Prior Probabilities and the Age Threshold Problem: First and Second Molar Development

Lyle W. Konigsberg,[1] Susan R. Frankenberg,[1] Valerie Sgheiza,1 and Helen M. Liversidge[2]*

ABSTRACT

Dental development has been used to assess whether an individual may be below or above an age that serves as a legal threshold. This study used development of the first and second mandibular molars from a large sample of individuals (N = 2,676) to examine the age threshold for minimum age of criminal responsibility. A bivariate ordered probit model was applied to dental scores following the Moorrees et al. (1963) system, with the addition of a crypt-absent/present stage. Then a 10-fold cross-validation within each of the sexes showed that the bivariate models produce unbiased estimates of age but are heteroskedastic (with increasing spread of the estimates against actual age). To address the age threshold problem, a normal prior centered on the threshold is assumed, and the product of the prior and the likelihood is integrated up to the age threshold and again starting at the age threshold. The ratio of these two integrals is a Bayes factor, which because the prior is symmetric around the threshold, can also be interpreted as the posterior odds that an individual is over versus under the age threshold. It was necessary to assume an unreasonably high standard deviation of age in the prior to achieve posterior odds that were well above "evens." These results indicate that dental developmental evidence from the first and second molars is of limited use in examining the question of whether an individual is below or over the minimum age of criminal responsibility. As the third molar is more variable in its development than the first two molars, the question of dental evidence regarding the age of majority (generally 18 years) remains problematic.

We have previously considered the legal age threshold problem (Konigsberg et al. 2019) in regard to the age of majority. This age, typically 18 years, often is the basis for deciding whether to grant asylum to an individual seeking refuge, with individuals younger than 18 years of age being granted asylum and those older than 18 being deported. We previously considered the developmental status of the third molar, the latest forming of the teeth, and demonstrated the importance of selecting an appropriate prior age distribution and accurately accounting for variability in age at attainment (Konigsberg et al. 2019). Here we consider a younger age threshold known as the minimum age of criminal responsibility (MACR). This age is variable around the globe and is subject to change under local legal systems. We thus expanded the previously presented method to address the more complex problem of evaluating age estimated from multiple teeth against an age

[1]Department of Anthropology, University of Illinois Urbana–Champaign, Urbana, Illinois, USA.

[2]Institute of Dentistry, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

*Correspondence to: Lyle Konigsberg, University of Illinois Urbana–Champaign, Department of Anthropology, Davenport Hall 109, MC-148, 607 S. Mathews Ave., Urbana, IL 61801 USA. E-mail: lylek@illinois.edu.

KEY WORDS: MINIMUM AGE OF CRIMINAL RESPONSIBILITY, DENTAL FORMATION, BAYES FACTOR, BIVARIATE ORDINAL PROBIT, N-FOLD CROSS-VALIDATION.

Human Biology, Winter 2021, v. 93, no. 1. doi: 10.13110/humanbiology.93.1.02.

threshold that varies by country. Cipriani (2009) thoroughly reviewed worldwide variation in the MACR, but treatment of MACR has changed considerably since that time. We therefore review some recent history and variation in the MACR.

The UN position on MACR has evolved considerably in the last 30 years. The 1989 UN Convention on the Rights of the Child stated that a minimum age should be established by individual nation states, but gave no recommended age (United Nations 1989). In 2007, CRC General Comment 10 provided a recommendation of a single age of at least 12 years, and in 2019 CRC General Comment 24 raised this to 14 years (UN Committee on the Rights of the Child 2007, 2019). Both documents commended countries with higher ages and opposed the use of multiple age thresholds, either for the severity of the crime or for comprehension of criminality by the child. In cases of unknown age, the UN suggests exhausting the search for documentation and use of interviews before resorting to developmental examination. Dental and skeletal markers are not recommended because they are "often inaccurate, due to wide margins of error" (UN Committee on the Rights of the Child 2019: para. 34). The 2019 UN Global Study on Children Deprived of Liberty takes a firmer position, stating that minimum age established by states "shall not be below 14 years of age" (UN General Assembly 2019, sec. 109) and that severity of the crime should never be used to lower this age.

In line with the UN minimum age for military recruitment and armed conflict of 15 years, Ursini (2015) recommends an international MACR of 15 years to resolve location-based inconsistencies in the prosecution of war crimes committed by child soldiers. Under current laws, countries must prosecute war crimes, but the MACR depends on the country. As a result, a child who would not face criminal charges in one country could be executed in another. An international MACR of 15 years acknowledges that children who are too young to be in the military should also be too young to commit war crimes, and that military recruitment of children under 15 is itself a war crime (Ursini 2015). Despite these recommendations, standards for MACR remain widely variable. As of July 2019, over 120 countries had an MACR under 14 years (UN General Assembly 2019). Some existing MACRs may be well below 14 years, as in England and Wales, where the Children and Young Person's Act of 1933 (see https://www.legislation.gov.uk/ukpga/Geo5/23-24/12) states in section 50 that "it shall be conclusively presumed that no child under the age of 10 years can be guilty of any offence."

Because in this study we focused on age thresholds younger than 18 years, we use earlier-forming teeth than in our previous study (Konigsberg et al. 2019). Specifically, here we consider the bivariate problem of development of the first and second mandibular molars. Bivariate, and more generally multivariate, problems must deal with the fact that typically there is residual correlation between various age indicators. In other words, once the effect of age is accounted for, a correlation remains between the indicators (Green 1961; Šešelj 2013). Several previous studies present parametric approaches to deal with this residual correlation. Boldsen et al. (2002) and Fieuws et al. (2016) describe an ad hoc method that accounts for the residual correlations. Konigsberg (2015) uses a Markov chain Monte Carlo approach to estimate the residual correlations. Hens and Godde (2020) uses the composite likelihood method implemented in the R package "mvord" (Hirk et al. 2018, 2020). Additionally, Braga et al. (2005) used a nonparametric Bayesian approach to find the proportion of individuals in a given age category who were in a particular pattern of dental development, so they had no need to estimate residual correlations. Because here our parametric approach considers only the two-tooth bivariate problem, we use methods that maximize the bivariate likelihood to estimate the single residual correlation directly.

Materials and Methods

Radiographic Sample

The initial sample consisted of mandibular panoramic radiographs showing the central lower incisors through third molars for 3,334 males and females from London. These individuals were classified as "white" or Bangladeshi. We removed from the sample 11 individuals with no molars observable, 4 individuals with unobservable first molars, 27 individuals with no second or third molars unobservable, and 585 individuals with unobservable third molars. Additionally, one girl age 9.64 years had a first molar at stage "root three-quarter

complete" and a second molar at stage "crown half complete." Both teeth were far too underdeveloped for the given age. We assume there was an error in the age or that the radiograph was from a different individual, so we omitted this individual. This left 2,706 individuals for whom all three molars were observable. Of these, 30 had second molars with roots greater than three-quarters length following Moorrees et al.'s (1963) scoring but without a crypt formed for the third molar. These individuals were omitted because Baba-Kawano et al. (2002) have shown that third molar agenesis is related to late tooth formation. This resulted in a final sample of 2,676 individuals (1,325 F, 1,351 M). Figure 1 shows the age distribution for these individuals.

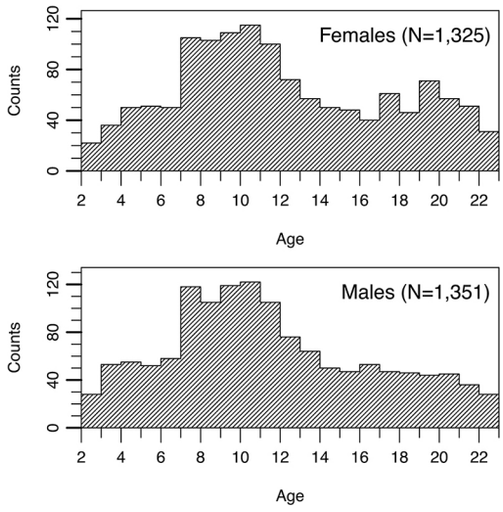

FIGURE 1. Age distribution of the females and males included in this study.

The first and second molars were scored by the fourth author using the Moorrees et al. (1963) system, with the addition of two stages (crypt absent and crypt present) prior to cusp initiation. The final scoring was thus 1 = no crypt, 2 = crypt, 3 = cusp initiation, 4 = cusp coalescence, 5 = cusp occlusal outline complete, 6 = crown half complete, 7 = crown three-quarter complete, 8 = crown complete, 9 = root initiation, 10 = root cleft formation, 11 = root one-quarter complete, 12 = root one-half complete, 13 = root three-quarter complete, 14 = root complete, 15 = root apex half complete, and 16 = root apex complete. In our sample, all 16 stages were observed for the second molar; for the first molar the earliest stage observed was stage 6 (crown half complete).

Testing the Univariate Probit Models

Konigsberg et al. (2016) describe a Lagrange multiplier test (Bera et al. 1984; Johnson 1996) that can be used to determine whether a univariate probit model adequately represents the age progression for an ordinal categorical trait. We used this test on the straight scale of age (not logged) to test whether the univariate probit gives an adequate goodness of fit for the first and second molars separately in females and males. We separated the sample into females and males because of the known sexual dimorphism in tooth development.

Bivariate Ordered Probit Model

Using the first and second molars, it is possible to fit a bivariate ordered probit model (Greene and Hensher 2010: 291–294). This model can be fit using BIOPROBIT (Sajaia 2008) in Stata (StataCorp 2019). BIOPROBIT has the advantage that it calculates the gradient vector (partial derivatives of the log-likelihood) and is consequently fast and accurate. We used BIOPROBIT to fit the entire data set to 10 different models, ranging from a single effect (age) to a three-way interaction among age, sex, and ethnicity, as well as all second order and main effects. The models consequently test for direct and indirect impacts of both sexual dimorphism and ethnic membership on age progression in tooth development. We then used Schwarz's (1978) Bayesian information criterion (BIC) to pick the preferred model. We also applied BIOPROBIT separately within each sex using three models (age, age and ethnicity, and age-by-ethnicity interaction with age and ethnicity main effects). Again, we used Schwarz's (1978) BIC to pick the preferred model within each sex.

Cross-validation of the Bivariate Ordered Probit Models

It is important to test the appropriateness of the bivariate ordered probit models for estimating ages for individuals. While a leave-one-out strategy often has been applied to assess appropriateness of regression-based analyses via the PRESS statistic (Allen 1974), the calculation of the bivariate ordered probit is too time-consuming to use this approach. We instead used a 10-fold cross-validation approach, which optimizes the bias/variance trade-off (Kohavi 1995). Employing maximum likelihood, we first fitted parameters of a bivariate ordered probit model using a training sample. We then found the age for an individual in

Table 1. Layout for the 10-fold Cross-validations

Group	Earliest Stages	Sample										Latest Stages
	M_1 = Cr.5, M_2 = no crypt	1	2	3	4	5	6	7	8	9	10	M_1 = A.c M_2 = A.c
Female	7	90	90	90	90	90	90	90	90	90	90	418
Male	12	98	98	98	98	98	99	99	99	99	99	354

Abbreviations: M_1, M_2, first and second molars; no crypt, stage 1; Cr.5, crown half complete (stage 6); A.c, root apex complete (stage 16).

[a]Samples 1–10 are randomized with respect to age. For example, for test sample 1 in females (n = 90), the training sample consisted of the 7 individuals in the earliest tooth stages, the 418 individuals in the latest tooth stages, and the individuals in training samples 2–10.

a testing sample that gave the highest probability of the appearance of the first and second molar stages. This is equivalent to a Bayesian estimate of age with a uniform prior.

Individuals in the earliest stages for both molars have posterior densities of age that rise to the left boundary, and individuals in the last stages for both molars similarly have posterior densities that rise to the right boundary. While we always included these individuals in the training sample, we excluded them from the testing sample because their maximum likelihood age estimates would go to a boundary. All other individuals were randomly permuted and then divided into 10 samples of (approximately) equal size. For the females, there were 7 individuals in the earliest stages of the two molars and 418 in the latest stage. All 10 of the cross-validation testing sets consequently had 90 individuals. For the males, there were 12 individuals in the earliest stages of the two molars and 354 in the latest stages. Five of the cross-validation testing sets consequently had 98 individuals and five had 99 individuals.

The cross-validations proceeded by taking nine of the sets and combining these individuals with the individuals in the earliest and latest tooth stages to form a training sample. The bivariate ordered probit model was then estimated by constrained maximum likelihood (with constraints being the ordering of the intercepts) using "constrOptim.nl" in the package "alabama" (Varadhan 2015) in R (R Core Team 2020). We used Fisher's z-transform (Bond and Richardson 2004) so as to not require constraints at –1 and 1 for the residual polychoric correlation between tooth stages. This transformation is $z = \text{arctanh}(r)$, which can be back converted to $r = \tanh(z)$. Finally, the training sample models formed from the nine data sets were applied to each of the 10 test data sets. Table 1 gives the layout for these cross-validations, as well as further explanation on the procedure.

Prior and Posterior Probabilities

As we point out in Konigsberg et al. (2019), the age threshold problem generally has an implicit prior probability density. Hillewig et al. (2013) used an explicit uniform prior of 16–26 years for an age threshold of 18 years. This is an informative prior, because (26–18)/(18–16) = 4.0, so the prior odds that an individual is between ages 18 and 26 (and thus older than age 18) versus between ages 16 and 18 (and thus younger than age 18) are 4.0. Sironi and Taroni (2015: 131) caution against the use of uniform priors, noting that "posterior odds on the chronological age are strongly biased by the uniform distribution because individuals in specific extreme age ranges would generally not be asked to be examined for forensic age assessment purposes." Sironi et al. (2017: e27) suggest selecting a prior distribution "by reasoning on the distribution of the ages of the persons for whom a medico-legal expert evaluation may be requested." This is a difficult proposition, as the individuals for whom evaluations are requested will not have known ages.

We chose to use a normal distribution of age centered on the threshold age for the prior density. As we are concerned with younger age thresholds, and thus a more readily discernable period of growth and development, we used a standard deviation of 0.5378, which places the 0.01% and 99.99% values of age at 2 years below and above the age threshold. This tighter range contrasts with Konigsberg et al.'s (2019) 0.01% and 99.99% values of age 5 years below and above the threshold for the age of majority. As the prior odds are "evens," the posterior odds that an individual is above versus below the age threshold T is

$$\frac{\Pr\left(a \geq T | \{M_1, M_2\}\right)}{\Pr\left(a < T | \{M_1, M_2\}\right)} = \frac{\int_{a=T}^{22} \pi(\{M_1, M_2\}|a) \times f(a) da}{\int_{a=2}^{T} \pi(\{M_1, M_2\}|a) \times f(a) da} \quad (1)$$

This is in the same format as Konigsberg et al.'s (2019: eq. 8) univariate (single tooth) equation. The brace notation $\{M_1, M_2\}$ indicates the stages for the first and second molars, and $\pi(\{M_1, M_2\}|a)$ indicates the probability from the bivariate ordered probit of being in the given stages at exact age a. The symbol $f(a)$is the normal probability density for exact age

a. The numerator integration is up to 22 years, as this is well past the upper tail of the prior density. Similarly, denominator integration starts at 2 years, as this is well below the lower tail of the prior density. Taking the posterior odds from Equation (1), the posterior probability that an individual is over the threshold age can be found from the usual equation $PO/(1+PO)$, where PO represents the posterior odds from Equation (1).

Results

The goodness-of-fit p-values from the Lagrange multiplier test in females are 0.8158 and 0.3305 for the first and second molars, respectively, and in males are 0.2169 and 0.2199. As all these p-values > 0.10 on the raw scale of age, it appears that the ages at transition between adjacent stages are normally distributed. Adopting a cumulative probit on the raw scale of age thus is appropriate. Table 2 shows the BIC for the 10 different bivariate ordered probit models of the effects of age, sex, and/or ethnicity on tooth stage. The total number of parameters for the model with no effects is 26, consisting of 10 intercepts for the first molar (to represent 11 stages), plus 15 intercepts for the second molar (to represent 16 stages), plus one correlation. The model with age as a single effect adds two slopes (one for each tooth), bringing the total to 28 parameters, as shown in the bottom row of Table 2. The other models listed in Table 2 have greater numbers of parameters because of the slopes added for each main effect and each interaction. The model with the lowest BIC (i.e., the "best" model) is the model that contains only the main effects of age and sex; this model has two more parameters (for the first and second molar slopes on sex) than the age model.

Table 3 shows similar BIC comparisons subdivided by sex. The table considers the model with only age as a main effect, the model with age and ethnicity as main effects, and the model with these main effects and an interaction between age and ethnicity. These results show that the model with only age as a main effect has the lowest BIC value and thus represents the best model for both males and females when analyzed separately by sex. This suggests that ethnicity does not measurably impact age at attainment for these tooth stages. These findings agree with those from Liversidge (2011), who showed sexual dimorphism in the timing of dental development but similar dental development in White and Bangladeshi samples of the same sex.

Table 2. Bayesian Information Criteria (BIC) for the Entire Sample

Model	Number of Parameters (k)	lnLK	BIC
Age*ethnicity*sex	40	–4310.7924	8937.268
Age+ethnicity*sex	34	–4320.8680	8910.067
Age*ethnicity+sex	34	–4319.4462	8907.223
Age*sex+ethnicity	34	–4323.7843	8915.899
Age+sex+ethnicity	32	–4325.8821	8904.311
Age*ethnicity	32	–4353.6817	8959.910
Age*sex	32	–4324.1296	8900.806
Age+ethnicity	30	–4359.5199	8955.802
Age+sex[a]	30	–4326.2032	8889.169
Age	28	–4359.8920	8940.762

BIC = $-2 \times \ln LK + \ln(2676) \times k$, where ln$LK$ is the log-likelihood of the model and 2,676 is the sample size.

[a]Age+sex gave the smallest BIC.

Table 3. Bayesian Information Criteria (BIC) for Females Only and Males Only

Model	Number of Parameters (k)	lnLK	BIC
Females (n = 1,325)			
Age*ethnicity	32	–2022.0711	4274.196
Age+ethnicity	30	–2023.1207	4261.916
Age[a]	28	–2026.5845	4254.466
Males (n = 1,351)			
Age*ethnicity	32	–2273.0752	4776.826
Age+ethnicity	30	–2280.2027	4776.663
Age[a]	28	–2281.7588	4765.358

BIC = $-2 \times \ln LK + \ln(n) \times k$, where ln$LK$ is the log-likelihood of the model and n is the sample size.

[a]Age gave the smallest BIC in both samples.

Table 4 shows the parameter values for females and males estimated from the age main effect model run separately by sex using BIOPROBIT in Stata. The parameters are listed following a typical transition analysis format, which gives the mean transition ages, or mean age at attainment for each stage, followed by the common standard deviation for each tooth by sex and the residual correlation coefficient between M_1 and M_2 for each sex. Using these parameters to calculate a bivariate integral requires dividing the transition means by their associated standard deviation. For example, consider a male who is in the crown half complete

Table 4. Bivariate Ordinal Probit Parameter Values for Females and Males by Tooth: Mean Age at Attainment (Years) for Each Stage

Parameter[a]	Female		Male	
	M_1	M_2	M_1	M_2
No crypt/crypt	—	2.264	—	2.296
Crypt/Cu.in	—	3.137	—	2.997
Cu.in/Cu.co	—	3.782	—	3.976
Cu.co/Cu.oc	—	4.601	—	4.881
Cu.oc/Cr.5	—	5.073	—	5.373
Cr.5/Cr.75	1.818	6.114	1.919	6.341
Cr.75/Cr.c	2.717	6.911	2.923	7.433
Cr.c/R.i	3.239	7.867	3.489	8.202
R.i/R.cl	3.761	8.481	3.998	8.861
R.cl/R.25	4.479	9.236	4.811	9.540
R.25/R.5	5.266	10.232	5.457	10.469
R.5/R.75	6.265	11.139	6.508	11.496
R.75/R.c	7.303	12.373	7.657	12.889
R.c/A.5	7.948	13.190	8.493	13.527
A.5/A.c	9.103	14.673	9.713	14.776
Standard deviation	0.948	0.978	0.924	0.989
Correlation	0.671		0.660	

Standard deviation indicates common standard deviation for each tooth by sex. Correlation indicates the first (M_1)–second (M_2) residual correlation for females and males, respectively.

[a]Stages: Cu.in, cusp initiation (stage 3 in this study); Cu.co, cusp coalescence (stage 4); Cu.oc, cusp outline complete (stage 5); Cr.5, crown half complete (stage 6); Cr.75, crown three-quarter complete (stage 7); Cr.c, crown complete (stage 8); R.i, root initiation (stage 9); R.cl, root cleft formation (stage 10); R.25, root one-quarter complete (stage 11); R.5, root half complete (stage 12); R.75, root three-quarter complete (stage 13); R.c, root complete (stage 14); A.5, root apex half complete (stage 15); and A.c, root apex complete (stage 16).

stage for the first molar and crypt formed for the second molar. This individual's limits of integration for the first molar would be negative infinity and 1.919/0.924; for the second molar they would be 2.296/0.989 and 2.997/0.989. The means for the bivariate normal would be the given age divided by each standard deviation (0.924 and 0.989), and the residual correlation would be 0.660.

As an example of how to apply the parameter values from Table 4, Figure 2A shows a posterior density of estimated age for a 10-year-old female with her first molar in the root apex complete stage and her second molar in the root one-quarter complete stage. The posterior density was found using a uniform prior, so the original likelihood function is divided by the integral of the likelihood function. The maximum likelihood estimate for the age is 9.92 years, close to the true age of 10 years. The 95% highest posterior density bounds, found using the "hpd" function in the R package "TeachingDemos" (Snow 2020), are 8.05–11.84 years. These calculations were done using the nine "folds" of the cross-validation that did not include the actual case.

Figure 2B is an example of how an outlier—an individual who is developmentally advanced or delayed for their chronological age—will fall outside the estimated age distribution. This individual, a female that had the second highest squared difference between the true and the estimated age in the 10-fold cross-validation, had a true age of 9.8 years, although her first molar had the root apex complete and her second molar had a complete root. The maximum likelihood age estimate (12.8 years) is above the true age of 9.8 years, and the 95% highest posterior density bounds of 10.9–14.8 years does not

FIGURE 2. (A) Posterior density of age for an actual 10-year-old female with the first molar in the root apex complete stage and the second molar in the root one-quarter complete stage. (B) Posterior density of age for an actual 9.8-year-old female with the first molar in the root apex complete stage and the second molar in the root complete stage. Dashed lines, actual age; solid lines, maximum likelihood estimate (MLE; highest posterior density); gray region, 95% highest posterior density.

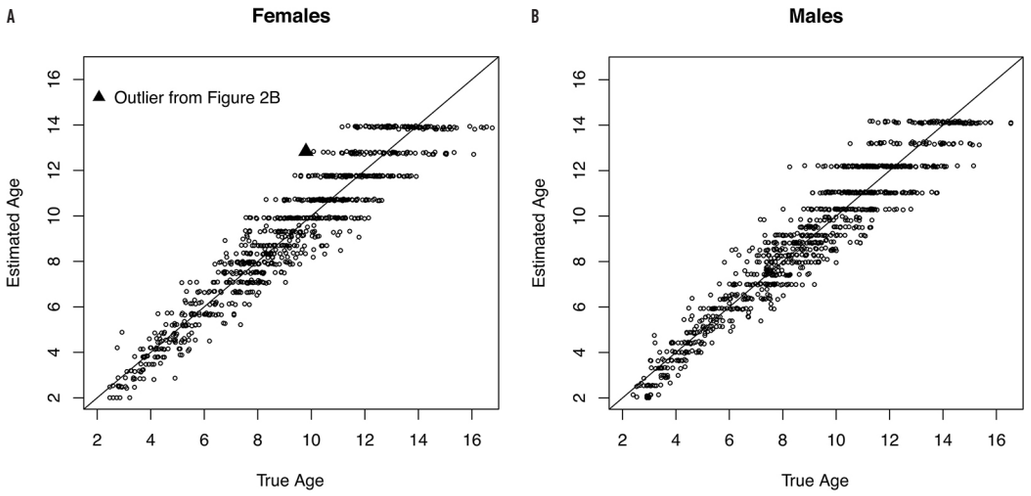

FIGURE 3. Estimated age plotted against true age for females (A) and males (B) using 10-fold cross-validation. The triangle in A shows the individual from Figure 2B; diagonal lines, lines of identity.

include the true age. Again, the model used here was from the nine "folds" of the cross-validation that did not include the actual case. Of the 900 females split into groups of 90 individuals each for the 10-fold cross-validation, 848 individuals had their actual ages within the estimated 95% highest posterior densities. This is 94.22% of individuals, close to the expected 95%. The comparable figure for the 985 males in the cross-validation study is 936 (95.02%) individuals within the estimated 95% highest posterior densities.

Figure 3 shows estimated ages against true ages using the 10-fold cross-validation for females and for males separately; Figure 3A indicates the outlier individual shown in Figure 2B. The figure shows that the estimates of age appear to be unbiased, because the lines of identity evenly divide the clusters of points. Figure 3 also shows that the age estimates are heteroskedastic, with the spread of residuals increasing with increasing age. This is to be expected because variation in acceleration or deceleration of growth or development increases after birth.

Figure 4 illustrates the posterior probability that an individual is older than 10 years under different scenarios. The posterior probability that an individual is over the threshold is calculated using Equation (1). Figures 4A and 4B use a prior density for age with a mean of 10 years and a standard deviation of 0.5378. As described in the methods, this standard deviation places the 0.01% and 99.99% values of age at 2 years above and below the threshold age, so from 8 to 12 years. Figure 4A shows the same individual as from Figure 2A, a 10-year-old female with her first molar in the root apex complete stage and her second molar in the root one-quarter complete stage. Because the prior density is symmetric around the threshold, the Bayes factor is also the posterior odds. This individual is 10 years old and has the dental development of a typical 10-year-old, so the posterior odds are near "evens" at 0.9311. Given a case like this, the calculated posterior probability that this individual is older than 10 years is 0.4822.

Figure 4B shows a 12-year-old male who had a first molar with the root apex complete and a second molar in the root three-quarters complete stage. The posterior odds that an individual at this developmental stage would be over 10 years old as opposed to under 10 years old is 4.7696, with a posterior probability of being older than 10 years equal to 0.8267. Figures 4C and 4D show the same case as in Figure 4B but with increasingly diffuse priors: Figure 4C uses a standard deviation of 0.8067 years (0.01% and 99.99% values of 7 and 13 years), whereas Figure 4D uses a standard deviation of 1.3444 (0.01% and 99.99% values of 5 and 15 years). With these increasingly diffuse priors, the posteriors odds are 8.587 (Figure 4C) and 18.3278 (Figure 4D). These posterior odds translate into calculated posterior probabilities of being greater than 10 years old of 0.8957 (Figure 4C) and 0.9483 (Figure 4D).

Table 5 summarizes the results from Figure 4, listing the actual age and sex of each case with their observed tooth stages, the prior standard deviations, the 0.01% and 99.9% boundary ages (in years) for the prior distribution, and the posterior

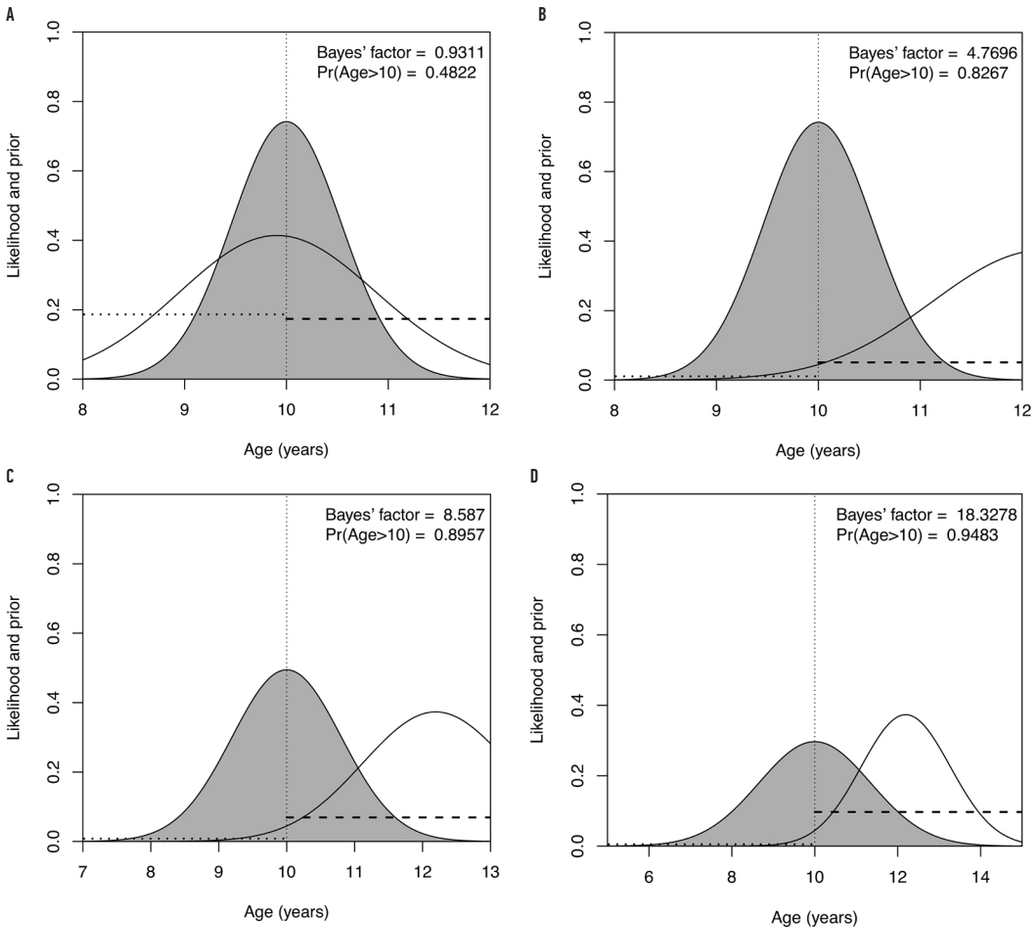

FIGURE 4. (A) Prior density (gray region) and likelihood curve for a female with an actual age 10 years, a first molar in the root apex complete stage, and a second molar in the root one-quarter complete stage, using a prior density for age with a mean of 10 years and a standard deviation of 0.5378. (B–D) A male with an actual age of 12 years, a first molar in stage root apex complete, and a second molar in the root three-quarter complete stage, with a prior that runs from age 8 to age 12 (B), from age 7 to age 13 (C), and from age 5 to age 15 (D). Dotted horizontal lines, integral of the prior times the likelihood up until age 10 years; dashed horizontal lines, integral past age 10 years.

Table 5. Summary of Scenarios for Figure 4A–D

Scenario	Sex	Actual Age	M_1	M_2	*Prior SD*	*Normal Range*[a]	*Posterior*[b]
Fig. 4A	Female	10	A.c	R.25	0.5378	8–12	0.4822
Fig. 4B	Male	12	A.c	R.75	0.5378	8–12	0.8267
Fig. 4C	Male	12	A.c	R.75	0.8067	7–13	0.8957
Fig. 4D	Male	12	A.c	R.75	1.3444	5–15	0.9483

[a]The 0.01% and 99.99% ages for a normal distribution centered on 10 years that serves as the prior distribution.

[b]The posterior probability that an individual is <ME>10 years of age.

probability that the case is at or older than 10 years. These posterior probabilities were obtained by converting the posterior odds ratios as described in the methods. The posterior probability that an individual is at or over 10 years of age from this table represents an evidentiary statement. The posterior probabilities also can be used in a decision theoretic framework (Sironi et al. 2020) to decide whether or not an individual is at or over 10 years of age. While this seems an attractive approach, the problem is that it requires specifying loss functions and the relative costs of misclassifications.

Discussion

The analytical approach we used in this study has distinct advantages over previously used methods to estimate age with respect to legally defined thresholds, but our results also highlight issues with relying solely on dental (and skeletal) age estimators. The principal advantages of our approach are that it is Bayesian based, generates clear statements of probabilities, and is flexible. The models used here allow specification of prior age distributions, selection of different relevant age thresholds, and the use of multiple traits or observations. This is particularly relevant to the minimum age of criminal responsibility (MACR), which continues to vary by country and which has not been addressed in the age estimation literature to the same extent as

has the legal age of majority. Despite considerable discussion about using third molar development to assess whether an individual has reached the legal age of majority (Acharya 2011; Akkaya et al. 2019; Cameriere et al. 2008; Corradi et al. 2013; De Luca et al. 2014; Galić et al. 2015; Liversidge and Marsden 2010; Lucas et al. 2016; Márquez-Ruiz et al. 2017; Sironi et al. 2018; Streckbein et al. 2014; Uys et al. 2018), comparatively little work has been done on dental development as a marker for whether or not an individual has reached MACR (Balla et al. 2019; Cameriere et al. 2018; Lucas et al. 2014; Ravi et al. 2020; Thomas et al. 2021; Yadava et al. 2011).

All six previous studies that used dental development to assess MACR, unlike the approach taken here, adopted methods influenced by the reference sample age distribution (i.e., the problem of age mimicry). The earliest of these studies (Yadava et al. 2011) focuses on the 10-year threshold but does not actually calculate the probability of being equal to or older than 10 years as opposed to less than 10 years. Instead, the study attempted to estimate ages using random-effects meta-analysis, in which each tooth stage observed in an individual is assigned a mean age-within-stage and a standard error of the mean based on a reference sample. Each mean is weighted by the inverse of the sum of the squared standard error and an estimated common variance based on divergence between the means. The mean age-within-stage thus depends on both development and the reference sample age distribution. Additionally, the study lacks a statistically sound way to evaluate estimated age ranges relative to an age threshold.

Lucas et al. (2014) also focused on the 10-year threshold but attempted to estimate the probability that an individual is at or older than 10 years. To do so, they integrated normal distribution functions up to 10 years using the means and standard deviations of age within tooth stages in a way that does not account for influence of the reference sample age distribution. Although not clearly described, they also introduced a series of different weightings and mishandled how they combined probabilities. The study first assumed that the probability values "are of equal importance" (Laird and Mosteller 1990: 20), which amounts to a simple averaging of the *p*-values. They also included zero probabilities from unobserved teeth in their averaging and ignored accepted methods for combining *p*-values (Heard and Rubin-Delanchy 2018), although the assumption of independence of *p*-values required by many of these methods is unlikely to be the case here. Lucas et al. (2014) also introduced a two-by-two tabling method similar to that seen in more recent studies that use dental development to assess MACR.

The four more recent studies (Balla et al. 2019; Cameriere et al. 2018; Ravi et al. 2020; Thomas et al. 2021) all used some form of two-by-two tabling of actual age at or above versus below a threshold against the number of individuals that fall above or below some cut-point for an age indicator. This two-by-two tabling approach suffers from three problems. First, it does not place a higher probability on being above the age threshold for individuals who are further above the cut-point value of some indicator. Second, the two-by-two tabling approach is specific to the age threshold. For age-of-majority problems, which are generally age 18, this is not a relevant problem. In contrast, we have seen that the MACR varies widely around the globe. Finally, as we previously pointed out (Konigsberg et al. 2019), the two-by-two tabling method largely depends on the age distribution of the reference sample. Thus, our study appears to be the first to properly address the problem of estimating from dentition probabilities of being at or above age thresholds relevant to the MACR.

Unlike the previous studies, the approach we take here avoids influence of the reference sample age distribution, frames probabilities in a way that treats dental development as a function of age, and appropriately deals with multiple indicators. However, our results also highlight issues with relying solely on dental (and skeletal) age estimators. Many of these issues stem from the less than perfect relationship between age estimators and actual age, due to individual variability in developmental rates, error associated with sampling and age estimation methods, and the potential impacts of sex, ethnicity/population membership, and other factors on developmental trajectories. Below we specifically consider these issues with respect to variability in dental development and age estimators, the call for population-specific standards, and the UN stance against using skeletal and dental indicators for assessing MACR.

One consequence of the imperfect relationship between estimated and actual age is that the

models presented here require broad prior probabilities of age in order to achieve high posterior probabilities from dental developmental data that an individual is over a given age threshold, much as we found previously (Konigsberg et al. 2019). It is conceivable that posterior densities of age could be narrowed, and likelihoods sharpened, by adding information from additional teeth. However, we also expect that the residual correlations between teeth after conditioning on age will be high, particularly for adjacent teeth (Garn et al. 1960; Garn and Smith 1980; Parner et al. 2002), such as those we used here. Controlling for this effect may severely limit the additional information to be gleaned from adding more teeth, particularly teeth within the same developmental field. In the present study we did attempt to control for extremes in individual variability in developmental rates by excluding individuals with third molar agenesis, as these individuals are generally on a slower developmental track (Baba-Kawano et al. 2002; Garn et al. 1963). Our methods and results consequently should not be applied to individuals with third molar agenesis.

Another consequence of the imperfect relationship between estimated and actual age is the call by many for population-specific age estimators or standards. For example, Noll (2016: 240) questions whether radiological age assessments should be considered "junk" science and raises the problem of "the absence of population-specific standards." The methods we present here consequently may produce different results when applied to different populations. The extent to which dental development is population specific rather than sample specific is debated (Corron et al. 2018), with some studies finding consistent developmental differences between groups (Liversidge 2008; Liversidge et al. 2017) and others seeing better results with pooled reference samples than with specific ones (Braga et al. 2005; Thevissen 2013). Sample specificity, unlike population specificity, is the result of sampling strategy rather than biological difference.

Sample specificity is the more immediate issue because a developmental difference between groups cannot be determined until the effect of the sampling strategies can be ruled out. What does consistently emerge from these studies is that the sample specificity of individual methods can be at least partially ameliorated via a large reference sample and a Bayesian approach (Braga et al. 2005; Corron et al. 2018). These two strategies, both taken here, address the issues of small sample effects, which may make two reference samples appear different even when drawn from the same population, and age distribution effects, which can produce apparent differences between samples even when the development of individuals at the same age is identical. Sample specificity, and possibly population specificity, also can be ameliorated by ensuring that the reference sample includes individuals from the same source population as the case of interest. Our sample is not globally representative, so we cannot rule out the possibility of population specificity.

As noted in the introductory remarks, skeletal and dental indicators are not recommended in a transnational policy context for determining whether an individual is at or above the MACR, because of the "wide margins of error" (UN Committee on the Rights of the Child 2019: para. 34). The results of our study support the view that the inherent variability in dental age estimators limits the utility of such approaches when used alone to determine whether an individual is below or above the MACR. Skeletal estimates of developmental age (Auf der Mauer et al. 2018; Aynsley-Green et al. 2012; Cameriere et al. 2015; Dedouit et al. 2012; Ekizoglu et al. 2015, 2016; Krämer et al. 2014; Lottering et al. 2017; O'Connor et al. 2014; Ottow et al. 2017; Pinchi et al. 2014; Vieth et al. 2018; Wittschieber et al. 2014) are also limited by the difference between chronological and developmental age. However, this is not an argument for abandoning skeletal and dental age estimation. We instead advocate for using an age estimation approach, like that presented here, to provide an additional or supporting line of evidence with respect to actual age. The models presented here are flexible in accommodating different informative prior age distributions, number of traits, and age threshold values, but they are statistically structured to produce results that can be incorporated into other models. Because of the sequential nature of the Bayes theorem, the age estimates and threshold evaluation results generated here can be combined with other forms of quantitative evidence to strengthen probabilistic arguments, provided that the other quantitative evidence is independent of dental development.

Received 28 September 2020; accepted for publication 1 June 2021.

LITERATURE CITED

Acharya, A. B. 2011. Accuracy of predicting 18 years of age from mandibular third molar development in an Indian sample using Demirjian's ten-stage criteria. *Int. J. Legal Med.* 125:227–233.

Akkaya, N., H. Ö. Yılancı, H. Boyacıoğlu et al. 2019. Accuracy of the use of radiographic visibility of root pulp in the mandibular third molar as a maturity marker at age thresholds of 18 and 21. *Int. J. Legal Med.* 133:1,507–1,515.

Allen, D. M. 1974. The relationship between variable selection and data argumentation and a method for prediction. *Technometrics* 16:125–127.

Auf der Mauer, M., D. Säring, B. Stanczus et al. 2018. A 2-year follow-up MRI study for the evaluation of an age estimation method based on knee bone development. *Int. J. Legal Med.* 133:205–215.

Aynsley-Green, A., T. Cole, H. Crawley et al. 2012. Medical, statistical, ethical and human rights considerations in the assessment of age in children and young people subject to immigration control. *Br. Med. Bull.* 102:17–42.

Baba-Kawano, S., Y. Toyoshima, L. Regalado et al. 2002. Relationship between congenitally missing lower third molars and late formation of tooth germs. *Angle Orthod.* 72:112–117.

Balla, S. B., S. S. Chinni, I. Galic et al. 2019. A cut-off value of third molar maturity index for indicating a minimum age of criminal responsibility: Older or younger than 16 years? *J. Forensic Leg. Med.* 65:108–112.

Bera, A. K., C. M. Jarque, and L. F. Lee. 1984. Testing the normality assumption in limited dependent variable models. *Int. Econ. Rev.* (*Philadelphia*) 25:563–578.

Boldsen, J. L., G. R. Milner, L. W. Konigsberg et al. 2002. Transition analysis: A new method for estimating age from skeletons. In *Paleodemography: Age Distributions from Skeletal Samples*, R. D. Hoppa and J. W. Vaupel, eds. New York: Cambridge University Press, 73–106.

Bond, C. F., and K. Richardson. 2004. Seeing the Fisher Z-transformation. *Psychometrika* 69:291–303.

Braga, J., Y. Heuze, O. Chabadel et al. 2005. Non-adult dental age assessment: Correspondence analysis and linear regression versus Bayesian predictions. *Int. J. Legal Med.* 119:260–274.

Cameriere, R., L. Ferrante, D. De Angelis et al. 2008. The comparison between measurement of open apices of third molars and Demirjian stages to test chronological age of over 18 year olds in living subjects. *Int. J. Legal Med.* 122:493–497.

Cameriere, R., A. Giuliodori, M. Zampi et al. 2015. Age estimation in children and young adolescents for forensic purposes using fourth cervical vertebra (C4). *Int. J. Legal Med.* 129:347–355.

Cameriere, R., L. A. V. Palacio, J. Pinares et al. 2018. Assessment of second (I_{2M}) and third (I_{3M}) molar indices for establishing 14 and 16 legal ages and validation of the Cameriere's I_{3M} cut-off for 18 years old in Chilean population. *Forensic Sci. Int.* 285:205.e1–205.e5.

Cipriani, D. 2009. *Children's Rights and the Minimum Age of Criminal Responsibility: A Global Perspective*. Burlington, VT: Ashgate.

Corradi, F., V. Pinchi, I. Barsanti et al. 2013. Probabilistic classification of age by third molar development: The use of soft evidence. *J. Forensic Sci.* 58:51–59.

Corron, L., F. Marchal, S. Condemi et al. 2018. A critical review of sub-adult age estimation in biological anthropology: Do methods comply with published recommendations? *Forensic Sci. Int.* 288:328.e1–328.e9.

Dedouit, F., J. Auriol, H. Rousseau et al. 2012. Age assessment by magnetic resonance imaging of the knee: A preliminary study. *Forensic Sci. Int*. 217:232.e1–232.e7.

De Luca, S., R. Biagi, G. Begnoni et al. 2014. Accuracy of Cameriere's cut-off value for third molar in assessing 18 years of age. *Forensic Sci. Int.* 235:102.e1–102.e6.

Ekizoglu, O., E. Hocaoglu, E. Inci et al. 2015. Forensic age estimation by the Schmeling method: Computed tomography analysis of the medial clavicular epiphysis. *Int. J. Legal Med.* 129:203–210.

Ekizoglu, O., E. Hocaoglu, E. Inci et al. 2016. Forensic age estimation via 3-T magnetic resonance imaging of ossification of the proximal tibial and distal femoral epiphyses: Use of a T2-weighted fast spin-echo technique. *Forensic Sci. Int.* 260:102.e1–102.e7.

Fieuws, S., G. Willems, S. Larsen-Tangmose et al. 2016. Obtaining appropriate interval estimates for age when multiple indicators are used: Evaluation of an ad-hoc procedure. *Int. J. Legal Med.* 130:489–499.

Galić, I., T. Lauc, H. Brkić et al. 2015. Cameriere's third molar maturity index in assessing age of majority. *Forensic Sci. Int.* 252:191.e1–191.e5.

Garn, S. M., A. B. Lewis, D. L. Polacheck. 1960. Interrelations in dental development. I. Interrelationships within the dentition. *J. Dent. Res.* 39:1,049–1,055.

Garn, S. M., A. B. Lewis, and J. H. Vicinus. 1963. Third molar polymorphism and its significance to dental genetics. *J. Dent. Res.* 42:1,344–1,363.

Garn, S. M., and B. H. Smith. 1980. Developmental communalities in tooth emergence timing. *J. Dent. Res.* 59:1,178.

Green, L. J. 1961. The interrelationships among height, weight and chronological, dental and skeletal ages. *Angle Orthod.* 31:189–193.

Greene, W. H., and D. A. Hensher. 2010. *Modeling Ordered Choices: A Primer.* New York: Cambridge University Press.

Heard, N. A., and P. Rubin-Delanchy. 2018. Choosing between methods of combining *p*-values. *Biometrika* 105:239–246.

Hens, S. M., and K. Godde. 2020. New approaches to age estimation using palatal suture fusion. *J. Forensic Sci.* 65:1,406–1,415.

Hillewig, E., J. Degroote, T. Van der Paelt et al. 2013. Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: Towards more sound age estimates. *Int. J. Legal Med.* 127:677–689.

Hirk, R., K. Hornik, and L. Vana. 2020. mvord: An R package for fitting multivariate ordinal regression models. *J. Stat. Softw.* 93:1–41.

Hirk, R., K. Hornik, L. Vana et al. 2018. mvord: Multivariate Ordinal Regression Models. R package version 0.3.1. https://CRAN.R-project.org/package=mvord.

Johnson, P. A. 1996. A test of the normality assumption in the ordered probit model. *METRON* 54:213–221.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI* 14:1,137–1,145.

Konigsberg, L. W. 2015. Multivariate cumulative probit for age estimation using ordinal categorical data. *Ann. Hum. Biol.* 42:368–378.

Konigsberg, L. W., S. R. Frankenberg, and H. M. Liversidge. 2016. Optimal trait scoring for age estimation. *Am. J. Phys. Anthropol.* 159:557–576.

Konigsberg, L. W., S. R. Frankenberg, and H. M. Liversidge. 2019. Status of mandibular third molar development as evidence in legal age threshold cases. *J. Forensic Sci.* 64:680–697.

Krämer, J. A., S. Schmidt, K.-U. Jürgens et al. 2014. Forensic age estimation in living individuals using 3.0 T MRI of the distal femur. *Int. J. Legal Med.* 128:509–514.

Laird, N. M., and F. Mosteller. 1990. Some statistical methods for combining experimental results. *Int. J. Technol. Assess. Health Care* 6:5–30.

Liversidge, H. 2008. Timing of human mandibular third molar formation. *Ann. Hum. Biol.* 35:294–321.

Liversidge, H. M. 2011. Similarity in dental maturation in two ethnic groups of London children. *Ann. Hum. Biol.* 38:702–715.

Liversidge, H. M., M. Kalaiarasu Peariasamy, P. I. Ngom et al. 2017. A radiographic study of the mandibular third molar root development in different ethnic groups. *J. Forensic Odontostomatol.* 35:97–108.

Liversidge, H., and P. Marsden. 2010. Estimating age and the likelihood of having attained 18 years of age using mandibular third molars. *Br. Dent. J.* 209:1–12.

Lottering, N., C. L. Alston-Knox, D. M. MacGregor et al. 2017. Apophyseal ossification of the iliac crest in forensic age estimation: Computed tomography standards for modern Australian subadults. *J. Forensic Sci.* 62: 292–307.

Lucas, V. S., M. Andiappan, F. McDonald et al. 2016. Dental age estimation: A test of the reliability of correctly identifying a subject over 18 years of age using the gold standard of chronological age as the comparator. *J. Forensic Sci.* 61:1,238–1,243.

Lucas, V. S., F. McDonald, M. Neil et al. 2014. Dental age estimation: The role of probability estimates at the 10 year threshold. *J. Forensic Leg. Med.* 26:61–64.

Márquez-Ruiz, A. B., M. C. Treviño-Tijerina, L. González-Herrera et al. 2017. Three-dimensional analysis of third molar development to estimate age of majority. *Sci. Justice* 57:376–383.

Moorrees, C. F. A., E. A. Fanning, and E. E. Hunt. 1963. Age variation of formation stages for ten permanent teeth. *J. Dent. Res.* 42:1,490–1,502.

Noll, G. 2016. Junk science? Four arguments against the radiological age assessment of unaccompanied minors seeking asylum. *Int. J. Refug. Law* 28:234–250.

O'Connor, J. E., J. Coyle, C. Bogue et al. 2014. Age prediction formulae from radiographic assessment of skeletal maturation at the knee in an Irish population. *Forensic Sci. Int.* 234:188.e1–188.e8.

Ottow, C., R. Schulz, H. Pfeiffer et al. 2017. Forensic age estimation by magnetic resonance imaging of the knee: The definite relevance in bony fusion of the distal femoral- and the proximal tibial epiphyses using closest-to-bone T1 TSE sequence. *Eur. Radiol.* 27:5,041–5,048.

Parner, E., J. M. Heidmann, I. Kjaer et al. 2002. Biological interpretation of the correlation of emergence times of permanent teeth. *J. Dent. Res.* 81:451–454.

Pinchi, V., F. De Luca, F. Ricciardi et al. 2014. Skeletal age estimation for forensic purposes: A comparison of GP, TW2 and TW3 methods on an Italian sample. *Forensic Sci. Int.* 238:83–90.

Ravi, V., A. S. SP, S. B. Balla et al. 2020. Evaluation of secondary dentin deposition in lower first molars to indicate a legal age threshold of 14 years using receiver operating characteristic curves. *Med. Sci. Law* 60:249–256.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Sajaia, Z. 2008. BIOPROBIT: Stata module for bivariate ordered probit regression. https://econpapers.repec.org/software/bocbocode/s456920.htm.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461–464.

Šešelj, M. 2013. Relationship between dental development and skeletal growth in modern humans and its implications for interpreting ontogeny in fossil hominins. *Am. J. Phys. Anthropol.* 150:38–47.

Sironi, E., S. Gittelson, S. Bozza et al. 2020. Minor or adult? Introducing decision analysis in forensic age estimation. *Sci. Justice* 61:47–60.

Sironi, E., V. Pinchi, F. Pradella et al. 2018. Bayesian networks of age estimation and classification based on dental evidence: A study on the third molar mineralization. *J. Forensic Leg. Med.* 55:23–32.

Sironi, E., and F. Taroni. 2015. Bayesian networks for the age classification of living individuals: A study on transition analysis. *J. Forensic Sci. Med.* 1:124–132.

Sironi, E., J. Vuille, N. Morling et al. 2017. On the Bayesian approach to forensic age estimation of living individuals. *Forensic Sci. Int.* 281:e24–e29.

Snow, G. 2020. TeachingDemos: Demonstrations for teaching and learning. R package version 2.12. https://CRAN.R-project.org/package=TeachingDemos.

StataCorp. 2019. *Stata Statistical Software: Release 16.* College Station, TX: StataCorp.

Streckbein, P., I. Reichert, M. A. Verhoff et al. 2014. Estimation of legal age using calcification stages of third molars in living individuals. *Sci. Justice* 54:447–450.

Thevissen, P. 2013. *Dental Age Estimation in Sub-adults: Striving for an Optimal Approach*. Leuven, BE: KU Leuven.

Thomas, L. A., L. R. Thomas, S. B. Balla et al. 2021. Above or below 14 years? An orthopantomographic study based on chronological course of eruption of mandibular premolars and second molars in a sample of south Indian children. *Leg. Med.* (*Tokyo*) 48:101814.

UN Committee on the Rights of the Child. 2007. *Convention on the Rights of the Child. General Comment No. 10: Children's Rights in Juvenile Justice.* Geneva: UN Committee on the Rights of the Child.

UN Committee on the Rights of the Child. 2019. *Convention on the Rights of the Child. General Comment No. 24 on Children's Rights in the Child Justice System.* Geneva: UN Committee on the Rights of the Child.

UN General Assembly. 2019. *UN Global Study on Children Deprived of Liberty.* Geneva: UN Human Rights, Office of the High Commissioner.

United Nations. 1989. *Convention on the Rights of the Child.* New York: UN Human Rights, Office of the High Commissioner.

Ursini, B. 2015. Prosecuting child soldiers: The call for an international minimum age of criminal responsibility. *St Johns Law Rev.* 89:1,023–1,048.

Uys, A., H. Bernitz, S. Pretorius et al. 2018. Estimating age and the probability of being at least 18 years of age using third molars: A comparison between Black and white individuals living in South Africa. *Int. J. Legal Med.* 132:1,437–1,446.

Varadhan, R. 2015. alabama: Constrained nonlinear optimization. R package version 2015.3-1. https://CRAN.R-project.org/package=alabama.

Vieth, V., R. Schulz, W. Heindel et al. 2018. Forensic age assessment by 3.0T MRI of the knee: Proposal of a new MRI classification of ossification stages. *Eur. Radiol.* 28:3,255–3,262.

Wittschieber, D., R. Schulz, V. Vieth et al. 2014. The value of sub-stages and thin slices for the assessment of the medial clavicular epiphysis: A prospective multi-center CT study. *Forensic Sci. Med. Pathol.* 10:163–169.

Yadava, M., G. J. Roberts, and V. S. Lucas. 2011. Dental age assessment (DAA): Reference data for British children at the 10-year-old threshold. *Int. J. Legal Med.* 125:651–657.