Int J Leg Med (1995) 107:183–186

ORIGINAL ARTICLE

Helene Borrman · Tore Solheim · Bengt Magnusson
Sigrid I. Kvaal · Wencke Stene-Johansen

Inter-examiner variation in the assessment of age-related factors in teeth

Received: 20 July 1994 / Received in revised form: 21 September 1994

Abstract Inter-observer variations in the registration of dental age-related characteristics have not previously been studied. Examination and registration were made by 6 dentists with varying experience in age estimation. A total of 13 age-related dental characteristics in 30 teeth extracted from adults were assessed macroscopically, by stereomicroscope and from radiographs. The measurements were analysed in a microcomputer, using the SPSS/PC+ statistical package. The results showed that, except for the score on surface roughness, significant differences were found between some of the observers for all types of measurements, when a paired *t*-test analysis was made. The correlation coefficients between the observers varied and were rather weak for the surface roughness score. The study revealed systematic differences among the observers as well as differences in interpreting the definitions of the scores for the different parameters. Thus age estimation using statistical methods is seen to be dependent upon the experience of the individual observer and interpretation. Care should therefore be taken not to rely too much upon the results of an odontological age estimation. The possible implications of these results for forensic work are also discussed.

Key words Forensic odontology · Age determination · Human teeth · Radiography · Microscopy

Zusammenfassung Subjektiv bedingte Bewertungsunterschiede zwischen verschiedenen Untersuchern bei altersabhängigen Zahnveränderungen sind bisher noch nicht untersucht worden. Die Untersuchungen und Aufzeichnungen wurden von sechs verschiedenen Zahnärzten mit unterschiedlicher Erfahrung in der Altersbestimmung durchgeführt. Dreizehn altersabhängige Kriterien wurden an 30 extrahierten Zähnen von Erwachsenen makroskopisch, stereomikroskopisch und radiographisch untersucht. Die Meßwerte wurden mit einem Mikrocomputer unter Anwendung des Statistikprogramms SPSS/PC+ analysiert. Die Ergebnisse zeigen, daß außer für den Wert der Oberflächenrauheit, signifikante Unterschiede zwischen den Meßwerten der einzelnen Untersucher, ermittelt mit Hilfe des gekoppelten *t*-Tests, auftraten. Die Korrelationskoeffizienten zwischen den Ergebnissen der einzelnen Untersucher variierten und waren beim Wert der Oberflächenrauheit am geringsten. Die Studie zeigte Differenzen in der Systematik der verschiedenen Untersucher und in der Interpretation der Definitionen altersabhäginger Parameter. Somit ist die Altersbestimmung mit Hilfe statistischer Methoden stets von der Erfahrung und Interpretation einzelner Untersucher abhängig. Daher sollte der Wert einer odontologischen Altersbestimmung nicht überschätzt werden. Konsequenzen dieser Studie für die forensische Praxis werden diskutiert.

Schlüsselwörter Forensische Odontologie · Altersbestimmung · Zähne · Radiographie · Microskopie

Introduction

In forensic odontology, remaining teeth may be used to establish identity. This is especially the case when only skeletonized remains of a deceased person are found. Several studies have demonstrated that dental characteristics are well suited for this purpose [1–3]. Regressive and/or pathological changes, such as periodontal disease, cementum apposition, root resorption and/or secondary dentine formation, may be basic elements for odontological identification. Furthermore, physiological changes like attrition and increasing root translucency can help in estimating the age of a person at the time of death as an indicator of identity. The ability of these parameters to aid in age estimation of a deceased person have mainly been analysed on sectioned teeth. However, unsectioned teeth have also been radiographically and macroscopically studied.

H. Borrman (✉) · B. Magnusson
Faculty of Odontology, Göteborg University, Medicinargatan 12, S-413 90 Göteborg, Sweden

T. Solheim · S. I. Kvaal · W. Stene-Johansen
Department of Oral Pathology, University of Oslo, Oslo, Norway

Techniques for estimating the age of a person mostly provide semi-quantitative assessments, but are still methods of choice [2–5]. Their semi-quantitative character implies that differences in scores by various observers may occur. So far, no systematic studies have been performed for analyzing the variation between different investigators in registrations of age-related changes. Thus the aim was to perform such a study, in order to evaluate the validity of estimating the age of individuals by means of dental characteristics.

Material and methods

Thirty teeth extracted from individuals of known age were examined, the mean being 54 years (range 10–90 years). The teeth were maxillary and mandibular canines and premolars from the tooth collection in the Department of Oral Pathology, Oslo. All the teeth had been extracted from a Caucasian population in the State of Washington, USA, and in Norway. The teeth had been stored in a 4% formaldehyde solution.

Examiners

The teeth were examined by 6 dentists with varying experience (1–20 years). Three were oral pathologists, one an oral radiologist, and 2 dental practitioners. All the examiners had previously been involved in forensic work and had experience assessing an individual's age from teeth.

Microscopical and visual examination

The attrition was analyzed according to a scoring system [2]. The size of the attrited facets was measured using a stereomicroscope with a grid with an accuracy of 1.0 mm^2 [6]. Using a pair of vernier callipers with an accuracy of 0.1 mm, the periodontal retraction was measured from the cemento-enamel junction to the uppermost coronal remaining periodontal fibres on 4 root surfaces, and the mean value was calculated. The length of apical translucent dentine was measured with a pair of vernier callipers with an accuracy of 0.1 mm, as described by Bang and Ramm [4]. The root surface was studied under a stereomicroscope and scores of the surface roughness were determined after comparison with stereophotographs of the typical scores 1–4 [7]. The colour of the crown was estimated in terms of a Biodent colour guide and the colour of the root dentine according to a scoring system [8].

Radiographic examination

The teeth were radiographed in a vestibulo-lingual projection at 60 kV and with 20 cm FSD (Focus Skin Distance) the teeth being placed in close contact with the film. The exposure time was 0.3 seconds and the film used was Kodak Ektaspeed^R. Each radiograph was coded and mounted in a plastic frame. The following parameters were assessed from the radiographs: secondary dentine, assessed according to the scoring system described by Johanson [2]; the tooth and pulp width measured at cemento-enamel junction in the stereomicroscope using a measuring eyepiece with an accuracy of 0.1 mm; the length of the whole tooth, the pulp and crown pulp, measured with a pair of vernier callipers with an accuracy of 0.1 mm.

Statistical analyses

Statistical analyses were performed with a Sanyo MBC-15 plus 5 microcomputer, using the SPSS/PC statistical program [9]. The subprogram Descriptives was used to calculate the mean and standard deviation for each variable and each observer. The subprogram Correlation was used to calculate the correlations with age, and the measurements were compared for two of the observers with paired *t*-tests using the T-test subprogram. The Pearson's correlations with age for each set of observations were also calculated. Finally the number of paired *t*-tests resulting in a significantly different mean observation at the level of $P < 0.01$ for each parameter was counted. The total number of paired *t*-tests for each parameter was 15.

Results

The results from the macroscopic and stereomicroscopic observations are shown in Table 1 giving the maximum and minimum values of the mean between the observers. The strongest correlations with age were found for attri-

Table 1 Macroscopic and stereomicroscopic scores of measurements made on 30 unsectioned teeth. The range of the mean between the observers and the units of the measurements are given. The range of Pearson's correlation (r) with age and the number of paired *t*-tests which showed a significant difference between pairs of observers are given ($n = 15$)

	Range of Mean	Unit	Range of r	*t*-test
Attrition				
Johanson:	2.5–3.3	score	0.6–0.8	4
Area:	5.5–9.0	square mm	0.5–0.8	3
Perio. fibres				
Mean	4.5–7.0	mm	0.1–0.5	7
Translucency ($n = 4$)				
Length:	4.9–6.9	mm	0.4–0.7	0 ($n = 6$)
Surface				
SRS:	2.3–2.7	score	0.2–0.7	0
Colour				
Biodent:	–	scale no.	0.2–0.6	–
Score:	1.8–3.3	score	0.4–0.6	7

– not applicable

Table 2 Stereomicroscopic scores and measurements made on radiographs from 30 teeth. Variations of the mean between observers and the unit of measurement are given. The variation in the Pearson correlation (r) with age and the number of paired *t*-tests which showed significant difference between pairs of observers are given ($n = 15$)

Secondary dentin	Range of Mean	Unit	Range of r	*t*-test
Johanson	3.4– 4.2	score	0.2– 0.6	–
Root width	4.1– 5.2	mm	–0.2–-0.4	4
Pulp width	0.3– 0.5	mm	–0.2–-0.6	1
Tooth length	22.2–23.2	mm	0.2– 0.3	4
Pulp length	15.4–16.4	mm	0.2– 0.3	3
Crown pulp length	0.6– 1.5	mm	0.2– 0.5	2

– not applicable

tion scores of Johanson [2] and were slightly weaker for the area of attrition measurements. The *t*-tests however, demonstrated that 4 pairs of observers had set scores which were significantly different from each other. For the retraction of the periodontal attachment the correlations with age were not found to be significant by some observers but 7 pairs of observers had also made significantly different measurements. Disappointingly, the colour score also significantly differed for 7 pairs of observers.

It was noted during the examination of the measurements and scores that the degree of precision varied considerably between the observers. There was a particularly large variation in measuring the attrition area under the stereomicroscope. Two observers had misinterpreted the instructions and estimated the length of the transluscent zone in the scores of Johanson [2] and these were omitted from the Table. The colour measurements were also difficult, as some observers had employed not the complete Biodent scale but only the colour shades related to the scores. Some of the observers replaced the scores by colour scales which had to be translated to the corresponding scores.

The measurements on the radiographs are shown in Table 2. No great mistakes in interpreting the type of measurements were revealed, and the measurements did not seem to have such great individual variation as those obtained directly from the teeth. Apparently, some observers had not fully used the advantage in accuracy that can be obtained by applying a measuring eyepiece. Except for the secondary dentine scores of Johanson [2], the measurements were obviously related to age. For all registrations the results of some observers showed no statistically significant relationship to age. It was amazing that the results indicated that the tooth and pulp length increased with age. The *t*-tests revealed that one or more pairs of observers deviated significantly from each other.

Discussion

It may be argued that the material might preferably have consisted of the same type of tooth. However, the aim of this experiment was to study how the instructions were acted upon and the variation with which scores and measurements were taken. Although several statistical methods exist to compare the concordance of 2 sets of paired observations, these are mostly correlation tests [9]. The coefficients may tell whether there is any concordance between the observations and, if so, to what extent. In this study there would always be a positive correlation, but the problem may be whether the differences in registrations are due to systematic differences in how the instructions are interpreted and how the scores and measurements are made. For this purpose a paired *t*-test may be most efficient in demonstrating systematic differences, even though the correlation between the results may be strong.

The observers, although not a homogeneous group, should all have had some experience in microscopy and with scientific work in general. One could have expected that they would have examined and followed the instructions more closely. In addition, published papers dealing with the technique could have been made available for the observers. However, it is a well known fact that people tend to make their own modifications in procedures which may even be an improvement, but in this study the modifications had negative influences and in some cases resulted in results which elude comparison.

Attrition was found to be the factor most closely related to age both for the scores and the area measurements. However, systematic differences between the observers were encountered for both types of quantitative estimate. It has been claimed that measurements are more objective than scores [4], but this investigation does not support that view.

Periodontal recession had a rather weak correlation with age. The same was found in another study, where it was shown that logarithmic transformation of the data were more closely related to age [5]. More unexpected was the high number of persons, 7 pairs, making systematically different measurements. Unclear demarcation of the uppermost coronal periodontal fibres may be one reason for this.

Translucency was one of the few factors where no systematic differences were found between the observers, but that was after the results from 2 of the observers were removed, as they had assessed this factor in a scoring system. The correlations with age were, however, weaker than that found by Johanson [2] for all teeth.

Root resorption has not shown statistically significant correlation with age, however surface roughness has, but the reproducibility was not found to be good [7]. It is surprising that this study did not reveal any systematic differences between the observers.

Colour assessments resulted in much weaker correlation with age than was found in another study [8]. The main reason for this might be different interpretations of how to assess the colour and the use of the scale. It indicates that caution should be used in the inclusion of colour in methods of age estimation. If it is included, an accurate reproduction of the scoring method is important. Furthermore, the differences might partly be explained by differences in the lighting conditions.

Radiographs of single teeth have not found much use in the practice of age estimation. The need to preserve the material in an unsectioned state, especially in archaeological collections, renders radiographs most valuable for assessing the size/shape of the dental pulp. This may also be so in forensic cases where it is important for religious and cultural reasons that the material is not damaged, and in cases involving living persons. A method for this has been presented [10] and work is in progress in the Scandinavian countries in the same direction [11].

The information on radiographs was used to assess the size of the dental pulp in addition to measuring the length and width of the teeth. Even though the variations in the mean between the observers was not significant in the measurements, in the visual examination of the teeth the *t*-test indicated systematic differences in some cases. One

reason for this may be that for some observers the measuring accuracy was 1 mm, while others used the nearest 0.1 mm.

A rather large inter-observer variation in the assessment of important variables used in age estimation has been revealed, suggesting that care should be taken not to rely too much upon the results of one single observer. Of course the correlation of the parameters with age that has been demonstrated may be true enough, but age estimates using certain methods may not be as reliable as the statistics indicate.

This study indicates that an inexperienced forensic odontologist, who may want to use age estimations, may find difficulties in following the instructions for the method. It would be advisable to follow a practical course and try the method under the supervision of an experienced forensic odontologist. Until more personal experience is acquired, one should be cautious in the interpretation of the calculated age from dental characteristics. An investigation with this design may serve as an assurance of quality in forensic procedures. The variation in the results shows the importance of obtaining a second opinion as an integrated part of the work.

Conclusions

Using the same methods and instructions, different examiners may obtain significantly different results in calculating the age from a tooth. Systematic differences between examiners might be due either to scores being interpreted differently or to definition of area/length measurements not being sufficiently clear. Unsystematic differences between examiners could be due to vague definitions of scores or to blurred edges of the area/length measurements.

References

1. Gustafson G (1950) Age determination on teeth. J Am Dent Assoc 41: 45–54
2. Johanson G (1972) Age determination from human teeth. Odont Rev 22:Suppl 21: 1–126
3. Solheim S (1993) A new method for dental age estimation in adults. Forensic Sci Int 59:137–147
4. Bang G, Ramm E (1970) Determination of age in humans from root dentin transparency. Acta Odontol Scand 28:3–35
5. Solheim T (1992) Recession of periodontal ligament as an indicator of age. J Forensic Odontostomatol 19:32–42
6. Solheim T (1988) Dental attition as an indicator of age. Gerodontics 4:299–304
7. Solheim T, Kvaal S (1993) Dental root surface structure as an indicator of age. J Forensic Odontostomatol 11: 9–21
8. Solheim T (1988) Dental color as an indicator of age. Gerodontics 4:114–118
9. Norusis MJ (1986) SPSS/PC+ for the IBM PC/XT/AT. Chicago:SPSS Inc.
10. Matsikidis G, Schulz P (1982) Altersbestimmung nach dem Gebiss mit Hilfe des Zahnfilms. Zahnarztl Mitt 72: 2524, 2527–2528
11. Kvaal SI, Solheim T (1994) A non-destructive dental method for age estimation. J Forensic Odontostomatol 12: 6–11