Title:

A New Multivariate Method for Determining Sex of Immature Human Remains using the Maxillary First Molar

Short Running Title:

A Method for Sexing Immature Remains using the Upper M1

Authors and Institutional Affiliations:

Christopher Aris[1,2], Pia Nystrom[2], Elizabeth Craig-Atkins[2]

[1]Human Osteology Lab, Skeletal Biology Research Centre, School of Anthropology and Conservation, University of Kent, Canterbury, UK

[2]Department of Archaeology, University of Sheffield, Sheffield, UK

Key Words:

Juvenile, Permanent Dentition, Sex Assessment, Dental Metrics

ABSTRACT

Objectives

This study investigated the use of sexually dimorphic metrics of the first permanent maxillary molar (M^1) to determine sex in adult and immature individuals within and between populations.

Methods

Ten M^1 dimensions were measured in 91 adults (19-55 years) and 58 immatures (5-18 years) from two English populations, one of documented sex (Spitalfields crypt) and another of morphologically-assigned sex (Black Gate). Preliminary statistical analysis was undertaken to explore bilateral differences and variation by age and sex, followed by multivariate analyses to predict sex from dental metrics.

Results

Both cross-validated linear discriminant analysis and binary logistic regression predicted biological sex consistent with known sex in 94.6% of adults and 90.9% of immatures. When functions extracted from the Spitalfields data were used to assign sex to Black Gate adults, consistency with morphological sex varied from 83.3% to 57.7%. A new function developed on Black Gate resulted in only a 4.8% increase in maximum accuracy but reduced bias. The immature cohort comprised 19 (52.8%) males and 17 (47.2%) females.

Conclusions

This study demonstrates substantial sexual dimorphism in a single tooth which is commonly preserved in archaeological and forensic contexts. It successfully assigns biological sex to immatures from five years of age with substantially greater accuracy than any other morphological or metric method. We

suggest that accurate cross-population functions based on dentition require a trade-off between accuracy and applicability, and that functions extracted from populations of documented sex can be used to assign sex to other archaeological and forensic remains.

INTRODUCTION

The inability to accurately determine the biological sex of immature individuals from their skeletal remains is a shortcoming affecting both physical and forensic anthropology. In consequence, important archaeological questions concerning individual identity, demography, disease prevalence and gendered patterns of behavior and experience have not been adequately explored among children and vital work to identify and repatriate modern remains is curtailed. Problematically, the immature skeleton presents limited sexual dimorphism, which is not considered to become fully apparent until after puberty. This hinders the development of both morphological and metric approaches (Parfitt, Travers, Rauch, and Glorieux, 2000; Seeman, 2001). Nonetheless, attempts to assign skeletal sex to immatures most frequently utilize the same skeletal traits known to be accurate for determining the sex of adults, including multiple features of the pelvis (Olivares and Aguilera, 2016; Rissech and Malgosa, 2005; Schutkowski, 1993; Sutter, 2003; Vlak, Roksandic, and Schillaci, 2008) and skull (Galdames, Matamala, and Smith, 2008; Loth and Henneberg, 2001; Molleson, Cruse, and Mays, 1998; Schutkowski, 1993). Morphological methods have been criticized for high rates of inter-observer error (Cardoso and Saunders, 2008; Luna, Aranda, and Santos, 2017; Sutter, 2008). Consequently, efforts have been made to develop metric techniques (Decker, Davy-Jow, Ford, and Hilbelink, 2011; Krüger, L'Abbé, and Stull, 2017; López-Costas, Rissech, Trancho, and Turbón, 2012; Luna et al., 2017; Stull, L'Abbé, and Ousley, 2017; Wilson, MacLeod, and Homphrey 2008; Wilson, Cardoso, and Humphrey, 2011). Even though reportedly more repeatable, metric methods require a good standard of skeletal preservation, which cannot be guaranteed, especially when dealing with the fragile remains of immature individuals. Shifting focus to permanent dentition may provide an alternative approach to the assignment of biological sex to immatures. Indeed, permanent tooth crowns develop early within the chronology of immature growth and do not undergo remodeling throughout life. Therefore, should they present observable and reliable sexual dimorphism in adults, they may also be able to be utilized to assign sex to immature individuals.

This project explored the potential for biological sex of adults and immatures to be assessed on the basis of sexual dimorphism of the first permanent maxillary molar (M^1). The specific approach investigated required the development and testing of multivariate functions based on adult individuals, followed by their application to immatures to assign biological sex. The objectives of this study were:

(1) to use a known-sex population of adults to extract functions for the accurate prediction of biological sex on the basis of metric features of M^1 (2) to test the function on immature individuals from the same population, and evaluate its accuracy in comparison to documented sex (3) to investigate the population-specificity of the function through application to a second skeletal population from a similar geographic location but different time period (4) to evaluate whether a new, population-specific function is able to assign sex with greater accuracy to this second population.

Sexual dimorphism of dentition

Teeth are hard and durable structures which, unlike bone, are resistant to most forms of destructive post-mortem influences and therefore regularly appear undamaged in the archaeological record. As a result, a number of studies have investigated the possibility of using observable sexual dimorphism of dentition in order to determine sex from human remains. Most of these studies have focused on adult populations and found multiple metric dimensions to be significantly larger in adult males than in females, with the permanent maxillary first molar, second molar and canines yielding the most consistently dimorphic results (Brace and Ryan, 1980; Cardoso, 2008; Kondo, Townsend, and Yamada, 2005; Schwartz and Dean, 2005; Sonika, Harshaminder, Madhushankari, and Kennath, 2011; Viciano, López-Lázaro, and Alemán 2013; Zorba, Moraitis, and Manolis, 2011). There have also been studies concerning immature samples, including those analyzing permanent (Cardoso, 2008) and deciduous teeth (Black, 1978; Cardoso 2010; Viciano et al., 2013; Żądzińska, Karasińska, Jedrychowska-Dańska, Watala, and Witas, 2008). Even though they have produced positive results, studies of deciduous dentition are limited in how widely they can be used or replicated due to frequent poor preservation of deciduous teeth and the early age at which these are exfoliated. Thus, despite the wide scope of these papers, the potential of discriminating functions based on permanent dentition which can be applied to both adult and immature individuals has yet to be adequately realized.

The M^1 presents particular opportunities for the exploration of sexual dimorphism in dentition. In addition to being regularly identified as sexually dimorphic in previous studies, the M^1 is also one of the earliest developing permanent teeth. Its four main cusps complete calcification by 42 weeks post conception (Butler, 1967). Between the ages of three to four years the M^1 forms its cemento-enamel junction with all crown diameters having formed by the age of five. At this point in development the process of eruption begins, with the M^1 reaching occlusion between the ages of five to six years (Al Qahtani, Hector, and Liversidge 2010). Full root apex closure occurs around the age of seven to eight years. The comparatively early development of M^1 means that any sexual dimorphism discernible in its dimensions both develops and becomes observable long before puberty and the associated development of skeletal dimorphism. Multi-rooted teeth such as the M^1 are also more frequency recovered among archaeological assemblages than bi- or single-rooted teeth, leading to a higher frequency of recovery and the best chance of obtaining large sample sizes. Thus, any successful method based on the features of the M^1 would allow for reliable sex assignments to be made for a major fraction of an immature skeletal population, for whom no accurate morphological methods currently exist.

Using dentition to assign sex is not unproblematic. Immature individuals from archaeological populations represent non-survivors as a result of their early deaths. Such individuals are disproportionally likely to have experienced chronic disease and periods of biological stress. Skeletal manifestations of biological stress have been suggested to include reduction in the size of permanent teeth (Conceição & Cardoso, 2011; Flores-Mir, Raul Mauricio, Fernanda Orellana, and Major, 2005; Stojanowski, Larsen, Tung, and McEwan, 2007; Żądzińska, Lorkiewicz, Kurek, and Borowska-Strugińska, 2015). However, the evidence presented to support this claim has been sporadic and inconsistent between populations (Cardoso, 2008; Stojanowski et al., 2007). Not all immature deaths would have been associated with a prolonged period of biological stress, as some will have been the result of acute disease or accident. It is apparent that the impact of mortality bias on patterns in dental dimensions among archaeological populations requires further investigation. As it has the potential

to confound attempts to determine sex among immature individuals, this issue is considered explicitly in this paper.

MATERIALS AND METHODS

This project assessed M^1 teeth from 37 adult and 22 immature individuals from the crypt of Christ Church, Spitalfields, London (17^{th}-19^{th} century AD) and 54 adult and 36 immature individuals from the Black Gate cemetery, Newcastle-upon Tyne (c. 8^{th}-12^{th} century AD). Immature individuals were defined as those 18 years or under at death, thereby representing a group for whom conventional morphological sex assessment is rarely applied as a result of low reported accuracy.

The 59 individuals from the Spitalfields collection were of documented biological sex and age at death, obtained from personalized coffin plates recovered during excavation (Molleson and Cox, 1993). All individuals with at least one fully-formed and unworn M^1 were included in this study. The adult sample comprised 15 males and 22 females, who died between the ages of 23 and 52 years. The immature sample comprised 10 males and 12 females, ranging from five to 18 years at death (Figure 1). All measurements were taken before records of sex and age were checked to remove any possible bias.

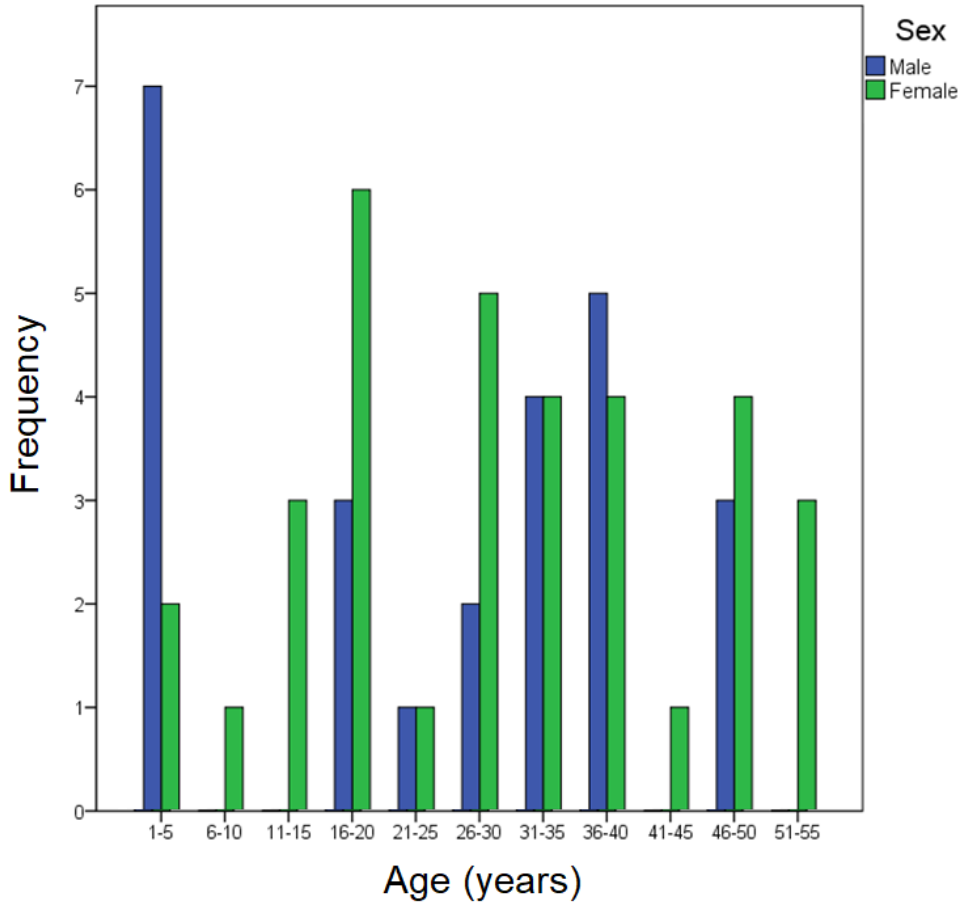

FIGURE 1. Demographic profile of the Spitalfields sample (n = 59)

The Black Gate adults comprised 31 males and 23 females, ranging from 19 to 55+ years of age at death. Data for biological sex were taken from the site archive, and had been obtained using a standard suite of sexually dimorphic traits of the skull and pelvis. Age at death was assigned using degeneration of the pubic symphysis and auricular surface and dental wear (Swales, 2012). The immature population comprised 36 individuals, ranging in age from five to 17 years (Figure 2). Age at death of the immature skeletons was also obtained from the site archive and was determined using epiphyseal fusion stages alongside dental development and eruption (Swales, 2012). Biological sex had not been assigned to the immatures.

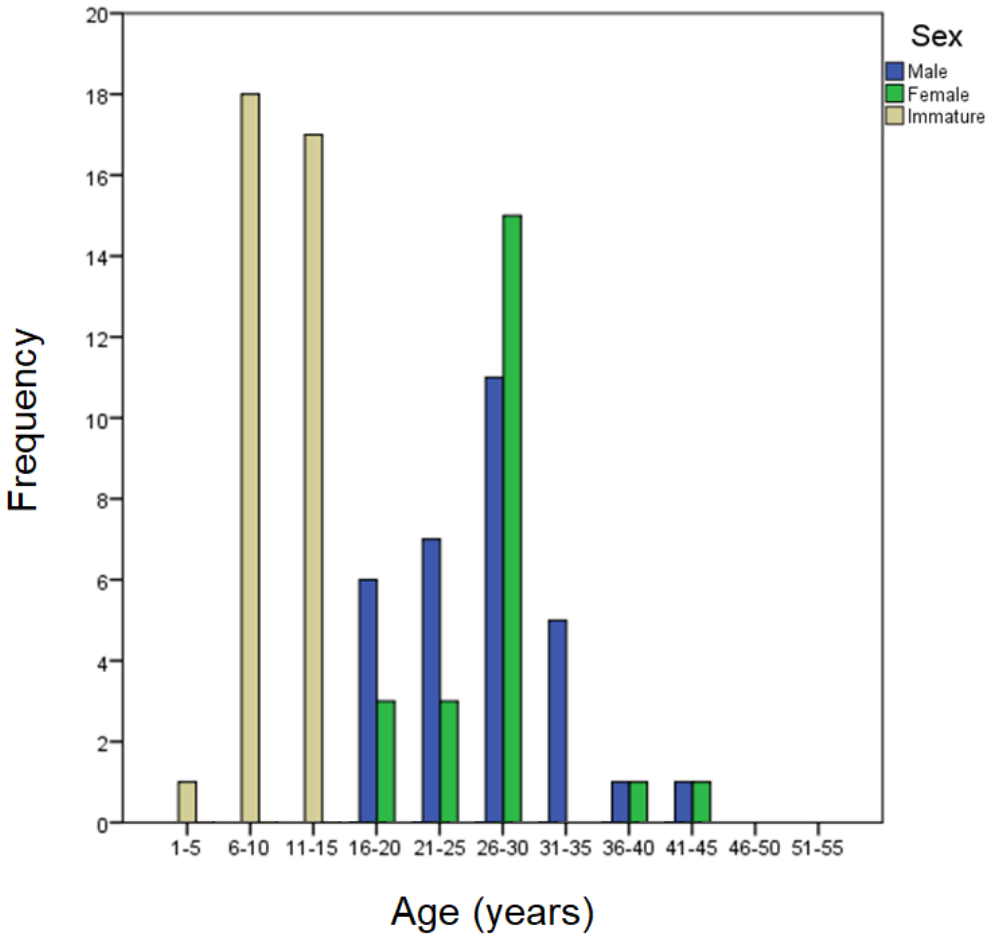

FIGURE 2. Demographic profile of the Black Gate sample (n = 90)

A total of 10 metric dimensions were measured across the crown and cervical neck of the first maxillary molars. These linear measurements were selected specifically for their previously-identified sexual dimorphism (Kondo et al., 2005; Schwartz and Dean, 2005; Zorba et al., 2011), and because they relate to features which fully mineralize early in childhood and are therefore unchanged in later life by root development (Garn, Lewis, Swindler, and Kerewsky, 1967; Schwartz and Dean, 2005). In younger individuals, where the molars had not fully erupted, the measurements were taken from teeth that were free from the dental crypt. The measurements comprised six diameters defined by Hillson et al. (2005) and four cusp lengths defined by Kondo et al. (2005) (Table 1).

Table 1. Definitions of the 10 dental metric features measured in this study.

Crown Measurements	
Buccallingual diameter (CrBL)	Maximum distance between the two parallel planes, one of which touches with the most buccal edge of the crown and the other the most lingual edge of the crown
Mesiodistal diameter (CrMD)	Maximum distance between the two parallel planes, one of which touches with the most mesial edge of the crown and the other the most distal edge of the crown
Mesiolingual-distobuccal diameter (CrMLDB)	Maximum distance between the mesiolingual corner of the crown and the distobuccal corner of the crown
Mesiobuccal-distolingual diameter (CrMBDL)	Maximum distance between the mesiobuccal corner of the crown and the distolingual corner of the crown
Cervical Measurements	
Buccallingual diameter (CvBL)	Maximum distance between the buccal and lingual surfaces of the cervical neck
Mesiodistal diameter (CvMD)	Maximum distance between the mesial and distal surfaces of the cervical neck
Cusp Measurements	
Metacone length (MeL)	Maximum diagonal distance between the central pit of the crown and the distobuccal corner of the crown associated with the metacone cusp
Paracone length (PaL)	Maximum diagonal distance between the central pit of the crown and the mesiobuccal corner of the crown associated with the paracone cusp
Hypocone length (HyL)	Maximum diagonal distance between the central pit of the crown and the distolingual corner of the crown associated with the hypocone cusp
Protocone length (PrL)	Maximum diagonal distance between the central pit of the crown and the mesiolingual corner of the crown associated with the protocone cusp

All measurements were taken using an electronic needle caliper calibrated to one hundredth of a millimeter (0.01mm) and accurate to one thousandth of a millimeter (0.001mm). The left M^1 was utilized wherever possible, with the right tooth only being used in instances where the left was missing, poorly preserved, too heavily worn or damaged. Measurements from a sub-sample of left and right teeth from both populations were assessed for bilateral asymmetry to investigate the impact

of this choice. Some measurements could not be taken from every tooth due to fragmentation, and in cases where dental attrition influenced the size of a dimension it was not recorded. In cases where wear was substantial, the tooth was excluded from analysis completely. In addition, some measurements were difficult to collect from teeth where they were still held within the alveolar bone adjacent to both the second premolar and second molar. While this was accommodated where possible by using the needle calipers, the mesiodistal crown diameter and mesiodistal cervical diameter presented particular difficulty with regards to positioning the beaks of the needle caliper in the correct position for measurement. For this reason, in a select few cases, the right tooth was used in place of the left where the right was free of the dental arcade. The samples included in this study (Spitalfields n=59, Black Gate n=90) represent all individuals from both populations for whom a complete set of measurements could be taken.

Statistical analysis

Statistical analyses were performed using SPSS 24.0. Paired-samples *t*-tests were used to test for differences between measurements of the left and right M^1, thereby ensuring that bias arising from asymmetry was not introduced through the substitution of right teeth in cases where the left was unmeasurable. In addition, two-sided independent samples *t*-tests and Pearson correlations were used to examine whether differences in dental dimensions existed between age cohorts. Any dimensions which were found to vary throughout the life course were excluded from further analysis of the associated population. Independent samples *t*-tests were also employed as a preliminary step in the exploration of sexual dimorphism to identify any significant differences in the mean values of the 10 dimensions of male and female individuals. The threshold for significance was set at $p<0.05$, with Bonferroni corrections applied to allow for the error incurred by conducting multiple independent statistical tests simultaneously (Field, 2009). Two measures of effect size were calculated to indicate the magnitude of differences between groups: *Pearson correlation coefficient* ® and *Cohen's d* (*d*). Cohen's (1988) guidelines for small (0.2), medium (0.5) and large (0.8) effects were followed. However, in response to the need to apply these guidelines tentatively and contextually, effect sizes were also compared within and between tests to provide a sense of comparative magnitude.

Two approaches were selected to explore multivariate relationships within the dental measurements and the extent of their association with biological sex: linear discriminant analysis (LDA) and binary logistic regression (BLR). These methods produced functions for the prediction of biological sex from dental metrics, while enabling both the relative contribution of each dental measurement to the prediction of sex to be assessed and the validity of the models to be quantified and compared (Krishan, Chatterjee, Kanchan, Kaur, Baryah, and Singh, 2016). For LDA, it is assumed that data have a multivariate normal distribution with equal covariance between variables whereas BLR does not make either of these assumptions. Should the data violate the assumptions of LDA, BLR would be expected to produce better results.

LDA has data requirements that must be considered further. Tests involving small sample sizes and large numbers of discriminating variables can result in overfitting and overly optimistic prediction accuracies (Concato, Feinstein, and Holford, 1993). The minimum sample size should be larger than

the number of independent variables used to group the sample (10 in the case of this study) and an ideal sample is around four or more times the number of independent variables (Poulsen and French, 2008). On this basis, sample sizes throughout this study must be considered small. In accordance with standard statistical procedures overfitting in LDA was addressed through Wilks' Lambda and the leave-one-out (LOO) cross-validation method (Subramanian and Simon, 2013). A stepwise approach was also used to select variables which made greatest contribution to prediction in the model, thereby providing a structured means of reducing the number of variables in the functions and guarding against overfitting. However, it is also possible for a model to perform poorly due to underfitting, which occurs when a model is overly simple and as such cannot adequately capture the underlying structure of the data (Everitt and Skrondal, 2010). As a key aim of this study was to examine the possibility of creating accurate functions which could be applied across populations, thus balancing over- and underfitting, a variety of functions extracted by both stepwise and non-stepwise approaches were evaluated. Overfitting in BLR was avoided by manual selection of subsets of variables that made greatest contribution to discrimination of the sexes.

First, each LDA and BLR was conducted on the adults of the Spitalfields collection to extract functions to predict sex from subsets of the ten dental dimensions. The adults from Spitalfields formed a 'trial' population of known sex from which to derive the function and test its accuracy. The same functions were then applied to the immature individuals from Spitalfields and compared to known sex to assess accuracy in a 'test' population.

The Black Gate adult sample was used to explore the population-specificity of the function. First, the Black Gate adult sample was assigned sex using the Spitalfields functions and these results compared to sex obtained from morphological sex assessment. Next, Black Gate-specific functions were extracted. The performance of the Spitalfields and Black Gate functions in the assignment of sex to the Black Gate adults and immatures was compared. Finally, the functions were applied to the Black Gate immature population to assign biological sex to these individuals for the first time.

Technical error of measurements

Before any statistical analysis was conducted, intra-observer error was quantified. A repeat set of measurements were made for 30 randomly-selected adults and immatures seven days after the initial data were produced. Technical error of measurements (TEM), relative %TEM and the coefficient of reliability® were calculated for all 10 linear measurements (Table S1). Intra-observer error for both the adults and immatures was extremely small, ranging from 0.02% (hypocone length, adults) to 0.65% (paracone length, immatures); additionally *R* was greater than 0.99 in all cases, meaning that less than 1% of the total variation in the data was due to observer error. This is well within an acceptable level, where any result of >0.95 indicates good quality control (Goto and Mascie-Taylor, 2006).

RESULTS

Developing the models: Spitalfields crypt

Descriptive statistics for the Spitalfields adults and immatures are presented in the supplementary files (Table S2, S3). Paired *t*-tests of 30 individuals from Spitalfields with two measurable M^1s indicated that there was no significant difference between right and left teeth for any of the 10 measurements (p=0.688-0.992) (Table 2). The dental dimensions of non-survivors (immatures) were not significantly smaller than those of survivors (adults), suggesting that any potential impact of biological stress in early life was negligible (p=0.114-0.997). In addition, there was a very low probability of there being correlations between age and crown metrics, and these correlations were weak (-0.300<r<0.300, eight of ten p>0.05) suggesting that once mineralized these dimensions did not change during any stage of life (Table 3). The means of all 10 variables were significantly larger in males than females. The effect size correlation was strong in all cases, and most dimensions displayed a difference of around two or three standard deviations between the means of the two sexes (Table 4).

Table 2. Comparison of difference in the mean values of dental metrics between the left and right M^1 teeth of Spitalfields adult individuals using paired sample t-tests.

Paired Samples Test						
	Paired Differences					
	N	Mean difference	t	r	d	Sig. (2-tailed)
CrMD	30	-0.03567	-0.188	0.02467	0.04871	0.851
CrBL	30	0.01467	0.077	0.01011	0.01996	0.939
CrMLDB	30	-0.01533	-0.091	0.01194	0.02341	0.928
CrMBDL	30	-0.00200	-0.010	0.00131	0.00272	0.992
CvBL	30	-0.04833	-0.230	0.03018	0.05803	0.819
CvMD	30	-0.01033	-0.060	0.00787	0.01530	0.952
MeL	30	-0.04267	-0.403	0.05284	0.10536	0.688
PaL	30	-0.00233	-0.019	0.00249	0.11851	0.985
HyL	30	0.00333	0.019	0.00249	0.00489	0.985
PrL	30	0.00433	0.034	0.00446	0.00866	0.973

Table 3. Statistical exploration of dental metrics in adult and immature individuals of the Spitalfields population. The t-tests compared the means of adults and immatures. Pearson correlations correlated dental measurements with median age at death. Significance for the tests was set at p<0.004.Significant results are marked in **bold.**

	N		t-test for Equality of Means					Pearson Correlations	
	Adult	Immature	Mean	t	r	d	Sig. (2-tailed)	Correlation	Sig. (2-tailed)
CrMD	37	22	10.25	0.488	0.0665	0.12619	.627	0.00	0.982
CrBL	37	22	10.76	-0.004	0.0005	0.00114	.997	-0.08	0.525
CrMLDB	37	22	10.53	0.151	0.0199	0.04415	.881	-0.05	0.685
CrMBDL	37	22	11.69	-0.005	0.0007	0.00128	.996	-0.08	0.538
CvBL	37	22	9.44	-1.481	0.1925	0.39096	.144	-0.27	0.370
CvMD	37	22	7.31	-1.620	0.2080	0.50874	.114	-0.29	0.230
MeL	37	22	4.94	-1.323	0.1715	0.43394	.194	-0.20	0.123
PaL	37	22	5.91	-0.045	0.0060	0.01521	.964	-0.11	0.409
HyL	37	22	5.56	-0.666	0.0875	0.20953	.510	-0.20	0.128
PrL	37	22	5.61	1.066	0.1392	0.27967	.293	0.04	0.735

Table 4. Results of independent t-tests to identify significant differences in the mean values of dental metrics between the male and female individuals from the Spitalfields population. Significance for this test was set at p<0.004. Significant results are marked in **bold.**

Independent Samples Test									
	N		t-test for Equality of Means						
	Males	Females	t	Max	Min	Mean	r	d	Sig. (2-tailed)
CrMD	15	22	8.672	11.81	9.15	10.25	**0.82607**	**3.58262**	**0.000**
CrBL	15	22	7.425	12.19	8.79	10.76	**0.78209**	**2.07601**	**0.000**
CrMLDB	15	22	6.935	11.83	9.00	10.53	**0.76078**	**2.31760**	**0.000**
CrMBDL	15	22	7.035	13.06	10.31	11.69	**0.76534**	**2.03785**	**0.000**
CvBL	15	22	7.110	10.97	7.10	9.44	**0.76869**	**1.86599**	**0.000**
CvMD	15	22	6.473	8.87	5.63	7.31	**0.73814**	**1.42315**	**0.000**
MeL	15	22	8.327	6.20	3.92	4.94	**0.81516**	**3.04181**	**0.000**
PaL	15	22	7.665	7.22	4.76	5.91	**0.79162**	**2.12139**	**0.000**
HyL	15	22	9.046	7.10	4.32	5.56	**0.83691**	**2.20388**	**0.000**
PrL	15	22	8.991	6.69	4.38	5.61	**0.83537**	**2.28466**	**0.000**

LDA and BLR analysis were performed on the 10 measurements from the adult sample (15 males, 22 females). The function extracted by LDA had valid discriminating ability (Wilks' lambda = 0.161; $p<0.0005$) and differentiated males (>0) from females (<0) as follows:

*Function Score = (CrMD*0.897)+(CrBL*0.344)+(CrMLDB*-0.221)+(CrMBDL*-0.787)+(CvBL*0.508)+(CvMD*0.507)+(MeL*1.784)+(PaL*0.648)+(HyL*0.894)+(PrL*0.936)-32.547*

The cross-validated results of the function were consistent with documented sex in 94.6% of individuals. The incorrect assignments were one male classified as female and one female classified as male, thus 93.3% of males and 95.3% of females were correctly assigned biological sex. The estimated probabilities of correct classification were lowest for the two incorrectly assigned individuals (55.0% and 92.3%, mean of correct classifications 100%) suggesting misclassification resulted from comparatively marginal sexual dimorphism.

Repeating the analysis with a stepwise approach assessed the potential for overfitting to have inflated the reported accuracy and enabled evaluation of the impact of various combinations of variables on the prediction of sex. Thresholds for minimizing Wilks' Lambda of $p(F)=0.05$ for entry and $p(F)=0.10$ for exclusion extracted the function:

(CrMD*1.669)+(HyL*1.684)-26.447

This had a cross-validated accuracy of 91.9%. Reduction of sensitivity to $p(F)=0.10$ for entry and $p(F)=0.15$ for exclusion produced the four-variable function:

(CrMD*1.087)+(MeL*1.528)+(HyL*1.066)+(PrL*1.137)-30.952

This function performed identically to that with ten variables (both equivalent accuracy and the same misclassified individuals). The simpler models are much less likely to overfit, so confirm the high accuracy of sex predicted from dental metrics.

In the next stage of analysis BLR functions were extracted for various combinations of variables. Functions with three or more variables produced complete discrimination, which is likely a negative consequence of small sample size and should be interpreted with extreme caution (Geyer 2009). A stepwise BLR retained the same two variables identified in stepwise LDA as follows:

Log (p/1-p) = *(CrMD*-4.819)+ (HyL*-6.413)+87.165*

This function also predicted sex accurately in 94.6% of the adults, however hypocone length did not make a significant contribution to the function (Wald chi-square = 2.306, $p=0.129$) suggesting that the crown mesiodistal diameter was sexually dimorphic in its own right. Indeed several dimensions (CrMD, CrMLDB, MeL, HyL) were able to discriminate between males and females with greater than 90% accuracy alone, and the remainder performed no worse than 81.1% accuracy. However, no single measurement predicted sex as accurately as the multi-variable functions.

When the three discriminant functions and BLR function were used to predict sex from M^1 measurements of the 22 immature individuals, 90.9% were correctly assigned biological sex in all cases. One female was misclassified as male and one male misclassified as female using the ten- and four-term discriminant functions and two males misclassified as females when the two-variable

discriminant and BLR functions were used. As with the adult data, misclassified individuals had a lower confidence assigned to the predictions than correctly classified individuals.

Testing the models for inter-population variation

To explore population-specificity, the functions extracted to distinguish between male and female individuals at Spitalfields were used to assign sex to the Black Gate adults (*n*=54). This process served to explore the magnitude of variation in the size and dimorphism of teeth between different populations from relatively similar geographic regions, and thereby whether this variation was substantial enough to affect cross-population application of the functions. As the Black Gate population is of unknown sex, the predictions were compared to morphologically-assigned biological sex to estimate accuracy.

Sex predicted from the 10-variable discriminant function agreed with morphological sex in 83.3% of cases. One individual morphologically assessed as male was predicted to be female and eight individuals morphologically assessed as females were predicted to be males. Therefore, the function performed exceptionally well on males (96.7%) but less well on females (65.2%). Sex predicted from the four-variable discriminant function agreed in fewer cases (78.8%), again performing better in males (87.1%) than females (66.7%). The two-variable functions performed worst with the two-variable discriminant function correctly assigning 57.7% individuals (58.1% males, 57.1% females) and the BLR function correctly assigning 63.5% (78.9% males, 45.5% females).

The performance of the Spitalfields function when applied to the Black Gate adults may be due to several factors. First, the relatively poor performance of the 2- and 4-variable functions suggests that the stepwise functions may be overfit to Spitalfields and that prediction of sex at Black Gate may depend on a different combination of dental metrics to those selected in the four- and two-variable functions. Second, the poorer performance of the functions on females at Black Gate may arise from inter-population variation in the magnitude of sexual dimorphism in the M^1. Both of these outcomes will have been influenced by variation in tooth size between the two populations. An alternative, but important, consideration is that morphological sex assessment has limitations to its own accuracy, and therefore we do not know whether the individuals inconsistently assigned sex were, in fact, assigned the wrong morphological sex. To explore this issue, we can generate new functions to predict sex based on the Black Gate individuals and consider whether consistency with biological sex improves when population-specific functions are utilized.

Developing new functions: Black Gate

Descriptive statistics for the Black Gate adults and immatures are presented in the supplementary files (Table S4, S5). Paired *t*-tests of individuals with two measurable M^1s (*n*=19-30) indicated that there were no significant differences between right and left teeth for the 10 measurements (p=0.063-0.904) (Table 5). The dental dimensions of non-survivors (immatures) were not significantly smaller than those of survivors (adults) for seven measurements (p=0.016-0.977). However, the hypocone and protocone lengths were significantly smaller in immatures and the paracone length significantly larger in immatures. These results were mirrored by correlations between the metric lengths and age with

the same three cusp lengths, which were moderate and significant (Table 6). As a result, the hypocone, protocone and paracone were discounted from further analysis of the Black Gate population. The means of all seven retained variables were larger in males than females, five of which achieved statistical significance. Of those five variables, all displayed only moderate Pearson's correlations and around one standard deviation in difference. The two variables which were not significantly different between the sexes displayed lower Pearson's correlations and less than 0.75 standard deviations difference between the means (Table 7).

Table 5. Comparison of difference in the mean values of dental metrics between the left and right M^1 teeth of Black Gate adult individuals using paired sample t-tests. Significant results marked in **bold.**

Paired Samples Test						
	Paired Differences					
	N	Mean difference	t	r	d	Sig. (2-tailed)
CrMD	29	0.01793	1.383	0.18018	0.08725	0.178
CrBL	29	0.00897	0.325	0.04333	0.01515	0.748
CrMLDB	29	-0.00276	-0.121	0.01602	0.00018	0.904
CrMBDL	26	0.01154	0.564	0.07797	0.10375	0.578
CvBL	29	0.03000	1.645	0.21469	0.06300	0.111
CvMD	30	0.05300	1.931	0.24577	0.11433	0.063
MeL	19	0.01000	0.604	0.09626	0.23199	0.553
PaL	19	-0.01211	-0.555	0.08852	0.08067	0.586
HyL	19	-0.00405	-0.306	0.04894	0.02962	0.763
PrL	20	0.05400	1.647	0.25200	0.11292	0.116

Table 6. Statistical exploration of dental metrics in adult and immature individuals of the Black Gate population. The t-tests compared the means of adults and immatures. Pearson correlations correlated dental measurements with median age at death. Significance for the tests was set at p<0.004. Significant results are marked in **bold.**

	N		t-test for Equality of Means					Pearson correlation	
	Adult	Immature	Mean	t	r	d	Sig. (2-tailed)	Correlation	Sig. (2-tailed)
CrMD	54	36	10.27	0.543	0.0579	0.12063	0.588	0.85	0.546
CrBL	54	36	10.82	-0.999	0.1058	0.21768	0.321	0.18	0.185
CrMLDB	54	36	10.90	-0.506	0.0539	0.10886	0.614	0.17	0.220
CrMBDL	54	36	12.08	-0.719	0.0764	0.15316	0.474	0.33	0.160
CvBL	54	36	10.14	1.209	0.1278	0.26046	0.230	0.16	0.246
CvMD	54	36	7.57	-0.266	0.0283	0.05636	0.791	0.01	0.907
MeL	54	36	5.19	-2.452	0.2533	0.53618	0.016	-0.18	0.188
PaL	54	36	6.30	-4.831	**0.4578**	**1.06497**	**0.000**	0.05	0.721
HyL	54	36	5.27	6.448	**0.5664**	**1.40437**	**0.000**	0.32	0.019
PrL	54	36	5.61	6.165	**0.5492**	**1.33735**	**0.000**	0.38	0.005

Table 7. Results of independent t-tests to identify significant differences in the mean values of dental metrics between the male and female adult individuals from the Black Gate population. Significance for this test was set at p<0.004. Significant results are marked in **bold.**

Independent Samples Test									
	N		t-test for Equality of Means						
	Males	Females	t	Max	Min	Mean	r	d	Sig. (2-tailed)
CrMD	31	23	2.455	12.00	8.82	10.30	0.3236	0.673365	0.017
CrBL	31	23	4.957	12.13	9.21	10.77	**0.5665**	**1.391337**	**0.000**
CrMLDB	31	23	3.139	12.02	9.78	10.88	**0.3964**	**0.86603**	0.**003**
CrMBDL	31	23	2.618	13.02	10.89	12.04	0.3396	0.714253	0.012
CvBL	31	23	6.623	11.18	8.98	10.20	**0.4114**	**1.801689**	**0.002**
CvMD	31	23	3.217	8.75	6.13	7.56	**0.4074**	**0.895361**	**0.000**
MeL	31	23	4,398	6.29	4.18	5.10	**0.5207**	**1.237633**	**0.000**

LDA was performed on the seven selected measurements from the adult sample (morphologically assigned as 31 males and 23 females). The function extracted from LDA had valid discriminating ability (Wilks' lambda = 0.351; p<0.0005) and differentiated males (score>0) from females (score<0) as follows:

*Function score = (CrMD*0.15)+(CrBL*1.422)+(CrMLDB*-0.639)+(CrMBDL*-0.824)+(CvBL*1.600)+(CvMD*0.235)+(MeL*1.688)-25.417*

The cross-validated results of the function were consistent with morphologically-assigned sex in 83.3% of individuals. Inconsistent assignments were five males classified as female and four females classified as male, thus 83.9% of males and 82.6% of females were consistently assigned biological sex. Stepwise LDA minimizing Wilks' Lambda with maximum p(F) for entry = 0.05 and minimum p(F) for exclusion = 0.10 extracted a function with two variables:

(CvBL*1.926)+(MeL*1.534)-27.466

This assigned sex correctly in 86.0% of individuals, however the highest consistency between predicted and biological sex was obtained when maximum p(F) for entry = 0.10 and minimum p(F) for exclusion = 0.15. Four of the seven features were included in this function, which predicted sex consistently with biological sex in 87.3% of individuals:

(CrBL*1.280)+(CrMDBL*-1.099)+(CvBL*1.667)+(MeL*1.571)-25.555

As a final stage of analysis we applied all of the functions reported in this paper to the unknown sex immature sample from Black Gate. While the accuracy of the assessments could not be quantified as no accurate comparative data exist, the utility of the functions we have extracted could still be evaluated by considering consistency of sex determinations made by the seven functions. A strong majority result (6/7 or 7/7 consistent results) was obtained for immature sex assessment in 67.7% (24/36) of individuals, with all but three of the remainder obtaining a 2:5 split in assessments of male and female (Table 8, S6). The most frequent inconsistencies included the assignment of male by the two-term Spitalfields functions (both LDA and BLR) to individuals otherwise consistently determined to be female (7/36, 36.8%) and in another six cases (6/36, 31.5%) assignment of male sex to individuals otherwise deemed to be female was made by the Spitalfields two-variable BLR function. This pattern is explained by the emphasis placed on both these functions on hypocone length, which was determined to be smaller (and therefore more likely to be assessed as female) among Black Gate immatures than adults. It is notable that the four-variable Spitalfields BLR, which includes three variables that were found to differ significantly between the sexes but not with age in addition to the hypocone, performed much more consistently with the Black Gate functions (34/36, 89.5%).

Table 8. Consistency of the seven different functions in sex assignment for the Black Gate immatures.

	Consistent assessments of sex				
	7	6	5	4	Total
Male	4 (21.1%)	6 (31.6%)	7 (36.8%)	2 (10.5%)	19 (100%)
Female	11 (64.7%)	3 (17.6%)	2 (11.8%)	1 (5.9%)	17 (100%)
Total	15 (41.7%)	9 (25%)	9 (25%)	3 (8.3%)	36 (100%)

DISCUSSION

Sexual dimorphism in the adult populations

The adult populations from the Spitalfields and Black Gate collections presented ten and five significantly sexually dimorphic morphological features of the M^1 respectively. These findings support claims made by other researchers that the dimensions of the M^1 are sexually dimorphic and can thereby be utilized to determine biological sex (Brace and Ryan, 1980; Cardoso, 2008; Colby, 1996; Garn, Cole, Wainwright, and Guire, 1977; Garn, Cole, and Wan Alstine, 1979; Hillson, FitzGerald, and Flinn, 2005; Kondo et al., 2005; Schwartz and Dean, 2005; Viciano et al., 2013; Zorba et al., 2011). Previously, cervical measurements have been reported to be more sexually dimorphic than crown measurements (Colby, 1996; Zorba et al., 2011), while lack of significant dimorphism has been detected in certain cusp lengths (Kondo et al., 2005) and diagonal crown diameters (Cardoso, 2008). Some of the previous findings were not supported by the data from Spitalfields, where all dimensions differed significantly between males and females, however at Black Gate one crown length and one diagonal crown diameter were not significant, supporting some previous findings (Cardoso, 2008; Zorba et al., 2001). A degree of variation in the extent of sexual dimorphism of M^1 between populations is to be expected as a result of genetic variability (Zorba et al., 2011), and was observed here; Black Gate presented less dimorphism overall than Spitalfields. However, the question we sought to address in this research was whether there is sufficient *shared* variation between adults and immatures and between different populations to enable accurate prediction of biological sex from the M^1.

Variation between the adult and immature metrics

A concern of this project was the possibility that mortality bias would affect the dimensions of the permanent dentition. If this were the case, the characteristics of sexual dimorphism presented by the adult and immature individuals would be different, and therefore incomparable. The majority of the M^1 dental features appear robust to environmental influences. This conclusion is consistent with the results of those osteological studies which found neither nutritional nor physiological stress evidenced in the skeleton to be related to the size of permanent dentition (Lewis, Shapland, and Watts, 2016a; Lewis, Shapland, and Watts, 2016b; Mays, 1995). However, three cusp measurements from Black Gate (paracone, hypocone and protocone length) were found to vary significantly between adults and immatures. Existing research regarding prenatal occlusal surface morphology indicates that at one year of age the M^1 cusps, while calcified, are still in the process of developing their final shape and relative dimensions (Butler, 1967). Peretz, Nevis, and Smith (1998) found that there was no significant correlation between the size of M^1 dimensions and the relationship of the cusps to each other, concluding that the growth of crown size and development of cusp morphology were independent in both pattern and rate. Therefore, it is possible that the central pit of the crown continues to vary in position with age even after the conclusion of diameter growth of the crown. One explanation for this may be that the development of root dentine and the pulp chamber could exert sufficient pressure on the occlusal portions of enamel to result in continued change in shape even after the occlusion of the tooth. There is no current evidence for this occurring on the M^1 tooth at an individual level, but it has been seen across the mandibular arcade (Demirjian, Goldstein, and Tanner, 1973). This developmental pressure may therefore explain why only cusp lengths of the M^1 varied to a significant level between the adult and immature Black Gate samples. Furthermore, the development of crown and cusp morphology of the M^1 has been found to be distinct between human populations (Gómez-Robles, Martinón-Torres, De Castro, Margvelashvili, Bastir, Arsuaga, and Martínez, 2007; Morris, 1986). If cusp lengths are found to vary with age in other populations then they should be excluded in future from any attempts to predict sex in immatures.

Statistical analysis: prediction of sex from dental metrics

Cross-validated discriminant functions extracted from combinations of 10 measurements of adult M^1 teeth from Spitalfields predicted sex consistent with documented sex in 94.6% of cases. When applied to immatures from the same population they assigned sex with 90.9% accuracy. Stepwise LDA resulted in limited loss of accuracy. A BLR model robust against violations of LDA's distributional assumptions was fit to the same data, resulting in the identification of the same variables in the model and an identical predictive accuracy. It was noted that the Spitalfields population presented clear dimorphism in the dentition, with a single tooth dimension predicting sex with comparable accuracy to the best morphological and metric methods currently available. These findings can be contrasted with previous studies based on cranial metrics which have found substantial loss of accuracy under cross validation (Nikita 2014), to support the argument that dental measurements make better predictors of sexual dimorphism in adults. The models' retention of high accuracy when applied to the immatures (thus tested on individuals who were not among the group from which the function was extracted) indicates the ability to accurately predict sex in individuals as young as five years of age within a population with considerably greater accuracy that any morphological or metric method currently available.

Sex predicted for adults from the Black Gate population using the 10-variable Spitalfields function achieved 83.3% consistency with morphologically assigned sex, but performed with a higher degree of consistency for males than females. Loss of consistency was seen when functions extracted by stepwise methods were applied, suggesting they were overfit to unique features of the Spitalfields population. A discriminant function extracted from the Black Gate adults achieved similar consistency to the 10-variable Spitalfields function. However, unlike the Spitalfields function, the Black Gate function assigned sex to Black Gate adults with a similar consistency for both males and females. A stepwise procedure revealed that consistency of a Black Gate-specific function could be increased to 87.3% through the extraction of a function with only four variables.

An investigation of the population-specificity of the Spitalfields functions indicated variable loss in accuracy when applied to a second population with functions containing more variables performing better than functions with fewer variables. A decrease of only 4.8% in accuracy was observed between the best-performing cross-population function and the best performing within-population function (83.3% vs 87.3%). However, a bias was identified towards the identification of males as females in the Black Gate population. This could indicate that dimorphism in teeth is sufficiently variable between these two geographically-similar but chronologically-distinct populations to introduce this bias. However, we cannot rule out whether there is a bias towards the assignment of female morphological sex to males at Black Gate that is influencing our assessment of accuracy. The latter is plausible, as inaccuracy and bias have both been reported in the assignment of sex from morphological traits of the pelvis and skull. Widely-reported accuracy of these methods is around 80% for the cranium, while pelvic features are often cited as having accuracies of over 90% (Ferembach, Schwidetsky, and Stloukal, 1980; Krogman and İşcan, 1986; Loth and Henneberg, 1996; Phenice, 1969). Bias has also been reported towards one or other sex (Ubelaker 2002). It remains possible that the assignment of sex from dentition of the Black Gate individuals is, in fact, more accurate than the assignment based on morphological methods that we compared it with.

The process of development and testing of functions for the accurate prediction of biological sex from metric features of M^1 enabled reflection on the impact of different statistical approaches. Comparison of LDA and BLR functions indicated limited increase in performance of the latter and suggested that the simpler method capture the variation present in the data. Stepwise methods, used to alleviate overfitting and over-inflation of predictive accuracy, actually resulted in equivalent or greater within-population prediction accuracy. This, combined with the fact that the function developed at Spitalfields using the stepwise method performed worse when applied to Black Gate than the full 10-variable function, suggests that stepwise methods increased the population-specificity of the models by identifying subsets of variables that very accurately predicted sex in one population but not in the other. While stepwise methods may avoid the potential pitfalls of spurious accuracy, preference for their usage in physical anthropology may be contributing to an exaggerated population-specificity of predictive models based on stepwise LDA. A different approach is needed whereby cross-population applicability is better balanced against within-population accuracy if we are going to evaluate the true potential of sex assessment from the dentition, or indeed any other morphological skeletal features, across populations. The findings here suggest that even a poorly-tailored function with multiple redundant variables may predict sex with greater accuracy than most commonly-used morphological methods and so functions with less accuracy but greater applicability may still be highly valuable to our field. Cross-referencing between different skeletal collections of documented sex would be very

useful to establish the extent to which the sexes can be differentiated in geographically- and chronologically-distinct populations, and when and how the functions should be deployed to best effect.

Conclusions

This project has tested the viability of utilizing measurements of early-developing permanent teeth to accurately predict biological sex in both adult and immature individuals. It has provided evidence for measurable sexual dimorphism in adults and immatures from the M^1, a single tooth commonly recovered and preserved in archaeological and forensic contexts. We have evaluated and tested two approaches to the assignment of sex using measurements of the M^1, one involving the cross-population application of a single function and the other based on the generation of population-specific functions from adults to apply to immatures. The first involved the extraction of functions to predict sex from a population of documented sex, which performed exceptionally well under cross-validation and when tested on the immatures of the same population. These functions assigned sex to immatures with substantially greater accuracy than that reported by any other morphological or metric method. Accuracy was reduced slightly when the same functions were used to assign sex to adults from a second population, but was still well within the range of commonly-used methods of assigning biological sex to skeletal remains and exceeded, for example, the reported accuracy of sex assignment from the skull of c. 80%.

The second approach circumvented issues of population-specificity, but at the expense of applicability. It involved the extraction of functions from adults to apply only to the immatures of the same population. While possible only for medium to large populations comprising sufficient adult males and females with accurate morphologically-assigned sex to extract the function, this approach could be fine-tuned to achieve high accuracy without excessive overfitting. Functions extracted from the adults performed excellently on immatures of the documented-sex collection from Spitalfields and showed great promise when applied to the archaeological Black Gate population, suggesting sex assessment of immatures can be achieved with greater accuracy using measurements of dentition than any other metric or morphological method currently available.

The findings suggest that that inter-population variability may not be an insurmountable obstacle to the prediction of biological sex in archaeological populations on the basis of dentition from a population of known sex, but that any function created for this purpose must balance the conflicting priorities of accuracy and applicability. A cross-population function would be particularly valuable for the assignment of sex to immatures, but also for adults in situations where the most accurate methods of morphological and metric sex assessment from the pelvis cannot be applied due to poor preservation, including archaeological and forensic situations where the skeleton is partial or fragmented, or only dentition is recovered.

ACKNOWLEDGEMENTS

The authors would like to thank the Department of Archaeology, University of Sheffield, for permitting access to the Black Gate skeletal collection, and the Natural History Museum for granting access to

the Spitalfields crypt skeletal collection. Further thanks go to Elissa Menzel and Petra Verlinden for their invaluable assistance in accessing collections and to Glynis Jones and Eleanor Stillman for advice on statistical methods. Thanks also go to the editor and two reviewers for their insightful comments that have enabled us to improve this paper significantly. This project was supported by a British Association of Biological Anthropology and Osteoarchaeology Academic Research Grant awarded to CJDA).

LITERATURE CITED

Al Qahtani, S. J., Hector, M. P., & Liversidge, H. M. (2010). Brief communication: the London atlas of human tooth development and eruption. *American Journal of Physical Anthropology*, *142*(3), 481-490.

Black, T. K. (1978). Sexual dimorphism in the tooth-crown diameters of the deciduous teeth. *American Journal of Physical Anthropology*, *48*(1), 77-82.

Brace, C. L., & Ryan, A. S. (1980). Sexual dimorphism and human tooth size differences. *Journal of Human Evolution*, *9*(5), 417-435.

Butler, P. M. (1967). The prenatal development of the human first upper permanent molar. *Archives of Oral Biology*, *12*(4), 551-563.

Cardoso, H. F. (2008). Sample-specific (universal) metric approaches for determining the sex of immature human skeletal remains using permanent tooth dimensions. *Journal of Archaeological Science*, *35*(1), 158-168.

Cardoso, H. F. (2010). Testing discriminant functions for sex determination from deciduous teeth. *Journal of Forensic Sciences*, *55*(6), 1557-1560.

Cardoso, H. F., & Saunders, S. R. (2008). Two arch criteria of the ilium for sex determination of immature skeletal remains: a test of their accuracy and an assessment of intra-and inter-observer error. *Forensic Science International*, *178*(1), 24-29.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Colby, G. R. (1996). Analysis of dental sexual dimorphism in two western Gulf of Mexico precontact populations utilizing cervical measurements. *American Journal of Physical Anthropology*, 22, 87.

Concato, J., Feinstein, A. R., & Holford, T. R. (1993). The risk of determining risk with multivariable models. *Annals of Internal Medicine*, *118*(3), 201-210.

Conceição, E. L. N., & Cardoso, H. F. V. (2011). Environmental effects on skeletal versus dental development II: further testing of a basic assumption in human osteological research. *American Journal of Physical Anthropology*, *144*(3), 463-470.

Decker, S. J., Davy-Jow, S. L., Ford, J. M., & Hilbelink, D. R. (2011). Virtual determination of sex: metric and nonmetric traits of the adult pelvis from 3D computed tomography models. *Journal of Forensic Sciences*, 56, 1107–1114.

Demirjian, A., Goldstein, H., & Tanner, J. M. (1973). A new system of dental age assessment. *Human Biology*, 211-227.

Everitt B.S. & Skrondal A. (2010). *Cambridge Dictionary of Statistics*. Cambridge University Press.

Ferembach, D., Schwidetsky, I., Stloukal, M. (1980). Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution,* 9, 517-549.

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications.

Flores-Mir, C., Raul Mauricio, F., Fernanda Orellana, M., & Major, P. W. (2005). Association between growth stunting with dental development and skeletal maturation stage. *The Angle Orthodontist*, *75*(6), 935-940.

Galdames, I., Matamala, D., & Smith, R. (2008). Blind Test of Mandibular Morphology with Sex Indicator in Subadult Mandibles. *International Journal of Morphology*, 26, 845–848

Garn, S. M., Cole, P. E., Wainwright, R. L., & Guire, K. E. (1977). Sex discriminatory effectiveness using combinations of permanent teeth. *Journal of Dental Research*, *56*(6), 697-697.

Garn, S. M., Cole, P. E., & Van Alstine, W. L. (1979). Sex discriminatory effectiveness using combinations of root lengths and crown diameters. *American Journal of Physical Anthropology*, *50*(1), 115-117.

Garn, S. M., Lewis, A. B., Swindler, D. R., & Kerewsky, R. S. (1967). Genetic control of sexual dimorphism in tooth size. *Journal of Dental Research*, *46*(5), 963-972.

Geyer, C. J. (2009). Likelihood inference in exponential families and directions of recession. *Electronic Journal of Statistics,* 3, 259–289.

Gómez-Robles, A., Martinón-Torres, M., De Castro, J. B., Margvelashvili, A., Bastir, M., Arsuaga, J. L., & Martínez, L. M. (2007). A geometric morphometric analysis of hominin upper first molar shape. *Journal of Human Evolution*, *53*(3), 272-285.

Goto, R., & Mascie-Taylor, C. N. (2007). Precision of measurement as a component of human variation. *Journal of Physiological Anthropology*, *26*(2), 253-256.

Hillson, S., FitzGerald, C., & Flinn, H. (2005). Alternative dental measurements: proposals and relationships with other measurements. *American Journal of Physical Anthropology*, *126*(4), 413-426.

Kondo, S., Townsend, G. C., & Yamada, H. (2005). Sexual dimorphism of cusp dimensions in human maxillary molars. *American Journal of Physical Anthropology*, *128*(4), 870-877.

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *261*, 165-e1.

Krogman, W. M., & İşcan M. Y. (1986). The human skeleton in forensic medicine, 2nd Ed. Springfield, IL: CC Thomas.

Krüger, G., L'abbe, E., & Stull, K. (2017). Sex estimation from the longbones of modern South Africans. *International Journal of Legal Medicine*, 131, 275–285.

Lewis, M., Shapland, F., & Watts, R. (2016a). On the threshold of adulthood: a new approach for the use of maturation indicators to assess puberty in adolescents from medieval England. *American Journal of Human Biology*, *28*(1), 48-56.

Lewis, M. E., Shapland, F., & Watts, R. (2016b). The influence of chronic conditions and the environment on pubertal development. An example from medieval England. *International Journal of Paleopathology*, *12*, 1-10.

López-Costas, O., Rissech, C., Trancho, G., & Turbón, D. (2012). Postnatal ontogenesis of the tibia. Implications for age and sex estimation. *Forensic Science International*, 214, 207.e1–211.

Loth, S. R., & Henneberg, M. (2001). Sexually dimorphic mandibular morphology in the first few years of life. *American Journal of Physical Anthropology*, *115*(2), 179-186.

Luna, L. H., Aranda, C. M., & Santos, A. L. (2017). New Method for Sex Prediction Using the Human Non-Adult Auricular Surface of the Ilium in the Collection of Identified Skeletons of the University of Coimbra. *International Journal of Osteoarchaeology*, 27, 898-911.

Mays, S. (1995). The relationship between Harris lines and other aspects of skeletal development in adults and juveniles. *Journal of Archaeological Science*, *22*(4), 511-520.

Molleson, T., Cruse, K., & Mays, S. (1998). Some sexually dimorphic features of the human juvenile skull and their value in sex determination in immature skeletal remains. *Journal of Archaeological Science*, *25*(8), 719-728.

Molleson, T. and Cox, M. 1993. *The Spitalfields project. Volume 2 – The anthropology.* York, Council for British Archaeology.

Morris, D. H. (1986). Maxillary molar occlusal polygons in five human samples. *American Journal of Physical Anthropology*, *70*(3), 333-338.

Nikita, E. 2014. Age-associated variation and sexual dimorphism in adult cranial morphology: Implications in anthropological studies. *International journal of Osteoarchaeology*, 24(5), 557-569.

Olivares, J. I., & Aguilera, I. A. (2016). Validation of the sex estimation method elaborated by Schutkowski in the Granada Osteological Collection of identified infant and young children: Analysis of the controversy between the different ways of analysing and interpreting the results. *International Journal of Legal Medicine*, 1-10.

Parfitt, A. M., Travers, R., Rauch, F., & Glorieux, F. H. (2000). Structural and cellular changes during bone growth in healthy children. *Bone*, *27*(4), 487-494.

Peretz, B., Nevis, N., & Smith, P. (1998). Morphometric analysis of developing crowns of maxillary primary second molars and permanent first molars in humans. *Archives of Oral Biology*, *43*(7), 525-533.

Phenice, T. W. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, *30*(2), 297-301.

Poulsen, J., & French, A. (2008) *Discriminant Function Analysis*. [Online]. Available: http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf

Rissech, C., & Malgosa, A. (2005). Ilium growth study: applicability in sex and age diagnosis. *Forensic Science International*, *147*(2), 165-174.

Schutkowski, H. (1993). Sex determination of infant and juvenile skeletons: I. Morphognostic features. *American Journal of Physical Anthropology*, *90*(2), 199-205.

Schwartz, G. T., & Dean, M. C. (2005). Sexual dimorphism in modern human permanent teeth. *American Journal of Physical Anthropology*, *128*(2), 312-317.

Seeman, E. (2001). Sexual dimorphism in skeletal size, density, and strength. *The Journal of Clinical Endocrinology & Metabolism*, *86*(10), 4576-4584.

Sonika, V., Harshaminder, K., Madhushankari, G. S., & Sri Kennath, J. A. (2011). Sexual dimorphism in the permanent maxillary first molar: A study of the Haryana population (India). *Journal of Forensic Odontostomatology*, *29*(1), 37-43.

Stojanowski, C. M., Larsen, C. S., Tung, T. A., & McEwan, B. G. (2007). Biological structure and health implications from tooth size at Mission San Luis de Apalachee. *American Journal of Physical Anthropology*, *132*(2), 207-222.

Stull, K. E., L'Abbé, E. N., & Ousley, S. D. (2017). Subadult sex estimation from diaphyseal dimensions. *American Journal of Physical Anthropology*, *163*(1), 64-74.

Subramanian, J., & Simon, R. (2013). Overfitting in prediction models–is it a problem only in high dimensions? *Contemporary Clinical Trials*, *36*(2), 636-641.

Sutter, R. C. (2003). Nonmetric subadult skeletal sexing traits: I. A blind test of the accuracy of eight previously proposed methods using prehistoric known-sex mummies from northern Chile. *Journal of Forensic Sciences*, *48*(5), 1-9.

Swales, D. L. M. (2012). *Life stress: a bio-cultural investigation into the later Anglo-Saxon population of the Black Gate Cemetery* (Doctoral dissertation). Department of Archaeology, University of Sheffield, Sheffield.

Ubelaker, D. H. & Volk, C. G. (2002). A test of the phenice method for the estimation of sex. *Journal of Forensic Science, 47*(1): 19-24.

Viciano, J., López-Lázaro, S., & Alemán, I. (2013). Sex estimation based on deciduous and permanent dentition in a contemporary Spanish population. *American Journal of Physical Anthropology*, *152*(1), 31-43.

Vlak, D., Roksandic, M., & Schillaci, M. A. (2008). Greater sciatic notch as a sex indicator in juveniles. *American Journal of Physical Anthropology*, *137*(3), 309-315.

Wilson, L. A., Cardoso, H. F., & Humphrey, L. T. (2011). On the reliability of a geometric morphometric approach to sex determination: A blind test of six criteria of the juvenile ilium. *Forensic science international*, *206*(1), 35-42.

Wilson, L. A., MacLeod, N., & Humphrey, L. T. (2008). Morphometric criteria for sexing juvenile human skeletons using the ilium. *Journal of Forensic Sciences*, *53*(2), 269-278.

Żądzińska, E., Karasińska, M., Jedrychowska-Dańska, K., Watala, C., & Witas, H. W. (2008). Sex diagnosis of subadult specimens from Medieval Polish archaeological sites: metric analysis of deciduous dentition. *HOMO-Journal of Comparative Human Biology*, *59*(3), 175-187.

Żądzińska, E., Lorkiewicz, W., Kurek, M., & Borowska-Strugińska, B. (2015). Accentuated lines in the enamel of primary incisors from skeletal remains: A contribution to the explanation of early childhood mortality in a medieval population from Poland. *American Journal of Physical Anthropology*, *157*(3), 402-410.

Zorba, E., Moraitis, K., & Manolis, S. K. (2011). Sexual dimorphism in permanent teeth of modern Greeks. *Forensic Science International*, *210*(1), 74-81.

FIGURE LEGENDS

Fig 1. Demographic profile of the Spitalfields sample (*n*=59)

Fig 2. Demographic profile of the Black Gate sample (*n*=90)