



Comparison of three methods to estimate dental age in children

Bianca Gelbrich¹ · Corinna Carl^{2,3} · Götz Gelbrich^{2,4}

Received: 19 September 2018 / Accepted: 7 October 2019 / Published online: 14 November 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Objectives In 1973, Demirjian and colleagues introduced a staging for tooth mineralization. Staging of the teeth of the left mandible generates a dental maturity score and an estimate of dental age. An update of this method was published in 2001 by Willems and colleagues. The London Atlas (2008) is another method used to assess dental age. This study compared the precision of these three methods for forensic age estimation.

Materials and methods A total of 500 panoramic images of 251 boys and 249 girls aged 6–16 years had dental maturity evaluated by the same observer using the London Atlas and the methods of Demirjian and Willems. Individuals with syndromes potentially affecting dental maturation were excluded. The three methods were compared with respect to their precision in estimating chronological age.

Results Age estimates using Willems' method were significantly more precise than those obtained using Demirjian's method, while estimates obtained using the London Atlas method were comparable with both the others. The highest precision was obtained using the average of age estimates from Willems' method and the London Atlas.

Conclusions To estimate age in individuals suspected of being below 16 years of age, we suggest first applying the London Atlas then computing age from the Demirjian stages of particular teeth using Willems' method, and finally using the average age estimate of both methods.

Clinical relevance Combining the London Atlas and Willems' scoring method provides more precise estimates of dental age than the current practice of applying a single preferred method.

Keywords Age estimation · Dental age · Dental mineralization · Demirjian stage · Willems' method · London Atlas

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00784-019-03109-2>) contains supplementary material, which is available to authorized users.

✉ Bianca Gelbrich
Bianca.Gelbrich@medizin.uni-leipzig.de

¹ Department of Orthodontics, University Hospital Leipzig, Liebigstraße, 10-14/Building 1, 04103 Leipzig, Germany

² Institute of Clinical Epidemiology and Biometry, University of Würzburg, Am Schwarzenberg 15/Building A15, 97078 Würzburg, Germany

³ Department of Orthodontics, University Hospital Würzburg, Pleicherwall 2, 97070 Würzburg, Germany

⁴ Clinical Trial Center Würzburg, University Hospital Würzburg, Josef-Schneider-Str. 2/Building D7, 97080 Würzburg, Germany

Introduction

Forensic scientists use age estimation to help discover the identity of unidentified dead individuals. In living persons, age estimation can be used, for example, to decide whether a perpetrator with unproven age (mostly young immigrants without documents) has reached the age of criminal responsibility and, if so, whether that age is adult or still minor, and therefore which criminal code is applicable. The maturation of bones and teeth is strongly correlated with chronological age in young individuals and can therefore be used to estimate age. Current recommendations in Germany [1] and in other countries suggest incorporating dental age into the age estimation procedure. Research in this area has focused on maximizing the precision of age estimation methods to keep the frequency of erroneous decisions in legal proceedings as low as possible. For an overview on age estimation in clinical dentistry, the interested reader is referred to a recent work by Kirschneck and Proff [2].

Several schemes of assessing dental maturity have been proposed. One of the most popular is the classification of mineralization stages introduced in 1973 by Demirjian and colleagues [3]. Each tooth from the first incisor to the second molar in the left mandible is assigned a stage from A (beginning mineralization) to H (apex closed). These stages are converted to maturity scores using one table, then another table is used to convert the sum score into an estimated age.

The Demirjian method was updated in 2001 by Willems and co-workers [4] using newer data from Belgian children. The updated method also applies Demirjian stages A–H, but the resulting sum score provides the estimated age, meaning that there is no second conversion step. Both the Demirjian and Willems methods are applicable up to 16 years of age because mineralization of the incorporated teeth has then finished in most individuals meaning that these methods cannot be used to discriminate ages beyond this threshold.

In 2010, an atlas method was proposed by a research group from London [5, 6]. The atlas consists of a series of schematic reference images of the whole dentition for certain ages. The expert performing age estimation determines which of the reference images best matches the panoramic radiograph of the individual being examined, and the age corresponding to that reference image is the estimated age assigned to the individual. After the first year of life, the London Atlas assigns age in 1-year ranges.

This study compared the performance of the three available methods (the “classical” Demirjian scheme, its update by Willems, and the London Atlas) for forensic age estimation.

Methods

Routine panoramic radiographs from children treated at the orthodontic department of the University Hospital Würzburg were assessed. Eligibility criteria were age 6–16 years, sufficient image quality and presence of teeth in order to apply the London Atlas and to perform Demirjian staging. Individuals were excluded if there were any notes in their patient records indicating that they had or might have had a systemic disorder that would potentially affect the speed of dental maturation.

Panoramic radiographs were digitalized using an X-ray scanner and stored as bitmaps. The examiner was free to apply any standard image processing procedures such as changing size, brightness or contrast so that they felt subjectively comfortable when rating dental age or stages of particular teeth. Ratings were carried out by the second author who had previous training and supervised by the last author during data capture. For each of the panoramic radiographs, age according to the London Atlas was rated first then Demirjian staging was done at least 14 days later.

The London Atlas [5] provides a series of schematic one-sided reference images of the deciduous and permanent

dentition for different ages (in steps of 1 year from age 1.5 years onwards). The observer had to select the reference image that best matched the panoramic radiograph under consideration. To assess the extent of matching, all teeth were included, and stages of mineralization, eruption, and root resorption (of deciduous teeth) were taken into account. The observer chose one of the reference images, and age corresponding to that image was the estimated age according to the London Atlas. In-between estimates were not used. The same series of reference images was applicable for both sexes.

For Demirjian staging, the observer used the schematic and radiographic reference images provided by Demirjian and colleagues [3] to assess incisors then through to the second molars in the left mandible. Each tooth was assigned one of eight stages of mineralization (from A through H). Again, in-between stages were not used. If any tooth in the left mandible was missing, the contra-lateral tooth in the right mandible was used as substitute. Stages were converted into scores based on the tables provided [3], and the sum score was converted into estimated age. The conversion tables were implemented as program syntax in statistical software (SPSS), so that the computation of estimated age from mineralization stages was done automatically. The same reference images were applicable for both sexes, but there were separate conversion tables for boys and girls. The same teeth mineralization stages were used for age estimation using Willems’ method. Again, stages were converted into scores, which provided the estimated age.

Repeat assessment of 30 randomly selected panoramic images was carried out to assess intra-observer agreement for all assessments.

Statistical analyses

Methods were compared for bias (systematic over- or under-estimation of age on average across all individuals) and mean absolute error (absolute difference between estimated and true age) using the one-sample *t* test for assessment of each method. Repeat measurement analysis of variance was used to compare the methods to each other; Cochran’s *Q* test was used for related samples to determine the frequency of absolute errors above 2 years. In addition, a case-wise comparison between each pair of methods was carried out by defining that for each two methods A and B and for each case X:

$$\begin{aligned} \text{Method A wins if } & |\text{AgeEstimate}(A,X) - \text{TrueAge}(X)| \\ & < |\text{AgeEstimate}(B,X) - \text{TrueAge}(X)| \\ \text{Method B wins if } & |\text{AgeEstimate}(A,X) - \\ & \text{TrueAge}(X)| > |\text{AgeEstimate}(B,X) - \text{TrueAge}(X)| \\ \text{Tie of A and B if } & |\text{AgeEstimate}(A,X) - \\ & \text{TrueAge}(X)| = |\text{AgeEstimate}(B,X) - \text{TrueAge}(X)| \end{aligned}$$

The null hypothesis of equal probabilities of winning of A and B was examined using the binomial test. As a fourth

method of age estimation, we computed the average of the age estimates obtained from the two newer methods (Willems’ and London Atlas), and compared them to those of the three methods described above. Two-sided *P* values of < 0.05 were considered statistically significant.

Sample size

It was calculated that 471 subjects would be required to achieve 95% power to detect a bias of ± 3 months for a method with a standard deviation of errors of 1.5 years. A sample of 327 subjects would provide 95% power to detect a superiority of 60:40 in the case-wise comparison of two methods. Therefore, sample size was set at 500.

Ethics

This study retrospectively evaluated anonymous X-ray images obtained in routine clinical practice. Therefore, no ethical evaluation or approval was required according to German law (written statement by the Ethics Committee of the Medical Faculty of the University of Würzburg, reference number 20170317-01).

Results

A total of 500 patients were included (Table 1). The intra-class correlation coefficient for intra-observer agreement was 0.95 for Demirjian’s method and 0.98 for the age estimates obtained by the London Atlas.

Performance data for the three methods and their comparisons are summarized in Table 2. Demirjian’s method had only a modest bias towards estimating age as too low, whereas Willems’ method systematically underestimated age by an

average of 4.5 months, and the London Atlas overestimated age by approximately 3.5 months. However, the mean absolute error of Willems’ method and the London Atlas was smaller than the mean absolute error of Demirjian’s method, but only the comparison between Willems’ and Demirjian’s method reached statistical significance. The same result was obtained when considering the number of individuals with an absolute age estimation error of > 2 years. In the case-wise comparison (Table 3, upper part), Willems’ method tended to perform better than the other two methods, but this trend did not reach statistical significance. The London Atlas was comparable with Demirjian’s method.

Bias almost disappeared when the average age estimates of Willems’ method and the London Atlas were used because these two methods had similar biases in opposite directions. Furthermore, compared with each of the three individual methods, the average method had a bias that was significantly closer to zero with a narrower confidence interval, and the mean absolute error and percentage of cases with an absolute error of > 2 years were significantly smaller (Table 2).

In the case-wise comparison, the average of the Willems’ and London Atlas methods provided a significantly more precise estimate of age than each of the three individual methods (Table 3, lower part). Thus, the average of two methods performed best with respect to all criteria considered.

Figure 1 displays the errors quoted in Table 2 separately for both sexes. The significant overall bias of Demirjian’s method was attributable solely to underestimation of age in boys, and there was no bias in girls. The underestimation of age by the Willems method was comparable for boys and girls, and the overestimation of age by the London Atlas method was considerably larger in girls than in boys. Notably, the average method had modest biases with different signs for boys and girls, but the difference between both small biases was significant. Mean absolute errors of all methods were comparable for both sexes. However, the rates of absolute errors of > 2 years were larger in boys than in girls for Demirjian’s method and larger in girls than in boys for the London Atlas, and were similar for both sexes using Willems’ and the average methods.

Figure 2 shows the case-wise comparisons from Table 3 carried out for boys and girls separately. On the horizontal axis, differences in the percentage of cases won by the first method minus those won by the second method are shown. The extent of superiority or inferiority of methods did not differ significantly between boys and girls, apart from more pronounced superiority of the average method over the Willems method in girls.

Figure 3 displays the distribution of errors of age estimates obtained by the average of two methods. The mean error (bias) was – 0.04 years (i.e., a mean underestimation by 2 weeks), and the standard deviation was 0.96 years.

Table 1 Age at panoramic imaging and sex of included subjects

Age (years)	Number of subjects		
	Boys	Girls	Total
6 to < 7	6	5	11
7 to < 8	9	14	23
8 to < 9	19	29	48
9 to < 10	27	35	62
10 to < 11	39	33	72
11 to < 12	43	41	84
12 to < 13	34	37	71
13 to < 14	35	24	59
14 to < 15	24	23	47
15 to 16	15	8	23
Total	251	249	500

Table 2 Comparison of precision of prediction of chronological age by dental age based on Demirjian's and Willems' scoring methods, the London Atlas, and the average of Willems' method and the London Atlas

Mean bias		<i>P</i> value for the comparison with ...		
Method	Bias (years)	Willems	London	Average
Demirjian	-0.11 (-0.23 to -0.00); <i>P</i> = 0.043	< 0.001	< 0.001	0.042
Willems	-0.38 (-0.47 to -0.29); <i>P</i> < 0.001	–	< 0.001	< 0.001
London	+0.30 (+0.20 to +0.39); <i>P</i> < 0.001	–	–	< 0.001
Average	-0.04 (-0.12 to +0.04); <i>P</i> = 0.348	–	–	–
Mean absolute error		<i>P</i> value for the comparison with ...		
Method	MAE (years)	Willems	London	Average
Demirjian	0.97 (0.90 to 1.04)	< 0.001	0.072	< 0.001
Willems	0.88 (0.82 to 0.94)	–	0.631	< 0.001
London	0.90 (0.84 to 0.96)	–	–	< 0.001
Average	0.76 (0.71 to 0.81)	–	–	–
Absolute error > 2 years		<i>P</i> value for the comparison with ...		
Method	Frequency	Willems	London	Average
Demirjian	53 (10.6%)	0.002	0.105	< 0.001
Willems	32 (6.4%)	–	0.519	< 0.001
London	38 (7.6%)	–	–	< 0.001
Average	13 (2.6%)	–	–	–

Bias and mean absolute error (MAE) are presented as means with 95% confidence intervals. The *P* value for the bias refers to the null hypothesis that the bias equals zero. *P* values for comparison refer to the null hypotheses that the two methods being compared have equal bias, equal mean absolute errors, or equal percentages of cases with an absolute error above 2 years, respectively

Discussion

Willems' update of Demirjian's age estimation method improved the absolute precision of the original method but increased the bias towards underestimation of age. The London Atlas was competitive with Demirjian's method, although its 1-year classification bands limit the precision of age estimation that can be achieved. Willems' method tended to be a bit

more precise than the London Atlas, but none of the comparisons achieved statistical significance. Beyond numerical performance, some technical advantages of the London Atlas should be mentioned. First, it is very easy to use and takes less time than staging of individual teeth. Second, the scoring method cannot be applied in the simultaneous absence of contralateral teeth in the left and right mandible. In particular, about 5% of the population have agenesis of at least one

Table 3 Case-wise comparison of Demirjian's (DS) and Willems' (WS) scoring methods, the London Atlas (LA), and the average of WS and LA (AM) to each other with respect to the precision of the prediction of chronological age by dental age

Methods first–second	Pairwise comparisons of the three methods		
	DS–WS	DS–LA	WS–LA
First method wins	231 (46%)	242 (48%)	262 (52%)
Second method wins	269 (54%)	245 (49%)	236 (47%)
Tied	0	13 (3%)	2 (<1%)
<i>P</i> value	0.098	0.928	0.263
Methods first–second	Comparisons of the three methods with the average method		
	DS–AM	WS–AM	LA–AM
First method wins	194 (39%)	215 (43%)	191 (38%)
Second method wins	306 (61%)	283 (57%)	307 (61%)
Tied	0	2 (<1%)	2 (<1%)
<i>P</i> value	< 0.001	0.003	< 0.001

DS, Demirjian score method; WS, Willems score method; LA, London Atlas method; AM, average of WS and LA methods. Data are the numbers and percentages of individuals for whom the first or the second method provided the more precise age estimate, and for whom the estimates were equally precise. *P* values refer to the null hypothesis that the percentages of won cases would be equal for the first and second method

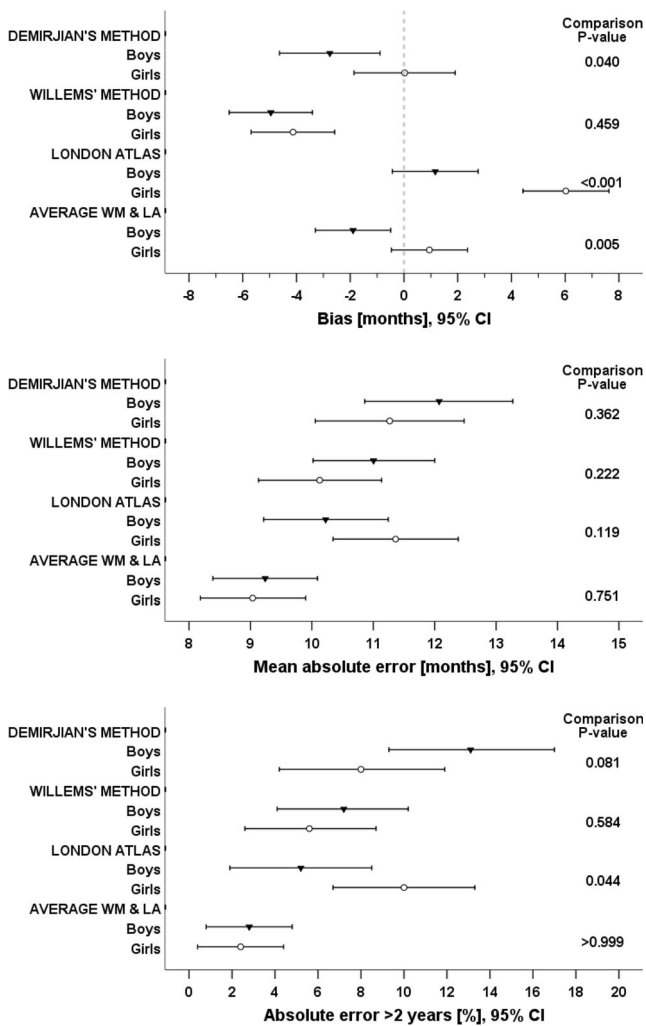


Fig. 1 Bias, mean absolute errors, and percentages of absolute errors of > 2 years in boys and girls for the three methods and the average of Willems' method and the London Atlas. Comparison *P* values refer to the null hypothesis that errors are equal in boys and girls

second premolar [7, 8], the mandibular teeth are more frequently affected, and both mandibular second premolars are often missing. Both Demirjian and Willems scores cannot be computed in these cases, while the Atlas method can be used because it can be applied by looking at the rest of the dentition. This is also the case when some teeth cannot be assessed due to problems with image quality.

To compare our results with other studies examining the accuracy of Willems' method, we refer to a recent meta-analysis incorporating 23 studies in various populations [9]. The mean biases reported ranged from underestimation by 8 months to overestimation by 7 months, compared with the underestimation by 4.5 months in our sample. There was a wide range of biases across studies and this did not appear to be associated with the region of origin (which might be explained by delayed mineralization in some populations due to malnutrition resulting from poverty). Therefore, we suggest that discrepancies in classification into Demirjian

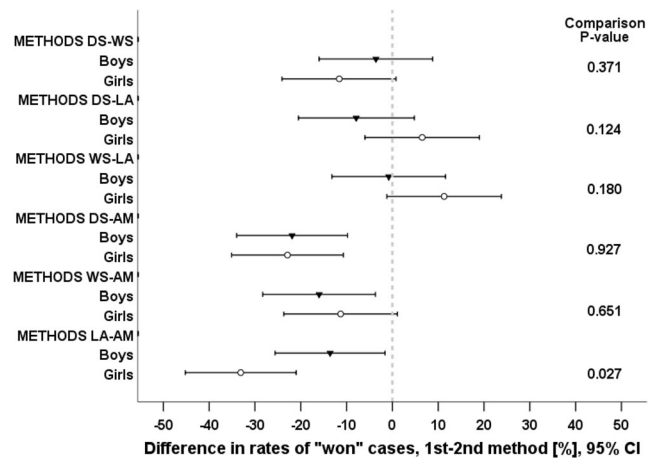


Fig. 2 Case-wise comparison of Demirjian's (DS) and Willems' (WS) scoring methods, the London Atlas (LA), and the average of WS and LA (AM) to each other in boys and girls. Positive/negative differences of rates of "won" cases mean that the first/s method is superior, respectively. Comparison *P* values refer to the null hypothesis that the differences of rates are the same in boys and girls

stages by observers from all over the world might be responsible for the wide variations reported.

Validation studies for the London Atlas are more consistent. Three recent studies [10–12] reported a slight overestimation of chronological age. Of those, data from a European population [10] showed a mean overestimation by 3.5 months, which is consistent with our finding.

Each of the methods has specific limitations. Discrimination of ages based on first incisors through second molars is poor from the age of 14 years onwards and impossible beyond age 16 years [13]. The scoring methods by Demirjian and Willems do not incorporate third molars, limiting their diagnostic power at the upper bound of the age tables. In addition, overestimation of age is the less likely as true age approaches 16 years (and impossible if this age has been reached). This results in age-dependent negative bias. On the other hand, tooth mineralization in girls is slightly ahead of that in boys when age is < 15 years (as confirmed in a recent study [13]). This probably explains the variation between sexes when using the London Atlas because this method is not sex specific. A combination of both methods might reduce the estimation errors occurring due to these limitations.

Indeed, our data showed that the best precision, with almost no bias, was achieved using the average of age estimates obtained from Willems' method and the London Atlas. This is probably because the information obtained using each of these methods adds to the other. Scoring of particular teeth might be considered more objective and more detailed, while the overall assessment may have a larger subjective component. On the other hand, the atlas incorporates all teeth while the Willems method scores only the teeth of the left mandible (or their substitutes in the right mandible). This means that

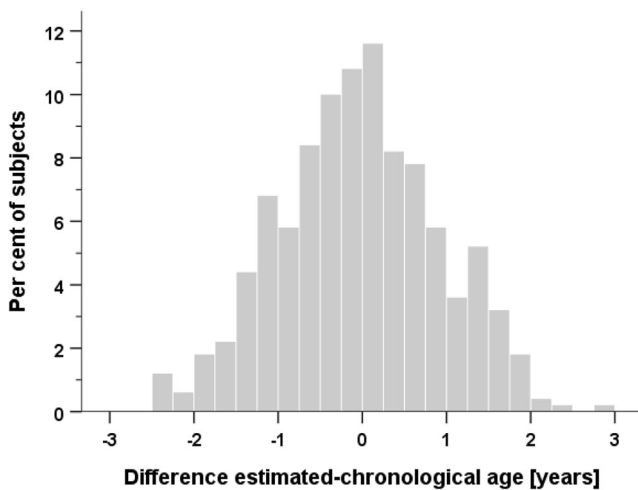


Fig. 3 Distribution of the age estimation errors of the average of Willems' method and the London Atlas

this method ignores asymmetric teeth development whereas an observer using the atlas could take this information into account. Furthermore, biases of Willems' method and the London Atlas are in opposite directions such that these balance each other out in the average method (particularly the large bias of the atlas in girls).

Based on our findings, we suggest applying both methods simultaneously and using their average for dental age estimation when age is < 16 years. This is a new approach compared with the current practice of using a single preferred method. When methods are combined, the London Atlas should be applied first followed by assessment of Demirjian stages in the lower left mandible. The rationale for this recommendation is that the overall view of the dentition when applying the atlas method will probably have little influence on subsequent Demirjian staging of individual teeth. Conversely, if Demirjian stages were determined first and used to calculate estimated age, this could bias subsequent interpretation of overall dentition using the London Atlas meaning that the two methods did not provide independent information, and the error variance seen in our study would not be achieved.

In forensic age estimation, the combination of several methods involving different anatomic structures is recommended [1]. When combining age estimates obtained from teeth and hand bones [14, 15], the weighted average of both estimated ages had been suggested for the common estimate, where the weights should be inversely proportional to the variances of the errors of both methods [16]. Assuming that the standard deviation for the Thiemann hand atlas method was 0.97 years [16] and that for the average of Willems' method and the London Atlas was 0.96 years, the weights for the two methods would be almost equal, and the common estimate would therefore

be the usual average of dental and skeletal age (with a standard deviation of errors of 0.68 years).

The use of an orthodontic patient sample should be mentioned as a limitation of this study. Although patients with disorders potentially affecting dental mineralization were excluded, the variance in the speed of tooth mineralization and of eruption in individuals with an orthodontic treatment indication might be larger than in the general population. Therefore, age estimates might be less precise, and error rates for the detection of age thresholds could be increased. As a consequence, the standard deviation of the errors of the average of Willems' method and the London Atlas in the general population might be less than the 0.95 years reported above. In the combined age estimation including skeletal age obtained from the hand, the teeth may then get a higher weight, and the standard deviation of the errors of the combined estimates might be less than the above-quoted 0.68 years. However, the validity of the comparison of several methods of age estimation should not be diminished by this limitation, because the higher biological variability in the sample would influence the results of all methods in the same manner, and therefore the property of one method being superior to another should not be altered.

In conclusion, the results of this study showed that a combination of the London Atlas and Willems' scoring method provides more precise estimates of dental age than the current practice of applying a single method.

Acknowledgements The authors wish to thank Professor Angelika Stellzig-Eisenhauer, director of the Department of Orthodontics of the University Hospital Würzburg, for granting access to the radiographic archive data of her polyclinic and providing the opportunity to extract anonymous image material for scientific research. The authors also wish to express their gratitude to the anonymous reviewers whose comments were very helpful in improving the presentation of this work.

Authors' contributions BG had the idea and developed the plan for the study. CC collected the data under the guidance and supervision of GG, and BG checked data for plausibility. GG performed the statistical analyses. BG and GG interpreted the results and prepared the first draft of the manuscript. CC critically revised the manuscript. All authors approved the submission and take responsibility for the integrity of the data and analyses.

Compliance with ethical standards

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

Ethical approval This study retrospectively evaluated anonymous radiographic images obtained in routine clinical practice. Therefore, no ethical evaluation or approval was required according to applicable German law. This was confirmed by the Ethics Committee of the Medical Faculty of the University of Würzburg (reference number 20170317-01).

Informed consent It was determined that no formal consent was required for this study.

References

- Schmeling A, Grundmann C, Fuhrmann A, Kaatsch HJ, Knell B, Ramsthaler F, Reisinger W, Riepert T, Ritz-Timme S, Rösing FW, Röttscher K, Geserick G (2008) Updated recommendations of the study group on forensic age diagnostics for age estimation in living individuals in criminal proceedings. *Rechtsmedizin* 18:451–453 [German]
- Kirschneck C, Proff P (2018) Age assessment in orthodontics and general dentistry. *Quintessence Int* 49:313–323
- Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. *Hum Biol* 45:211–227
- Willems G, van Olmen A, Spiessens B, Carels C (2001) Dental age estimation in Belgian children: Demirjian's technique revisited. *J Forensic Sci* 46:893–895
- AlQahtani SJ, Liversidge HM, Hector MP (2010) Atlas of tooth development and eruption. *Am J Phys Anthropol* 142:481–490
- AlQahtani SJ, Hector MP, Liversidge HM (2014) Accuracy of dental age estimation charts: Schour and Massler, Ubelaker and the London Atlas. *Am J Phys Anthropol* 154:70–78
- Kenrad JB, Christensen IJ, Kjær I (2013) Gender differences in patterns of second premolar agenesis observed in 4,756 individuals. *Eur Arch Paediatr Dent* 14:397–403
- Gelbrich B, Hirsch A, Dannhauer KH, Gelbrich G (2015) Agenesis of second premolars and delayed dental maturation. *J Orofac Orthop* 76:338–350
- Mohd Yusof MYP, Wan Mokhtar I, Rajasekharan S, Overholser R, Martens L (2017) Performance of Willem's dental age estimation method in children: a systematic review and meta-analysis. *Forensic Sci Int* 280:245.e1–245.e10
- Pavlović S, Palmela Pereira C, Vargas de Sousa Santos RF (2017) Age estimation in Portuguese population: the application of the London atlas of tooth development and eruption. *Forensic Sci Int* 272:97–103
- Baylis S, Bassed R (2017) Precision and accuracy of commonly used dental age estimation charts for the New Zealand population. *Forensic Sci Int* 277:223–228
- Ghafari R, Ghodousi A, Poordavar E (2018) Comparison of the accuracy of the London atlas and smith method in dental age estimation in 5-15.99-year-old Iranians using the panoramic view. *Int J Legal Med*. <https://doi.org/10.1007/s00414-018-1808-6>
- Hofmann E, Robold M, Proff P, Kirschneck C (2017) Age assessment based on third molar mineralisation: an epidemiological-radiological study on a Central-European population. *J Orofac Orthop* 78:97–111
- Thiemann HH, Nitz I, Schmeling A (2006) Radiographic atlas of the normal hand in childhood, 3rd edn. Thieme, Stuttgart [German]
- Greulich WW, Pyle SI (1959) Radiographic atlas of skeletal development of the hand and wrist. Stanford University Press, Stanford
- Gelbrich B, Frerking C, Weiss S, Schwerdt S, Stellzig-Eisenhauer A, Tausche E, Gelbrich G (2015) Combining wrist age and third molars in forensic age estimation: how to calculate the joint age estimate and its error rate in age diagnostics. *Ann Hum Biol* 42: 389–396

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.