

Andreas Olze · Dominique Bilanz · Sven Schmidt ·
Klaus-Dieter Wernecke · Gunther Geserick ·
Andreas Schmeling

Validation of common classification systems for assessing the mineralization of third molars

Received: 11 May 2004 / Accepted: 9 September 2004 / Published online: 6 November 2004
© Springer-Verlag 2004

Abstract One major criterion for dental age estimation is the evaluation of third molar mineralization. There are various methods for evaluating tooth mineralization based on classification by stages. The aim of the present work is to assess the validity of the common classification systems. To this end, we analyzed 420 conventional orthopantomograms of German females aged 12–25 years old. The mineralization status of tooth 38 was determined using the stages defined by Gleiser and Hunt, Demirjian et al., Gustafson and Koch, Harris and Nortje and Kullman et al., respectively. Of the methods tested, the most accurate results were obtained with Demirjian et al.'s classification system, which performed best not only for observer agreement but also for the correlation between estimated and true age. It is argued that this is due to the fact that Demirjian et al.'s classification is based on a sufficient number of stages which are defined independently of speculative estimations of length. This leads to the conclusion that the method devised by Demirjian et al. should be used for evaluating the mineralization of third molars for purposes of forensic age determination.

Keywords Forensic age estimation · Dental age · Mineralization · Third molars

Introduction

Today forensic age diagnostics is an established research sector of legal medicine in its own right (Ohtani et al. 2003; Olze et al. 2004; Ritz-Timme et al. 2003; Schmeling et al. 2004; Takasaki et al. 2003). In recent years it has become increasingly important to determine, in particular, the age of living persons (Schmeling et al. 2001b). From a legal perspective, such age estimates are carried out to determine whether a suspect without valid identification documents has reached the age of criminal responsibility and whether general criminal law in force for adults is to be applied. In many countries the age thresholds of relevance to criminal prosecution lie between 14 and 18 years (Dünkel et al. 1997).

In line with recommendations drawn up by the international interdisciplinary Study Group on Forensic Age Diagnostics (<http://www.charite.de/rechtsmedizin/agfad/index.htm>), a forensic age diagnosis for the purpose of criminal investigations should consist of a clinical examination, including the recording of body measurements and an evaluation of signs of sexual maturity, an X-ray examination of the left hand, and a dental examination which records dentition status and evaluates an orthopantomogram (Schmeling et al. 2001a). One major criterion for dental age estimation is the evaluation of third molar mineralization.

Various classifications have been devised for evaluating tooth mineralization (Gleiser and Hunt 1955; Nolla 1960; Haavikko 1970; Liliequist and Lundberg 1971; Demirjian et al. 1973; Gustafson and Koch 1974; Nortje 1983; Harris and Nortje 1984; Kullman et al. 1992; Köhler et al. 1994). Since the validity of an age estimate crucially depends on the classification method used, the most appropriate one should be selected. The present work examines the validity of five different stage-based systems.

A. Olze · D. Bilanz · S. Schmidt · G. Geserick ·
A. Schmeling (✉)
Institute of Legal Medicine (Charité), Humboldt University of
Berlin,
Hannoversche Strasse 6,
10115 Berlin, Germany
e-mail: andreas.schmeling@charite.de

K.-D. Wernecke
Institute of Medical Biometry (Charité), Humboldt University
of Berlin DRK-Kliniken Westend,
Haus 31,
13344 Berlin, Germany

Materials and methods

We analyzed 420 conventional orthopantomograms of German females aged 12–25 years old and each age group was represented by X-rays from at least 30 females. This study could be limited to female probands, as there are no statistically significant gender-related differences to the chronology of wisdom tooth mineralization (Olze et al. 2003). The mineralization status of the left mandibular third molar (tooth 38) was determined for all persons included in the study using five different methods, namely the classifications by Gleiser and Hunt (1955), Demirjian et al. (1973), Gustafson and Koch (1974), Harris and Nortje (1984) and Kullman et al. (1992).

The methods of Gleiser and Hunt (1955), Demirjian et al. (1973), Harris and Nortje (1984) and Kullman et al. (1992) are described in both text and in diagrammatic form (Figs. 1, 2, 3, and 4).

Gustafson and Koch (1974) only provided a verbal description and defined the following 4-stage classification:

- Stage 1: commencement of mineralization
- Stage 2: completion of crown
- Stage 3: eruption when the cusp(s) penetrate the gingiva
- Stage 4: completion of root(s).

For reasons of space we have omitted the verbal descriptions of the remaining classifications and kindly ask readers to refer to the original works.

Each of the five methods included in the study was evaluated by two independent observers (A, and B), with observer A examining all X-ray images twice (A1 and A2) where 6 months elapsed between evaluations A1 and A2.

A 2-factor analysis of variance was used to test the differences between these individual evaluations and methods; in order to account for the fact that observer values $X_{m,i,j}(k)$ were measured on a category scale, it was carried out non-parametrically (Brunner et al. 2002). Significance was assessed at $p < 0.05$.

Fig. 1 Diagrammatic representation of the stage classification by Gleiser and Hunt (1955)

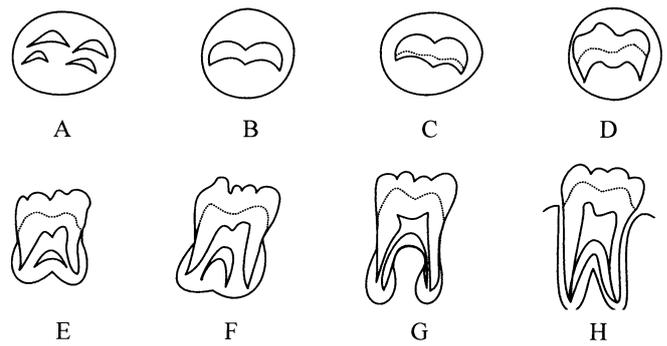
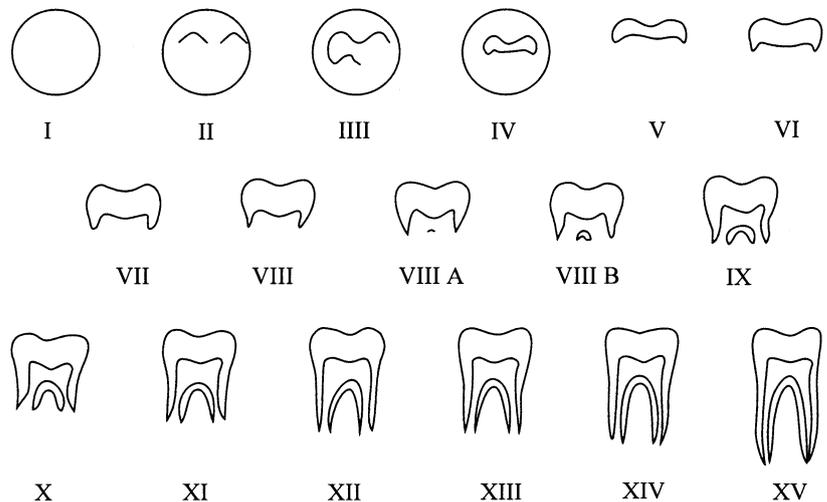


Fig. 2 Diagrammatic representation of the stage classification by Demirjian et al. (1973)

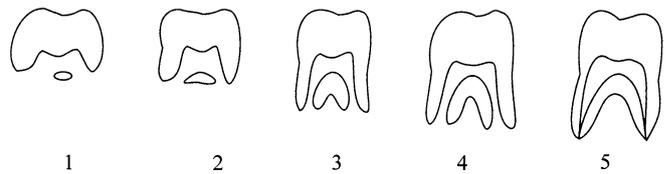


Fig. 3 Diagrammatic representation of the stage classification by Harris and Nortje (1984)

Both inter-observer and intra-observer agreement were determined using the weighted kappa coefficient (Fleiss 1981), taking two (scaled) measurements at a time (intra-class correlation) and the 95% confidence intervals were calculated for each kappa value. The method that yields the best inter-observer and intra-observer agreement, i.e. where assessment variability is low, is assumed to be the most appropriate method.

Correlation between true age and estimated age based on the results obtained from the various methods was evaluated by directly comparing the stage determined on a category scale with the age determined on an interval scale using the eta coefficient. Eta is a coefficient of categorical-by-interval association (association refers to measures of strength of relationship in which at least one of the variables is dichotomous, nominal, or ordinal). For linear relationships, eta equals the Pearson's correlation coefficient.

Fig. 4 Diagrammatic representation of the stage classification by Kullman et al. (1992)



cient (Siegel 1956). The method providing maximum correlation was considered to be the most appropriate.

Results

Table 1 shows the *p*-values of the analysis of variance for a comparison between all three observations across all five methods (test 1), a comparison of observer A1 versus observer B for all methods (test 2), a comparison of observer A2 versus observer B for all methods (test 3) and a comparison of observer A1 versus observer A2 for all five methods (test 4). No differences were observed between observers, but significant differences were observed between methods, and there was a significant interaction between methods and observers. In this context, significant interaction means that a difference between methods is not independent of the observers, i.e., it is particularly characteristic of certain observers.

For a quantification of these systematic influences, an intra-class coefficient (ICC) was determined together with 95% confidence intervals (Table 2).

The highest intra-class coefficient was displayed by Demirjian et al.'s method for evaluation A2-B. An intra-class coefficient of >0.95 was obtained for the Gleiser and Hunt method and for Kullman et al. (A2-B). The methods developed by Gustafson and Koch (A1-B, A1-A2) and by Harris and Nortje (A1-B) yielded intra-class coefficients of <0.90 .

In order to assess correlation between age estimated on the basis of the stages defined in the various methods and true age, eta coefficients were determined for all methods and all observers (Table 3).

The method of Demirjian et al. shows the highest eta coefficient (eta coefficient for A1=0.883), followed by the methods of Kullman et al. (eta coefficient for A1=0.880) and of Gleiser and Hunt (eta coefficient for A1=0.879). An eta coefficient of <0.8 was calculated for the method developed by Gustafson and Koch.

Hence, the method developed by Demirjian et al. can be considered the best of the five methods reviewed here, producing the highest values for both observer agreement and correlation between true age and age estimated on the basis of the stages defined. Good results were also achieved with the methods developed by Gleiser and

Hunt and by Kullman et al. The methods developed by Gustafson and Koch and by Harris and Nortje performed noticeably less well.

Discussion

Various classifications are available for evaluating tooth mineralization (Gleiser and Hunt 1955; Nolla 1960; Haavikko 1970; Liliequist and Lundberg 1971; Demirjian et al. 1973; Gustafson and Koch 1974; Nortje 1983; Harris and Nortje 1984; Kullman et al. 1992; Köhler et al. 1994). They differ with regard to the number of stages, the definition of each stage and the presentation.

Gustafson and Koch, like Harris and Nortje, defined 4–5 stages, Kullman et al. offered 7 stages, and Demirjian et al. as well as Nortje established 8 stages. The remaining classifications use 10–16 stages. Apart from Gustafson and Koch, where the method was only described verbally, all the authors described their stages in both text and diagram form. Apart from Demirjian et al., all the authors used stages based on fractions of the future length of crown or root. Demirjian et al.'s stages are defined by changes in form and do not depend on speculative estimates of length.

As some of these methods are very similar, our review was confined to five basic types of classification. We selected the methods by Gleiser and Hunt (1955), Demirjian et al. (1973), Gustafson and Koch (1974), Harris and Nortje (1984) and Kullman et al. (1992).

A number of authors have already compared the validity of different stage-based methods. Hägg and Matsson (1985) examined 300 Swedish children aged 3.5–12.5 years old using the multi-stage classifications developed by Liliequist and Lundberg (1971), Demirjian et al. (1973) and Gustafson and Koch (1974).

Accuracy tests were performed to establish mean deviations and the increase in linear correlation between estimated and real age. A high rate of agreement was obtained from Demirjian et al.'s method with children aged 3.5–6.5 years old, but for older children Demirjian et al.'s method displayed less accuracy. When applying the method by Liliequist and Lundberg (1971), accuracy was low across all age groups with a general tendency to

Table 1 *P*-values of the analysis of variance for comparison between observers and methods

Analysis of variance	Test 1	Test 2	Test 3	Test 4
Systematic differences between observers	0.73958	0.63824	0.99388	0.62295
Systematic differences between methods	0.00000	0.00000	0.00000	0.00000
Interaction between observers and methods	0.00000	0.00012	0.01904	0.00000

Table 2 Intra-class coefficients (ICC) with 95% confidence intervals (CI)

Method	Observer	ICC	95% CI
Demirjian et al. 1973 (interind.)	A1-B	0.953	0.9440–0.9612
Demirjian et al. 1973 (interind.)	A2-B	0.989	0.9846–0.9926
Demirjian et al. 1973 (intra-ind.)	A1-A2	0.957	0.9488–0.9659
Gustafson and Koch 1974 (interind.)	A1-B	0.873	0.8380–0.9088
Gustafson and Koch 1974 (interind.)	A2-B	0.979	0.9662–0.9908
Gustafson and Koch 1974 (intra-ind.)	A1-A2	0.885	0.8505–0.9191
Gleiser and Hunt 1955 (interind.)	A1-B	0.954	0.9314–0.9772
Gleiser and Hunt 1955 (interind.)	A2-B	0.979	0.9569–0.9991
Gleiser and Hunt 1955 (intra-ind.)	A1-A2	0.975	0.9696–0.9810
Kullman et al. 1992 (interind.)	A1-B	0.923	0.9098–0.9426
Kullman et al. 1992 (interind.)	A2-B	0.979	0.9725–0.9862
Kullman et al. 1992 (intra-ind.)	A1-A2	0.941	0.8994–0.9593
Harris and Nortje 1984 (interind.)	A1-B	0.833	0.7831–0.8827
Harris and Nortje 1984 (interind.)	A2-B	0.931	0.9111–0.9505
Harris and Nortje 1984 (intra-ind.)	A1-A2	0.902	0.8727–0.9307

Table 3 Eta coefficients for all methods and all observers

Method	Observer A1	Observer A2	Observer B
Demirjian et al. 1973	0.883	0.865	0.856
Gustafson and Koch 1974	0.745	0.770	0.773
Gleiser and Hunt 1955	0.879	0.871	0.855
Kullman et al. 1992	0.880	0.859	0.846
Harris and Nortje 1984	0.863	0.846	0.811

underestimate. For the method described by Gustafson and Koch (1974), accuracy was only high for male subjects.

Gustafson and Koch's method (1974) displayed greater inter-observer and intra-observer error than the methods devised by Demirjian et al. (1973) and Liliequist and Lundberg (1971).

Pöyry et al. (1986) investigated the mineralization of teeth 31–38 using 96 orthopantomograms from 48 Finnish boys aged 3–14 years old. They compared Haavikko's (1970) modified version of the Gleiser and Hunt classification (1955) (method A) with the classification defined by Demirjian et al. (1973) (method B). The estimates were performed by two observers assessing each X-ray twice. For method A the mean intra-observer error was 20.1% and for method B, 10.2%. However, the inter-observer error was 31.0% for method A and 19.2% for method B. Stages which proved difficult to assess were Cco (coalescence of cups), Cr1/2 (crown 1/2 complete) and Crc (crown complete) in method A and stage G in method B.

Staaf et al. (1991) scrutinized the validity of age estimation methods based on the classifications set out by Demirjian et al. (1973), Haavikko (1970) and Liliequist and Lundberg (1971). They examined the orthopantomograms of 541 Swedish children of both genders aged 5.5–15.5 years old and 37 randomly selected cases were evaluated a second time. When using the method developed by Demirjian et al. (1973), the age of both boys and girls was overestimated by 6–10 months. The methods drawn up by Haavikko (1970) and by Liliequist and Lundberg (1971) resulted in a systematic underestimation

or overestimation by 6–7 and 7 months, respectively. Intra-observer error was low for all three methods and not significant. In discussing why the method by Demirjian et al. (1973) should prove less valid, the authors proposed possible ethnic differences between the original reference population and the sample with which they were working.

A study by Mörnstad et al. (1995) considered 197 orthopantomograms from Swedish children aged 5, 6, 9 and 12 years old. A total of 13 independent observers assessed the mineralization status of the 7 mandibular teeth excluding the third molar using the classifications by Demirjian et al. (1973), Gustafson and Koch (1974), Liliequist and Lundberg (1971) and Haavikko (1970). As Hägg and Matsson (1985) and Staaf et al. (1991) had already reported overestimated age in Scandinavian subjects following the use of Demirjian et al.'s stages, the authors incorporated reference values from a study by Kataja et al. (1989), where the criteria defined by Demirjian et al. (1973) had been applied to a Finnish population. The highest accuracy compared with real age was obtained from the method by Demirjian et al. (1973), although applying the reference values gained from a Canadian sample to a Swedish population resulted in an overestimation across both genders and all age groups of 0.4–1.8 years. The overestimation was much smaller (0.1–0.8 years) when using the Scandinavian reference values reported by Kataja et al. (1989).

Reventlid et al. (1996) published intra-observer and interobserver error rates from the random sample investigated by Mörnstad et al. (1995). All 197 orthopantomograms were assessed by 13 independent observers with experience in estimating children's ages: 28 orthopantomograms were analyzed again by 12 observers after an interval of about 2 years. Mean intra-observer error was 0.03–0.20 years. Statistically significant differences emerged in the 6–9 year age group when using the method by Haavikko (1970) and in the 12 year age group when using the classifications by Gustafson and Koch (1974) and Demirjian et al. (1973). No systematic overestimation or underestimation was displayed by one

or more observers. For interobserver error there were significant differences between estimated age across all observers, age groups and classification systems. The scatter for mean estimated age was 0.8–1.2 years for all observers and methods. The range for individual estimated ages varied from 3.2 to 8 years, being particularly low using the method by Haavikko (1970) and particularly high using the method by Gustafson and Koch (1974).

Existing studies on the validity of different classification systems for assessing dental mineralization are of limited value in that the stages defined for each method were based either on the reference population used for the original work or on some other population dissimilar to the sample selected for the study. It is impossible, therefore, to dismiss the risk that differences in sample size, age group, age distribution across the group, ethnic origin or the health status of the subjects might have exerted a substantial influence on the comparative studies published to date. The work by Mörnstad et al. (1995) revealed, for example, that applying Demirjian et al.'s stages to different reference populations will result in different deviations between estimated and real age.

The procedure adopted by the present study resolves this methodological problem, to the authors' knowledge for the first time, by using weighted kappa coefficients (Fleiss 1981) and eta coefficients (Siegel 1956) to analyze the various stages independently of reference populations.

Of the various methods examined, Demirjian et al.'s classification achieved the highest values for both observer agreement and for correlation between the stages as defined by the method and true age. It can, therefore, be regarded as the best method. Good results were also obtained by the classifications according to Gleiser and Hunt and Kullman et al. Noticeably poorer results were derived from the methods by Gustafson and Koch and Harris and Nortje. The small number of stages evidently exerts a negative impact here, because a large age interval between stages means that guessing a wrong stage will result in a correspondingly greater error in estimated age. One particular advantage of the classification by Demirjian et al. is that stages are not defined on the basis of speculative estimates of length.

The authors conclude that Demirjian's stages should be used to evaluate third molar mineralization for forensic age estimates.

Acknowledgments This study was supported by a grant from the Deutsche Forschungsgemeinschaft (GE 968/3–1).

References

- Brunner E, Domhof S, Langer F (2002) Nonparametric analysis of longitudinal data in factorial experiments. John Wiley & Sons, New York
- Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. *Hum Biol* 45:221–227
- Dünkel F, Kalmthout A van, Schüler-Springorum H (1997) Entwicklungstendenzen und Reformstrategien im Jugendstrafrecht im europäischen Vergleich. Forum, Mönchengladbach
- Fleiss JL (1981) Statistical methods for rates and proportions. Wiley, New York
- Gleiser I, Hunt EE (1955) The permanent mandibular first molar; its calcification, eruption and decay. *Am J Phys Anthropol* 13:253–284
- Gustafson G, Koch G (1974) Age estimation up to 16 years of age based on dental development. *Odontol Rev* 25:297–306
- Haavikko K (1970) The formation and the alveolar and clinical eruption of the permanent teeth. *Suom Hammaslaak Toim* 66:103–170
- Hägg U, Matsson L (1985) Dental maturity as an indicator of chronological age: the accuracy and precision of three methods. *Eur J Orthod* 7:25–34
- Harris MJP, Nortje CJ (1984) The mesial root of the third mandibular molar. A possible indicator of age. *J Forensic Odontostomatol* 2:39–43
- Kataja M, Nyström M, Aine L (1989) Dental maturity standards in southern Finland. *Proc Finn Dent Soc* 85:187–197
- Köhler S, Schmelzle R, Loitz C, Püschel K (1994) Die Entwicklung des Weisheitszahnes als Kriterium der Lebensalterbestimmung. *Ann Anat* 176:339–345
- Kullman L, Johanson G, Akesson L (1992) Root development of the lower third molar and its relation to chronological age. *Swed Dent J* 16:161–167
- Liliequist B, Lundberg M (1971) Skeletal and tooth development: a methodologic investigation. *Acta Radiol (Diagn)* 11:97–112
- Mörnstad H, Reventlid M, Teivens A (1995) The validity of four methods for age determination by teeth in Swedish children. *Swed Dent J* 19:121–130
- Nolla CM (1960) The development of the permanent teeth. *J Dent Child* 27:254–266
- Nortje CJ (1983) The permanent mandibular third molar. *J Forensic Odontostomatol* 1:27–31
- Ohtani S, Ito R, Yamamoto T (2003) Differences in the D/L aspartic acid ratios in dentin among different types of teeth from the same individual and estimated age. *Int J Legal Med* 117:149–152
- Olze A, Schmelting A, Rieger K, Kalb G, Geserick G (2003) Untersuchungen zum zeitlichen Verlauf der Weisheitszahnmineralisation bei einer deutschen Population. *Rechtsmedizin* 13:5–10
- Olze A, Schmelting A, Taniguchi M, Maeda H, Niekerk P Van, Wernecke K-D, Geserick G (2004) Forensic age estimation in living subjects: the ethnic factor in wisdom tooth mineralization. *Int J Legal Med* 118:170–173
- Pöyry M, Nyström M, Ranta R (1986) Comparison of two tooth formation rating methods. *Proc Finn Dent Soc* 82:127–133
- Reventlid M, Mörnstad H, Teivens AA (1996) Intra- and inter-examiner variation in four dental methods for age estimation of children. *Swed Dent J* 20:133–139
- Ritz-Timme S, Laumeier I, Collins M (2003) Age estimation based on aspartic acid racemization in elastin from the yellow ligaments. *Int J Legal Med* 117:96–101
- Schmelting A, Kaatsch H-J, Marré B et al. (2001a) Empfehlungen für die Altersdiagnostik bei Lebenden im Strafverfahren. *Rechtsmedizin* 11:1–3
- Schmelting A, Olze A, Reisinger W, Geserick G (2001b) Age estimation of living people undergoing criminal proceedings. *Lancet* 358:89–90
- Schmelting A, Schulz R, Reisinger W, Mühler M, Wernecke K-D, Geserick G (2004) Studies on the time frame for ossification of medial clavicular epiphyseal cartilage in conventional radiography. *Int J Legal Med* 118:5–8
- Siegel S (1956) Nonparametric statistics for the behaviour sciences. McGraw-Hill, New York
- Staa V, Mörnstad H, Welander U (1991) Age estimation based on tooth development: a test of reliability and validity. *Scand J Dent Res* 99:281–286
- Takasaki T, Tsuji A, Ikeda N, Ohishi M (2003) Age estimation in dental pulp DNA based on human telomere shortening. *Int J Legal Med* 117:232–234