Legal Medicine 20 (2016) 75–79

Contents lists available at ScienceDirect

Legal Medicine

journal homepage: www.elsevier.com/locate/legalmed

Combining dental and skeletal evidence in age classification: Pilot study in a sample of Italian sub-adults

Vilma Pinchi [a,*], Federica De Luca [a], Martina Focardi [a], Francesco Pradella [a], Giulia Vitale [a], Federico Ricciardi [b], Gian-Aristide Norelli [a]

[a] Department of Health Sciences, Section of Medical Forensic Sciences, University of Florence, Largo Brambilla 3, 50134 Florence, Italy
[b] Department of Statistical Science, University College London, 1-19 Torrington Place, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Received 24 October 2015
Received in revised form 24 April 2016
Accepted 25 April 2016
Available online 26 April 2016

Keywords:
Age estimation
Skeletal age
Dental age
Forensic odontology
Age assessment
Linear Discriminant Analysis

ABSTRACT

Background: Dental and skeletal maturation have proved to be reliable evidence for estimating age of children and prior studies and internationally accredited guidelines recommend to evaluate both evidence in the same subject to reduce error in age prediction. Nevertheless the ethical and legal justification of procedures that imply a double exposition of children stands as a relevant issue. This study aims to evaluate the accuracy of age estimation provided by a combination of skeletal and dental methods applied in the same sample of children.

Materials and methods: The sample consisted of 274 orthopantomographies and left hand-wrist X-rays of Italian children, (aged between 6 and 17 years) taken on the same day. Greulich and Pyle's (GP), Tanner-Whitehouse's version 3 (TW3) and Willems' (W) and the Demirjian's (D) methods were respectively applied for estimating skeletal and dental age. A combination of skeletal and dental age estimates through Linear Discriminant Analysis (LDA) is proposed to obtain a classifier respect to an age threshold.

Results: The combination of D and TW3 obtained an improvement of accuracy in classifying female subjects respect to the 12 years threshold respect to the original methods (from about 77% using either original methods to 83.3% combining TW3 + D) as well as a consistent reduction of false positives rate (from 17% to 21% for original methods to 5.6% with TW3 + D). For males the LDA classifier (based on TW3 and W) enable a small improvement in accuracy, whilst the decreasing of false positives was as noticeable as for females (from 17.6 to 14.1% for original methods to 6.2% combining TW3 + W).

Conclusions: Although the study is influenced by the limited size and the uneven age distribution of the sample, the present findings support the conclusion that age assessment procedures based on both dental and skeletal age estimation can improve the accuracy and reduce the occurrence of false positives.

1. Introduction

In forensic practice, age estimation assessments in living children and adolescents have increased due to irregular immigration, asylum seeker proceedings, criminality in adolescence and adoption procedures [1]. Legal requirements depend on whether age estimation is requested for criminal proceedings or for civil or administrative purposes. Moreover, legal age thresholds of criminal responsibility vary considerably worldwide [2]. Dental calcification and wrist-hand bone maturation have proved to be reliable tools for estimating chronological age or classifying individuals with respect to an age threshold. Many studies [3–16] have concluded that an age assessment for legal or forensic purposes should rely on estimation of both dental and skeletal age. Nevertheless very few researchers had the opportunity to compare the estimation of dental age with bone age in the same sample of sub-adults because of the difficulties in collecting samples of known age, who underwent wrist X-rays and orthopantomography in the same period of time [9–11].

In the past dentists requested dental pantomographies and wrist-hand radiographies before orthodontic treatments of children. Nowadays orthodontists can obtain the requested auxological stage from cervical vertebrae development on lateral X-rays of the skull used for cephalograms, thus avoiding the additional exposition due to wrist radiography [12–14]. This difficulty in collecting appropriate sample of dental and wrist radiographies explains the low number of papers dealing with these observations. The aim of this study is to evaluate the accuracy of age estimation provided by different skeletal and dental methods applied

* Corresponding author.
E-mail address: pinchi@unifi.it (V. Pinchi).

http://dx.doi.org/10.1016/j.legalmed.2016.04.009
1344-6223/

in orthopantomography and wrist X-rays taken from the same children on the same day. Additionally, a procedural combination of skeletal and dental age estimates by means of Linear Discriminant Analysis (LDA) is experimented in order to verify whether the accuracy of final age assessment improves when a dental method is combined with a skeletal one.

2. Materials and methods

A sample of the left wrist-hand radiographies and ortopantomographies from 274 individuals (139 males and 135 females) 6–17 years of age (Table 1) was selected using the following inclusion criteria:

- dental pantomography and left hand and wrist radiography taken on the same day;
- Unremarkable medical history;
- Known gender, date of birth and date of X-ray examination

A specialist in forensic medicine examined the left hand and wrist X-rays and estimated skeletal age according to the atlas method of Greulich and Pyle (GP) [17] and the score method of Tanner and Whitehouse, RUS, version 3 (TW3) [18]. GP and TW3 were selected since GP is largely used in Italy and TW3 proved to be reliable for bone age estimation in Italian children [7], whilst other methods (Gisland and Ratib digital atlas, e.g.) are not used in Italy [19,20]. A forensic odontologist provided estimations of dental age based on teeth development applying Willems' (W) [21] and Demirjian's (D) methods [22,23]. For the present study, the age of 12 is selected as the age threshold of interest, since it is relevant in some countries and because it is the most suitable to be analysed according to the age distribution of the sample (Table 1).

The reproducibility of the methods was evaluated using the Intra-class Correlation Coefficient (ICC). Inter-rater variability was investigated by submitting 27 (about 10% of the sample) randomly selected X-rays to two different experts provided by similar scientific background and experience of the original operators.

Finally we experimented the use of LDA to combine dental and bone age estimates. LDA is a statistical methodology used to find a linear combination of variables that is adequate to separate two (or more) classes of objects or events. In this context it aims to express a dependent variable (the classification of a subject with respect to the age threshold) as a function of a set of explicatory variables, which are the different estimates of the age provided by the different applied methods (TW3, GP, W and D).

The best dental and skeletal method were selected by comparing the mean error (both *overall* and *per gender*), accuracy, false positive and false negative rates, when classifying subjects with respect to the age threshold of 12 years. LDA was then applied to the selected methods and the performance of the LDA classifier [24] was compared to each original method. Since the estimated ages were assumed as independent and continuous variables, and in LDA independent variables must be normal, Shapiro-Wilk normality tests were performed and a graphical overview of normality is given by means of Q-Q plots.

In order to validate the approach based on LDA, the sample was randomly divided using the R statistical computing environment and individuals were assigned to a **training sample**, which was composed of 60% of the original sample, and to a **test sample**. The predictive performance of the LDA model was then evaluated using test sample. This is a quite common technique for assessing how the results of a statistical analysis will generalize to another dataset.

For a preliminary comparison among methods, we considered the difference, express through Mean Error, between estimated age and chronological age. The mean error gives a measure of differences between chronological and estimated ages that is not affected by the sign (positive or negative) of the difference, thus providing a more easily interpretable measure of the distance between "reality" and "guess".

The Correct Classification rate (also referred to as "Accuracy" throughout the paper) of each method indicates the proportion of individuals that are correctly classified with respect to the age threshold of interest.

The false positive rate (children younger than 12 misclassified as older) and the false negative rate (individuals older than 12 misclassified as younger) have been assessed.

3. Results

The ICC values resulted to be 0.93 for GP, 0.90 for TW3, 0.85 for W and 0.82 for D. Hence all applied methods demonstrated to be highly reproducible since all these values are higher than the minimum satisfactory level, traditionally fixed at 0.80 [25].

Table 2 shows the mean error of estimations for each method calculated for the whole sample and separately for each gender. Skeletal methods yielded slightly lower mean error than dental methods in both genders. Furthermore TW3 and GP produced similar mean error (0.922 GP and 0.916 TW3) whilst, among dental methods, D was associated with lower overall errors, even if W performed better for male subjects.

The performance of the methods in the classification of individuals with respect to the age threshold of 12 is displayed in Table 3.

For accuracy, it emerged that TW3 performs better for male individuals (83.5%), whilst other methods (GP, D, W) obtained similar values for both males and females (ranging from 75% to 79%). The Willems' and the Demirjian's methods were consistently much more prone to age underestimation for both genders than skeletal methods, leading to higher occurrence of false negatives. Regarding the false positive rate, our data revealed that W was the safer method and that GP was affected by an impressively high percent-

Table 1
Composition of the sample.

Age	Females	Males	Total
6	1	0	1
7	1	3	4
8	4	4	8
9	35	8	43
10	31	38	69
11	24	32	56
12	17	27	44
13	15	12	27
14	3	10	13
15	3	1	4
16	0	3	3
17	1	1	2
Total	135	139	274

Table 2
Overall (left), Females (middle) and Males (right) Mean Errors (ME) related to the four methods considered.

Method	Mean error		
	Overall	Females	Males
GP	0.922	0.945	0.901
TW3	0.916	0.950	0.882
W	1.187	1.267	1.110
D	1.108	1.081	1.134