by Villotte et al. (2010) placed the majority of individuals into the same age category (60+, see Table 1). Although using the documented ages increases the accuracy of our models, it causes a limitation for the application of this model to archaeological remains for which age-at-death is not known from associated records.

2.2.2.1 Testing the new models: individual entheses

Individual entheses represent individual muscles (or collections, in the case of the common extensor and flexor origins) and therefore should provide the most specific indicator of the type of activity undertaken, e.g. extension or flexion of elbow. The majority of bioarchaeological inferences to activities have focussed on this approach (e.g. Weiss et al. 2012). In the current study the presence of changes to individual entheses was recorded (Villotte et al., 2010). The entheses incorporated were the insertions of the subscapularis, supraspinatus, infraspinatus, and teres minor. Common extensor, flexor and anconeus origins were also studied. Sample sizes for each of these can be found in Table 2.

2.2.2.2 Testing the new models: joint complexes

Joint complexes were created to study joint use. This is less specific than studying each individual enthesis, but provides a more specific model of activity than studying upper limb use, as done in the Villotte et al. (2010) model. Three joints of the upper limb were studied: shoulder, elbow and hand/wrist. The shoulder consists of the complex of subscapularis, supraspinatus, infraspinatus and teres minor insertions. The elbow of the biceps and triceps brachii insertions. Finally, the hand/wrist of the common extensor and flexor origins. The anconeus enthesis was not incorporated into a joint complex because its footprint is sometimes absent and its muscle fibres blend with those of the triceps brachii (Molinier et al., 2011).

For each of these complexes if an EC was present in any one enthesis it was considered present for the joint complex under study. This approach was used for the GLM model because frequencies (number of EC present per enthesis in the joint complex) were non-normally distributed, rendering the use of a linear GLM inappropriate. Odds ratios were calculated on the total number of EC present for all entheses in the joint complex as this is a better representation of the joint complex rather than individual entheses. For example, for the hand/wrist complex the common extensor origin is more frequently affected by EC than the common flexor origin. However, odds ratios comparable to the GLM models were also calculated for comparison.

As will be discussed in section 3.1, GLM were not used to compare boneformers to non-boneformers, because the resulting sample sizes per group were too small, which in turn led to the quasi-complete separation of the data. However, the confounding factor of boneforming was included to compare frequencies of EC presence and the effect of age.

3.1 Results

3.2 Results of the Villotte et al. (2010) GLM method

Figures 1 and 2 show the model predictions for EC frequency above the frequency found in the sample. The frequencies for the right side for the two oldest age categories and the left nonmanual category closely match the predicted model, as does the second youngest (3-39) age category for the left side. Where the model does not accurately predict the outcome frequency, can, in part be explained by small sample sizes, particularly evident in the youngest age category (nonmanual n=2, manual n=1). This is partly caused by the sample size which over-inflates or under-inflates frequencies of EC present. The latter is evident in the 40-49 category for the nonmanual group (n=7). Tables 2, 4 and 5 demonstrate the small sample sizes involved, prior to subsampling by age category while Table 1 demonstrates the small numbers of individuals in each age category except the oldest (n=24).

3.3 Testing the new models

Generalised linear models were created specifically for the sample examined in the present paper to test the impact of multiple factors on EC frequency, particularly given the small sample sizes under study. During the analysis of the data it became clear that the large number of predictors (three variables and their pairwise interactions) was causing computational problems (specifically quasi-complete separation in the data) and generated invalid results. For this reason, the tables presented in this section, only show the results for the models that did not exhibit computational problems. For those where computational problems exist, only odds ratios are presented. In addition, in order to minimize such problems, the models included both the main effects and the two-way interactions between predictors, as well as only the main effect of each predictor. In this way, the number of parameters in the model was reduced, which improved the model outcomes for small sample sizes.

It must be stressed that many of the above results ceased to be statistically significant when a Holm-Bonferroni correction for multiple comparisons was used (see Tables 2, 4 and 5). Note that the fact that a statistically significant effect was identified by one or more predictors only in very few cases, is very likely also due to the small sample sizes being analyzed, that is, the samples are too small to allow for the identification of a significant effect even if one is present.

3.3.1 Testing the new model: individual entheses

In the case of the humeri (Table 2), only the main effect of each predictor could be explored using GLM, since all models which simultaneously incorporate the main effect and the interaction between predictors exhibited computational errors due to small sample sizes. The only exception was the right anconeus (see below). It can be seen that for the right humerus age is statistically significant in the case of the subscapularis ($p = 0.002$), while its p-value is relatively close to statistical significance for the common extensor origin ($p = 0.067$) and the common flexor origin ($p = 0.063$). In addition, bone size is statistically significant for the supraspinatus ($p = 0.004$) and anconeus ($p = 0.043$). In contrast, the type of activity (manual/nonmanual) does not have a significant effect for any enthesis on the right side. For the left side, subscapularis is significantly affected by the type of activity ($p = 0.031$), age ($p = 0.001$) and bone size ($p = 0.012$), whereas no other enthesis appears to be significantly influenced by any of the examined factors. In the case of the right anconeus, for which the interaction between variables could also be incorporated in the model without computational issues, none of the variables exhibited a significant impact on EC presence (p always > 0.05).

The interaction between variables could be taken into consideration in the GLM along with the main effect of each predictor in most cases for the entheses of the radius and ulna. Table 4 demonstrates that on the right side, the type of activity is statistically significant for the triceps ($p = 0.05$), as is size ($p = 0.03$), whereas for the biceps it is only age that has an effect ($p = 0.005$). In contrast, no factor has a significant effect on the entheses of the left side. In respect to the interaction between predictors, only the interaction between age and size is significant in the case of the right triceps ($p = 0.016$), while it is very close to the significant level for the right biceps ($p = 0.055$). When the interaction between predictors is removed from the model and only the main effects are examined, the only significant effect is that of age for the left biceps ($p = 0.029$), while the type of activity for the right triceps is also very close to the level of statistical significance ($p = 0.053$).

3.3.2 Testing the new models: joint complexes

When multiple entheses are combined in joint complexes, although the sample sizes increase slightly, computational restrictions remain. Table 5 shows that, in the model including main effects and interactions, the type of activity and age are significant for the right shoulder ($p = 0.047$ and $p < 0.001$, respectively), as is their interaction ($p = 0.003$), but no other factor or interaction between factors appears to have a significant effect on any of the joints under study. When only the main effect of each predictor is explored, age is significant for the right shoulder ($p = 0.003$) and bone size for the left shoulder ($p = 0.036$). None of the GLM comparable odds ratios are significant (Table 5). However, for the joint complex taken as a whole (Table 5) the odds ratio for the right elbow shows a difference between manual and nonmanual occupations ($p=0.001$).

3.4. Boneformers

Computational problems meant that it was not possible to use GLM to determine whether age, body size, occupation category or the nature of boneforming was the primary cause of EC in those individuals classified as boneformers. For these individuals it became apparent that they were older than the rest of the sample (Table 6), but were a similar size (unbiased d is lowest for the right vertical head diameter of the humerus is 0.06 and highest for right antero-posterior diameter of the radius at 0.38). Boneformers had a much higher EC frequency for most entheses (Table 6). An age-matched control group was created to test whether the primary effect on EC presence was age. This was created by using non-boneformer individuals of the same or ±1 year difference to the boneformer sample, the sample was also balanced in terms of occupation classification with an odds ratio of 0.95 for the difference in occupation categories between the two groups. No large differences in EC frequency were found between the two age-matched samples. This indicates that age is likely to have been the primary factor in this difference. A comparison between occupation types was inappropriate due to the small sample size.

4.1 Discussion

The first aim of this paper was to test a model developed using GLM (Villotte et al. 2010) on identified skeletal samples with a sample size approximately typical of archaeological assemblages. The model performed badly for some age categories caused by sample sizes creating an abnormal spread of EC frequencies, e.g. the range of 0 to 100% EC presence for the right side for nonmanual workers (n=2), whereas the model predicts a frequency of 26% (Fig. 2). The difference in frequency is less of a problem than the fact that the shape of the model and values do not completely overlap. This is also a reflection of sample size which gives individuals or individual entheses a greater impact on pooled EC frequency than would occur in a much larger sample. This is also an effect of the age categorisation. Increasing the age range for each age category could improve sample size for small samples and the model clearly shows a dramatic increase in EC frequency between the ages of 40 and 50 (Figs. 1 and 2) as has been discussed elsewhere (Villotte et al., 2010). The original data should be used to develop such a model, which would be more useful for archaeological samples for which age categories are harder to determine accurately. Such a model should also be tested on small sample sizes to determine whether it is appropriate.

The second aim was to test whether new models can be effectively generated using GLM when the samples of the material under study are small. The models in the current study differed from the one by Villotte et al. (2010) in that they took into account body size and focused on individual entheses as well as joint complexes. For the present models real age, rather than age categories were used and z-scores were employed to standardise body sizes. However, the small sample size meant that in many cases the assumptions of GLM were violated. Where those assumptions were not violated, no single factor was found to systematically affect ECs. It is noteworthy that activity-pattern was only found to be a significant factor for one joint (right shoulder) and no entheses, while even this one case did not appear to be significant after a Holm-Boferroni correction was used. Age and body size were found to have a significant effect in certain cases, but these were very few, especially after the Holm-Bonferroni correction. Previous studies, using a related statistical method, logistic regression, have demonstrated that ageing and size play an important role in EC frequency and enthesis size (Alves Cardoso and Henderson, 2013; Nolte and Wilczak 2013). These studies were undertaken on larger sample sizes, so sample size is likely the key factor in the findings of this study. The most important observation from this analysis is that the significance of each factor differs when interactions are included in the model. This highlights the importance of assessing multiple predictors simultaneously in the study of phenomena with a multifactorial aetiology, such as ECs, since the impact of each individual predictor is affected by that of the remaining ones. However, the present study also demonstrates that taking into account multiple predictors is very difficult when small sample sizes are available due to the quasi-complete separation in the data, which causes computational errors and often fails to identify a statistically significant effect even if one is present.

What is important to note is that the odds ratios, which compared manual and nonmanual workers, do not entirely mirror the results of the GLM (example triceps brachii in Table 4) indicating the importance of considering the other aetiological factors (e.g. age and body size) in EC presence. This further demonstrates the importance of using models which can take into account multiple effects. Nevertheless it is important to present odds ratios to enable comparisons with other samples for meta-analyses (Henderson, 2013a) and where assumptions are violated or sample sizes are too small. For this study, the effect of boneforming was not analysed using GLM for these reasons. Boneformers were found to have a higher frequency of ECs than the rest of the sample, using odds ratios. However, the difference in age profile is likely the cause of this, based on the odds ratios of the age-matched sample. However, multiple effects could not be studied, nor could their interaction, using this method. The impact of boneforming is an area which does require further study using larger sample sizes.

The study was limited by sample size which also impacted on the range of occupations represented. The conglomeration of four separate samples provided a means to create a sample which presented a more diverse range of occupations. This was also used in the original paper which tested whether this impacted on the results: it did not (Villotte et al. 2010). However, it does raise concerns regarding the heterogeneity of the sample geographically and temporally and the socio-cultural implications which this may have particularly on occupations and non-professional activities. This is exemplified by the tailors from Fewston who are known from documentary evidence to have been engaged in farming activities (Henderson et al., 2013). This heterogeneity may not be found in normal archaeological single cemetery samples. This is a factor which should be considered when developing and testing models. However, the nature of a model should mean that it is applicable outside the original population, therefore the impact of heterogeneity is unlikely to be a serious limitation in this study.

While the model prediction closely mirrored the real results for the older age categories, the problem of using ten year age categories, both in terms of reducing sample sizes and due to limitations of osteological ageing methods, mean that the model is not yet widely applicable. Neither are the single and pooled joint GLM models created here. Consequently, they cannot currently be recommended for use on archaeological collections. Further work is needed to develop a model which can be used on small sample sizes, particularly the need to recognise that some age categories are often under-represented archaeologically, as they are here (Table 1). This may be possible to achieve, for example, by creating larger age categories. Nevertheless, the statistical approaches used here, should be considered for archaeological analyses when studying phenomena of multifactorial aetiology. It is also important to present the data in a way which enables comparisons between studies, e.g. using descriptive statistics, odds ratios or effect sizes.

5.1 Conclusions

Generalised linear models offer a method to test the cause of phenomena with a multifactorial aetiology. They are particularly appropriate for biological phenomena where the effects are often measured in very different ways. The aim of this study was to determine whether a previously developed GLM method could be applied to a typically-sized archaeological sample, and to determine whether testing the interaction of body size with ageing and activity-pattern would create a better model.

The outcomes indicate that the size of the sample affects the frequencies of the ECs observed causing the original model to fail to accurately predict EC frequencies in this sample. The effect of small sample size was exacerbated by dividing the sample into ten-year age categories. It is recommended that, for archaeological studies, this approach should be avoided and larger age categories are created possibly based on a division between 40 and 50 years of age. The new model which was created demonstrated that body size and age should be taken into account, but that there is no clear pattern of interaction between EC presence, activity-pattern, age, and body size. It is, therefore, recommended, that further studies, using a larger sample size should be undertaken to test these effects using this statistical approach with the aim of creating a model which can be applied to archaeological sample sizes and on individuals whose age-at-death and occupation are not documented.

Acknowledgements
The authors would like to thank the reviewers for their helpful comments.

The first author would like to thank the Museum of London and St. Bride's Church Fleet Street for access to the skeletal collections in London, with particular thanks to Jelena Bekvalac and Dr Rebecca Redfern. The first author would also like to extend their thanks to Malin Holst of York Osteoarchaeology Ltd., John Buglass of John Buglass Archaeological Services, Washburn Heritage Centre and Fewston and Blubberhouses Parochial Church Council who have granted access to the skeletons. The Fenwick human osteology laboratory, Department of Archaeology, Durham University provided workspace and access to this collection. CH's contribution to this research was funded by Portuguese national funds through FCT, the Foundation for Science and Technology postdoctoral grant SFRH/BPD/82559/2011 and their funding for the research centre of CIAS - Research Centre for Anthropology and Health (FCT/PEst-OE/SADG/UI0283/2013). FCT are supported by POHP/QREN (Programa Operacional Potencial Humano/Quadro de Referência Estratégico Nacional) co-funded by the Portuguese Government and the European Social Fund of the European Union.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

Agresti A (2002) Categorical Data Analysis. Wiley, New York

Alves Cardoso F, Henderson CY (2010) Enthesopathy Formation in the Humerus: Data from Known Age-at-Death and Known Occupation Skeletal Collections. Am J Phys Anthropol 141:550-560

Alves Cardoso F, Henderson CY (2013) The Categorisation of Occupation in Identified Skeletal Collections: A Source of Bias? Int J Osteoarchaeol 23:186-196

Buikstra JE, Ubelaker DH (1994) Standards for Data Collection from Human Skeletal Remains, Arkansas Archeological Survey Research Series, Fayetteville, Arkansas

Caffell A, Holst M (2010) Osteological Analysis The Church of St Michael and St Lawrence, Fewston, North Yorkshire. York Osteoarchaeology Ltd. 1210. doi: 10.5284/1025701

Cardoso FA (2008) A Portrait of Gender in Two 19th and 20th Century Portuguese Populations: A Palaeopathological Perspective. Doctoral thesis, Durham University, Durham

Cowie R, Bekvalac J, Kausmally T (2008) Late 17th to 19th century burial and earlier occupation at All Saints, Chelsea Old Church, Royal Borough of Kensington and Chelsea. Museum of London Archaeology Services, London

Chrisletta P, Spini D (2004) An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. J Educ Behav Stat 29:421-437

Henderson CY (2008) When hard work is disease: the interpretation of enthesopathies. In: Brickley, M., Smith, M. (eds.) Proceedings of the Eighth Annual Conference of the British Association for Biological Anthropology and Osteoarchaeology, British Archaeological Reports: International Series, Oxford, pp. 17-25

Henderson CY (2009) Musculo-skeletal stress markers in bioarchaeology : indicators of activity levels or human variation. Doctoral thesis, Durham University, Durham.
Henderson, C., 2013a. Subsistence strategy changes: The evidence of entheseal changes. HOMO J Comp Hum Biol 64:491-508. doi: 10.1016/j.quaint.2013.07.032

Henderson CY (2013b) Technical note: Quantifying size and shape of entheses, Anthropol Sci 121:63-73

Henderson CY, Alves Cardoso F (2013) Preface to Special Issue Entheseal Changes and Occupation: Technical and Theoretical Advances and Their Applications. Int J Osteoarchaeol 23:127-134. doi: 10.1002/oa.2298

Henderson CY, Craps DD, Caffell AC, Millard AR, Gowland R (2013) Occupational mobility in nineteenth century rural England: the interpretation of entheseal changes. Int J Osteoarchaeol 23:197-210

Jurmain RD, Alves Cardoso F, Henderson C, Villotte S, (2012) Bioarchaeology's Holy Grail: the reconstruction of activity. In: Grauer, A. (ed.) Companion to Paleopathology, Wiley-Blackwell, Chicester, England, pp. 531-552

Kuorinka I, Forcier L (1995) Work related musculoskeletal disorders (WMSDs): a reference book for prevention. Taylor and Francis, London

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13-22

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edition. Chapman and Hall, London

Milella M, Belcastro MG, Zollikofer CP, Mariotti V (2012) The effect of age, sex, and physical activity on entheseal morphology in a contemporary Italian skeletal collection, Am J Phys Anthropol 148:379-388

Milella M, Alves Cardoso F, Assis S, Perréard Lopreno G, Speith N (2014) Exploring the relationship between entheseal changes and physical activity: a multivariate study. Am J Phys Anthropol doi: 10.1002/ajpa.22640

Miles A, White W, Tankard D (2008) Burial at the site of the parish church of St Benet Sherehog before and after the Great Fire. Excavations at 1 Poultry, City of London. Museum of London Archaeological Service publications, London

Molenberghs G (2010) Generalized Estimating Equations Notes on the Choice of the Working Correlation Matrix. Method Inform Med 49:419–420

Molinier F, Laffosse JM, Bouali O, Tricoire JL, Moscovici J (2011) The anconeus, an active lateral ligament of the elbow: new anatomical arguments. Surg Radiol Anat 33: 617-621

Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. Biolog Rev 82:591-605

Niinimäki S (2012) The Relationship Between Musculoskeletal Stress Markers and Biomechanical Properties of the Humeral Diaphysis. Am J Phys Anthropol 147:618-628

Niinimäki S, Niskanen M, Niinimäki J, Nieminen M, Tuukkanen J, Junno J-A (2013) Modeling skeletal traits and functions of the upper body: Comparing archaeological and anthropological material. J Anthropol Archaeol 32:347-351 doi: 10.1016/j.jaa.2012.01.004

Nikita E (2014) The use of generalized linear models and generalized estimating equations in bioarchaeological studies. Am J Phys Anthropol 153:473-483

Nolte M, Wilczak C (2013) Three-dimensional Surface Area of the Distal Biceps Enthesis, Relationship to Body Size, Sex, Age and Secular Changes in a 20th Century American Sample. Int J Osteoarchaeol 23:163-174

Scheuer JL, Bowman JE (1995) Correlation of Documentary and Skeletal Evidence in the St. Bride's Crypt Population, in Saunders, S.R. and Herring, A. (Eds.), Grave Reflections: Portraying the Past through Cemetery Studies. Canadian Scholars' Press Inc, Toronto, pp. 49-70

Shaw HM, Benjamin M (2007) Structure-function relationships of entheses in relation to mechanical load and exercise. Scand J Med Sci Sport 17:303-315

Villotte S, (2006) Connaissances médicales actuelles, cotation des enthésopathies: nouvelle méthode. Bulletins et Mémoires de la Société d'Anthropologie de Paris 18:65–85

Villotte S (2008) Enthésopathies et Activités des Hommes Préhistoriques-Recherche Méthodologique et Application aux Fossiles Européens du Paléolithique Supérieur et du

Mésolithique. Doctoral thesis. Université Sciences et Technologies-Bordeaux I

Villotte S, Castex D, Couallier V, Dutour O, Knüsel CJ Henry-Gambier D (2010). Enthesopathies as occupational stress markers: evidence from the upper limb. Am J Phys Anthropol 142:224–34

Weiss E, Corona L, Schultz B (2012) Sex Differences in Musculoskeletal Stress Markers: Problems with Activity Pattern Reconstructions. Int J Osteoarchaeol 22:70–80

Wilczak CA (1998) Consideration of sexual dimorphism, age, and asymmetry in quantitative measurements of muscle insertion sites. Int J Osteoarchaeol 8:311–325

Table 1. Individuals recorded with occupation, occupation category, age, age category and whether they are boneformers. Question marks indicate individuals whose occupation is not certain. “X” indicates which individuals were used in the age-matched sample to compare EC frequency between boneformers and non-boneformers.

Site code	Skeleton	Age	Age Category	Occupation	Occupation category	Disease presence	Used in boneformer age-matched sample
Chelsea Old Church	OCU00_35	35	B	Proprietor Chelsea Bun House	Nonmanual		
Chelsea Old Church	OCU00_198	44	C	Gentleman	Nonmanual		
Chelsea Old Church	OCU00_462	61	E	Gentleman (with chambers in Temple)	Nonmanual		X
Chelsea Old Church	OCU00_147	67	E	Gentleman	Nonmanual	Boneformer	Boneformer
Chelsea Old Church	OCU00_713	68	E	Gentleman	Nonmanual	Boneformer	Boneformer
Chelsea Old Church	OCU00_701	78	E	Brick layer	Manual		
Chelsea Old Church	OCU00_681	84	E	Butcher, beadle of the parish	Manual	Boneformer	Boneformer
Chelsea Old Church	OCU00_622	84	E	Proprietor Chelsea Bun House	Nonmanual		X
Fewston	SLF09 342	26	A	Grocer's apprentice, then farm labourer	Manual		
Fewston	SLF09 119	38	B	Farmer	Manual		
Fewston	SLF09 339	41	C	Tailor	Manual		
Fewston	SLF09 351	63	E	Stone mason and registrar	Manual		X
Fewston	SLF09 130	66	E	Farmer	Manual		X
Fewston	SLF09 360	67	E	Farmer	Manual		X
Fewston	SLF09 366	76	E	Farmer	Manual		
Fewston	SLF09 307	78	E	Farmer	Manual	na	na
Fewston	SLF09 408	78	E	Tailor and farmer	Manual	na	na
Fewston	SLF09 226	84	E	Farmer	Manual		X
Fewston	SLF09 138B	na	na	Farmer	Manual	na	na
St. Benet Sherehog	ONE94_761	35	B	Gentleman?	Nonmanual		
St. Benet Sherehog	ONE94_356	39	B	Merchant, Mayor	Nonmanual		
St. Benet Sherehog	ONE94_387	46	C	Licensee of The Green Man	Nonmanual		
St. Benet Sherehog	ONE94_601	48	C	na	Nonmanual		
St. Bride's	SB50/57	22	A	Land surveyor	Nonmanual		
St. Bride's	SB51/50	25	A	Gentleman	Nonmanual		
St. Bride's	SB15/12	34	B	Brass founder	Manual		
St. Bride's	SB239/103	35	B	Gentleman	Nonmanual		
St. Bride's	SB14/10	36	B	Clerk in council office	Nonmanual		
St. Bride's	SB64/85	41	C	Licensed victualler	Nonmanual		
St. Bride's	SB191/156	42	C	Licensed victualler	Nonmanual		
St. Bride's	SB224/106	45	C	Jeweller	Manual		
St. Bride's	SB100/60	46	C	Lord Mayor of London; Merchant	Nonmanual		
St. Bride's	SB127/28	51	D	Coal merchant	Manual		
St. Bride's	SB233/64	53	D	Late ward Beadle	Nonmanual		
St. Bride's	SB181/164	55	D	Surgeon?	Nonmanual		
St. Bride's	SB118/34	56	D	Pastry cook?	Manual		
St. Bride's	SB138/70	60	E	Corn factor?	Nonmanual	Boneformer	Boneformer
St. Bride's	SB169/116	60	E	Secretary of Albion Fire and Life Insurance Co.	Nonmanual		X
St. Bride's	SB240/65	62	E	Baker?	Manual		X
St. Bride's	SB183/131	62	E	Gentleman	Nonmanual	Boneformer	Boneformer
St. Bride's	SB47/96	63	E	Lottery office keeper?	Nonmanual		X
St. Bride's	SB149/133	63	E	Venetian blind maker?	Manual	Boneformer	Boneformer
St. Bride's	SB84/47	63	E	Gentleman	Nonmanual		X
St. Bride's	SB188/184	63	E	Gold beater	Manual	Boneformer	Boneformer
St. Bride's	SB92/31	64	E	Gentleman	Nonmanual		X
St. Bride's	SB20/177	64	E	Governor of the Bank of England	Nonmanual	Boneformer	Boneformer
St. Bride's	SB231/91	64	E	Isinglass merchant	Nonmanual		X
St. Bride's	SB112/170	65	E	Jeweller	Manual		X
St. Bride's	SB71/21	68	E	Skinner	Manual		X
St. Bride's	SB158/180	70	E	Merchant	Nonmanual		X
St. Bride's	SB244/	71	E	Printer and novelist	Nonmanual	Boneformer	Boneformer
St. Bride's	SB166/149	72	E	Solicitor?	Nonmanual		X
St. Bride's	SB131/52	75	E	Vicar	Nonmanual		
St. Bride's	SB216/90	77	E	Packing case maker	Manual		
St. Bride's	SB58/172	77	E	Sheriff of London	Nonmanual		
St. Bride's	SB243/76	80	E	Book seller and church warden	Nonmanual		X
St. Bride's	SB136/17	82	E	Shoemaker	Nonmanual	Boneformer	Boneformer
St. Bride's	SB105/110_111	na	na	Farrier	Manual		

Table 2. Entheses of the humerus, descriptive statistics, odds ratios and GLM models. Odds ratios present the difference between the nonmanual (used as the control) and manual workers. Odds ratios and p-values (including those for GLM) are marked in bold are those which are statistically significant ($p<0.05$). GLM models presented are those without interactions.

			Supraspinatus	Subscapularis	Infraspinatus	Teres minor	Common Extensor Origin	Common Flexor Origin	Anconeus
Right	Nonmanual	EC present (n)	2	6	1	2	9	3	2
		N	9	13	9	7	20	13	20
	manual	EC present (n)	3	6	2	1	7	2	3
		N	10	11	8	5	12	10	11
	odds ratio		1.50	1.40	2.67	0.63	1.71	0.83	3.38
	p-value		0.715	0.696	0.473	0.749	0.476	0.869	0.229
	GLM model	AICC	24.137	20.166	23.281	-	36.269	22.351	29.875
		(Intercept)	0.11	0	0.401	-	0.08	0.029	0.005
		Type of activity	0.31	0.12	0.892	-	0.171	0.179	0.685
		Age	0.158	**0.002***	0.773	-	0.067	0.063	0.086
		Size	**0.004***	0.313	0.106	-	0.147	0.088	**0.043**
Left	Nonmanual	EC present (n)	6	6	4	3	4	2	1
		N	13	12	7	6	14	11	10
	manual	EC present (n)	2	6	2	2	4	2	2
		N	7	8	6	3	10	8	8
	odds ratio		0.47	3.00	0.38	2.00	1.67	1.50	3.00
	p-value		0.46	0.28	0.40	0.65	0.57	0.73	0.42
	GLM model	AICC	27.135	15.782	24.236	31.146	23.006	-	25.348
		(Intercept)	0.079	0.001	0.165	0.436	0.096	-	0.921
		Type of activity	0.124	**0.031**	0.37	0.725	0.965	-	0.069
		Age	0.077	**0.001***	0.199	0.389	0.115	-	0.559
		Size	0.257	**0.012***	0.095	0.339	0.352	-	0.665

**Remains statistically significant after Holm-Bonferroni correction*

Table 3. Measurements for size standardisation demonstrating differences between manual and nonmanual (used as control) occupations and between left (used as control) and right sides. Measurements in mm. Age presented by individuals, not side. Effect sizes 0.50 to 0.79 are considered medium, effect sizes over 0.80 are considered large (both marked in bold).

			Age	Humerus maximum length	Humeral vertical head diameter	Humeral transverse head diameter	Condylar width	Radius maximum length	Radial A-P diameter	Radial M-L diameter
Right	n	nonmanual	27	10	12	9	20	8	21	21
		manual	19	7	16	10	14	8	14	14
	mean	nonmanual	53.15	338.0	45.54	43.98	44.60	245.75	12.38	16.22
		manual	60.68	326.1	46.75	43.87	46.66	242.00	13.19	17.05
	std	nonmanual	17.13	22.8	1.96	2.20	2.30	17.81	1.21	1.66
		manual	17.11	12.3	3.34	2.26	4.39	13.04	1.52	2.10
	unbiased d		-0.43	**0.59**	-0.42	0.05	**-0.61**	0.2	**-0.59**	-0.44
Left	n	nonmanual	na	11	13	9	13	6	20	20
		manual	na	6	11	8	14	9	15	15
	mean	nonmanual	na	327.9	45.32	42.80	44.40	233.0	12.43	16.00
		manual	na	329.3	46.65	43.02	46.67	238.8	12.49	15.98
	std	nonmanual	na	19.1	1.90	2.96	2.34	9.9	0.88	1.72
		manual	na	17.2	3.54	2.51	3.56	11.6	1.03	2.66
	unbiased d		na	-0.08	-0.46	-0.07	**-0.73**	-0.50	-0.06	0.01

Table 4. Entheses of the radius and ulna, descriptive statistics, odds ratios and GLM models. Odds ratios present the difference between the nonmanual (used as the control) and manual workers. Odds ratios and p-values (including those for GLM) are marked in bold.

		Right			Left		
		Biceps b.	**Brachialis**	**Triceps b.**	**Biceps b.**	**Brachialis**	**Triceps b.**
Nonmanual	EC present (n)	12	11	2	9	11	3
	N	23	21	23	18	20	23
manual	EC present (n)	10	11	7	10	9	2
	N	15	11	12	14	15	14
odds ratio		1.83	**21.00**	**14.70**	2.50	1.23	1.11
p-value		0.385	**0.043**	**0.004**	0.228	0.780	0.922
GLM model with interactions	AICC	49.218	-	36.912	-	52.04	31.273
	(Intercept)	0.013	-	0.634	-	0.097	0.282
	Type of activity	0.089	-	**0.05**	-	0.96	0.181
	Age	**0.005***	-	0.239	-	0.083	0.074
	Size	0.074	-	**0.03**	-	0.978	0.267
	Type of activity * Age	0.08	-	0.105	-	0.708	0.153
	Type of activity * Size	0.606	-	0.398	-	0.538	0.584
	Age * Size	0.055	-	**0.016***	-	0.906	0.243
GLM model without interactions	AICC	47.099	-	31.298	36.767	-	28.397
	(Intercept)	0.13	-	0.395	0.046	-	0.342
	Type of activity	0.983	-	0.053	0.687	-	0.812
	Age	0.092	-	0.748	**0.029**	-	0.697
	Size	0.651	-	0.363	0.441	-	0.247

** Remains statistically significant after Holm-Bonferroni correction*

Table 5. Pooled entheses results: descriptive statistics, odds ratios and GLM models. Odds ratios present the difference between the nonmanual (used as the control) and manual workers. Odds ratios and p-values (including those for GLM) in bold indicate statistical significance (p<0.05). GLM was not undertaken on the entheses pooled using the Villotte method.

		Right				Left			
		Shoulder*	**Elbow****	**Hand/wrist*****	**Villotte pooling method*******	**Shoulder***	**Elbow****	**Hand/wrist*****	**Villotte pooling method*******
Nonmanual	EC present (n)	11	25	12	33	19	23	6	31
	N	38	67	33	87	38	61	25	75
manual	EC present (n)	12	28	9	30	12	21	6	26
	N	34	38	22	66	24	43	18	53
odds ratio		1.34	**4.70**	1.21	1.36	1.00	1.58	1.58	1.37
p-value		0.576	**0.001**	0.747	0.355	1.000	0.262	0.513	0.394
odds ratio comparable to GLM		1.31	1.94	1.86	na	0.83	2.36	1.83	na
p-value		0.15	0.367	0.342		0.101	0.474	0.335	
GLM model with interactions	AICC	36.909	52.089	46.574	na	-	-	-	na
	(Intercept)	0	0.517	0.229		-	-	-	
	Type of activity	**0.047**	0.915	0.861		-	-	-	
	Age	**<0.001†**	0.135	0.199		-	-	-	
	Size	0.258	0.806	0.532		-	-	-	
	Type of activity * Age	**0.044**	0.962	0.966		-	-	-	
	Type of activity * Size	0.484	0.29	0.175		-	-	-	
	Age * Size	0.197	0.972	0.626		-	-	-	
GLM model without interactions	AICC	28.945	44.787	39.377	na	26.146	-	-	na
	(Intercept)	0.005	0.469	0.071		0.256	-	-	
	Type of activity	0.612	0.899	0.265		0.584	-	-	
	Age	**0.003†**	0.121	0.079		0.148	-	-	
	Size	0.665	0.362	0.207		**0.036**	-	-	

**Consists of the insertions of supra- and infraspinatus, subscapularis and teres min.*

***Consists of the insertions of biceps b. brachialis and triceps b.*

****Consists of the common extensor and flexor origins*

*****Consists of the insertions of the supra- and infraspinatus, subscapularis, common extensor and flexor origins, and biceps b. No GLM created.*

†Remains statistically significant after Holm-Bonferroni correction

Table 6. Comparison of boneformers and non-boneformers (used as control) for enthesis presence: descriptive statistics and odds ratios. Odds ratios and p-values in bold indicate statistical significance (p<0.05)

			Right Shoulder	**Right Elbow**	**Right Hand/wrist**	**Right Side Villotte pooling method**	**Left Shoulder**	**Left Elbow**	**Left Hand/wrist**	**Left Side Villotte pooling method**
All data*	Non-Boneformer	EC present (n)	22	46	19	30	41	10	57	53
		N	71	98	53	61	98	41	147	124
	Boneformer	EC present (n)	17	20	6	14	21	7	26	28
		N	23	26	12	18	28	13	39	38
	odds ratio		**6.31**	**3.77**	1.79	**3.62**	**4.17**	3.62	**3.16**	**3.75**
	p-value		**0.001**	**0.009**	0.373	**0.038**	**0.003**	0.053	**0.003**	**0.001**
Age matched**	Non-Boneformer	EC present (n)	11	18	10	20	20	6	26	34
		N	24	31	21	27	37	17	52	54
	Boneformer	EC present (n)	17	20	6	14	21	7	26	28
		N	23	26	12	18	28	13	39	38
	odds ratio		3.35	2.41	1.10	1.23	2.55	2.14	2.00	1.65
	p-value		0.05	0.14	0.90	0.79	0.09	0.32	0.11	0.29

* Age unbiased d = 0.81
** Age unbiased d = 0.05

Figure 1. Plot of Villotte model created from the sample versus the frequency of EC in the sample for the left side. Boxplots represent the frequency from the sample and the 1st and 3rd interquartile ranges. Grey starts represent the model prediction with points representing the lower and squares the upper 95% confidence intervals. Age categories: A= 20-29, B=30-39, C=40-49, D=50-59, E=60+. Occupation categories 0=nonmanual, 1=manual.

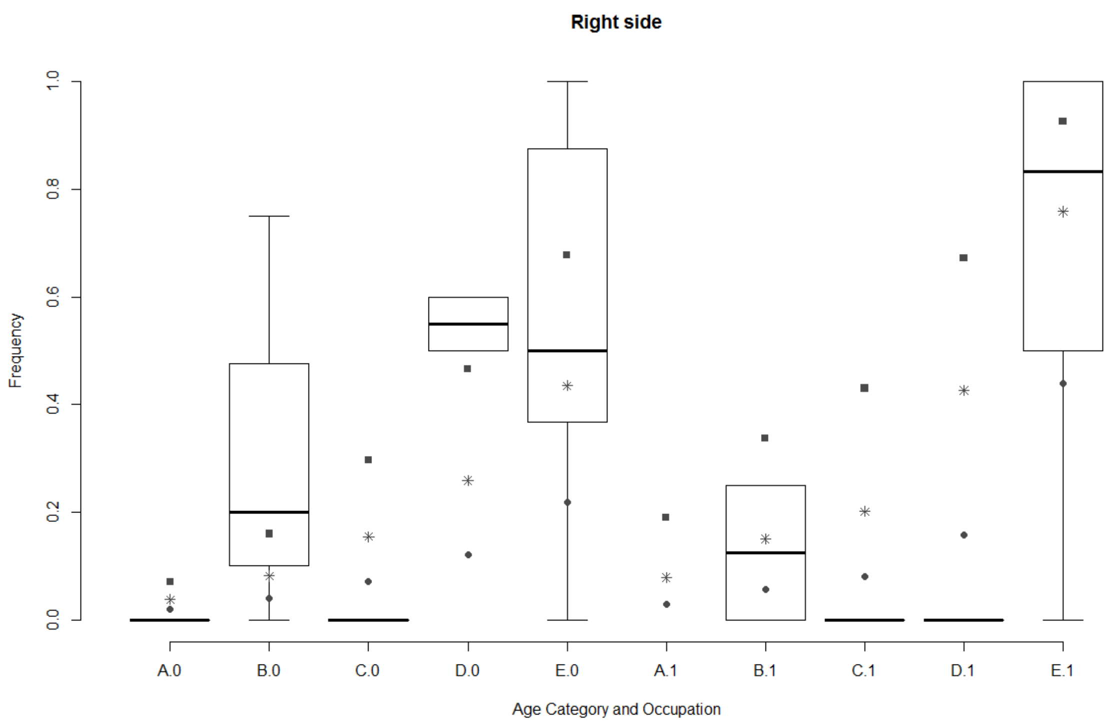

Figure 2. Plot of Villotte model created from the sample versus the frequency of EC in the sample for the right side. Boxplots represent the frequency from the sample and the 1st and 3rd interquartile ranges (black circle represents an outlier outside the interquartile range). Grey starts represent the model prediction with points representing the lower and squares the upper 95% confidence intervals. Age categories: A= 20-29, B=30-39, C=40-49, D=50-59, E=60+. Occupation categories 0=nonmanual, 1=manual.