Inter-observer agreement in the radiographic interpretation of Demirjian's developmental stages in the mandibular second and third molars – A comparative study

Jayasankar P. Pillai[1], Debesh Nilendu[2], Namitha Thomas[3], Sugandha Nagpal[4], Lakshmi Sai Sneha Nedunari[5]

[1]Department of Oral and Maxillofacial Pathology, Government Dental College and Hospital, Ahmedabad, [2]Medical Officer, Health India Insurance TPA, Vadodara, Gujarat, [3]Forensic Odontologist & Consultant Dental Surgeon, Kannur, Kerala, [5]Scientific Officer, Sherlock Institute of Forensic Science, [5]General Forensic Assistant, Delhi Police Crime Branch, New Delhi, India

Abstract

Background: The developmental stages of the teeth in the radiographs are graded on an ordinal scale. The present study was conducted using 123 digital orthopantomograms from individuals in the age group of 5 to 22 years and to analyze and evaluate the inter-observer agreement in grading the developmental stages of second and third molars. Four observers with different levels of practical experience in age estimation participated in the study. The development stages of both the molars in the lower left quadrant (3rd quadrant) were assigned based on the Demirjian's 10 stage chart. The percentage agreement and Kappa statistics were used to test the agreement between the observers. The data of the observer 1 were used as the standard for the comparison.

Results: There was 70.0%–75.6% agreement among the observers in staging second molar and 52%–68.3% for the third molars. There was an excellent agreement ($k > 080$) between observer 1 and observer 2 and a good agreement ($k = 060$–0.79) between observer 1 and the other two observers for both the molars. The Fleiss Kappa revealed moderate to a good overall agreement for both the molars ($k = 0.51$–0.66). The Freidman's test revealed a significant difference in the grading of third molars between all the raters ($\chi^2 = 25.02$, df 3, $P < 0.001$) and for the second molar the difference was not significant ($\chi^2 = 3.89$, df 3, $P > 0.05$). The stage-wise overall agreement was fair for Stage 3 in the second molar and Stage 9 and Stage 4 in the third molar.

Conclusions: In conclusion, proper training in the radiographic interpretation of developmental stages may minimize the errors during the age estimation methods.

Keywords: Demirjian's method, dental age estimation, inter-observer agreement, kappa statistics, orthopantomograms, percentage agreement

Address for correspondence: Dr. Jayasankar P. Pillai, Department of Oral and Maxillofacial Pathology, Government Dental College and Hospital, Ahmedabad - 380 016, Gujarat, India.
E-mail: jppillaigdch@gmail.com

INTRODUCTION

In forensic dentistry, the age estimation from radiographs for medico-legal purposes is one of the main domains and panoramic radiographs are often used for this purpose. The

Access this article online

Quick Response Code:

Website:
www.jomfp.in

DOI:
10.4103/jomfp.jomfp_85_21

reliability and acceptability of such age estimation methods based on graded interpretation of the dental radiographs depend on the repeatability and the agreement between observers when a particular radiograph is graded by more than one observer (inter-observer) or repeatedly by the same observer (intra-observer). Hence, studies testing the reliability and agreement are also important. The quality of such studies depends on the amount of error inherent to scoring or grading the developmental stages of teeth as observed in the radiographs. Using the radiographs, dental maturity, the degree of mineralization, and the extent of root formation are assessed based on certain preexisting standards.[1-3] The eight stages (A-H) standards developed by Demirjian *et al.* and later modified into 10 stages numeric scale (0–10) by Chaillet and Demirjian are the most widely accepted methods worldwide.[4] These above-mentioned scales developed by Demirjian *et al.* also formed the basis for developing the population-specific and tooth-specific age estimation methods.[5-9] As ordinal scales are used to grade the developmental stages of teeth, there exists a chance for subjectivity in grading by different raters or observers. The radiographic studies based on heterogeneous groups of observers are thus the need of the hour. Discrepancies between observers in the range of ± one stage have been observed while grading the teeth using radiographs. Studies have found significant effects of intra- and inter-observer variations on age estimation from the assessment of the stage of development, which is greater between observers, with the majority confined to one development stage.[10] Furthermore, the level of agreement can be improved by using reference radiographs for calibration and by prior and repeated calibrations of the observers. The level of practical experience by the observers also plays an important role while interpreting the developmental stage of teeth for age estimation purposes. In the dental undergraduate curriculum, dental students are exposed to the theoretical and practical training on dental age estimation methods using dental models and X-rays.[11] This may be an academic approach to teach dental age estimation methods, but when dealing with real-life forensic cases, the forensic odontologists or any qualified dentists need to be highly trained to confidently and accurately handle the age estimation cases. The dental age estimation by radiographic methods is one of the sought after topics in several fellowships and hands-on training programs in forensic odontology. However, the issue of subjectivity or inter-observer variations in grading the developmental stages is often encountered both during research and practical applications. The level of agreement between observers when dealing with ordinal data is usually tested using Kappa statistics.[12] The kappa values of <0.20 suggest poor agreement, 0.21–0.40 fair agreement,

0.41–0.60 moderate agreement, 0.61–0.80 good agreement, and >0.81 excellent agreement.[13] The purpose of this study was to investigate how consistently multiple observers assign the dental developmental stages for mandibular second and third molars using the digital orthopantomograms (OPGs). The present study was conducted with four independent observers, with different levels of experience in dental age estimation. The first observer was an experienced forensic odontologist with additional training in dental age estimation and who has practically performed dental age estimation on several medico-legal cases and conducted hand-on training in dental age estimation as a trainer in several continuing dental education programs. The second and the third observers were dental graduates with postgraduate training in forensic odontology. Both had different levels of training on dental age estimation before the start of the study. The observer 2 had both theoretical and hands-on practical training on age estimation, whereas the observer 3 hands only theoretical training, without much exposure to practical hands-on. The fourth observer was a general dentist without any basic training on dental age estimation methods.

METHODS

The sample comprised of digital panoramic radiographs from a total of 123 individuals (63 females, 60 males) aged 5 to 22 years with a mean age of 13.79 years ± 3.62. Participants with similar socioeconomic status and ethnic origin were included in the study. The radiographs were archived from the departments of orthodontia and pedodontia of the institute and comprised of patients from diverse regions of Gujarat seeking orthodontic and pediatric dental treatments. The radiographs were pre-treatment in nature and belonged to healthy individuals with no obvious developmental anomalies, especially in the mandibular second and third molar regions. The development stages of mandibular second and third molars on the lower left quadrant were evaluated based on the Demirjian and Chaillet's 10 stage chart and description.[4] The radiographs were randomly selected by one of the authors and were sent to the remaining four observers, who were blinded about the sex and age of the patients. Thus, the four observers independently assigned the developmental stages of both the molars. The grading of observer 1 was considered as the reference standard for comparison. The chronological age was calculated from the date of birth and the date of X-ray using the Excel sheet (Microsoft®, Redmond, Washington, U.S). To assess the intra-observer error in grading the stages of teeth, a set of 20 OPGs from the existing set of OPGs were re-evaluated by all the four observers for both the molars.

Statistical methods

The inter-observer reliability was determined using Kappa statistics, based on the data from each observer. The degree of agreement between each pair of observers was assessed by calculation of weighted Kappa (κ) statistics. The overall agreement in the grading of stages among all the four raters was tested using Fleiss Kappa. The intra-observer difference in grading the stages was tested using the Wilcoxon signed-rank test. The overall difference in the grading between the observers for all the four teeth was tested using the Freidman's non-parametric test, which is an alternative to the one-way ANOVA with repeated measures. This test compares the mean ranks between the related groups and indicates how the groups differ. The Statistical Package for the Social Studies (SPSS) software version 26 (IBM Corp. Armonk, NY: IBM Corp.) $P < 0.05$ was set as a significance threshold. The Microsoft Office Excel® 2007 (Microsoft®, Redmond, Washington, U.S) was used for the descriptive statistics and to calculate the percentage agreement between the pairs of raters.

RESULTS

The intra-observer agreement in grading the test and retest samples showed a significant moderate to a near excellent agreement ($k = 0.512$–0.876; $P < 0.001$) for both the molars by all the observers [Table 1]. The Wilcoxon signed-rank test also showed an insignificant difference ($P > 0.05$) in the grading between the test and the retest samples by all the observers for both the teeth. Tables 2 and 3 illustrate the frequency distribution of the second and the third molar respectively according to their developmental stages by all the four observers. The percentage agreement in grading the stages between the pair of observers with observer 1's data as reference ranged between 70.7% and 75.6% in the second molars and between 52% and 68.3% in the third molars [Tables 4 and 5]. The disagreement was maximum for stage 7 in the second molar and stage 3 in the third molar. Cohen's weighted kappa statistics revealed a good ($k = 0.61$–0.80) to an excellent agreement ($k \geq 0.80$) between the reference data set of observer 1 and the other three observers for both the molars [Table 6]. The overall agreement tested using the Fleiss kappa statistics between the data of all the observers' kappa (k) value of 0.66 and 0.51 for second and third molars, respectively [Table 7]. The nonparametric tests revealed a significant difference in the grading among all the observers for the third molar, whereas the second molar grading did not show any significant difference [Table 8]. The pair-wise comparison of grading by observer 1 (reference data) with that of other observers done using the Wilcoxon signed-rank test revealed a significant difference with observer 3

and observer 4 for the third [Table 9]. The Fleiss Kappa agreement on individual stage categories among all the

Table 1: Intra-observer agreement in staging the development stages of #37 and #38 tested using Cohen's kappa test and Wilcoxon signed rank test

Observer	#37				#38			
	Cohen's Kappa test		Wilcoxon signed-rank test		Cohen's Kappa test		Wilcoxon signed rank test	
	Kappa	P*	Z	P**	Kappa	P*	Z	P**
Observer 1	0.799	0.000	−0.577	0.564	0.876	0.000	−1.414	0.157
Observer 2	0.608	0.000	−1	0.317	0.814	0.000	−0.816	0.414
Observer 3	0.535	0.000	−1.265	0.206	0.512	0.000	−1.508	0.132
Observer 4	0.552	0.000	−1.318	0.187	0.608	0.000	−0.647	0.518

*Significant at $P < 0.005$, **Significant at $P < 0.05$

Table 2: The frequency distribution of the second molar (#37) based on their developmental stages by all the observers

Stage	#37				
	Observer 1, n (%)	Observer 2, n (%)	Observer 3, n (%)	Observer 4, n (%)	Total, n (%)
3	0	1 (0.8)	1 (0.8)	0	2 (0)
4	1 (0.8)	2 (1.6)	2 (1.6)	2 (1.6)	7 (1)
5	11 (8.9)	12 (9.8)	12 (9.8)	14 (11.4)	49 (10)
6	18 (14.6)	20 (16.3)	20 (16.3)	18 (14.6)	76 (15)
7	21 (17.1)	9 (7.3)	11 (8.9)	14 (11.4)	55 (11)
8	31 (25.2)	43 (35.0)	42 (34.1)	38 (30.9)	154 (31)
9	41 (33.3)	36 (29.3)	35 (28.5)	37 (30.1)	149 (30)
Total	123 (100)	123 (100)	123 (100)	123 (100)	492 (100)

Table 3: The frequency distribution of the third molars (#38 and #48) based on their developmental stages by all the observers

Stage	#38				
	Observer 1, n (%)	Observer 2, n (%)	Observer 3, n (%)	Observer 4, n (%)	Total, n (%)
0	1 (0.8)	2 (1.6)	2 (1.6)	2 (1.6)	7 (1)
1	9 (7.3)	11 (8.9)	9 (7.3)	8 (6.5)	37 (8)
2	16 (13)	10 (8.1)	10 (8.1)	6 (4.9)	42 (9)
3	14 (11.4)	10 (8.1)	16 (13.0)	14 (11.4)	54 (11)
4	23 (18.7)	14 (11.4)	19 (15.4)	14 (11.4)	70 (14)
5	18 (14.6)	32 (26.0)	28 (22.8)	38 (30.9)	116 (24)
6	18 (14.6)	12 (9.8)	17 (13.8)	24 (19.5)	71 (14)
7	16 (13)	20 (16.3)	15 (12.2)	10 (8.1)	61 (12)
8	8 (6.5)	10 (8.1)	7 (5.7)	6 (4.9)	31 (6)
9	0	2 (1.6)	0	1 (0.8)	3 (1)
Total	123 (100)	123 (100)	123 (100)	123 (100)	492 (100)

Table 4: The percentage agreement in the grading of development stages of second molar (#37) between the observer 1 and other observers

Tooth number	Stage	Count (%)					
		Observer 1 versus observer 2		Observer 1 versus observer 3		Observer 1 versus observer 4	
#37	4	1	100.0	1	100.0	1	100.0
	5	9	82.0	9	81.8	11	100.0
	6	14	77.8	16	88.9	13	72.2
	7	9	42.9	9	42.9	9	42.9
	8	27	87.1	25	80.6	22	71.0
	9	33	80.5	30	73.2	31	75.6
Total agreement (%)		93	75.6	90	73.2	87	70.7

observers in the second molar revealed an insignificant poor agreement for Stage 3. Stages 4 and 5 showed an excellent significant agreement, whereas Stages 7 and 8 showed a moderate agreement. For the third molar, stage 4 showed fair agreement. Stage 0 showed a good agreement and the remaining stages showed moderate agreement among the observers [Table 10].

DISCUSSION

In the context of dental age estimation using the ordinal scores, it is very important to ensure that individual assessors can assign the tooth development stage reliably and consistently. The Demirjian's method of age estimation has been considered as a landmark in radiographic age assessment techniques globally, and gradually, researchers have come up with modifications with repeated use in different ancestries. However, this method leaves room for subjectivity when assigning the developmental stages. However, the pioneers of the method did judge the impact, the agreement in the results between observers could have on the accuracy of the estimated age. The inter-examiner variability in their study was <10% with a difference of only one stage. Furthermore, the examiners were well trained by one of them which gave the results better agreeability.[3,4] Thus, the extent of the agreement between observers could have a direct effect on the accuracy of the estimated age. The agreement is based on the closeness of the stages between two observers.[14] An extensive work by Levesque *et al.* in 1980 proved the accuracy rate of 80% in assigning the tooth development stages, and in the remaining 20%, the staged assessment between observers varied by plus or minus one stage.[10] The inter-observer agreement while assigning the ordinal scales, like the one in Demirjian's method is usually assessed using the Kappa statistics, which tests the closeness or equality of grades between the observers. Cohen's kappa coefficient compares the observed probability of disagreement to the probability of disagreement expected by chance. The kappa coefficients vary between -1 and 1. Perfect agreement ($k = 1$) is obtained when no disagreement is observed. A value of zero ($k = 0$) indicates that the probability of disagreement is only to be expected by chance while negative values express that the observed probability of disagreement is larger than what is expected by chance.[15] In this present study, the data of observer 1 was used as a standard for comparisons. Although there was a good agreement with the data of all the observers were compared together, the pair-wise comparisons revealed an excellent agreement ($k > 0.80$) between the reference data (observer 1) and observer 2, who had undergone hands-on training in

Table 5: The percentage agreement in the grading of development stages of third (#38) between the observer 1 and other observers

Tooth number	Stage	Count (%)		
		Observer 1 versus observer 2	Observer 1 versus observer 3	Observer 1 versus observer 4
#38	0	1 (100.0)	1 (100.0)	1 (100.0)
	1	8 (88.9)	9 (100.0)	3 (33.3)
	2	8 (50.0)	8 (50.0)	5 (31.3)
	3	7 (50.0)	5 (35.7)	7 (50.0)
	4	12 (54.3)	8 (34.8)	7 (30.4)
	5	18 (100.0)	17 (94.4)	17 (94.4)
	6	14 (77.8)	11 (61.1)	14 (77.8)
	7	11 (68.75)	9 (56.25)	7 (43.75)
	8	5 (62.5)	3 (37.5)	3 (37.5)
	9	0	0	0
Total agreement (%)		84 (68.29)	71 (57.7)	64 (52.03)

Table 6: The results of the weighted Kappa statistics between pairs of observers

Tooth number	Pairwise observers	Weighted Kappa	Asymptotic			95% asymptotic CI	
			SE	Z	Significant	Lower bound	Upper bound
#37	Observer 1 versus observer 2	0.81	0.03	13.30	0.000	0.74	0.88
	Observer 1 versus observer 3	0.78	0.04	12.83	0.000	0.71	0.86
	Observer 1 versus observer 4	0.79	0.03	12.79	0.000	0.72	0.86
	Observer 2 versus observer 3	0.83	0.04	13.48	0.000	0.76	0.90
	Observer 2 versus observer 4	0.80	0.04	13.03	0.000	0.73	0.87
	Observer 3 versus observer 4	0.78	0.04	12.56	0.000	0.70	0.85
#38	Observer 1 versus observer 2	0.85	0.02	15.33	0.000	0.80	0.89
	Observer 1 versus observer 3	0.78	0.03	14.17	0.000	0.72	0.84
	Observer 1 versus observer 4	0.70	0.04	13.20	0.000	0.63	0.78
	Observer 2 versus observer 3	0.81	0.02	14.93	0.000	0.76	0.86
	Observer 2 versus observer 4	0.73	0.04	13.76	0.000	0.66	0.81
	Observer 3 versus observer 4	0.69	0.04	13.08	0.000	0.61	0.77

CI: Confidence interval, SE: Standard error

grading and had a better experience than the remaining observers at the time of the study. The percentage

Table 7: The results of the Fleiss Kappa statistics among all the four observers

Overall agreement between all the 4 raters	Kappa	Asymptotic		Asymptotic 95% CI	
		SE	Z	Significant	Lower bound	Upper bound
#37	0.661	0.019	33.987	0.000	0.623	0.699
#38	0.511	0.014	35.279	0.000	0.482	0.539

CI: Confidence interval, SE: Standard error

Table 8: The results of the Freidman's nonparametric test comparing the developmental stages recorded by all the four observers (tooth-wise)

Tooth	n	χ²	df	Significant
#37	123	3.89	3	0.274
#38		25.02		0.000

Table 9: The pairwise comparisons of the grading of developmental status of molars between the reference observer 1 and other observers

Pair	Tooth	Ranks	n (%)	Z	Significant
Observer 2 - Observer 1	#37	Negative ranks	20 16.26)	−1.753	0.080
		Positive ranks	10 8.13)		
		Ties	93 75.61)		
	#38	Negative ranks	15 12.20)	−1.331	0.183
		Positive ranks	24 19.51)		
		Ties	84 68.29)		
Observer 3 - Observer 1	#37	Negative ranks	19 15.45)	−1.054	0.292
		Positive ranks	14 11.38)		
		Ties	90 73.17)		
	#38	Negative ranks	10 8.13)	−4.053	0.000
		Positive ranks	42 34.15)		
		Ties	71 57.72)		
Observer 4 - Observer 1	#37	Negative ranks	22 17.89)	−1.348	0.178
		Positive ranks	14 11.38)		
		Ties	87 70.73)		
	#38	Negative ranks	24 19.51)	−2.064	0.039
		Positive ranks	35 28.46)		
		Ties	64 52.03)		

agreement with the observer 1 was least for observer 4 who neither had theoretical nor practical training in the radiographic dental age estimation method. Reliability study in grading the stages of both the upper and lower third molars was earlier reported.[16] Their study compared the inter and intra-observer agreement levels in four different age estimation methods. Their report shows that Demirjian's method had a good agreement when compared to other methods and the mandibular third molars showed the highest agreement between observers. The present study assigned the stages of only the mandibular second and third molars. These two molars are considered as valuable age indicators in adolescents and young adults.[17] The overall interobserver agreement was higher for the second molar. Inter-observer variability of one stage was reported in Norwegian children.[18] In their study, the two observers differ by one earlier stage in 10.5% of the scores and by one later stage in 11.6% of the cases. In the present study, it was observed that while grading the stage of third molars, observers 2, 3, and 4 selected an advanced stage (positive ranks) in 19.5%–43.09% of cases when compared to the observer 1. The study also points out that the agreement or near-perfect identification of the correct stage could be possible with additional hands-on training in identifying the correct stage using the standard reference. Between observer 1 and observer 2, it was observed that the observer 2 gave an advanced grading in the second molar, whereas in the third molar, he gave a lesser grading when compared with observer 1. In another comparative study comparing the results of application of Demirjian's, and Nolla's methods, the overall inter-observer agreement was 0.98 for both the methods.[19] The present study compared only the Demirjian's staging in second and third molars and found the overall inter-observer agreement 0.66

Table 10: The overall Fleiss Kappa agreement on individual stage categories in both the teeth

Tooth number	Developmental stage	Kappa	Asymptotic			Asymptotic 95% CI	
			SE	Z	Significant	Lower bound	Upper bound
#37	3	−0.004	0.037	−0.111	0.912	−0.076	0.068
	4	0.855	0.037	23.229	0.000	0.783	0.927
	5	0.811	0.037	22.035	0.000	0.739	0.883
	6	0.751	0.037	20.402	0.000	0.679	0.823
	7	0.502	0.037	13.635	0.000	0.430	0.574
	8	0.572	0.037	15.526	0.000	0.499	0.644
	9	0.708	0.037	19.233	0.000	0.636	0.780
#38	0	0.662	0.037	17.980	0.000	0.590	0.734
	1	0.601	0.037	16.316	0.000	0.528	0.673
	2	0.531	0.037	14.437	0.000	0.459	0.604
	3	0.487	0.037	13.227	0.000	0.415	0.559
	4	0.356	0.037	9.671	0.000	0.284	0.428
	5	0.601	0.037	16.339	0.000	0.529	0.674
	6	0.567	0.037	15.391	0.000	0.494	0.639
	7	0.445	0.037	12.084	0.000	0.373	0.517
	8	0.461	0.037	12.514	0.000	0.388	0.533
	9	0.217	0.037	5.907	0.000	0.145	0.290

CI: Confidence interval, SE: Standard error

and 0.51 for second and third molars, respectively. The intra-observer agreement in grading the development stages in molars ranged from 0.51 to 0.876 for all the four observers. Studies have shown an intra-observer agreement from 0.79 to 0.94.[20,21] The interobserver variability in staging the development stages of teeth using the intraclass coefficient revealed a Cronbach's alpha value of 0.759 and 0.259 for second and third molars, respectively.[22] In the present study too, the interobserver agreement in third molar was comparatively lesser when compared to the second molar. Commonly in various studies, transitory stages of the Demirjian and Chaillet's methods like stage D and E have been found to have maximum observer agreement.[17] This fact has been consistent with the present study. Surprisingly, when there has been a high rate of discrepancies in scoring the beginning or end stages: Stage A or Stage G or H, there has been inter-observer reliability at Stage 8 in the present study. Furthermore, if the traits of various stages of the method used are more defined and dichotomized, observers find it helpful to consistently give better and accurate scores. The scientific community recommends the use of different methods and repetitive observations to enhance the reproducibility among examiners. The use of more reference radiographs together with assisted magnification and digital images have proven to provide fairly higher precision due to the calibration of the readings.[23] The third molars exhibit marked differences in terms of formation, eruption morphology, and agenesis compared to the second molar. That could be attributed to the significant difference in scoring and a lesser agreement by all the observers in this study. Conversely, it forms the basis of many other certified methods. This again necessitates the use of more projections for calibration and a better outcome. Therefore, in the case of third molars, a scoring method should comprise the entirety of the maturation sequence with clear-cut demarcations. A sound statistical approach in conjunction with a practical one that not only renders appropriate estimation but also evades misassumptions is, hence imperative for age evaluation practices.[24] More studies such as the present one with interdisciplinary observer's error assessment shall be considered welcome to improve the approach towards age estimation.[25] A properly applied statistical approach and the experience of the forensic odontologist are essential for the success of any age estimation method.[26] The more the stages in a scoring method, the better is its precision, owing to well-defined objective measures. However, after a point, it decreases the reproducibility as a result of increased confusion among observers. Thus, a balance between detailed and sufficient staging along with practical feasibility with fair consistency and precision is always desirable for best results. This makes any method straight forward enough to eliminate stage overlaps and facilitate better learning among forensic odontologists.[24] The good inter-observer agreement is also essential in other age estimation methods where radiographs are not applicable.[26] The budding forensic odontologists may undergo exclusive training on the radiographic grading of the tooth development for age estimation from experienced forensic odontologist. From legal perspective, the age estimation methods need to the assessment that are accurate and reproducible with minimal or nil interobserver variability. This is possible by the correct and uniform identification of the development stages of teeth by the forensic odontologists irrespective of their practical experience in age estimation.

CONCLUSIONS

Based on the above study, it can be concluded that:
- The inter-observer agreement in staging the developmental stages of second and third molars was significantly different among observers for the third molar and insignificantly different for second molars
- There was moderate to a substantial agreement between the reference data and the data from the other three observers for the second molar and the third molars, there was fair to a substantial agreement between the reference data and the data from the other three observers
- Proper hand-on training in radiographic interpretation of developmental stages may improve the chance of minimizing the errors during the age estimation methods.

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Nolla CM. The development of permanent teeth. J Dent Child 1960;27:254-66.
2. Moorrees CF, Fanning EA, Hunt EE Jr. Age variation of formation stages for ten permanent teeth. J Dent Res 1963;42:1490-502.

3. Demirjian A, Goldstein H, Tanner JM. A new system of dental age assessment. Hum Biol 1973;45:211-27.

4. Chaillet N, Demirjian A. Dental maturity in South France: A comparison between Demirjian's method and polynomial functions. J Forensic Sci 2004;49:1059-66.

5. Mincer HH, Harris EF, Berryman HE. The ABFO study of third molar development and its use as an estimator of chronological age. J Forensic Sci 1993;38:379-90.

6. Willems G, Van Olmen A, Spiessens B, Carels C. Dental age estimation in Belgian children: Demirjian's technique revisited. J Forensic Sci 2001;46:893-5.

7. Acharya AB. Age estimation in Indians using Demirjian's 8-teeth method. J Forensic Sci 2011;56:124-7.

8. Moness Ali AM, Ahmed WH, Khattab NM. Applicability of Demirjian's method for dental age estimation in a group of Egyptian children. BDJ Open 2019;5:2.

9. Yassin SM. Accuracy of Demirjian's four methods of dental age estimation in a sample of Saudi Arabian population. Aust J Forensic Sci 2020;27:1-4.

10. Levesque GY, Demirjian A. The inter-examiner variation in rating dental formation from radiographs. J Dent Res 1980;59:1123-6.

11. Pillai JP, Chokkalingam TS, Aasaithambi B, Nuzzolese E. Establishment of the forensic odontology department: A proposed model for the basic infrastructure and forensic odontology kit. J Forensic Dent Sci 2019;11:64-72.

12. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, *et al.* Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud 2011;48:661-71.

13. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. John Wiley & Sons, Hoboken. 2013.

14. Stine W. Interobserver relational agreement. Psychol Bull 1989;106:341-7.

15. Vanbelle S. A new interpretation of the weighted kappa coefficients. Psychometrika 2016;81:399-410.

16. Dhanjal KS, Bhardwaj MK, Liversidge HM. Reproducibility of radiographic stage assessment of third molars. Forensic Sci Int 2006;159 Suppl 1:S74-7.

17. Lee SS, Byun YS, Park MJ, Choi JH, Yoon CL, Shin KJ. The chronology of second and third molar development in Koreans and its application to forensic age estimation. Int J Legal Med 2010;124:659-65.

18. Nyka¨nen R, Espeland L, Kvaal SI, Krogstad O. Validity of the Demirjian method for dental age estimation when applied to Norwegian children. Acta Odont Scand 1998;56:238-44.

19. Paz Cortés MM, Rojo R, Alía García E, Mourelle Martínez MR. Accuracy assessment of dental age estimation with the Willems, Demirjian and Nolla methods in Spanish children: Comparative cross-sectional study. BMC Pediatr 2020;20:361.

20. Günen Yılmaz S, Harorlı A, Kılıç M, Bayrakdar İŞ. Evaluation of the relationship between the Demirjian and Nolla methods and the pubertal growth spurt stage predicted by skeletal maturation indicators in Turkish children aged 10-15: investigation study. Acta Odontol Scand 2019;77:107-13.

21. Tomás LF, Mónico LS, Tomás I, Varela-Patiño P, Martin-Biedma B. The accuracy of estimating chronological age from Demirjian and Nolla methods in a Portuguese and Spanish sample. BMC Oral Health 2014;14:160.

22. Stella A, Jeevarathan J, Ennamuel DS, Selvi T. Analyzing the Interobserver Variability in Stages of Tooth Development with Orthopantomogram (OPG). Int J Recent Technol Eng 2019;7:256-9.

23. Willems G, Moulin-Romsee C, Solheim T. Non-destructive dental-age calculation methods in adults: Intra-and inter-observer effects. Forensic Sci Int 2002;126:221-6.

24. De Tobel J, Phlypo I, Fieuws S, Politis C, Verstraete KL, Thevissen PW. Forensic age estimation based on development of third molars: A staging technique for magnetic resonance imaging. J Forensic Odontostomatol 2017;35:117-40.

25. Lynnerup N, Belard E, Buch-Olsen K, Sejrsen B, Damgaard-Pedersen K. Intra-and interobserver error of the Greulich-Pyle method as used on a Danish forensic sample. Forensic Sci Int 2008;179:242e1-6.

26. Borrman H, Solheim T, Magnusson B, Kvaal SI, Stene-Johansen W. Inter-examiner variation in the assessment of age-related factors in teeth. Int J Legal Med 1995;107:183-6.