Int J Legal Med (2005) 119: 260–274
DOI 10.1007/s00414-004-0494-8

ORIGINAL ARTICLE

J. Braga · Y. Heuze · O. Chabadel · N. K. Sonan · A. Gueramy

# Non-adult dental age assessment: correspondence analysis and linear regression versus Bayesian predictions

Received: 16 July 2003 / Accepted: 21 September 2004 / Published online: 8 December 2004


**Abstract** This study focuses on the evaluation of factors influencing the quality (accuracy and reliability) of non-adult dental age assessment from radiographic stages of permanent teeth (excluding the third molar). We used four distinct cross-sectional samples of 1,528 healthy children: 3 of known geographic origin (Ivory Coast, Iran and France) and 1 additional sample of children whose grandparents originated from a different continent. Two different methods of calculations are compared: the correspondence analysis combined with linear regression (CAR) and Bayesian predictions (with no independence assumption). Our results indicate that the quality of age assessment does not seem to depend predominantly on the use of geographic-specific standards. In the case of Bayesian predictions, we observed a clear trend in favour of significantly higher accuracy and reliability levels when using non-geographic-specific standards. One of the main advantage of Bayesian predictions over maximum likelihood methods of estimation is an overall increase in accuracy with high levels of reliability on a fraction of the test sample and, importantly, across all age categories (contrary to methods based on regression analysis). Importantly, in the case of Bayesian non-adult predictions, and contrary to age estimation techniques based on regression, a better quality does not depend on age.

**Keywords** Age · Estimation · Accuracy · Reliability · Geographic-specific · Bayesian

**Abbreviations** CAR: correspondence analysis and regression · DMS: dental mineralisation sequence · GMI: global maturity index · SEE: standard error of the estimates.

Funded by the French Ministry of Research ("Action Concertée Incitative").

J. Braga (✉) · Y. Heuze
Laboratoire d'Anthropologie,
UMR 5199 CNRS,
Université Bordeaux 1,
Avenue des Facultés,
33405 Talence Cedex, France
e-mail: j.braga@anthropologie.u-bordeaux.1fr
Tel.: +33-05-40002546
Fax: +33-05-40002545

O. Chabadel
Service d'Odontologie, CHU,
Montpellier

N. K. Sonan
Service d'Orthopédie Dento-Faciale,
Université d'Abidjan,
Côte d'Ivoire

A. Gueramy
Department of Orthodontics,
Shiraz University of Medical Sciences,
Islamic Republic of Iran

## Introduction

Dental age is one of the measures of physiological development that is theoretically applicable from birth through late adolescence even if the interrelationships between the somatic, dental, skeletal and sexual maturity are not fully understood. Dental age assessment is common in orthodontic and pedodontic practice in order to plan the treatment of different types of malocclusions in relation to maxillofacial growth. Dental age is also used as a maturity indicator in pediatrics, orthopedic surgery, forensic science and physical anthropology. Dental age is estimated by comparing the dental mineralisation status in a person of known, or unknown, chronological age with dental developmental surveys, standard charts compiled from a large number of persons of known age and in a well defined geographic region (i.e., a reference sample). This involves the critical presumption that the growth rate of the reference sample is representative of the sex, ethnic, and regional background of the person being evaluated (for whom dental age is estimated). Several methods of dental age assessment in non-adults have been used: the atlas method of Schour and Massler (1940), the specific standards from the assessment of radiographic stages by Demirjian et al. (1973), the diagram of Gustafson and Koch

(1974) and the length and weight regression equations of Deutsch et al. (1984). There is very little information of within and between human population variability in crown and root mineralisation timing and patterns. As a result, some research groups challenge the existence of a significant variability within and between the extant human populations sampled so far. For example, Scheuer and Black (2000) consider that "some of the conclusions that claim population, and other differences are more likely to be due to the use of statistical treatments, or to sampling effects, rather than real differences between samples". In this case, large variations in relative dental mineralisation are considered to more likely reflect methodological problems rather than human diversity.

## Geographic-specific standards: accuracy and reliability

Even if true biological differences exist between samples, methodological disparity between studies is undoubtedly an important factor influencing the quality (mainly accuracy and reliability; see methods section) of dental age assessment (see Liversidge et al. 1998 for review). Most studies emphasising differences both within and among populations, and concluding that dental age assessment is dependent on the sample on which it is to be tested, consider the regional background, sex, chronological age distribution, and "ethnicity" as the main variation factors (Hägg and Matsson 1985; Staaf et al. 1991; Davis and Hägg 1994; Koshy and Tandon 1998). As regards other possible sources of variation, opinions diverge and most other putative variation factors are not sufficiently known. For example, cultural or environmental (e.g., socio-economic status, nutrition, dietary habits) but also endocrine factors have been considered of no effect (e.g., Garn et al. 1965a,b; Voors 1973) or, on the contrary, of considerable effect (e.g., Gulati et al. 1990; Bhargava 2000) on the rate of an individual's progress toward maturity. Therefore, many studies reach the central conclusion that no universal system for dental age assessment has been achieved and that maturity standards should be geographic-specific, i.e., based on studies made on the same geographic population for which they are going to be used. Are quality criteria, mainly the accuracy (the mean difference between dental and chronological age) and reliability (the range and percentage of the confidence limits), of a method for dental age estimation solely, or predominantly, dependent on the use of geographic-specific standards in which regional background, sex, chronological age distribution, and "ethnicity" would be the main controlling factors? This question represents the first issue that we aim to examine in this paper. This is problem number one.

## Dental age calculation methods

Comparisons between growth studies of differing design and methods are beset with problems. Significant differences between dental age and chronological age may be the result of differing methodology rather than a true reflection of the populations (Staaf et al. 1991; Liversidge 1994). Various chronologies and radiographic stage methods have been used for dental age estimation of non-adults using deciduous and/or permanent teeth (Gleiser and Hunt 1955; Garn et al. 1958, 1959; Nolla 1960; Fanning 1961; Moorees et al. 1963a,b; Haataja 1965; Nanda and Chawla 1966; Wolanski 1966; Haavikko 1970, 1974; Fanning and Brown 1971; Liliequist and Lundberg 1971; Demirjian et al. 1973; Gustafson and Koch 1974; Anderson et al. 1976; Demirjian and Goldstein 1976; Nyström et al. 1977, 1986). In this paper, we focus our attention on the pioneer study of Demirjian et al. (1973), assessing maturity of children of known age by means of eight radiographic stages. Each tooth having a stage is converted into a numerical score using a conversion table. The scores of all teeth are then added together to give the global maturity index (GMI) which can be converted directly into a dental age using an appropriate table of standards. At this stage, two important methodological issues should be stressed:

- The conversion tables are obtained by Demirjian et al. (1973) by means of a weighted score (reviewed by Demirjian 1986) based on the principle of skeletal age by assessing maturity from wrist ossification status by Tanner (1962).Other conversion tables are obtained by Demirjian and Goldstein (1976) by using the mathematical model proposed by Healy and Goldstein (1976). However, some scholars consider that this mathematical model is not satisfactory because it gives significantly higher weights to advanced stages of maturation (Proy and Gauthier 1985).
- The tables of standards can be obtained by using the linear regression analysis (e.g., Koshy and Tandon 1998) or centile distributions (Demirjian et al. 1973) to derive the relationship between the GMI and chronological age. The use of regression analysis for ordinal data (GMIs are derived from radiographic stages and therefore represent a series of discrete values) may be responsible for some problems as already emphasised by Lucy et al. (1996) who consider that there is not "any rationally justifiable reason" for the use of regression analysis "in situations that involve categorical or ordinal data." Among the five assumptions of regression analysis listed by Lucy et al. (1996), we should expect partial correlations between the radiographic stages for each tooth to approach zero, when controlled for age, all the teeth contributing the same amount of information about age. Moreover, the scores should vary continuously with age. In fact, the stages adopted by Demirjian et al. (1973) and Demirjian and Goldstein (1976) do not approximate the continuous phenomenon of dental mineralisation. We deal with an ordinal variable. The use of balancing mathematical models or of statistical procedures not adapted to categorical data may be responsible for some of the problems encountered in dental age assessment. Methodological alternatives do exist and this represents the second issue

that we aim to examine in this paper. This is problem number two.

### Objectives

This study has two distinct objectives. The first objective is to focus on the question of whether the regional background (or "ethnicity", see material section), sex, chronological age distribution of the sample sets, and statistical procedure represent major factors controlling accuracy and reliability in non-adult dental age assessment.

Besides the study of the possible factors influencing (with their relative contributions) non-adult dental age assessment, the second objective of this study is to apply a Bayesian statistical procedure and to provide clear information about the quality level of our trials. In the context of this study, the procedure for the evaluations and comparisons of quality of trials is based on performance (defined in the methods section), accuracy and reliability. Our "individual age assessment" approach should be distinguished from the techniques of estimating population age structure in palaeodemography (in particular, when selecting prior probabilities, see methods section). We focus on the Bayesian method because it represents a statistical alternative well adapted to the analysis of dependent attributes, or categorical data (radiographic stages of dental development). We aim to discuss clearly the advantages and drawbacks of the Bayesian predictions, with no independence assumption, used in this study.

## Materials and methods

### Material

Our 3 geographic population samples consist of cross-sectional standardised orthopantomographs of the teeth of 262 children aged 49 to 194 months (108 boys and 154 girls) from The Ivory Coast, 393 children aged 69 to 197 months (136 boys and 257 girls) from Iran, and 454 children aged 45 to 192 months (212 boys and 257 girls) from France. All the children are of known chronological age (in months; from the birth date to the time of taking of the radiograph) and, importantly, "ethnicity" (their grandparents originated from the same country and/or the same geographic area). An additional sampling was made in France. Most of the children from this sample were born in France and have one or more of their grandparents not originating from Europe (originating mainly from north or west Africa). This additional sample comprises 404 children aged 49–191 months (170 boys and 234 girls). Age and sex distribution of our three geographic samples and the additional sample are shown in Table 1. The orthopantomographs were collected from dental records at (i) the Service d'Orthopédie Dento-Faciale, Université d'Abidjan, Ivory Coast, (ii) the Department of Orthodontics, Shiraz University of Medical Sciences, Islamic Republic of Iran, (iii) the Service d'Odontologie des CHU de Bordeaux et de Montpellier, and private consultancy practices (Dr Frapier in Montpellier and Dr De Brondeau in Bordeaux). We obtained the consent from all the parents of the children included in this study. For the French sample, the consent of the Commission Nationale de l'Informatique et des Libertés (CNIL; an organisation dedicated to information technology and chronological rights in France) was obtained. We selected orthopantomographs only from children apparently free from any abnormal developmental parameters and who had a complete mandibular permanent developing dentition (children having even one tooth missing were removed from the study). Two examiners rated the left permanent mandibular teeth (excluding the third molar, the more variable, to increase the accuracy of estimation) on the radiographs, according to the eight radiographic stages defined by Demirjian et al. (1973). One examiner randomly selected orthopantomographs for reassessment. Disagreement occurred in less than 5% of films and, at the most, by one stage, mainly for the anterior teeth and premolars. For all statistical procedures, the samples were divided into males and females. The investigator did not know the chro-

**Table 1** Case numbers in the samples of girls and boys of known geographic origin (Ivory Coast, Iran and France) and one additional sample of children whose grandparents originated from a different continent

| Age categories | France | | France | | Iran | | Ivory Coast | |
|---|---|---|---|---|---|---|---|---|
| (months) | F | M | F | M | F | M | F | M |
| ≤48–54< | 1 | 1 | 3 | 0 | 0 | 0 | 0 | 1 |
| ≤54–60< | 1 | 0 | 1 | 3 | 0 | 0 | 0 | 0 |
| ≤60–66< | 0 | 3 | 1 | 3 | 0 | 0 | 1 | 2 |
| ≤66–72< | 2 | 1 | 4 | 1 | 1 | 0 | 4 | 4 |
| ≤72–78< | 0 | 1 | 7 | 6 | 1 | 1 | 7 | 4 |
| ≤78–84< | 1 | 3 | 5 | 7 | 9 | 4 | 9 | 11 |
| ≤84–90< | 2 | 3 | 3 | 6 | 8 | 3 | 17 | 7 |
| ≤90–96< | 10 | 9 | 11 | 5 | 13 | 12 | 13 | 10 |
| ≤96–102< | 14 | 7 | 7 | 9 | 15 | 11 | 21 | 10 |
| ≤102–108< | 13 | 12 | 16 | 10 | 16 | 9 | 20 | 4 |
| ≤108–114< | 11 | 11 | 13 | 13 | 17 | 6 | 12 | 9 |
| ≤114–120< | 14 | 18 | 24 | 8 | 20 | 10 | 5 | 8 |
| ≤120–126< | 24 | 16 | 23 | 15 | 11 | 10 | 5 | 6 |
| ≤126–132< | 21 | 15 | 20 | 8 | 10 | 11 | 6 | 10 |
| ≤132–138< | 28 | 19 | 20 | 14 | 14 | 7 | 6 | 3 |
| ≤138–144< | 26 | 13 | 18 | 16 | 12 | 9 | 9 | 3 |
| ≤144–150< | 20 | 19 | 11 | 14 | 18 | 5 | 5 | 3 |
| ≤150–156< | 15 | 17 | 11 | 6 | 16 | 10 | 2 | 2 |
| ≤156–162< | 11 | 14 | 10 | 9 | 20 | 10 | 6 | 3 |
| ≤162–168< | 13 | 7 | 11 | 5 | 9 | 6 | 2 | 4 |
| ≤168–174< | 6 | 11 | 3 | 4 | 10 | 5 | 1 | 4 |
| ≤174–180< | 2 | 4 | 6 | 2 | 10 | 3 | 2 | 0 |
| ≤180–186< | 4 | 5 | 2 | 1 | 6 | 2 | 0 | 0 |
| ≤186–192< | 3 | 3 | 4 | 5 | 13 | 2 | 0 | 0 |
| ≤192–198< | 0 | 0 | 0 | 0 | 8 | 0 | 1 | 0 |
| Total | 242 | 212 | 234 | 170 | 257 | 136 | 154 | 108 |

Individuals are classified into 25 chronological age categories representing 6-month intervals and ranging from ≤48–54<to ≤192–198<months.

nological age of the children when assessing the radiographs.

## Methods

In order to investigate the relative influence of the calculation method, for each sex, our three geographic (European, Asian and African) samples were alternatively used as reference (training) and target (test) samples (e.g., once trained, the statistical model based on reference sample data for girls from Iran, will be used to test the age of girls from France; Fig. 1). The quality (Fig. 1) of these trials (using different combinations of training and test samples) was compared using two different statistical procedures: the widely used correspondence analysis and regression (CAR) method and the Bayesian method (Table 2). Quality is based on accuracy, reliability and performance. Quality is considered as globally high if it is satisfactory for most (if not all) test samples (all possible datasets available for this study). Moreover, in order to assess the influence of geographic-specific standards (training and test sample of the same geographic origin), we compared the quality of Bayesian predictions using non-geographic-specific and geographic-specific standards (Table 3). In the case of trials with geographic-specific standards, we used a jackknife resampling technique (Tukey 1977), for each sex within our three geographic samples (by leaving each individual out to test the dental age). In order to evaluate the global quality of the Bayesian method, we also used a larger data set and a jackknife resampling technique, by combining in one training sample, for each sex, our three geographic population samples and one additional sample from France. Each individual was left out to test the dental age. The quality of these trials for girls and boys (Fig. 2) was compared to those using geographic-specific standards (Table 4).

Each individual was assigned to 25 chronological age categories representing 6-month intervals: from =48–54< to =192–198< months (Table 1). This assignment is characterised by a 6-month precision. For any individual from the test sample, each dental age assessment (by either of the two methods compared in this study) was compared to the initial chronological age assignment and subsequently given a rating (Fig. 1). This rating is based on where the dental age assessment falls with regard to the chronological age assignment. For example, if the individual dental age assessment falls in the chronological age assignment, it is classified as 0. If the individual dental age assessment is one

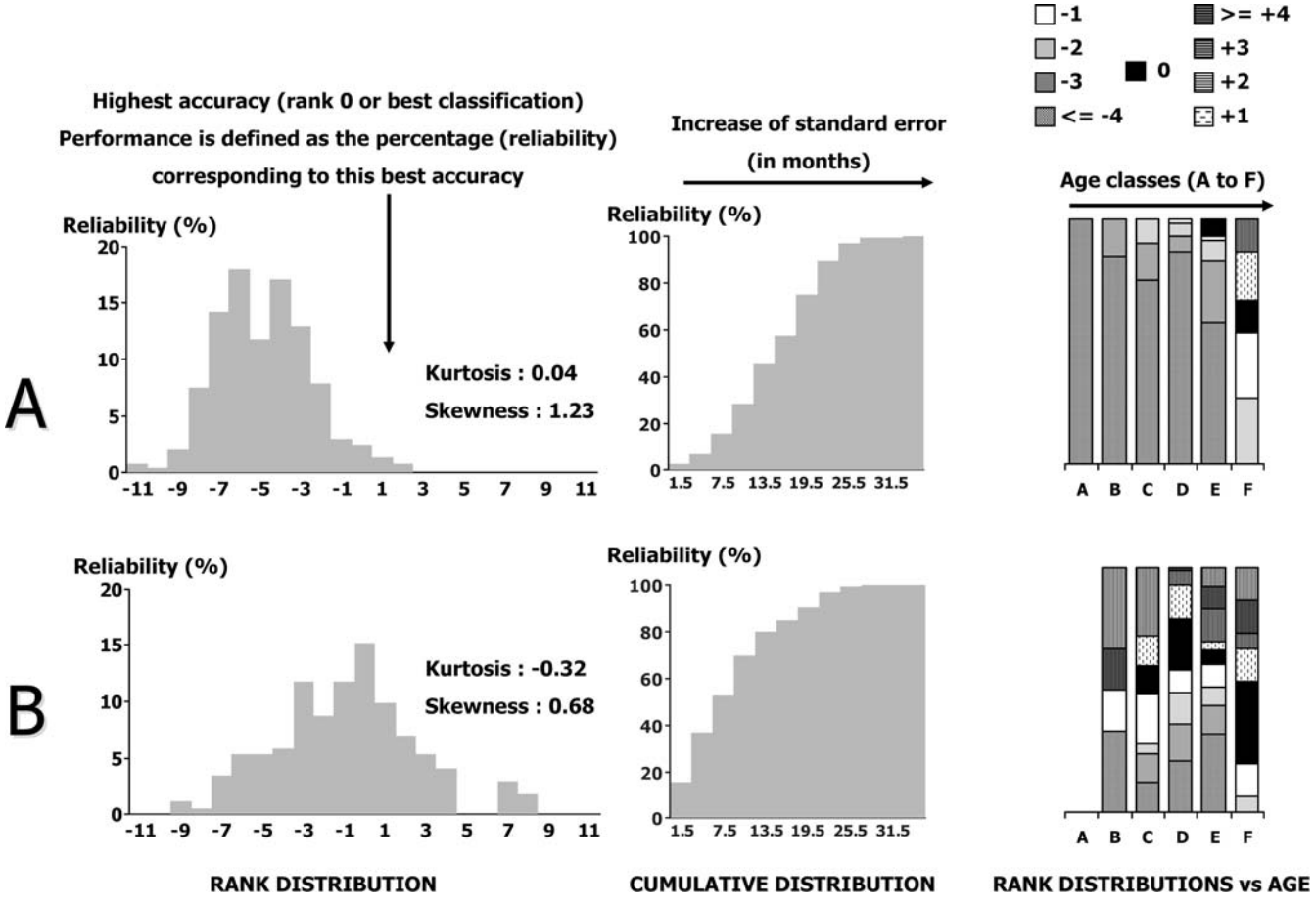


**Fig. 1** An example of differences in the quality of non-adult dental age assessment using two different methods of calculation (A: CAR; B: Bayesian): rank distributions, cumulative distributions and rank distributions versus age classes

**Table 2** Comparisons of quality levels in non-adult dental age assessment using two different methods of calculation (CAR and Bayesian)

| Test samples | Criteria | Training samples girls | | | | | | Criteria | Training samples boys | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Iran (N=257) | | Ivory Coast (N=154) | | France (N=242) | | | Iran (N=136) | | Ivory Coast (N=108) | | France (N=212) | |
| | | CAR | Bayes | CAR | Bayes | CAR | Bayes | | CAR | Bayes | CAR | Bayes | CAR | Bayes |
| Iran | 1 | | | 46.8 | | 19.4 | | 1 | | | 44.4 | | 20.1 | |
| | 2 | | | 98.8 | 64.2 | 99.2 | 65.4 | 2 | | | 96.3 | 47.8 | 100 | 61.0 |
| | 3 | | | 22.5(98.8) | 19.5(95.2) | 22.5(97.6) | 22.5(98.2) | 3 | | | 19.5 (96.9) | 28.5 (96.9) | 16.5 (97.1) | 19.5 (96.4) |
| | 4 | | | 0.2764* | 0.9016 | 0.4892* | | | | | 0.0696* | 0.1088* | 0.1278* | 0.2800* |
| | 5 | | | 17.3 | 15.8 | 16.5 | 13.1 | 5 | | | 14.5 | 9.2 | 18.4 | 15.7 |
| | 6 | | | 0.6747 | | 0.3432 | | | | | 0.2975 | | 0.6061 | |
| | 7 | | | 1.07 | −1.21 | 0.30 | −1.25 | 7 | | | 0.02 | 1.12 | −0.59 | −1.38 |
| | 8 | | | 1.23 | 0.36 | 1.07 | 0.55 | 8 | | | 1.00 | 1.30 | 0.93 | 0.30 |
| Ivory Coast | 1 | 22.6 | | | | 19.4 | | | 29.4 | | | | 20.1 | |
| | 2 | 99.4 | 50.6 | | | 98.7 | 39.0 | 2 | 95.4 | 48.1 | | | 100 | 43.5 |
| | 3 | 16.5 (98.0%) | 22.5 (97.4%) | | | 13.5 (96.1) | 19.5 (96.7) | 3 | 16.5 (95.1) | 19.5 (96.1) | | | 16.5 (97.2) | 19.5 (95.7) |
| | 4 | 0.0189* | 0.2641* | | | 0.0079* | 0.3179* | 4 | 0.2107* | 0.5455* | | | 0.2587* | 0.3535* |
| | 5 | 14.4 | 20.5 | | | 23.0 | 15.0 | 5 | 11.7 | 19.2 | | | 15.7 | 23.4 |
| | 6 | 0.2344 | | | | 0.1943 | | | 0.2017 | | | | 0.2543 | |
| | 7 | −1.42 | 0.73 | | | −0.41 | −0.86 | 7 | −1.62 | −0.29 | | | −0.79 | 0.52 |
| | 8 | 0.50 | 1.21 | | | 0.86 | 0.88 | 8 | 0.35 | 0.69 | | | 0.76 | 1.07 |
| France | 1 | 22.6 | | 46.8 | | | | 1 | 29.4 | | 44.4 | | | |
| | 2 | 99.2 | 70.7 | 97.5 | 61.2 | | | 2 | 99.1 | 62.7 | 95.3 | 39.2 | | |
| | 3 | 25.5 (96.7%) | 22.5 (96.5) | 13.5 (95.3) | 19.5 (97.3) | | | 3 | 40.5(100) | 28.5 (96.2) | 19.5(99.5) | 34.5(95.2) | | |
| | 4 | 0.1397* | 0.0000 | 0.0019 | 0.1578* | | | 4 | - | 0.0000 | 0.0000* | 0.0068* | | |
| | 5 | 2.5 | 15.2 | 10.6 | 16.9 | | | 5 | 4.3 | 12.0 | 17.8 | 9.6 | | |
| | 6 | 0.0000 | | 0.0743 | | | | 6 | 0.0072 | | 0.0824 | | | |
| | 7 | 0.04 | −0.32 | −0.89 | 0.12 | | | 7 | 4.08 | 0.05 | −1.08 | 1.06 | | |
| | 8 | 1.23 | 0.68 | 0.68 | 1.03 | | | 8 | 1.63 | 1.07 | 0.64 | 1.36 | | |

*1* Percentage of single dental mineralisation sequences.
*2* Percentage of response.
*3* Standard error for the 95th percentile (or a higher reliability level, as close as possible to 95%) of the cumulative distribution.
*4* Level of significance of differences between the standard errors at the 95% level.
*5* Performance.
*6* Level of significance of differences between performances.
*7* Kurtosis.
*8* Skewness.

**Table 3** Comparisons of quality levels in Bayesian non-adult dental age assessment using geographic-specific and non geographic-specific standards

| Geographic samples tested | Criteria | Training samples girls | | | | Criteria | Training samples boys | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Iran vs Ivory Coast | | Iran vs France | | | Iran vs Ivory Coast | | Iran vs France | |
| | | Geo. specific | Ivory Coast | Geo. specific | France | | Geo. specific | Ivory Coast | Geo. specific | France |
| Iran | 1 | 22.6 | 46.8 | 22.6 | 19.4 | 1 | 28.7 | 44.4 | 28.7 | 20.1 |
| | 2 | 77.4 | 64.2 | 77.4 | 65.4 | 2 | 71.3 | 47.8 | 71.3 | 61.0 |
| | 3 | 22.5 (96.0) | 19.5 (95.2) | 22.5 (96.0) | 22.5 (98.2) | 3 | 22.5 (96.9) | 28.5 (96.9) | 22.5 (96.9) | 19.5 (96.4) |
| | 4 | 0.3959 | 0.0321 | 0.2110 | | 4 | 0.0944* | 0.5280* | 0.5734* | 0.0214 |
| | 5 | 3.5 | 15.8 | 3.5 | 13.1 | 5 | 10.3 | 9.2 | 10.3 | 15.7 |
| | 6 | 0.0001 | | 0.0007 | | 6 | 0.8216 | | 0.2835 | |
| | 7 | 1.48 | −1.21 | 1.48 | −1.25 | 7 | −1.18 | 1.12 | −1.18 | −1.37 |
| | 8 | 1.36 | 0.36 | 1.36 | 0.55 | 8 | −0.32 | 1.30 | −0.32 | 0.10 |
| | | Ivory Coast vs Iran | | Ivory Coast vs France | | | Ivory Coast vs Iran | | Ivory Coast vs France | |
| | | Geo. specific | Iran | Geo. specific | France | | Geo. specific | Iran | Geo. specific | France |
| Ivory Coast | 1 | 46.8 | 22.6 | 46.8 | 19.4 | 1 | 44.4 | 29.4 | 44.4 | 20.1 |
| | 2 | 53.2 | 50.6 | 53.2 | 39.0 | 2 | 55.6 | 48.1 | 55.6 | 43.5 |
| | 3 | 16.5 (95.1) | 22.5 (97.4) | 16.5 (95.1) | 19.5 (96.7) | 3 | 16.5 (95.0) | 19.5 (96.2) | 16.5 (95.0) | 19.5 (95.7) |
| | 4 | 0.3060 | 0.6708* | 0.3102* | 0.6449* | 4 | 0.2813* | 0.635* | 0.5386* | 0.5939* |
| | 5 | 4.9 | 20.5 | 4.9 | 15.0 | 5 | 21.7 | 19.2 | 21.7 | 23.5 |
| | 6 | 0.0028 | | 0.0388 | | 6 | 0.7503 | | 6 | |
| | 7 | −0.75 | 0.73 | −0.75 | −0.86 | 7 | 1.59 | −0.29 | 1.59 | 0.52 |
| | 8 | 1.00 | 1.21 | 1.00 | 0.88 | 8 | 1.21 | 0.69 | 1.21 | 1.07 |
| | | France vs Ivory Coast | | France vs Iran | | | France vs Ivory Coast | | France vs Iran | |
| | | Geo. specific | Ivory Coast | Geo.specific | Iran | | Geo.specific | Ivory Coast | Geo.specific | Iran |
| France | 1 | 19.4 | 46.8 | 19.4 | 22.6 | 1 | 19.3 | 44.4 | 19.3 | 29.4 |
| | 2 | 80.6 | 61.2 | 80.6 | 70.7 | 2 | 80.7 | 39.2 | 80.7 | 62.7 |
| | 3 | 19.5 (95.9) | 19.5 (97.3) | 19.5 (95.9) | 22.5 (96.5) | 3 | 22.5 (98.8) | 34.5 (95.2) | 22.5 (98.8) | 28.5 (96.2) |
| | 4 | 0.4847 | | 0.0272 | 0.9667 | 4 | 4 | 0.0109* | 0.0006 | 0.0154* |
| | 5 | 13.3 | 16.9 | 13.3 | 15.2 | 5 | 6.4 | 9.6 | 6.4 | 12.0 |
| | 6 | 0.3589 | | 0.6089 | | 6 | 0.3623 | | 6 | |
| | 7 | −1.20 | 0.12 | −1.20 | −0.32 | 7 | −0.22 | 1.06 | −0.22 | 0.05 |
| | 8 | 0.70 | 1.03 | 0.70 | 0.68 | 8 | 0.85 | 1.36 | 0.85 | 1.07 |

*1* Percentage of single dental mineralisation sequences.
*2* Percentage of response.
*3* Standard error for the 95th percentile (or a higher reliability level, as close as possible to 95%) of the cumulative distribution.
*4* Level of significance of differences between the standard errors at the 95% level.
*5* Performance.
*6* Level of significance of differences between performances.
*7* Kurtosis.
*8* Skewness.

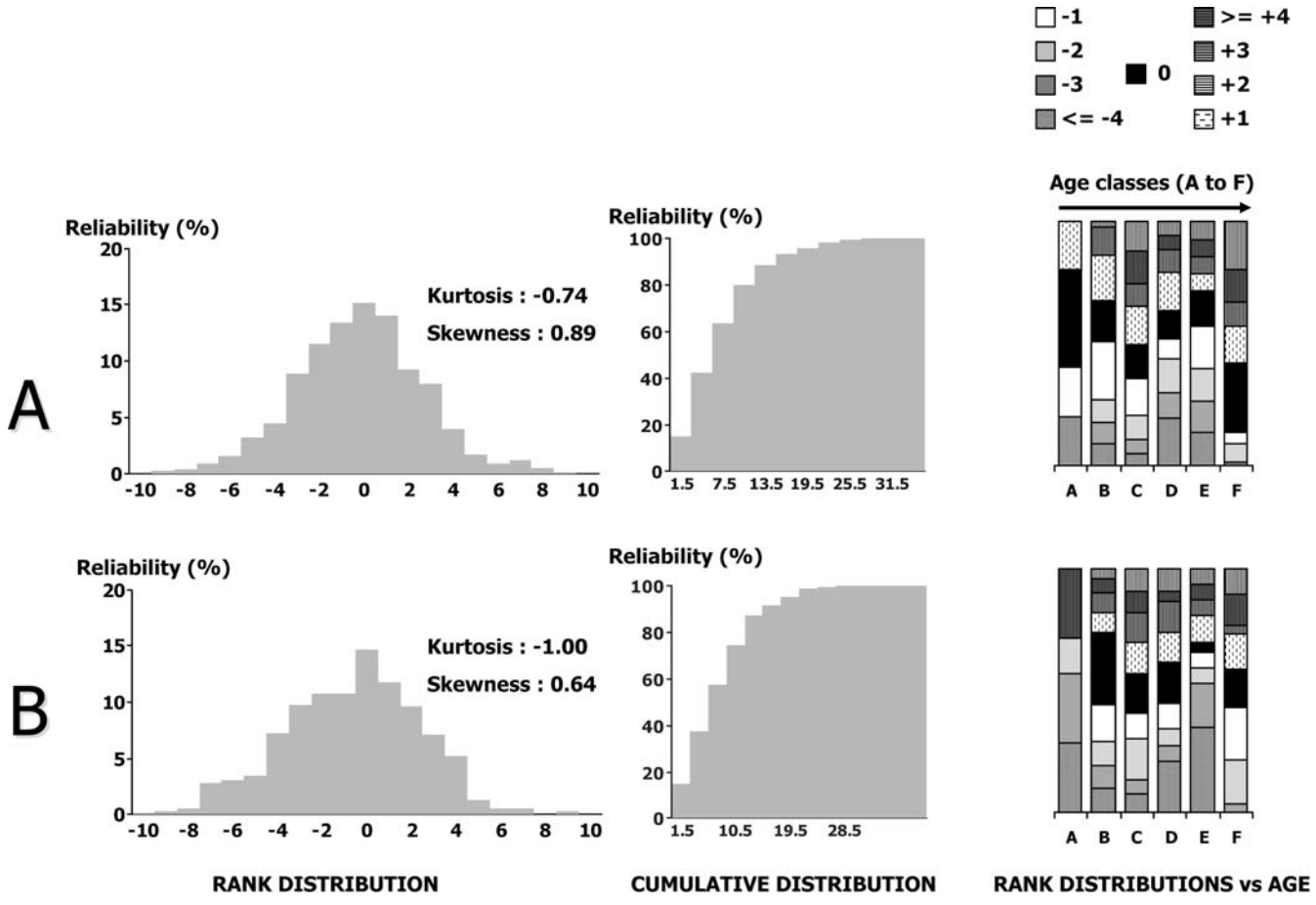


**Fig. 2** Quality of Bayesian (with no independence assumption) non-adult dental age assessment (A: test on 902 girls; B: test on 626 boys): rank distributions, cumulative distributions and rank distributions versus age classes

age category (6 months) below or above the chronological age assignment, it is classified as −1 and +1, respectively. This rank-ordering approach leads to a rank distribution (Fig. 1) and allows a comparison of the accuracy, reliability and performance obtained by the two methods used in this study: first, the CAR method, and second the Bayesian method.

A classification into 0 corresponds to the highest accuracy, the best classification, i.e. the better agreement between age assessment and the previous chronological age assignment (a maximum deviation of ±3 months). The percentage of individuals classified into the rank 0 is a measure of the reliability associated with this highest accuracy. In the context of this study, the term "performance" is used to represent the proportion of individuals with a rating of 0 (Fig. 1). The classification into rank −1 or rank +1 corresponds to the same accuracy level (the highest after that of rank 0 with, respectively, −3 to −9 months, and, +3 to +9 months). If we now consider the cumulative proportion of individuals classified with decreasing accuracy levels (or increasing standard errors), we obtain a cumulative distribution, f(x/y) (Fig. 1) where *x* represents the standard error (e.g., 4.5 months correspond to a classification into the range of ranks −1 to +1) and *y* represents its associated reliability (represented by an area of the cumulative distribution). A 95% confidence interval will be calculated by considering the standard error corresponding to the 95th percentile of the cumulative distribution, and then by multiplying this standard error by 2. As the cumulative distribution is not continuous, in order to provide the standard error at a 95% level, we will use the 95th percentile or a higher reliability level (as close as possible to 95%). The performances are compared by using $\chi^2$ or Fisher exact tests. Significance levels for comparisons are given (Tables 2, 3 and 4). The significance levels are also given for the differences between cumulative distributions at the 95% reliability threshold. What does it mean? This means that the 95% reliability threshold is often obtained with a lower standard error in 1 of the 2 trials to be compared, and a higher standard error in the other (Tables 2, 3 and 4). When comparing the reliabilities associated with these lower and higher standard errors, we obtain a significance level. If the significance level is equal to or lower than 0.05, the associated standard error is, in one of the two trials to be compared, associated with a significantly higher reliability (necessarily higher than 95%). We complemented the evaluation of global quality of our trials by analysing the shape of the rank distributions with the help of both kurtosis and skewness. The standard "normal" distribution has a skewness and kurtosis of zero and corresponds to a better

**Table 4** Comparisons of quality levels in Bayesian non-adult dental age assessment using the total samples (girls and boys separately) and geographic-specific standards

| Criteria | Samples for girls tested using jackknife | | | | | |
|---|---|---|---|---|---|---|
| | Total | Iran | Total | Ivory Coast | Total | France |
| 1 | 12.8 | 22.6 | 12.8 | 46.8 | 12.8 | 19.4 |
| 2 | 87.2 | 77.4 | 87.2 | 53.2 | 87.2 | 80.6 |
| 3 | 19.5 (95.7) | 22.5 (96.0) | 19.5 (95.7) | 16.5 (95.1) | 19.5 (95.7) | 19.5 (95.9) |
| 4 | 0.0002 | 0.1135 | 0.7769 | 0.5121 | 0.8923 | |
| 5 | 15.1 | 3.5 | 15.1 | 4.9 | 15.1 | 13.3 |
| 6 | 0.0000 | | 0.0115 | | 0.5373 | |
| 7 | −0.74 | 1.48 | −0.74 | −0.75 | −0.74 | −1.20 |
| 8 | 0.89 | 1.36 | 0.89 | 1.00 | 0.89 | 0.70 |
| Criteria | Samples for boys tested using jackknife | | | | | |
| | Total | Iran | Total | Ivory Coast | Total | France |
| 1 | 14.6 | 28.7 | 14.6 | 44.4 | 14.6 | 19.3 |
| 2 | 85.4 | 71.3 | 85.4 | 55.6 | 85.4 | 80.7 |
| 3 | 19.5 (95.1) | 22.5 (96.9) | 19.5 (95.1) | 16.5 (95.0) | 19.5 (95.1) | 22.5 (98.8) |
| 4 | 0.0010 | 0.2232 | 0.5945 | 0.3313 | 0.1601 | 0.8090 |
| 5 | 14.7 | 10.3 | 14.7 | 21.7 | 14.7 | 6.4 |
| 6 | 0.2441 | | 0.1541 | | 0.0041 | |
| 7 | −1.00 | −1.18 | −1.00 | 1.59 | −1.00 | −0.22 |
| 8 | 0.64 | −0.32 | 0.64 | 1.21 | 0.64 | 0.85 |

*1* Percentage of single dental mineralisation sequences.
*2* Percentage of response.
*3* Standard error for the 95th percentile (or a higher reliability level, as close as possible to 95%) of the cumulative distribution.
*4* Level of significance of differences between the standard errors at the 95% level.
*5* Performance.
*6* Level of significance of differences between performances.
*7* Kurtosis.
*8* Skewness.

performance (both better accuracy and reliability levels). Kurtosis is a measure of the pointedess of the rank distribution. A peaked distribution has negative kurtosis, while a flatter curve would have positive kurtosis (Fig. 1). If the left hand tail is longer, skewness will be negative. If the right hand tail is longer, skewness will be positive. Global reliability and accuracy levels will be considered different between two trials when both kurtosis and skewness will be closer to 0 in one trial, as compared to the other (Tables 2, 3 and 4).

In order to assess the possible influence of age distributions (in the training and test samples) on age assessment, we examined and compared graphically the rank distributions across age classes (representing 2 years) corresponding to age categories grouped together (to simplify graphic representation). These 2-year age classes are defined as follows: A=48–72< months, B=72–96< months, C=96–120< months, D=120–144< months, E=144–168< months, F=168–192< months, the last age category =192–198< months (individuals older than 16 years) being not represented because of too few children available.

### *The CAR method*

The eight radiographic stages of Demirjian et al. (1973) (A–H from the first appearance of calcified points to the end of root growth) are converted into a numerical score using a conversion table. To do so, the mathematical model proposed by Healy and Goldstein (1976) and advocated by Tanner et al. (1975), was used by Demirjian et al. (1973). This model is considered as highly suspect by Proy and Gauthier (1985) because of its over-weighting of advanced stages of maturation. Therefore Proy and Gauthier (1985) recommended the use of correspondence analysis, a multivariate analysis procedure applicable only to categorical features without any other limitation (Greenacre 1993). The important points to know about this procedure are that (i) it shows the patterns of relationships among many categorical variables, (ii) it plots the relationships between these multiple variables on a smaller number of dimensions according to variance explained, and (iii) the distance between variables on each dimension is standardised (i.e., items that are close together are likely to be related). In the CAR method, correspondence analysis is used to calculate the conversion table. Correspondence analysis standardises the distance between variables, i.e. each radiographic stage, on each dimension. The conversion table for each sample is obtained by replacing each variable by its coordinates on the first dimension of the correspondence analysis, which explains the higher percentage of variance. The correspondence analysis gives a set of numerical scores, one for each stage of each tooth: the conversion table (Table 5). The scores of all seven permanent left mandibular teeth are then added together to give the global maturity index (GMI). The chronological ages and the GMIs were assessed statistically by using the simple regression analysis with the GMIs taken as the independent variable and the chronological age as the dependent variable, for girls and boys separately. The dental age is the calculated age from the regression equations (Table 5), using 95% confidence limits. The standard error of the estimates (SEE; equal to the square root of the average squared deviation from the regression line) is a good measure of accuracy of predictions made in that way (Table 5). Indeed, the SEE (occasionally termed the "standard deviation of prediction errors in linear regression" or "error of prediction") is an overall indication of the accuracy with which the fitted regression function predicts the dependence of the chronological age on the GMI. We should note the similarity between the SEE (bivariate data) and the standard deviation calculated for univariate data. All the regressions and SEEs are given at the 95% confidence level. Therefore, a SEE of 14 months means that 95% confidence intervals of about ±28 months or even more (Giles and Klepinger 1988) have to be considered in age estimation.

**Table 5** Weighted scores for the radiographic stages defined by Demirjian et al. (1973) on seven mandibular permanent teeth (left side) in girls and boys from three samples of known geographic origin

| Tooth | Origin | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|---|
| Boys | | | | | | | | | |
| $I_1$ | IV | | | | | | 1.0814 | 0.3112 | −1.0875 |
| | IR | | | 0.657 | −0.9643 | −1.3233 | −1.2024 | −0.9856 | 0.6701 |
| | FR | | | | 2.1764 | 2.1868 | 2.1544 | 1.1686 | −0.418 |
| $I_2$ | IV | | | | | 0.7040 | 1.2414 | 0.1079 | −1.1017 |
| | IR | | | | −0.9643 | −1.3344 | −1.2711 | −0.8179 | 0.8038 |
| | FR | | | | 2.1719 | 2.1904 | 1.8006 | 0.7709 | −0.688 |
| C | IV | | | | 1.1885 | 0.8519 | −0.4590 | −0.8810 | −1.8359 |
| | IR | | | | −1.1438 | −1.2207 | −0.3823 | 1.0305 | 1.1939 |
| | FR | | | 2.2097 | 2.1551 | 1.4743 | 0.7967 | −0.602 | −0.935 |
| $P_3$ | IV | | | | 1.3418 | 0.4940 | −0.5843 | −0.8813 | −1.8557 |
| | IR | | | −0.9643 | −1.217 | −1.1446 | −0.0973 | 1.0199 | 1.1793 |
| | FR | | | 2.17577 | 2.10661 | 1.2235 | 0.67314 | −0.6053 | −0.9438 |
| $P_4$ | IV | | | 0.3846 | 1.3302 | 0.2935 | −0.6696 | −1.1292 | −1.4574 |
| | IR | | −1.295 | −0.9643 | −1.2067 | −0.8367 | 0.1598 | 1.1438 | 1.1893 |
| | FR | 2.2432 | 2.1441 | 2.1424 | 1.7101 | 0.9369 | 0.0869 | −0.802 | −0.9282 |
| $M_1$ | IV | | −0.6688 | | | 0.4887 | 0.6761 | 0.4590 | −1.1617 |
| | IR | | | | | −0.9643 | −1.295 | −0.9933 | 0.7235 |
| | FR | | | | | 2.1820 | 2.0246 | 0.8289 | −0.7583 |
| $M_2$ | IV | | 0.5800 | | 0.9381 | −0.1559 | −0.9542 | −1.0869 | −1.5360 |
| | IR | | −0.9643 | −1.3187 | −1.0898 | −0.6014 | 0.3213 | 0.9884 | 1.1844 |
| | FR | 2.2367 | 2.2408 | 2.1551 | 1.1871 | 0.5465 | −0.4913 | −0.8647 | −0.8991 |
| Girls | | | | | | | | | |
| $I_1$ | IV | | | | | | 0.8578 | 0.4785 | −0.7101 |
| | IR | | | | 1.2772 | 1.5081 | 1.3732 | 1.2647 | −0.6424 |
| | FR | | | | | 2.7017 | 2.3460 | 1.6949 | −0.3154 |
| $I_2$ | IV | | | | | 1.0014 | 0.9988 | 0.3014 | −0.8682 |
| | IR | | | | 1.2772 | 1.5081 | 1.3956 | 1.095 | −0.6968 |
| | FR | | | | | 2.7017 | 2.3460 | 1.1575 | −0.5700 |
| C | IV | | | | 1.1263 | 0.7741 | 0.2294 | −0.9963 | −1.8686 |
| | IR | | | 1.2544 | 1.3989 | 1.4379 | 1.0668 | −0.3901 | −0.9314 |
| | FR | | | | 2.5332 | 2.2279 | 1.4596 | −0.1437 | −0.7950 |
| $P_3$ | IV | | | | 1.1093 | 0.6965 | −0.2004 | −1.1650 | −1.7635 |
| | IR | | | 1.2772 | 1.4001 | 1.4178 | 0.8288 | −0.4136 | −0.929 |
| | FR | | | | 2.3325 | 1.8567 | 0.9147 | −0.4546 | −0.8137 |
| $P_4$ | IV | | | 0.7493 | 0.9615 | 0.5325 | −0.3844 | −1.3832 | −1.6033 |
| | IR | | | 1.2772 | 1.0125 | 1.2916 | 0.2925 | −0.7252 | −0.9558 |
| | FR | | | 2.1589 | 1.7517 | 1.3651 | 0.2811 | −0.6767 | −0.8232 |
| $M_1$ | IV | | | | | | 0.7678 | 0.5977 | −1.0410 |
| | IR | | | | | 1.2772 | 0.623 | 1.2181 | −0.596 |
| | FR | | | | | 2.7017 | 2.2818 | 1.1395 | −0.6102 |
| $M_2$ | IV | | 0.8206 | | 0.7714 | −0.0611 | −0.5159 | −1.4903 | −1.5810 |
| | IR | | | 1.2544 | 1.2222 | 0.8384 | −0.2405 | −0.8246 | −0.9414 |
| | FR | | 2.7017 | 2.1155 | 1.4393 | 0.8561 | −0.3140 | −0.7613 | −0.7772 |

*IV* Ivory Coast.
*IR* Iran.
*FR* France.

*The Bayesian method*

The basic concept of Bayesian statistics is very straightforward in dental age prediction. For an example of adult age estimation from morphological changes in the structure of teeth see Lucy et al. (1996). Using this concept (see Vieland 1998 for a review), we calculated the posterior probability, or conditional probability, of an event "*Ai*" after taking into account both prior information (not derived from the sample but corresponding to a prior knowledge)

and observed evidence from the sample, noted event "*B*". The posterior probability is noted "*P (Ai|B)*", and is calculated by the Bayes' theorem, a simple mathematical formula (Phillips 1973; Hartigan 1983) given as follows (Eq. 1), and whose significance was first appreciated by Bayes (1764):

$$\frac{P(Ai)\times P(B|Ai)P}{P(A_1)\times P(B|A_1)+P(A_2)\times P(B|A_2)+\ldots+P(A_k)\times P(B|A_k)} \quad \text{for } i = 1, 2, \ldots \text{ or } k \tag{1}$$

In Bayes' theorem, the posterior probability is a function of the marginal probability (denominator) and the joint probability (numerator). The joint probability is the probability of two events "*A* and *B*" happening together, multiplied by the prior probability of "*A*". The marginal probability is the sum of all possible joint probabilities, in other words the probability of an event "*A*", for all possible values of another event "*B*".

With the example of the calculation of the probability of an individual to belong to an age category "*i*" (noted event "*Ai*") from the radiographic stages of its permanent mandibular teeth (noted event "*B*"), the Bayes' theorem (Eq. 1) can be expressed as follows (Eq. 2):

$$\frac{P\left(\frac{\text{Radiographic stages}}{\text{Age category i}}\right)\times P_{\text{prior}}(\text{Age category i})}{\text{Sum of}\left[p\left(\frac{\text{Radiographic stages}}{\text{Age category i}}\right)\times P_{\text{prior}}(\text{Age category i})\right]} \tag{2}$$

across all age categories (for *i*=1, 2, ... or *k*).

Crucial to this approach is the selection of appropriate prior probabilities ("priors") for each age category, and here there are several options. As samples rarely exhibit similar age distributions, because they are usually constructed by "availability" sampling, we will not use them as priors. Another option is to select priors based on demographic sources. However, such data, are hardly comparable between populations. Moreover, our goal was individual age assessment rather than the estimation of population age structure in large series (Konigsberg and Frankenberg 1992). In the absence of any specific knowledge, we assumed the unbiased, and uniform, frequency distribution of our 25 chronological age categories. Each individual has the same prior probability of belonging to 1 of the 25 chronological age categories (i.e., 1/25). This is a so-called uninformative prior (Konigsberg and Frankenberg 1992). In this study, uninformative priors are appropriate in the calculation of the posterior probabilities because the default presumption that individuals are derived from approximately balanced age categories allows direct comparisons of results from different comparative samples (in terms of size, composition and number of chronological age categories). In the Bayes' theorem (Eq. 2), the uniform priors appear in both numerator and denominator. Consequently, by factorisation, the uniform priors have no influence on the calculation of the posterior probabilities.

The term p (Radiographic stages|Age category i) represents, for an age category "*i*", the observed proportion (or percentage) of individuals evincing in our sample, the sequence of radiographic stages under study (i.e., the radiographic stages seen in the individual for who dental age is estimated). Importantly, the radiographic stages, representing age changes in the developing permanent teeth, correspond to a sequence of events (or pattern) even if the true nature of the relationships, or associations, existing in this dataset is complex. Indeed, the developing permanent teeth do not give totally independent information about age. We know that the radiographic stages seen in an individual change systematically with age even if, so far, it is still not clear how a radiographic change in one tooth can directly cause change in another tooth. Therefore, as we would not expect partial correlations between the age changes to approach zero when controlled for age, the radiographic stages are considered as dependent variables. We acknowledge that taking radiographic stages dependencies into consideration, and therefore considering a dental mineralisation sequence (DMS) rather than seven independent radiographic stages, represents an additional difficulty. Indeed, as we consider seven permanent teeth, the Bayesian calculations require a important number of observations because each combination of the variables (DMS) should be represented in the data set. If not, i.e. when a DMS represented in the test sample is not found in the training sample (called single DMS), no Bayesian prediction is given (no "response"). Even if it seems logical, we will evaluate the influence of sample size on this proportion of absence of prediction (noted "% non-response", which is directly proportional to the proportion of single DMS). We will also evaluate the influence of sample size on the quality of the method. Another option would be the so-called naïve Bayes or simple Bayes classification, which circumvents the dependency problem. In this case, the teeth and their radiographic stages are assumed to be independent of one another. It is made to simplify the computations and in this sense, it is considered to be naïve. Even if this assumption seems to be widely used in age estimation (e.g., Lucy et al. 1996), it is unrealistic in the case of dental mineralisation, and become increasingly (exponentially) not tenable when the number of teeth scored (called "attributes") multiplied by the number radiographic stages (the values of each "attribute") increase (Domingos and Pazzani 1997). Moreover, more comparisons need to be done to understand the data characteristics which may affect the performance of naïve Bayes estimations. Therefore, at this stage, we do not make the assumption of independence. Instead, as the sizes of our samples are still limited, we use the jackknife resampling technique (Tukey 1977) when assessing the performance of the Bayesian method (the second objective of this study, see introduction section).

Let us now consider the example of a girl from Iran who is 86 months of age. Her left mandibular central (I1) and lateral (I2) incisors, canine (C), first (P3) and second (P4) premolars, first (M1) and second (M2) molars are rated at the following radiographic stages: 6; 6; 6; 5; 5; 7; 4, respectively. This rating represents the conditional "event

B" of Eq. 1 and corresponds to the DMS that can be noted as follows: I1=6 & I2=6 & C=6 & P3=5 & P4=5 & M1=7 & M2=4. Applying Bayes' theorem (Eq. 1) and Eq. 2, the posterior (or conditional) probability for this girl belonging to 1 out of the 25 chronological age categories (denoted "*i*") defined in this study, is given by Eq. 3:

$$\frac{\begin{array}{l}\mathrm{P}(\text{Age category i}/\mathrm{I1}=6\&12=6\&\mathrm{C}=6\&\mathrm{P3}=5\&\mathrm{P4}=5\&\mathrm{M1}=7\&\mathrm{M2}=4)= \\ P(\mathrm{I1}=6\&\mathrm{I2}=6\&\mathrm{C}=6\&\mathrm{P3}=5\&\mathrm{P4}=5\&\mathrm{M1}=7\&\mathrm{M2}=4/\text{Age category i})\times 1/25\end{array}}{\text{Sum of }[\mathrm{p}(\mathrm{I1}=6\&\mathrm{I2}=6\&\mathrm{C}=6\&\mathrm{P3}=5\&\mathrm{P4}=5\&\mathrm{M1}=7\&\mathrm{M2}=4/\text{Age category i})\times 1/25]} \quad (3)$$

across all 25 age categories (for *i*=1, 2, ... or 25) where 1/25 is the prior probability of any individual belonging to 1 out of the 25 chronological age categories (see methods section).

A total of 25 posterior probabilities, each corresponding to 1 out of our 25 chronological age categories, will then be computed, to lead to a distribution of posterior probabilities. In our example, p (=90–96| I1=6& I2=6& C=6& P3=5 & P4=5 & M1=7 & M2=4) is the posterior probability for the girl from Iran belonging to the =90–96< age category, knowing that her I1, I2, C, P3, P4, M1, M2 are at radiographic stages 6; 6; 6; 5; 5; 7 and 4, respectively. In any training sample of girls and its =90–96< age category, we count the number of individuals with I1=6 & I2=6 & C=6 & P3=5 & P4=5 & M1=7 & M2=4. Then we scale this number by the total number of girls in the =90–96< age category. This scaled value is the relative frequency or initial probability estimate (or occurrence), denoted "P (I1=6 & I2=6 & C=6 & P3=5 & P4=5 & M1=7 & M2=4 | Age category i )" in Eq. 3. This initial estimate is not a true probability because of finite sample size. This initial estimate is then multiplied by 1/25, the prior probability of the girl belonging to 1 of the =90–96< chronological age category (see methods section). The result of this multiplication is the numerator of Eq. 3 and corresponds to the joint probability for the =90–96< age category. The denominator, or marginal probability, corresponds to the sum of all 25 joint probabilities, across all 25 age categories. Each of these 25 joint probabilities is calculated, as for the numerator of Eq. 3, by multiplying the initial probability estimate, or relative frequency of individuals with I1=6 & I2=6 & C=6 & P3=5 & P4=5 & M1=7 & M2=4, observed in each age category, and the prior probability of 1/25 corresponding to an unbiased frequency distribution of age categories. Using Bayes' theorem and Eq. 3, we obtain the posterior probability for the girl from Iran of belonging to the =90–96< age category in the training sample, knowing that her I1, I2, C, P3, P4, M1, M2 are at radiographic stages 6; 6; 6; 5; 5; 7 and 4, respectively, by dividing the joint probability for this age category by the marginal probability. In our example, no posterior probability can be computed because the DMS is single (not found in the training samples). When we consider another 86-month-old girl from Iran with a slightly different DMS (7,7,6,5,5,7,4) and the training sample of girls from France, a posterior probability different from zero is obtained for only 2 out of the 25 chronological age categories representing a 6-month interval. This girl from Iran, in our example, falls in the =90–96< and =96–102< months age categories with posterior probabilities (reliabilities) of 0.412 and 0.588, respectively. This is a posterior probability distribution and the higher probability is obtained for the =96–102< age category. In comparison to the chronological age, this prediction is given the rank −2, so that the results of Bayesian prediction can be compared with the CAR data (see methods section). The rank −1 is given to the =90–96< months age category. When summing the posterior probabilities obtained for the ranks −2 and −1 (the =90–96< and =96–102< months age categories), a reliability of 100% is obtained. Therefore, the accuracy of this prediction is −15 to −3 months (classification ranging from −2 to −1) and its reliability is 100%.

Our table of raw data has *n* rows, each for an individual is set. The first column is the known age category of the individual, seven more columns follow where each represent a particular tooth. The entry for a tooth is the radiographic stage rated between 1 and 8. In order to automate the management of this database and, more specifically, to determine the relative frequencies of dental mineralisation sequences seen in individuals, within each chronological age category, we created a Microsoft Excel 2000 Visual Basic macro with the help of Dr. F. Houët (engineer, Laboratoire d'Anthropologie, Université Bordeaux 1). In our Microsoft Excel 2000 Visual Basic macro, using the relative frequencies of combinations of radiographic stages and applying the Bayes' theorem, we computed the marginal, the joint, and the posterior probabilities. Our Microsoft Excel 2000 Visual Basic macro is available upon request to JB or YH.

## Results

The "geographic-specific" conversion tables obtained for our three samples are given in Table 5. Right away, we note that among the youngest age classes, there are insufficient numbers to provide good estimates of maturity indices and, therefore, estimates of dental age. The estimated dental age can be calculated by using the linear regression equations given in Table 6. We acknowledge that most of the regression coefficients are negative contrary to the one obtained for the subsample of boys from Iran. This might be due to the lack of the youngest individuals, in this latter subsample, for which the sign of the weighted scores is opposite, for the earliest dental stages (Table 5). At first glance, we observe that the SEE ranges from 13.6 to 18.4 months (Table 6). In order to compare the quality levels obtained between the two methods of calculation, 6 trials using non-geographic-specific standards, for each method

**Table 6** CAR method: regression equations and standard errors of the estimates of the linear regressions between the chronological age and the GMI (global maturity index) given by the multiple correspondence analysis

| Sub-samples | Linear regression equations | Standard error of the estimates |
|---|---|---|
| Girls from Ivory Coast | Chronological age=109.15 –4.950×GMI | 13.56 (R2=0.740; $p$<0.005) |
| Boys from Ivory Coast | Chronological age=108.60 –4.685×GMI | 17.91 (R2=0.610; $p$<0.005) |
| Girls from Iran | Chronological age=136.37 –4.982×GMI | 18.44 (R2=0.699; $p$<0.005) |
| Boys from Iran | Chronological age=128.63 +4.4498×GMI | 15.16 (R2=0.716; $p$<0.005) |
| Girls from France | Chronological age=132.13 –3.833 ×GMI | 14.13 (R2=0.676; $p$=0.000) |
| Boys from France | Chronological age=131.76 –4.432×GMI | 14.16 (R2=0.761; $p$=0.000) |

*R2* Coefficient of determination.
*p* Significance level

and for each sex (24 trials and 12 comparisons in total), have been done (Table 2). In 2 comparisons out of 12 (training sample data for girls from Iran used to test the age of girls from France and training sample data for boys from Iran used to test the age of boys from France), we observed a significant difference in the performance levels of the Bayesian predictions. This increase ranged from 7.7 to 12.7%. In Fig. 1, we illustrate the comparison of trials with the training sample data for girls from Iran used to test the age of girls from France (A is for the CAR method and B is for the Bayesian predictions). In 6 comparisons out of 12, we observed a significant difference at the 95% reliability threshold. This means that a lower standard error was observed in one of the two methods and is associated with a significantly higher reliability (necessarily higher than 95%): in four comparisons the standard error is lower with the CAR method; in two comparisons the standard error is lower with the Bayesian method. Both global reliability and accuracy levels are different in 2 comparisons out of 12. A skewness and kurtosis closer to zero (corresponding to better global accuracy and reliability levels) was found for Bayesian predictions in one comparison and for the CAR method in the other.

In order to compare the quality levels obtained between trials using either geographic-specific (jackknife) or non-geographic-specific standards, six comparisons have been done for each sex (Table 3). In 4 comparisons out of 12 we observed a significant difference in the performance levels with an increase in the case of the use of non-geographic-specific standards (tests of sample data representing girls from Ivory Coast and Iran). This increase ranged from 9.6% to 15.6%. In 4 comparisons out of 12 we observed a lower standard error associated with a significantly higher reliability (necessarily higher than 95%): in 3 comparisons the standard error was lower using geographic-specific standards; in 1 comparison the standard error was lower using non-geographic-specific standards. Both global reliability and accuracy levels were different in 6 comparisons out of 12, in 5 of these comparisons, a skewness and kurtosis closer to zero was found using non-geographic-specific standards.

In order to evaluate the global quality of Bayesian predictions and their possible undermining due to high proportions of single DMS (see methods section), we compared the Bayesian predictions using a larger dataset (see methods section) with our trials using geographic-specific standards. Three comparisons have been done for each sex (Table 4). In three out of six comparisons we observed a significant difference in the performance levels with an increase in the case of the use of a larger sample. This increase ranged from 8.3 to 11.6%. In two out of six comparisons, we observed a lower standard error associated with a significantly higher reliability, only in the case of the use of a larger sample. Moreover, in three out of six comparisons, both global reliability and accuracy levels were better (skewness and kurtosis closer to zero) in the case of the use of a larger sample.

## Discussion

The first issue examined in this paper is the use of methods solely based on geographic-specific standards in order to increase the global quality of age assessment. Our results indicate that in the case of Bayesian predictions, the use of geographic-specific standards does not guarantee a better quality. In fact, the quality of age assessment does not seem to depend predominantly on the use of geographic-specific standards. When we evaluate quality differences between trials using geographic-specific standards and non-geographic-specific standards, from multiple comparisons we observed a clear trend in favour of higher accuracy and reliability levels when using non-geographic-specific standards (Tables 3 and 4). Moreover, from the samples available in this study, we observed that the method of calculation is also an important source of variation, as compared to other possible factors. Indeed, when evaluating quality for age assessment using the same training samples but different methods of calculation (Table 2), variations in quality can be as high as those obtained when using different training samples but identical methods (Table 3). However, besides methodology, differences in quality between trials should be explained and other possible sources of variation should be explored. Among these possible sources of variation, the age distribution of the sample has been evoked in the literature. A higher accuracy and reliability of estimation is predictable in early childhood than in older children and adolescents because (i) the variability of tooth formation increases with chronological age, and (ii) more indicators of short duration (more developing deciduous and permanent teeth) are available in the former group. As a consequence, we should find a better quality with the trials using training samples with higher proportions of young children and, on the contrary, with lower proportions of adolescents. In the samples of girls, the numbers of children below the age

of 7 years (or 84 months) were 5 (2.0%), 11 (4.3%) and 21 (13.6%) for France, Iran and The Ivory Coast, respectively (Table 1). Conversely, the proportions of "adolescents" (above 13 years old, i.e. 156 months) were 39 (16.1%), 75 (29.2%) and 12 (7.8%), respectively (Table 1). These absolute and relative frequencies may well explain why, in the case of Bayesian predictions, when comparing trials using geographic-specific and non-geographic-specific standards, a lower standard error was found in the trial (with the jackknife method) using the sub-sample of girls from The Ivory Coast (Table 3). However, we cannot explain from these proportions only, why the performance was better for the trial (jackknife) using the sub-sample of girls from France (Table 3) or why the performance was even better when using a training sample of girls from Iran, rather than a training sample from The Ivory Coast (Table 3). Moreover, in order to assess the possible influence of age distributions on the quality of age assessment, we can examine and compare graphically the rank distributions within and between age classes corresponding to age categories grouped together and representing 2 years, as shown in Fig. 1. Figure 1A corresponds to the quality of the trial with the training sample data for girls from Iran used to test the age of girls from France, with the CAR method. Figure 1B provides the same results when using Bayesian predictions. When comparing the rank distributions between the 2-years age classes, important differences are noticeable. For example, in the case of the CAR method, the rank 0 is represented only in the older age classes (E and F) (Fig. 1A). Moreover, when comparing the rank distributions across age categories, between the two methods, an important difference arises. Almost systematically, in comparison to the CAR results, in Bayesian predictions, a wider distribution of ranks (a larger number of ranks) is represented within any defined age category. This phenomenon, corresponding to an age apparent dependent systematic error, reduces an important bias in age assessment and has already been identified in regression analysis: estimated ages are too old for young individuals and too young for old individuals. A correction called "classical calibration" has been proposed to solve the problem but it appears to be "less efficient" in terms of accuracy (Aykroyd et al. 1997). The age-dependent systematic error occurring when using regression analysis is illustrated in the CAR results (Fig. 1A). On the contrary, in our Bayesian predictions (with no independence assumption) (Fig. 1B), a better accuracy does not depend on age. However, the Bayesian method also raises some difficulties. The low reliability levels seen in some trials (Tables 2 and 3) may be due, at least in part, to a small size and, consequently (i) to a low variability in terms of sequences and timing of dental mineralisation within each age category and (ii) to a high proportion of single DMS.

The second issue examined in this paper is the use of an alternative dental age estimation method over the CAR approach. The Bayesian method represents this alternative but clear information about the quality level of trials should be provided. The Bayesian method appears to provide different results with, in some cases (not systematically), an overall significant increase in performance (the highest accuracy associated with a better reliability). This difference with the CAR method has been noticed when using rather "small" training samples and we know the limitation of Bayesian predictions due to the use of dependent attributes (see methods sections) rather than the independence assumption. In order to evaluate the global quality (i.e., for any or most data sets) of the Bayesian method, but also to investigate the effect of the proportion of single DMS, we also used a larger data set by combining in one training sample for each sex, our three geographic population samples and our additional sample from France (Table 4). The proportion of single DMS is then reduced to 12.8% and 14.6% for girls and boys, respectively. In both sexes, the standard error associated with the 95% reliability threshold is 19.5 months (Table 4). This value seems quite high in comparison to the standard errors published by Ritz-Timme et al. (2000). However, it is generally lower than most of the values obtained in our trials using different training samples (Table 3) or methods (Table 2). Moreover, even if further trials are necessary (using training samples with lower proportions of single DMS), we note that in comparison to a maximum likelihood method of estimation, the Bayesian method offers the possibility to assess age with both high levels of accuracy and reliability. If not, the precision can be reduced (by classifying into larger age classes), to meet the accuracy and reliability needed. Doing so, we increase the numbers of children in each of the new age classes (which number will be reduced). This constitutes a more flexible way to deal with dental age estimation. Moreover, results illustrated in Fig. 2 confirm our observation that in the case of Bayesian predictions, and contrary to age estimation techniques based on regression, a better accuracy does not depend on age. For example, in the Bayesian predictions on our total samples of girls (Fig. 2A), there is always a fraction of the sample classified into rank 0 whatever the age class considered. This proportion ranges from 11.9% for the age class D to 40.0% for the age class A. This has never been observed in our trials using the CAR method. Importantly, to explain such differences, we should again note that the use of uninformative priors in the Bayesian method (see methods section) corresponds to the default presumption that individuals are derived from approximately balanced age categories and, therefore, allows direct comparisons of results from different comparative samples in terms of age distribution. Moreover, the absence of the independency assumption in our Bayesian predictions may also explain the absence of the age dependent systematic error. Indeed, the variability across age classes is more likely higher for a given radiographic stage observed on an isolated tooth than for a given mineralisation sequence from which this isolated tooth is derived. Individuals will be distributed in a larger age range when considering a single tooth with its radiographic stage than when considering the same tooth within a given DMS. Globally, the age range will decrease when the number of teeth represented in the DMS will increase. As a consequence, in the case of a dependent approach, within any defined age groups or classes, age estimation will be based on less variability of the criterion (DMS) used to assign age.

## Conclusion

Our results indicate that in the case of non-adult Bayesian age predictions using dental mineralisation sequences of permanent teeth (excluding the third molar), geographic-specific standards do not guarantee better quality levels. In fact, the quality of Bayesian age assessment does not seem to depend predominantly on the use of geographic-specific standards. Indeed, we observe a clear trend in favour of higher accuracy and reliability levels when using non-geographic-specific standards. Moreover, from the samples available in this study, we observe that the method of calculation is also an important source of variation, as compared to other possible factors.

After multiple trials using four cross-sectional subsamples from Europe, Africa and Asia (a total of 902 girls and 626 boys), the Bayesian method appears to offer new possibilities in predictions of non-adult dental age. One of the main advantage over maximum likelihood methods of estimation is an overall increase in accuracy with high levels of reliability on a fraction of the test sample and, importantly, across all age categories (contrary to methods based on regression analysis). This is of importance because there is an increasing need for accurate methods for age estimation in forensic practise (Ritz-Timme et al. 2000). In Bayesian predictions, depending on the needs and on the problem to be solved, accuracy and reliability levels can be chosen and discussed with some flexibility. One of the purposes of this study is to raise awareness of a Bayesian approach in non-adult dental age assessment. We do not propose to systematically replace classical statistics with Bayesian methods. We simply emphasise that the latter can be particularly useful because, in the case of multiple attributes (teeth) with a large number of scoring values (radiographic stages), the independence assumption (often used in "simple" or "naïve" Bayesian prediction) of the regression analysis (CAR method) is violated and leads to large errors which still need to be evaluated. The use of Bayesian predictions with no independence assumption is therefore recommended for further studies aiming to explore possible sources of variation in non-adult dental age assessment.

**Acknowledgements** The authors would like to thank Prof. G. Dasgupta, from the Department Civil Engineering and Engineering Mechanics, Columbia University in New York; Prof. B. Brinkmann and anonymous reviewers, whose suggestions greatly improved the manuscript. We thank Dr Frapier (Montpellier), Dr De Brondeau (Bordeaux) and the Services d'Odontologie des CHU de Bordeaux et de Montpellier, for access to the orthopantomographs. We also gratefully acknowledge the important support of the Ministère de la Recherche and the Centre National de la Recherche Scientifique for funding and the expertise of Dr F. Houët of the Laboratoire d'Anthropologie, Université Bordeaux 1.

## References

Anderson DL, Anderson GW, Popovich F (1976) Age of attainment of mineralization stages of the permanent dentition. J Forensic Sci 21:191–200

Aykroyd RG, Lucy D, Pollard AM, Solheim T (1997) Technical note: regression analysis in adult age estimation. Am J Phys Anthropol 104:259–265

Bayes T (1764) An essay toward solving a problem in the doctrine of chances. Philos Trans R Soc Lond 53:370–418

Bhargava A (2000) Modeling the effects of maternal nutritional status and socioeconomic variables on the anthropometric and psychological indicators of Kenyan infants from age 0–6 months. Am J Phys Anthropol 111:89–104

Davis PJ, Hägg U (1994) The accuracy and precision of the "Demirjian system" when used for age determination in Chinese children. Swedish Dent J 18:113–116

Demirjian A (1986) Dentition. In: Falkner F, Tanner JM (eds) Human growth. A comprehensive treatise, vol. II. Plenum, New York, pp 269–298

Demirjian A, Goldstein H (1976) New systems for dental maturity based on seven and four teeth. Ann Hum Biol 3:411–421

Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. Hum Biol 45:211–227

Deutsch D, Pe'er E, Gedalia I (1984) Changes in size, morphology and weight of human anterior teeth during the fetal period. Growth 48 :74–85

Domingos P, Pazzani M (1997) On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning 29: 103–130

Fanning EA (1961) A longitudinal study of tooth formation and root resorption. N Z Dent J 57:202–217

Fanning EA, Brown T (1971) Primary and permanent tooth development. Aust Dent J 16:41–43

Garn SM, Lewis AB, Koski PK, Polacheck DL (1958) The sex difference in tooth calcification. J Dent Res 37:561

Garn SM, Lewis AB, Polacheck DL (1959) Variability of tooth formation. J Dent Res 38:135

Garn SM, Lewis AB, Blizzard RM (1965a) Endocrine factors in dental development. J Dent Res 44:243–258

Garn SM, Lewis AB, Kerewsky RS (1965b) Genetic, nutritional and maturational correlates of dental development. J Dent Res 44: 228–242

Giles E, Klepinger LL (1988) Confidence intervals for estimates based on linear regression in forensic anthropology. J Forensic Sci 33:1218–1222

Gleiser I, Hunt EE (1955) The permanent mandibular first molar: its calcification, eruption and decay. Am J Phys Anthropol 13:253

Greenacre MJ (1993) Correspondence analysis in practice. Academic Press, London

Gulati A, Taneja JR, Chopra SL, Madan S (1990) Interrelationship between dental, skeletal and chronological ages in well-nourished and mal-nourished children. J Indian Soc Pedo Prev Dent 8:19–23

Gustafson G, Koch G (1974) Age estimation up to 16 years of age based on dental development. Odontol Revy 25:297–306

Haataja J (1965) Development of the mandibular permanent teeth of Helsinki children. Proc Finnish Dent Soc 61:43–53

Haavikko K (1970) The formation and the alveolar and clinical eruption of the permanent teeth. An orthopantographic study. Proc Finnish Dent Soc 66:101–170

Haavikko K (1974) Tooth formation age estimated on a few selected teeth: a simple method for clinical use. Proc Finnish Dent Soc 70:15–19

Hägg U, Matsson L (1985) Dental maturity as an indicator of chronological age: the accuracy and precision of three methods. Eur J Orthod 7:24–34

Hartigan JA (1983) Bayes theory. Springer, New York

Healy MJR, Goldstein H (1976) An approach to the scaling of categorized attributes. Biometrika 63:219–229

Konigsberg LW, Frankenberg SR (1992) Estimation of age structure in anthropological demography. Am J Phys Anthropol 89:235–256

Koshy S, Tandon S (1998) Dental age assessment: the applicability of Demirjian's method in south Indian children. Forensic Sci Int 94:73–85

Liliequist B, Lundberg M (1971) Skeletal and tooth development. A methodologic investigation. Acta Radiol 11:97–112

Liversidge HM (1994) Accuracy of age estimation from developing teeth of a population of known age (0–5.4 years). Int J Osteoarchaeol 4:37–45

Liversidge HM, Herdeg B, Rösing FW (1998) Dental age estimation of non-adults. A review of methods and principles. In: Alt WK, Rösing FW, Teschler-Nicola M (eds) Dental anthropology. Fundamentals, limits, and prospects. Springer, Wien New York, pp 419–442

Lucy D, Aykroyd G, Pollard AM, Solheim T (1996) A Bayesian approach to adult human age estimation from dental observations by Johanson's age changes. J Forensic Soc 41:189–194

Moorees C, Fanning E, Hunt E (1963a) Age variation of formation stages for ten permanent teeth. J Dent Res 42:1490–1502

Moorees C, Fanning E, Hunt E (1963b) Formation and resorption of three deciduous teeth in children. Am J Phys Anthropol 21: 205–213

Nanda RS, Chawla TN (1966) Growth and development of dentitions in Indian children. I. Development of permanent teeth. Am J Orthod 52:837

Nolla CM (1960) The development of the permanent teeth. J Dent Child 27:254–266

Nyström M, Kilpinen E, Kleemola-Kujala E (1977) A radiographic study of the formation of some teeth from 0.5 to 3.0 years of age. Proc Finnish Dent Soc 73:167–172

Nyström M, Haataja J, Kataja M, Evälahti M, Peck L, Kleemola-Kujala E (1986) Dental maturity in Finnish children, estimated from the development of seven permanent mandibular teeth. Acta Odont Scand 44:193–198

Phillips LD (1973) Bayesian statistics for social scientists. Thomas Nelson, London

Proy E, Gauthier N (1985) Maturation dentaire: constitution de tables. Rev Orthop Dento Faciale 19:523–534

Ritz-Timme S, Cattaneo C, Collins MJ, Waite ER, Schütz HW, Kaatsch H-J, Borrman HIM (2000) Age estimation: the state of the art in relation to the specific demands of forensic practice. Int J Legal Med 113:129–136

Scheuer L, Black S (2000) Developmental juvenile osteology. Academic Press, London

Schour I, Massler M (1940) Studies in tooth development: the growth pattern of human teeth. Parts I and II. J Am Dent Assoc 27:1778; 1918

Staaf V, Mornstad H, Welander U (1991) Age estimation based on tooth development: a test of reliability and validity. Scand J Dent Res 99:281–286

Tanner JM (1962) Growth at adolescence. Blackwell, Oxford

Tanner JM, Whitehouse RM, Marshall WA, Healy MJR, Goldstein H (1975) Assessment of skeletal maturity and prediction of adult height: TW2 method. Academic Press, London

Tukey J (1977) Exploratory data analysis. Addison-Wesley, Reading

Vieland V (1998) Bayesian linkage analysis, or: How I learned to stop worrying and love the posterior probability of linkage. Am J Hum Genet 63:947–954

Voors AW (1973) Can dental development be used for assessing age in underdeveloped communities? J Trop Pediatr Environ Child Health 19:242

Wolanski N (1966) A new method for the evaluation of tooth formation. Acta Genet (Basel) 16:186–197