scientific reports

OPEN

Validity of age estimation methods and reproducibility of bone/dental maturity indices for chronological age estimation: a systematic review and meta-analysis of validation studies

V. Marconi[1,6], M. Iommi[2,3,6], C. Monachesi[4], A. Faragalli[2,3], E. Skrami[2,3✉], R. Gesuita[2,3], L. Ferrante[2,3] & F. Carle[2,3,5]

Several approaches have been developed to estimate age, an important aspect of forensics and orthodontics, using different measures and radiological examinations. Here, through meta-analysis, we determined the validity of age estimation methods and reproducibility of bone/dental maturity indices used for age estimation. The PubMed and Google Scholar databases were searched to December 31, 2021 for human cross-sectional studies meeting pre-defined PICOS criteria that simultaneously assessed the reproducibility and validity. Meta-estimates of validity (mean error: estimated age-chronological age) and intra- and inter-observer reproducibility (Cohen's kappa, intraclass correlation coefficient) and their predictive intervals (PI) were calculated using mixed-effect models when heterogeneity was high ($I^2 > 50\%$). The literature search identified 433 studies, and 23 met the inclusion criteria. The mean error meta-estimate (mixed effects model) was 0.08 years (95% CI − 0.12; 0.29) in males and 0.09 (95% CI − 0.12; 0.30) in females. The PI of each method spanned zero; of nine reported estimation methods, Cameriere's had the smallest (− 0.82; 0.47) and Haavikko's the largest (− 7.24; 4.57) PI. The reproducibility meta-estimate (fixed effects model) was 0.98 (95% CI 0.97; 1.00) for intra- and 0.99 (95% CI 0.98; 1.00) for inter-observer agreement. All methods were valid but with different levels of precision. The intra- and inter-observer reproducibility was high and homogeneous across studies.

Age estimation, an important aspect of forensics and orthodontics, is often used when chronological age cannot be determined[1]. Indeed, estimating dental age in children is useful in several situations such as orthodontic treatment planning, forensic dentistry, and other clinical scenarios[2,3]. In living individuals, age estimation is a crucial and increasing forensic practice method due to widespread increases in individuals without identification documents and whose real age must clarified for criminal, civil, asylum, or old-age pension proceedings[4–9]. Age estimation is increasingly requested by judicial authorities to determine if the adult penal law should be applied according to legally-relevant age ranges[10]. Age estimation has also been used in professional sports, where age falsification could provide athletes with significant competitive advantages[11].

Different methods are used to determine age using different measures and radiological examinations[10], with the teeth and the hand-wrist commonly assessed. Teeth are one of the strongest structures in the human body

[1]Postgraduate School of Medical Statistics and Biometry, Department of Biomedical Sciences and Public Health, Università Politecnica delle Marche, 60126 Ancona, Italy. [2]Center of Epidemiology, Biostatistics and Medical Information Technology, Università Politecnica delle Marche, 60126 Ancona, Italy. [3]Department of Biomedical Sciences and Public Health, Università Politecnica delle Marche, 60126 Ancona, Italy. [4]Department of Pediatrics, Università Politecnica delle Marche, 60123 Ancona, Italy. [5]National Centre for Healthcare Research and Pharmacoepidemiology, 20126 Milan, Italy. [6]These authors contributed equally: V. Marconi and M. Iommi. ✉email: e.skrami@staff.univpm.it

and, together with the skeletal system, pass through a series of developmental changes that represent valid indices for age determination[12–16]. Skeletal maturity is based on radiography of specific structures such as the medial clavicular epiphysis cartilage[17–19], pubic symphysis[20], and the left hand-wrist area[10]. However, methods based on skeletal maturity are more variable and susceptible to error than methods based on tooth maturation[21–23]. Dental methods identify the stages of tooth mineralization in radiographs and code them according to predetermined scores[24] or continuous measures[13,25]. The most common method for age estimation was published in 1973 by Demirjian, Goldstein, and Tanner[24] and was subsequently modified by other authors. Demirjian's method is based on eight developmental stages ranging from crown and root formation to apex closure of the seven left permanent mandibular teeth. A score is assigned at each stage and then the sum of the scores represents the subject's dental maturity score (DMS). From this seminal paper, the DMS was used in regression equations to estimate the age of a subject.

Over the years, several different methods have been developed to increase the accuracy of age estimation. Technological developments in radiology have allowed more specific measurements to be made, increasing the accuracy of dental/skeletal maturation indicators[26–29]. There has also been a focus on refining age estimation methods to better predict chronological age[13,25,30].

To consider a method "valid", it is necessary to proceed with its validation. Validation refers to the process of applying the age estimation method to a sample other than the one used to calibrate the method[31]. The sample can be external or a test set obtained by splitting the study sample into training and test sets. To evaluate the method's validity, the distribution of errors between chronological and estimated age are then evaluated on this external sample or test set.

Inter-observer reliability is defined as the agreement between two or more observers, while intra-observer reliability is defined as the agreement of the same evaluator at two or more different time points. Cohen's K statistic is commonly used for reliability assessments of categorical scales, while the intraclass correlation coefficient (ICC) or the concordance correlation coefficient (CCC) statistics are appropriate for continuous scales[32].

Reference studies on forensic age estimation should report sex and ethnicity, two well-known factors associated with individual dental/skeletal maturity[33–35], in addition to chronological age, bone age, the difference between bone age and chronological age, and intra-observer and inter-observer reproducibility[36]. While several literature reviews and meta-analyses have compared different age estimation methods[3,37–41], to our knowledge there has yet to be a first meta-analysis also comparing validation and reproducibility. We aimed to assess the validity of age estimation methods based on bone or dental maturity indices and the reproducibility of these maturity indices, through meta-analysis of validation and reproducibility studies. Therefore, the Review questions are "What is the level of validity of age estimation methods based on bone and dental maturity indices? What is the level of reproducibility of bone and dental maturity indices?"

Results

Study selection. The literature search returned 51 articles from PubMed and 382 from Google Scholar (total 433). After removing duplicates (28 articles), the titles and abstracts were separately screened by two authors (VM and CM) to leave 75 eligible articles. After reading the full text, 59 articles were excluded because 31 articles did not validate the age estimation method; 10 articles focused only on assessing a threshold for 14- or 18-year-old subjects; and 18 articles had incomplete or unusable data.

Sixteen studies were therefore included in the qualitative synthesis, and seven studies, which complied with the inclusion criteria, were also included after further examination of previous meta-analyses or systematic reviews to provide a total of 23 articles (Fig. 1).

Characteristics of included articles. The characteristics of the 23 selected studies are detailed in Table 1. All studies adopted a cross-sectional design and were conducted in both university and hospital settings in 15 countries (Bosnia-Herzegovina[42,43], Brazil[44], China[45–47], Colombia[48], Egypt[49], India[50–52], Iran[53], Italy[54], South Korea[55], Macedonia[56], Malaysia[57,58], Saudi Arabia[59], Spain[60,61], Sri Lanka[62], and Turkey[63,64]). Sample sizes ranged from 70[52] to 2641[61] subjects who underwent orthopantomography (21 articles[42–51,53,55–63]), or wrist and hand X-rays (2 articles[52,54]). The age range was from 1[52] to 24 years[57], and most studies (17 out of 23[42–44,46,48–51,53–56,58–60,63,64]) enrolled subjects aged 16 years or younger.

Nine different age estimation methods were used, with a clear predominance of the Demirjian approach or its modification (16 out of 23[42,45–47,50,53–56,58–64]) and Willems (13 out of 23[43,44,46,47,49,50,53,55–58,60,62]). Other methods were used less or only once (Cameriere, 5 out of 23[43,48,49,53,58]; Haavikko, 4 out of 23[43,50,58,64]; Smith, 1 out of 23[53]; Nolla 7 out 23[50,51,58,60,61,63,64]; Chaillet, 1 out of 23[55]; Blenkin and Evans, 1 out of 23[62]; Greulich and Pyle, 2 out of 23[52,54]).

Sixteen studies provided complete data for both mean errors and examiner agreements, while eight studies report mean errors in age estimation without complete or usable data regarding the intra-or inter-observer agreement. The precision of the estimation methods was highly variable, with a mean error ranging from a maximum precision of −0.02 years using the Cameriere method applied to males[43] to a minimum of −2.96 years using the Haavikko method applied to females[50]. The inter-examiner agreement ranged between 0.73 and 1 for Cohen's k/Fleiss' k and between 0.84 and 1 for ICC; similarly, the intra-examiner agreement ranged between 0.82 and 0.99 for Cohen's k and between 0.80 and 1 for ICC.

Study quality assessment (qualitative synthesis). The risk of bias assessment for the selected studies is presented in Table 2 and illustrated in Fig. 2. All studies accurately described the patient selection procedure except for El Bakary et al.[49] and Javadinejad et al.[53], in which the procedure was not clearly explained, and Franco et al.[44], in which the criteria were not reported, so these studies were classified as "unclear". With respect to
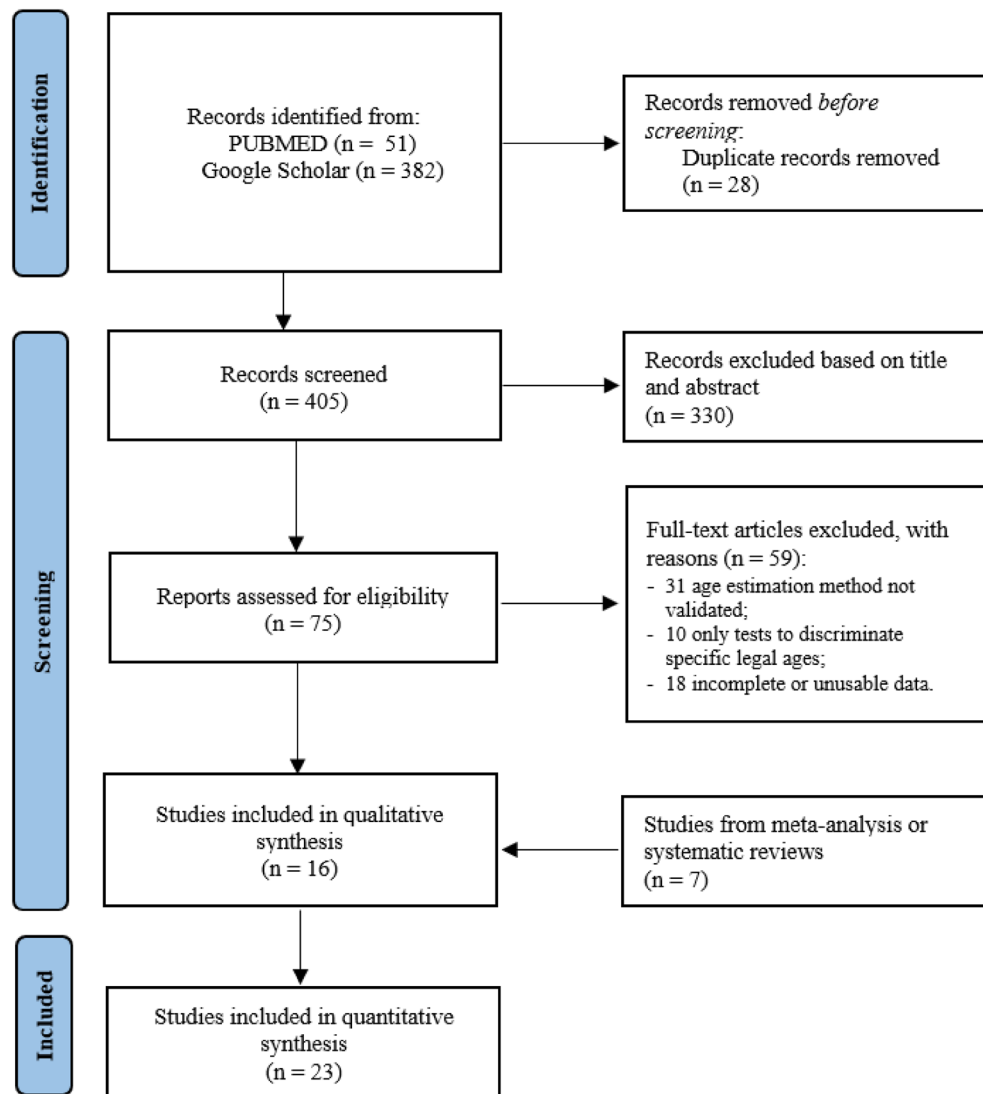
Figure 1. PRISMA flow diagram of the search results from the databases.

the index text, we considered any study that clearly expressed the method of analysis of the radiographs or the experience or number of observers making the measurements as "low" risk. Three studies[49,53,59] did not provide enough information, while another study was not completely specific[63]. Four studies[44,57,59,63] did not report how the chronological age was assessed (the reference standard in Fig. 2), and this was interpreted as a risk of bias since a person could be confused or lie about his age. All studies provided good information on flow and timing.

Despite the possibility of bias, no study had applicability concerns. All articles met the minimum criterion of regularity in the procedures, as defined by the PICOS/PECOS strategy[66], and therefore were included in the analysis.

Meta-analysis of age estimation validity. Since we found only two studies based on bone maturation indices, we did not produce a meta-estimate of the mean error. Concerning the age estimation validity based on dental maturation indices, significant heterogeneity was found for both males and females (males: $I^2 = 99.6\%$ [95% CI 99.6%; 99.7%]; $\tau^2 = 0.54$ [95% CI 0.38; 0.86]; females: $I^2 = 99.6\%$ [95% CI 99.5%; 99.6%]; $\tau^2 = 0.56$ [95% CI 0.38; 0.88]) due to the large sample size and the precision of the included studies. As a result, a mixed-effects model was applied to calculate the pooled mean error of age estimation by sex. The pooled male mean error of the age prediction was 0.08 years (95% CI − 0.12; 0.29), and the pooled female mean error was 0.09 years (95% CI − 0.12; 0.30). Figure 3 shows the stratification by age estimation methods, which are also summarized in Supplementary Methods 1.

Studies that implemented Nolla's method had a mean error closest to zero with a slight overestimation: mean male age prediction error of 0.02 (95% CI − 0.37; 0.41) and mean female age prediction error of 0.03 (95% CI − 0.34; 0.41). Haavikko's method was a less accurate method, with a mean error of − 1.12 (95% CI − 2.29; 0.06) and − 1.33 (95% CI − 2.54; − 0.13) for males and females, respectively. Cameriere's method also underestimated

Author (year)[reference]	Study site	Sample size (male/female)	Age range (years)	Age estimation method	Total mean error (SD)	Male mean error (SD)	Female mean error (SD)	Inter-examiner agreement estimate (variance§)	Intra-examiner agreement estimate (variance§)
Ambarkova et al. (2013)[56]	Macedonia	N = 966 (481/485)	6.05–13.96	Demirjian	1.07 (0.96)	1.02 (1.02)	1.12 (0.9)	ICC = 0.89 (0.010)	ICC = 0.97 (0.001)
				Willems	0.42 (0.86)	0.52 (0.87)	0.33 (0.83)	ICC = 0.94 (0.003)	ICC = 0.97 (0.001)
Baghdadi et al. (2012)[59]	Saudi Arabia	N = 176 (91/85)	4–14	Demirjian		0.31 (0.93)	0.56 (0.81)	Inappropriate estimation method	
El-Bakary et al. (2010)[49]	Egypt	N = 286 (134/152)	5–16	Willems	0.15 (0.62)	0.29 (0.48)	0.14 (0.74)	Inappropriate estimation method	
				Cameriere	− 0.29 (1.04)	− 0.49 (1.03)	− 0.26 (1.21)		
Franco et al. (2013)[44]	Brazil	N = 462 (205/257)	5–16	Willems		0.24 (0.97)	0.04 (0.97)	Cohen's k = 0.93	Cohen's k = 0.9
Galić et al. (2010)[42]	Bosnia-Herzegovina	N = 1106 (509/597)	5–14	Demirjian		1.46 (1.26)	1.27 (1.27)		Cohen's k = 0.82
Galić et al. (2011)[43]	Bosnia-Herzegovina	N = 1089 (498/591)	6–13	Cameriere		− 0.02 (0.71)	0.10 (0.71)	Cohen's k = 1	Cohen's k = 0.97
				Haavikko		− 0.09 (0.79)	− 0.23 (0.73)	Cohen's k = 0.85	Cohen's k = 0.98
				Willems		0.42 (0.77)	0.25 (0.89)	Cohen's k = 0.81	Cohen's k = 0.97
Javadinejad et al. (2015)[53]	Iran	N = 537 (264/273)	3.9–14.5	Demirjian	0.87 (1.00)	0.90 (1.01)	0.85 (0.98)		
				Willems	0.36 (0.87)	0.43 (0.82)	0.31 (0.91)		Cohen's k = 0.96
				Cameriere	− 0.19 (0.86)	− 0.27 (0.85)	− 0.11 (0.87)		
				Smith	0.06 (0.63)	0.12 (0.83)	0.00 (0.81)		
Jayaraman et al. (2012)[45]	China	N = 266 (133/133)	2–21	Demirjian		− 0.25 (1.43)	− 0.23 (1.37)		Cohen's k = 0.88
Kırzıoğlu & Ceyhan (2012)[64]	Turkey	N = 425 (212/213)	7–13	Nolla	− 0.54 (0.93)	− 0.53 (0.95)	− 0.57 (0.91)	ICC = 0.98 (0.008)	ICC = 0.95 (0.002)
				Haavikko	− 0.58 (0.80)	− 0.60 (0.80)	− 0.56 (0.81)		
				Demirjian	0.64 (0.89)	0.52 (0.86)	0.75 (0.90)		
Kumaresan et al. (2014)[58]	Malaysia	N = 426 (179/247)	5–16	Demirjian	− 0.97 (1.19)	− 0.98 (1.29)	− 0.97 (1.12)	Inappropriate estimation method	Inappropriate estimation method
				Willems	− 0.54 (1.28)	− 0.55 (1.40)	− 0.53 (1.20)		
				Nolla	− 0.54 (1.31)	− 0.50 (1.31)	− 0.57 (1.31)		
				Haavikko	1.31 (1.1)	0.94 (1.03)	1.59 (1.08)		
				Cameriere	0.41 (1.08)	0.44 (1.14)	0.39 (1.03)		
Lee et al. (2011)[55]	Korea	N = 1483 (754/729)	5–16	Demirjian	0.30 (0.81)	0.29 (0.75)	0.31 (0.87)	Inappropriate estimation method	Inappropriate estimation method
				Willems	− 0.17 (0.65)	− 0.15 (0.58)	− 0.19 (0.72)		
				Chaillet	− 0.35 (0.68)	− 0.38 (0.61)	− 0.31 (0.75)		
Melo & Ata-Ali (2017)[61]	Spain	N = 2641 (1322/1319)	7–21	Demirjian		0.99 (0.39)	0.72 (0.56)	ICC = 1 (0)	ICC = 1 (0)
				Nolla		− 0.27 (0.50)	− 0.16 (0.23)	ICC = 1 (0)	ICC = 1 (0)
Mohammed et al. (2015)[50]	India	N = 660 (330/330)	6–16	Demirjian	0.10 (1.63)	− 0.23 (1.87)	0.43 (1.27)	ICC = 0.9 (0.008)	ICC = 0.8 (0.026)
				Haavikko	− 2.90 (1.41)	− 2.84 (1.60)	− 2.96 (1.18)	ICC = 0.9 (0.008)	ICC = 0.8 (0.026)
				Nolla	0.47 (0.83)	0.32 (0.91)	0.62 (0.71)	ICC = 0.9 (0.008)	ICC = 0.8 (0.026)
				Willems	− 0.40 (1.53)	− 0.69 (1.69)	− 0.11 (1.30)	ICC = 0.9 (0.008)	ICC = 0.8 (0.026)
Nur et al. (2012)[63]	Turkey	N = 673 (342/331)	5–16	Demirjian	0.86 (1.26)	0.84 (1.36)	0.89 (1.15)	Inappropriate estimation method	Inappropriate estimation method
				Nolla	− 0.54 (1.4)	− 0.50 (1.38)	− 0.57 (1.43)		
Paz Cortés et al. (2020)[60]	Spain	N = 604 (302/302)	4–13	Willems	0.26 (0.91)	0.17 (0.88)	0.35 (0.93)		
				Demirjian	0.70 (0.95)	0.73 (0.94)	0.68 (0.95)	Cohen's k = 0.98	Cohen's k = 0.99
				Nolla	− 0.63 (0.97)	− 0.82 (0.98)	− 0.44 (0.93)	Cohen's k = 0.98	Cohen's k = 0.99
Ranasinghe et al. (2019)[62]	Sri Lanka	N = 668 (333/335)	8–17	Demirjian	0.19 (0.87)	0.18 (0.81)	0.21 (0.93)	Cohen's k = 0.83	Cohen's k = 0.92
				Willems	− 0.38 (0.84)	− 0.38 (0.85)	− 0.38 (0.84)		
				Blenkin & Evans	− 0.55 (1.04)	− 0.53 (1.02)	− 0.56 (1.05)		
Rivera et al. (2017)[48]	Colombia	N = 457 (240/217)	6–14	Cameriere		0.08 (0.68)	− 0.25 (0.65)	ICC = 0.96 (0.001)	ICC = 0.99 (0.001)
Santoro et al. (2012)[54]	Italy	N = 535 (243/292)	7–15	Greulich-Pyle*		− 0.1 (1.3)	0.4 (1.0)	Inappropriate estimation method	Inappropriate estimation method
				Demirjian		1 (1.5)	1.1 (1.6)		
Singh et al. (2020)[51]	India	N = 900 (458/442)	10–16	Nolla	− 0.15 (0.46)	− 0.21 (0.53)	− 0.09 (0.35)	Fleiss' k = 0.78	Fleiss' k = 0.84
Tiwari et al. (2020)[52]	India	N = 70 (37/33)	1–19	Greulich-Pyle*	− 0.56 (1.33)	− 0.75 (1.53)	− 0.36 (1.04)	Inappropriate estimation method	Inappropriate estimation method
Continued									

Author (year)[reference]	Study site	Sample size (male/female)	Age range (years)	Age estimation method	Total mean error (SD)	Male mean error (SD)	Female mean error (SD)	Inter-examiner agreement estimate (variance§)	Intra-examiner agreement estimate (variance§)
Ye et al. (2014)[46]	China	N = 941 (410/531)	7–14	Demirjian		1.68 (1.29)	1.28 (1.17)	Cohen's k = 0.89	Cohen's k = 0.89
				Willems		0.36 (1.19)	− 0.02 (1.18)		
Yusof et al. (2014)[57]	Malaysia	N = 1403 (691/712)	4–24	Willems	0.45 (1.39)	0.58 (1.33)	0.32 (1.43)	Cohen's k = 0.73	Cohen's k = 0.98
Zhai et al. (2016)[47]	China	N = 1004 (392/612)	11–18	Demirjian	− 0.57 (1.25)	− 0.47 (1.21)	− 0.63 (1.27)	Inappropriate estimation method	Inappropriate estimation method
				Willems	− 0.83 (1.28)	− 0.54 (1.37)	− 1.01 (1.19)		

Table 1. The studies included in the meta-analysis. All studies reported the type of examination as "orthopantomography" except two[52,54]. *Wrist and hand X-ray; §The ICC variance was estimated using the formula reported in Noble et al.[65].

Table 2. Quality assessment performed using the QUADAS-2 instrument.

the chronological age and was the only method with a higher absolute mean error for males than females (males: − 0.22 [95% CI − 0.44; 0.00]); females: − 0.17 [95% CI − 0.34; − 0.01]). Generally, Demirjian's and Willems's methods tended to overestimate chronological age in both males (Demirjian: 0.59 [95% CI 0.28; 0.91]; Willems: 0.07 [95% CI − 0.17; 0.31]) and females (Demirjian: 0.64 [95% CI 0.38; 0.90]; Willems: 0.09 [95% CI − 0.13; 0.31]).

We included three studies in the "others" category[53,55,62] for age estimation based on dental maturity (Blenkin & Evans, Chaillet and Smith). These methods underestimated chronological age for both sexes (males: mean = − 0.26; 95% CI [− 0.65; 0.12], females: mean = − 0.29; 95% CI [− 0.61; 0.02]).

For both males and females, the PI overlapped zero for all methods, rendering the difference between estimated and chronological ages not statistically significant. For both genders, Cameriere's method showed the smallest PI, while the Haavikko and other methods had the widest intervals.

Meta-analysis of intra- and inter-examiner agreement. It was not possible to obtain a pooled Cohen's k (or Fleiss' k) due to a lack of information on the standard error or variance in the examined studies. Therefore, we compared only studies with ICCs and the studies reporting only the global reliability without
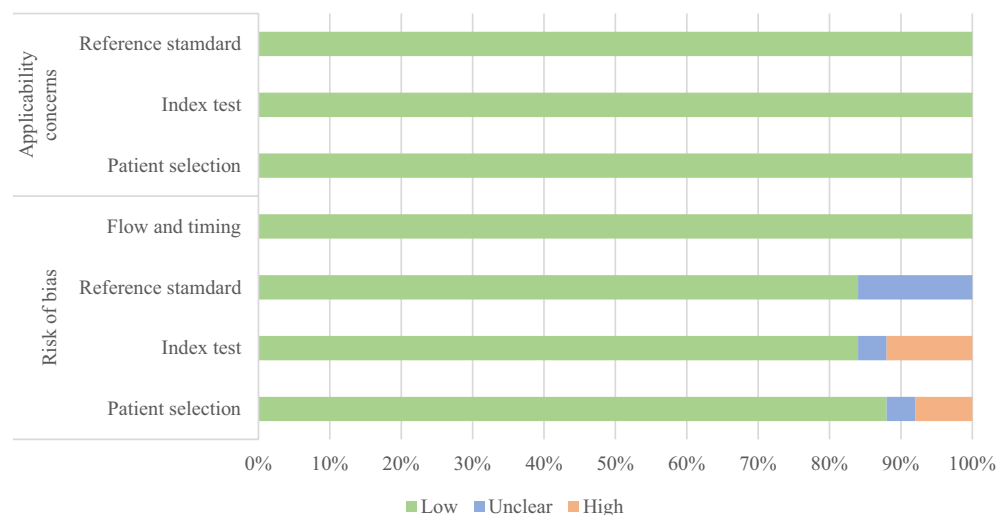
Figure 2. Quality assessment obtained using the QUADAS-2 instrument for the 23 selected studies.

stratification by gender. The meta-analytic pooled estimates of inter-examiner and intra-examiner agreement are summarized in Fig. 4.

No heterogeneity was observed in inter-examiner (heterogeneity: Q = 5.78, p = 0.888) and intra-examiner (heterogeneity: Q = 9.11, p = 0.611) agreement, so a fixed-effects model was used. For inter-examiner agreement, the ICCs ranged from 0.89 to 0.99, and the meta-analytic pooled ICC was 0.98 (95% CI 0.97; 1.00), which was close to perfect reliability. Concerning intra-examiner agreement, the ICCs ranged from 0.90 to 1.00, and the meta-analytic pooled ICC was 0.99 (95% CI 0.98; 1.00), which was also close to perfect reliability.

Discussion

Age estimation represents one of the most important aspects of dental/skeletal analysis and forensic anthropology, playing a key role in human identification, both in living subjects and to establish identity in human remains[1,2,29]. This meta-analysis provides a comprehensive overview of the current literature on the validity of age estimation methods and reproducibility of maturity indices, in particular those based on dental maturation. Although bone age has been widely used, we found only 2 validation studies on methods based on bone maturity indices that met our inclusion criteria. This low frequency could be due to the evidence that bone maturity indices suffer more from environmental factors than dental ones[23] and therefore it could be proper to validate each index only in the population in which it is built. The 21 studies on dental maturity indices identified were conducted in different countries with the aim of validating specific methods of age estimation in specific populations. Although the age estimation methods were applied to different populations, the meta-analysis results, stratified by gender and methods, showed similar accuracy. In fact, for both males and females, the prediction intervals obtained for each method spanned zero, indicating that, despite the different prediction intervals and different target populations, all methods can be considered accurate. Significant heterogeneity between studies was observed for both genders as a consequence of the large sample size of the studies and hence of the high level of precision of error estimates. Using a meta-regression model, we investigated whether this heterogeneity might be further explained by differences in characteristics of the studies or study populations such as type of method, publication year, ethnicity, mean age of the study sample, and impact factor of the journal; the I^2 index still remain very high (99.2%) for both genders (data not shown). The strategies adopted to take into account the heterogeneity between the studies are the estimation of random-effect models and the estimation of prediction intervals to detect a range in which the validity of further studies is expected to be included based on current evidence[67].

The studies that validated Nolla's method had a mean age estimation error closest to zero for both males (0.02 years) and females (0.04 years), while Cameriere's method had the narrowest prediction interval (male PI [−1.07; 0.63]; female PI [−0.82; 0.47]). Of the selected studies, Demirjian's method and its revisited version by Willems were the most frequently used methods for age estimation due to their ease of use, high reproducibility, and accuracy. Both methods tended to overestimate chronological age in males and females, but Willems' method had a narrow prediction interval, between −0.95 and 1.09 in males and −0.81 and 0.99 in females, compared with Demirjian (male prediction interval [−0.83; 2.01]; female prediction interval [−0.54; 1.81]).

The Haavikko method had the highest variability, with a prediction interval ranging from −6.88 to 4.65 for males and from −7.24 to 4.57 for females. This might be due to the variability in dental maturation among subjects of different ethnic origin[1], since Haavikko's method is calibrated on Finnish children, whose dental maturation seems to occur earlier[68]. Recently, Butti et al.[69] and Mohammed et al.[50] reached the same conclusion that Haavikko's method is unsuitable for both Italian and Indian children.

With respect to method reliability, our results showed pooled estimates of reproducibility values close to perfect reliability (about unity), indicating that the methods are highly repeatable by expert examiners. This high reproducibility might be due to positive publication bias, as studies reporting good reliability are more frequently available in the literature than studies reporting poor or no reliability[70].

(A)

Study	Sample size	Mean	Mean error	95% CI	Weight (fixed)	Weight (random)
Cameriere's method						
El-Bakary, 2010	134		-0.49	[-0.66; -0.32]	0.4%	2.1%
Galic, 2011	498		-0.02	[-0.08; 0.04]	2.9%	2.1%
Javadinejad, 2015	264		-0.27	[-0.37; -0.17]	1.1%	2.1%
Kumaresan, 2014	179		-0.44	[-0.61; -0.27]	0.4%	2.1%
Rivera, 2017	240		0.08	[-0.01; 0.17]	1.5%	2.1%
Fixed effect model	1315		-0.09	[-0.14; -0.05]	6.2%	--
Random effects model			-0.22	[-0.44; 0.00]	--	10.4%
Prediction interval				[-1.07; 0.63]	--	--
Heterogeneity: $I^2 = 94\%$, $\tau^2 = 0.059$, $p < 0.01$						
Demirjian's method						
Ambarkova, 2013	481		1.02	[0.93; 1.11]	1.3%	2.1%
Baghdadi, 2012	91		0.31	[0.12; 0.50]	0.3%	2.1%
Galic, 2010	509		1.46	[1.35; 1.57]	0.9%	2.1%
Javadinejad, 2015	264		0.90	[0.78; 1.02]	0.7%	2.1%
Jayaraman, 2012	133		-0.25	[-0.49; -0.01]	0.2%	2.0%
Kirzioglu, 2012	212		0.52	[0.40; 0.64]	0.8%	2.1%
Kumaresan, 2014	179		0.98	[0.79; 1.17]	0.3%	2.1%
Lee, 2011	754		0.29	[0.23; 0.34]	3.9%	2.1%
Melo, 2017	1322		0.99	[0.97; 1.01]	25.1%	2.1%
Mohammed, 2015	330		-0.23	[-0.43; -0.03]	0.3%	2.1%
Nur, 2012	342		0.84	[0.70; 0.98]	0.5%	2.1%
Paz Cortés, 2020	302		-0.73	[-0.84; -0.62]	1.0%	2.1%
Ranasinghe, 2019	333		0.18	[0.09; 0.27]	1.5%	2.1%
Santoro, 2012	243		1.00	[0.81; 1.19]	0.3%	2.1%
Ye, 2014	410		1.68	[1.56; 1.80]	0.7%	2.1%
Zhai, 2016	392		0.47	[0.35; 0.59]	0.8%	2.1%
Fixed effect model	6297		0.83	[0.81; 0.84]	38.7%	--
Random effects model			0.59	[0.28; 0.91]	--	33.3%
Prediction interval				[-0.83; 2.01]	--	--
Heterogeneity: $I^2 = 99\%$, $\tau^2 = 0.410$, $p = 0$						
Haavikko's method						
Galic, 2011	498		-0.09	[-0.16; -0.02]	2.3%	2.1%
Kirzioglu, 2012	212		-0.60	[-0.71; -0.49]	1.0%	2.1%
Kumaresan, 2014	179		-0.94	[-1.09; -0.79]	0.5%	2.1%
Mohammed, 2015	330		-2.84	[-3.01; -2.67]	0.4%	2.1%
Fixed effect model	1219		-0.56	[-0.61; -0.51]	4.1%	--
Random effects model			-1.12	[-2.29; 0.06]	--	8.3%
Prediction interval				[-6.88; 4.65]	--	--
Heterogeneity: $I^2 = 100\%$, $\tau^2 = 1.433$, $p < 0.01$						

Figure 3. Forest plots showing the pooled mean errors of the age predictions for males (**A**) and females (**B**) by method of age estimation.

Figure 3. (continued)

(B)

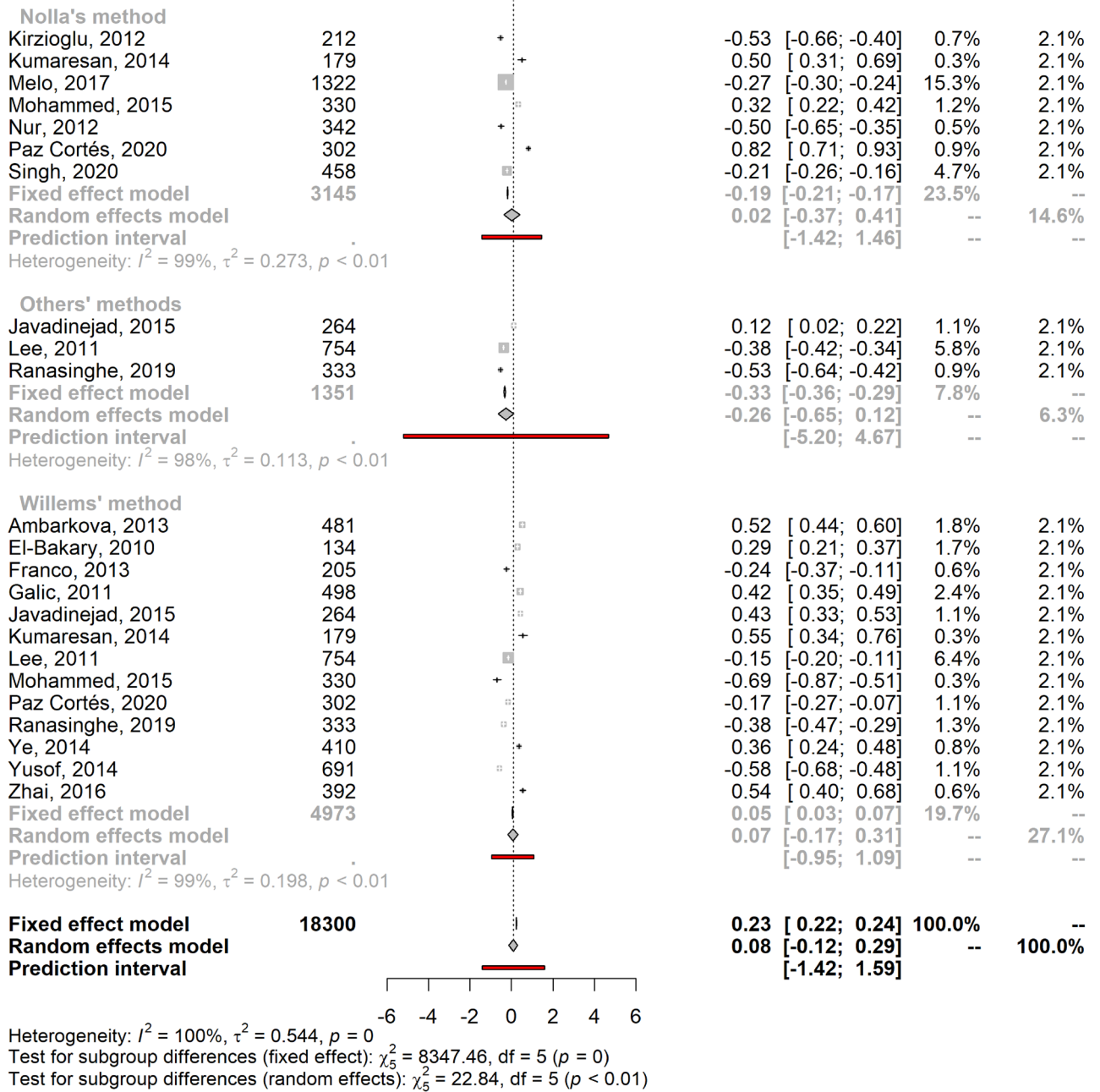| Study	Sample size	Mean	Mean error	95% CI	Weight (fixed)	Weight (random)
Cameriere's method						
El-Bakary, 2010	152		-0.26	[-0.45; -0.07]	0.2%	2.1%
Galic, 2011	591		0.10	[0.04; 0.16]	2.3%	2.1%
Javadinejad, 2015	273		-0.11	[-0.21; -0.01]	0.7%	2.1%
Kumaresan, 2014	247		-0.39	[-0.52; -0.26]	0.5%	2.1%
Rivera, 2017	217		-0.25	[-0.34; -0.16]	1.0%	2.1%
Fixed effect model	1480		-0.07	[-0.11; -0.03]	4.6%	--
Random effects model			-0.17	[-0.34; -0.01]	--	10.4%
Prediction interval				[-0.82; 0.47]	--	--
Heterogeneity: $I^2 = 95\%$, $\tau^2 = 0.034$, $p < 0.01$						
Demirjian's method						
Ambarkova, 2013	485		1.12	[1.04; 1.20]	1.2%	2.1%
Baghdadi, 2012	85		0.56	[0.39; 0.73]	0.3%	2.1%
Galic, 2010	597		1.27	[1.17; 1.37]	0.7%	2.1%
Javadinejad, 2015	273		0.85	[0.73; 0.97]	0.6%	2.1%
Jayaraman, 2012	133		-0.23	[-0.46; 0.00]	0.1%	2.0%
Kirzioglu, 2012	213		0.75	[0.63; 0.87]	0.5%	2.1%
Kumaresan, 2014	247		0.97	[0.83; 1.11]	0.4%	2.1%
Lee, 2011	729		0.31	[0.25; 0.38]	1.9%	2.1%
Melo, 2017	1319		0.72	[0.69; 0.75]	8.1%	2.1%
Mohammed, 2015	330		0.43	[0.29; 0.57]	0.4%	2.1%
Nur, 2012	331		0.89	[0.77; 1.01]	0.5%	2.1%
Paz Cortés, 2020	302		-0.68	[-0.79; -0.57]	0.6%	2.1%
Ranasinghe, 2019	335		0.21	[0.11; 0.31]	0.8%	2.1%
Santoro, 2012	292		1.10	[0.92; 1.28]	0.2%	2.1%
Ye, 2014	531		1.28	[1.18; 1.38]	0.8%	2.1%
Zhai, 2016	612		0.63	[0.53; 0.73]	0.7%	2.1%
Fixed effect model	6814		0.68	[0.66; 0.70]	17.7%	--
Random effects model			0.64	[0.38; 0.90]	--	33.3%
Prediction interval				[-0.54; 1.81]	--	--
Heterogeneity: $I^2 = 99\%$, $\tau^2 = 0.282$, $p < 0.01$						
Haavikko's method						
Galic, 2011	591		-0.23	[-0.29; -0.17]	2.1%	2.1%
Kirzioglu, 2012	213		-0.56	[-0.67; -0.45]	0.6%	2.1%
Kumaresan, 2014	247		-1.59	[-1.72; -1.46]	0.4%	2.1%
Mohammed, 2015	330		-2.96	[-3.09; -2.83]	0.5%	2.1%
Fixed effect model	1381		-0.78	[-0.83; -0.74]	3.6%	--
Random effects model			-1.33	[-2.54; -0.13]	--	8.3%
Prediction interval				[-7.24; 4.57]	--	--
Heterogeneity: $I^2 = 100\%$, $\tau^2 = 1.506$, $p = 0$						

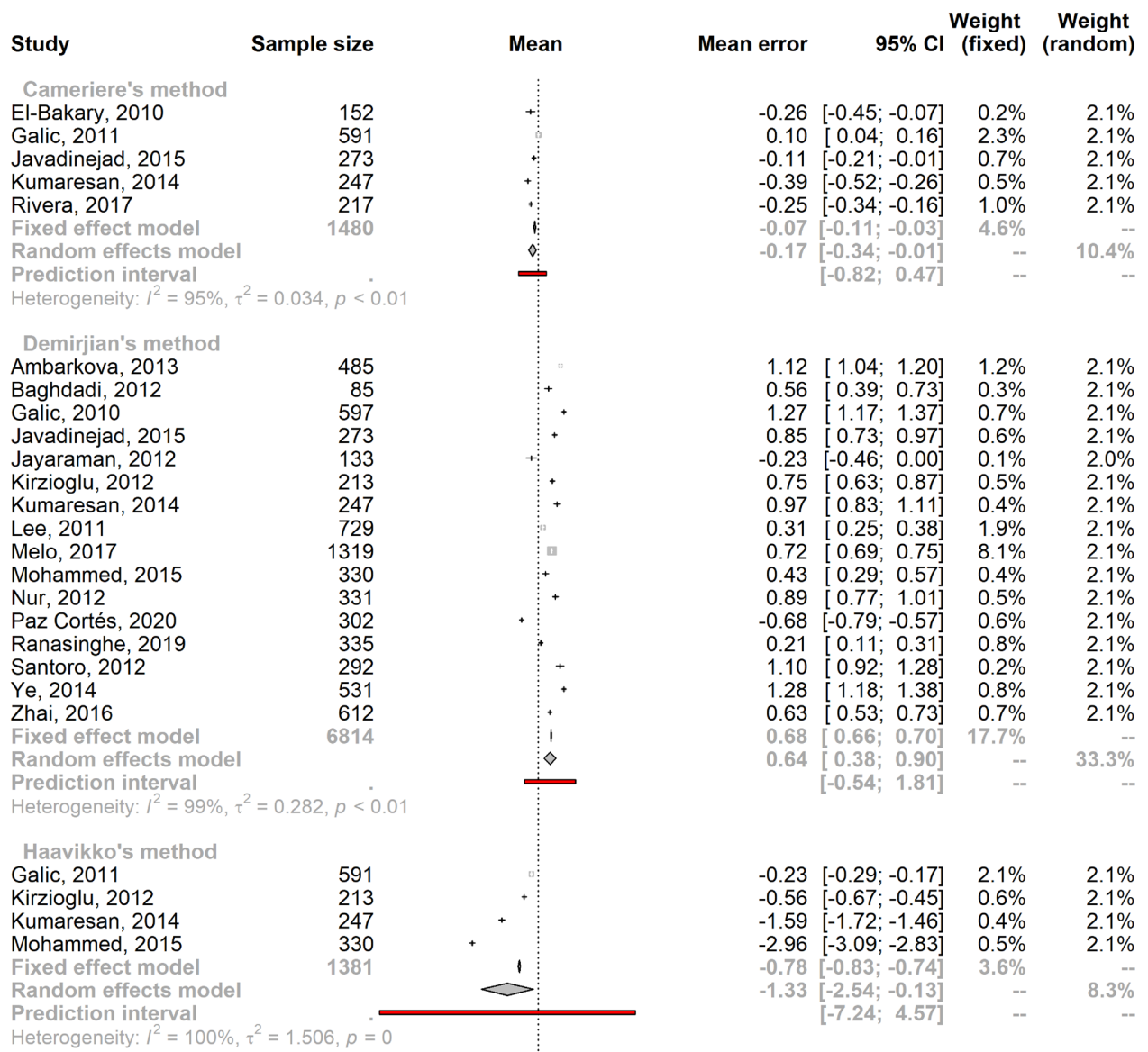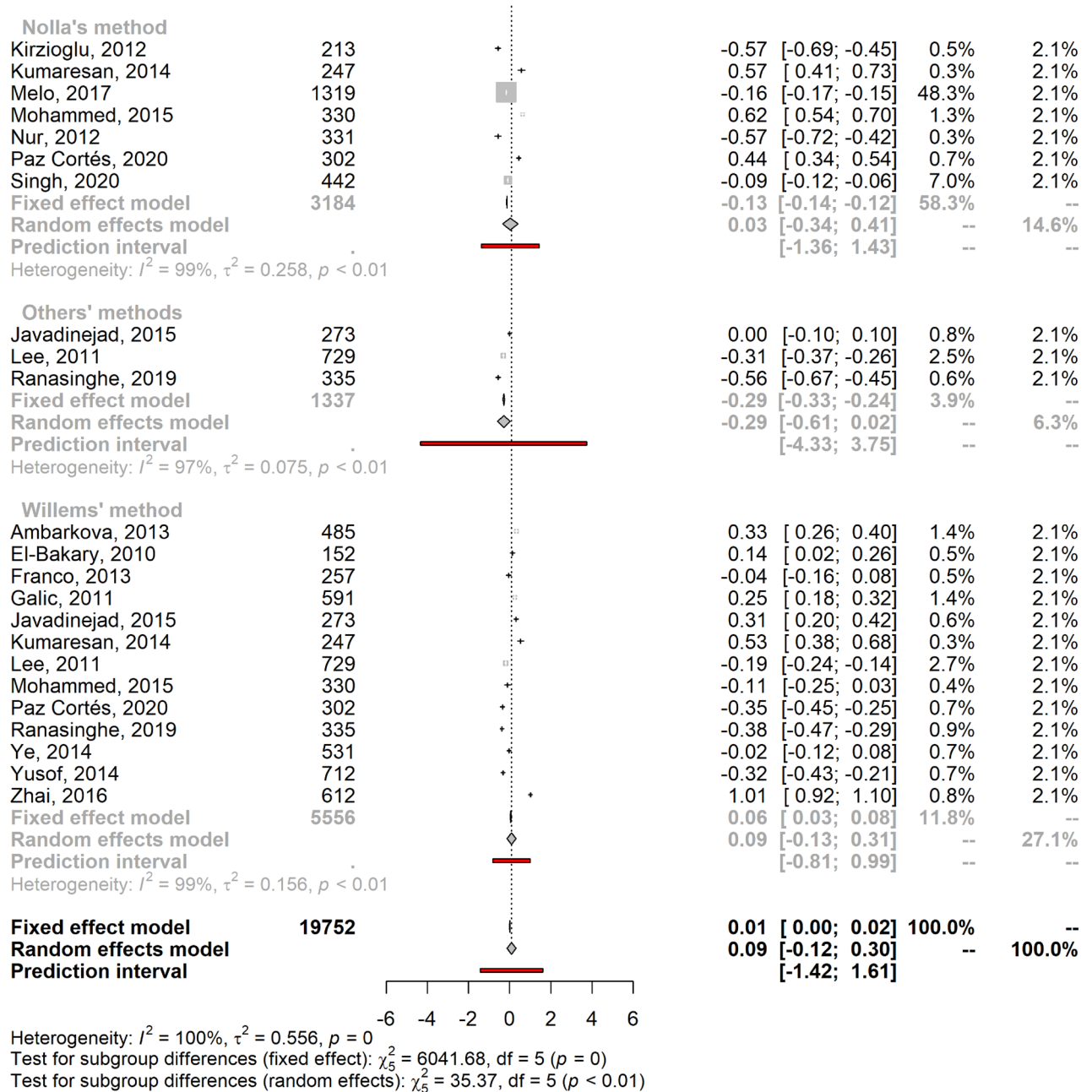**Figure 3.** (continued)

Figure 3. (continued)

The strengths of our research are the adequate number of studies included, the precision of pooled mean errors, and the comprehensive evaluation of all methods and indices based on dental maturity for which, respectively, the validity and reproducibility measures were available in literature. To our best knowledge, this is the first meta-analysis that simultaneously evaluated the validity of dental age estimation methods and reproducibility of maturity dental indices, thereby allowing more informed and safer choices in all medical and legal fields requiring these methods. Finally, the quality assessment of the selected studies was very high: only 10% of studies had an unclear risk or high risk of bias without any concerns about applicability.

However, our evaluation also has some limitations and it shows a partial picture of validity and reproducibility of age estimation methods, due to the strict exclusion criteria applied in order to provide unbiased meta-estimates. We excluded articles without information on both validity and reproducibility outcomes, articles not written in English or Italian, and those where it was impossible to obtain pooled reproducibility estimates of Cohen's kappa or the ICC due to a lack of information on the variability measure. In addition, some studies used inappropriate methods to estimate reproducibility, as discussed in Ferrante et al.[71]. Lastly, after reading the full texts, we excluded several studies (31 out of 75) with the word "validation" in the title or abstract but that used an inadequate approach to validate the method or no validation at all.

(A)

(B)
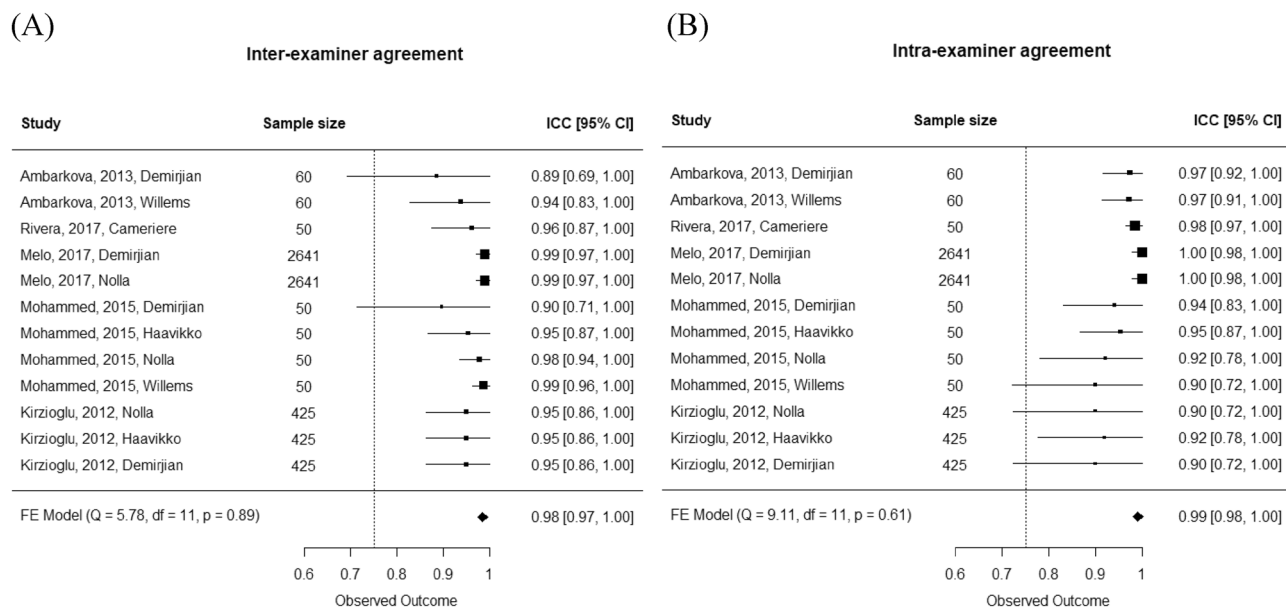
Figure 4. Forest plots showing the pooled inter-examiner (**A**) and intra-examiner (**B**) agreement.

In conclusion, since only two studies based on bone maturity indices reported the validation and reproducibility analysis, it was not possible to perform a meta-analysis for them. All studies reporting methods based on dental maturity indices, which underwent a validation process, were considered in this review and for each method the difference between estimated and chronological age was not significantly different from zero years, highlighting a high validity. Nevertheless, there was a high degree of variability in the precision of the prediction intervals (research focus 1; Supplementary material "Methods"). Furthermore, a high intra- and inter-observer reproducibility of dental maturity indices was observed (research focus 2; Supplementary material "Methods"). The Nolla and Cameriere methods might be recommended as preferred approaches, although the Cameriere method was validated on a smaller sample size than Nolla's and it requires further testing on additional populations to better assess the mean error estimates by sex. In the development of new methods of age estimation, it will be important to apply rigorous validation and publish a minimum dataset that ensures comparability of validity and reliability between different studies.

Data availability
All data is available from the included articles and in the Table 1.

References
1. de Donno, A., Angrisani, C., Mele, F., Introna, F. & Santoro, V. Dental age estimation: Demirjian's versus the other methods in different populations. A literature review. *Med. Sci. Law.* **61**(1_suppl), 125–129. https://doi.org/10.1177/0025802420934253 (2021).
2. Cunha, E. Aging the death: the importance of having better methods for age at death estimation of old individuals. *Ann. Med.* **53**(sup1), S1–S1. https://doi.org/10.1080/07853890.2021.1893522 (2021).
3. Franco, A. *et al.* Assessment of dental age estimation methods applied to Brazilian children: A systematic review and meta-analysis. *Dentomaxillofac. Radiol.* **50**(2), 20200128. https://doi.org/10.1259/dmfr.20200128 (2021).
4. Sironi, E., Gittelson, S., Bozza, S. & Taroni, F. Minor or adult? Introducing decision analysis in forensic age estimation. *Sci. Justice* **61**(1), 47–60. https://doi.org/10.1016/j.scijus.2020.09.004 (2021).
5. Márquez-Ruiz, A. B. *et al.* Three-dimensional analysis of third molar development to estimate age of majority. *Sci. Justice* **57**(5), 376–383. https://doi.org/10.1016/j.scijus.2017.04.002 (2017).
6. Balla, S. B. *et al.* Validation of third molar maturity index (I 3M) for discrimination of juvenile/adult status in South Indian population. *J. Forensic Legal Med.* **49**, 2–7. https://doi.org/10.1016/j.jflm.2017.05.003 (2017).
7. Balla, S. B. *et al.* A cut-off value of third molar maturity index for indicating a minimum age of criminal responsibility: Older or younger than 16 years?. *J. Forensic Leg. Med.* **65**, 108–112. https://doi.org/10.1016/j.jflm.2019.05.014 (2019).
8. Quispe Lizarbe, R. J., Solís Adrianzén, C., Quezada-Márquez, M. M., Galić, I. & Cameriere, R. Demirjian's stages and Cameriere's third molar maturity index to estimate legal adult age in Peruvian population. *Leg. Med.* **25**, 59–65. https://doi.org/10.1016/j.legalmed.2017.01.003 (2017).
9. Lupariello, F. *et al.* Minor migrants' age estimation: Comparison of two dental methods. *Forensic Sci. Int.: Synergy.* **3**, 100134. https://doi.org/10.1016/j.fsisyn.2021.100134 (2021).
10. de Tobel, J. *et al.* Dental and skeletal imaging in forensic age estimation: Disparities in current approaches and the continuing search for optimization. *Semin. Musculoskeletal Radiol.* **24**(05), 510–522. https://doi.org/10.1055/s-0040-1701495 (2020).
11. Timme, M., Steinacker, J. M. & Schmeling, A. Age estimation in competitive sports. *Int. J. Legal Med.* **131**(1), 225–233. https://doi.org/10.1007/s00414-016-1456-7 (2017).
12. Bommannavar, S. & Kulkarni, M. Comparative study of age estimation using dentinal translucency by digital and conventional methods. *J. Forensic Dent. Sci.* **7**(1), 71. https://doi.org/10.4103/0975-1475.150323 (2015).

13. Olze, A., Reisinger, W., Geserick, G. & Schmeling, A. Age estimation of unaccompanied minors. *Forensic Sci. Int.* **159**, S65–S67. https://doi.org/10.1016/j.forsciint.2006.02.018 (2006).

14. Cunha, E. *et al.* The problem of aging human remains and living individuals: A review. *Forensic Sci. Int.* **193**(1–3), 1–13. https://doi.org/10.1016/j.forsciint.2009.09.008 (2009).

15. Rai, B. Coronal displacement of cementum in impacted teeth: As age marker. *J. Forensic Leg. Med.* **16**(1), 5–6. https://doi.org/10.1016/j.jflm.2008.05.009 (2009).

16. Raju, G. S. *et al.* Cementum as an age determinant: A forensic view. *J. Forensic Dent. Sci.* **8**(3), 175. https://doi.org/10.4103/0975-1475.195121 (2016).

17. Buckley, M. B. & Clark, K. R. Forensic age estimation using the medial clavicular epiphysis: A study review. *Radiol. Technol.* **88**(5), 482–498 (2017).

18. Schmeling, A. *et al.* Studies on the time frame for ossification of the medial clavicular epiphyseal cartilage in conventional radiography. *Int. J. Legal Med.* **118**(1), 5–8. https://doi.org/10.1007/s00414-003-0404-5 (2004).

19. Cameriere, R. *et al.* Reliability of Schmeling's stages of ossification of medial clavicular epiphyses and its validity to assess 18 years of age in living subjects. *Int. J. Legal Med.* **126**(6), 923–932. https://doi.org/10.1007/s00414-012-0769-4 (2012).

20. Warrier, V., Kanchan, T., Shedge, R., Krishan, K. & Singh, S. Computed tomographic age estimation from the pubic symphysis using the Suchey–Brooks method: A systematic review and meta-analysis. *Forensic Sci. Int.* **325**, 110811. https://doi.org/10.1016/j.forsciint.2021.110811 (2021).

21. Valsecchi, A., Irurita Olivares, J. & Mesejo, P. Age estimation in forensic anthropology: methodological considerations about the validation studies of prediction models. *Int. J. Legal Med.* **133**(6), 1915–1924. https://doi.org/10.1007/s00414-019-02064-7 (2019).

22. Schmitt, A., Murail, P., Cunha, E. & Rougé, D. Variability of the pattern of aging on the human skeleton: Evidence from bone indicators and implications on age at death estimation. *J. Forensic Sci.* **47**(6), 1203–1209 (2002).

23. Mays, S. The effect of factors other than age upon skeletal age indicators in the adult. *Ann. Hum. Biol.* **42**(4), 332–341. https://doi.org/10.3109/03014460.2015.1044470 (2015).

24. Demirjian, A., Goldstein, H. & Tanner, J. M. A new system of dental age assessment. *Hum. Biol.* **45**(2), 211–227 (1973).

25. Cummaudo, M. *et al.* Age estimation in the living: A scoping review of population data for skeletal and dental methods. *Forensic Sci. Int.* **320**, 110689. https://doi.org/10.1016/j.forsciint.2021.110689 (2021).

26. Gassenmaier, S., Schaefer, J. F., Nikolaou, K., Esser, M. & Tsiflikas, I. Forensic age estimation in living adolescents with CT imaging of the clavicula—Impact of low-dose scanning on readers' confidence. *Eur. Radiol.* **30**(12), 6645–6652. https://doi.org/10.1007/s00330-020-07079-y (2020).

27. Štern, D., Payer, C. & Urschler, M. Automated age estimation from MRI volumes of the hand. *Med. Image Anal.* **58**, 101538. https://doi.org/10.1016/j.media.2019.101538 (2019).

28. Lu, T. *et al.* Forensic age estimation based on magnetic resonance imaging of the proximal humeral epiphysis in Chinese living individuals. *Int. J. Legal Med.* **135**(6), 2437–2446. https://doi.org/10.1007/s00414-021-02653-5 (2021).

29. de Tobel, J. *et al.* Multi-factorial age estimation: A Bayesian approach combining dental and skeletal magnetic resonance imaging. *Forensic Sci. Int.* **306**, 110054. https://doi.org/10.1016/j.forsciint.2019.110054 (2020).

30. Ubelaker, D. H. & Khosrowshahi, H. Estimation of age in forensic anthropology: historical perspective and recent methodological advances. *Forensic Sci. Res.* **4**(1), 1–9. https://doi.org/10.1080/20961790.2018.1549711 (2019).

31. Khorsan, R. & Crawford, C. External validity and model validity: A conceptual approach for systematic review methodology. *Evid. Complement. Alternat. Med.* **2014**, 1–12. https://doi.org/10.1155/2014/694804 (2014).

32. Sainani, K. L. Reliability statistics. *PM&R* **9**(6), 622–628. https://doi.org/10.1016/j.pmrj.2017.05.001 (2017).

33. Olze, A. *et al.* Forensic age estimation in living subjects: the ethnic factor in wisdom tooth mineralization. *Int. J. Legal Med.* **118**(3), 170–173. https://doi.org/10.1007/s00414-004-0434-7 (2004).

34. Meinl, A., Tangl, S., Huber, C., Maurer, B. & Watzek, G. The chronology of third molar mineralization in the Austrian population—A contribution to forensic age estimation. *Forensic Sci. Int.* **169**(2–3), 161–167. https://doi.org/10.1016/j.forsciint.2006.08.014 (2007).

35. Demirjian, A., Buschang, P. H., Tanguay, R. & Patterson, D. K. Interrelationships among measures of somatic, skeletal, dental, and sexual maturity. *Am. J. Orthod.* **88**(5), 433–438. https://doi.org/10.1016/0002-9416(85)90070-3 (1985).

36. Schmeling, A. *et al.* Criteria for age estimation in living individuals. *Int. J. Legal Med.* **122**(6), 457–460. https://doi.org/10.1007/s00414-008-0254-2 (2008).

37. Hostiuc, S., Diaconescu, I., Rusu, M. C. & Negoi, I. Age estimation using the Cameriere methods of open apices: A meta-analysis. *Healthcare.* **9**(2), 237. https://doi.org/10.3390/healthcare9020237 (2021).

38. Esan, T. A., Yengopal, V. & Schepartz, L. A. The Demirjian versus the Willems method for dental age estimation in different populations: A meta-analysis of published studies. *PLoS ONE* **12**(11), e0186682. https://doi.org/10.1371/journal.pone.0186682 (2017).

39. Mohd Yusof, M. Y. P., Wan Mokhtar, I., Rajasekharan, S., Overholser, R. & Martens, L. Performance of Willem's dental age estimation method in children: A systematic review and meta-analysis. *Forensic Sci. Int.* **280**, 245.e1-245.e10. https://doi.org/10.1016/j.forsciint.2017.08.032 (2017).

40. Sehrawat, J. S. & Singh, M. Willems method of dental age estimation in children: A systematic review and meta-analysis. *J. Forensic Leg. Med.* **52**, 122–129. https://doi.org/10.1016/j.jflm.2017.08.017 (2017).

41. Serinelli, S., Panetta, V., Pasqualetti, P. & Marchetti, D. Accuracy of three age determination X-ray methods on the left hand-wrist: A systematic review and meta-analysis. *Leg. Med.* **13**(3), 120–133. https://doi.org/10.1016/j.legalmed.2011.01.004 (2011).

42. Galić, I. *et al.* Dental age estimation among children aged 5–14 years using the Demirjian method in Bosnia-Herzegovina. *Acta Stomatol Croat.* **44**(1), 17–25 (2010).

43. Galić, I. *et al.* Accuracy of Cameriere, Haavikko, and Willems radiographic methods on age estimation on Bosnian-Herzegovian children age groups 6–13. *Int. J. Legal Med.* **125**(2), 315–321. https://doi.org/10.1007/s00414-010-0515-8 (2011).

44. Franco, A., Thevissen, P., Fieuws, S., Souza, P. H. C. & Willems, G. Applicability of Willems model for dental age estimations in Brazilian children. *Forensic Sci. Int.* **231**(1–3), 401.e1-401.e4. https://doi.org/10.1016/j.forsciint.2013.05.030 (2013).

45. Jayaraman, J., Roberts, G. J., King, N. M. & Wong, H. M. Dental age assessment of southern Chinese using the United Kingdom Caucasian reference dataset. *Forensic Sci. Int.* **216**(1–3), 68–72. https://doi.org/10.1016/j.forsciint.2011.08.019 (2012).

46. Ye, X., Jiang, F., Sheng, X., Huang, H. & Shen, X. Dental age assessment in 7–14-year-old Chinese children: Comparison of Demirjian and Willems methods. *Forensic Sci. Int.* **244**, 36–41. https://doi.org/10.1016/j.forsciint.2014.07.027 (2014).

47. Zhai, Y. *et al.* Dental age assessment in a northern Chinese population. *J. Forensic Leg. Med.* **38**, 43–49. https://doi.org/10.1016/j.jflm.2015.11.011 (2016).

48. Rivera, M. *et al.* Measurement of open apices in tooth roots in Colombian children as a tool for human identification in asylum and criminal proceedings. *J. Forensic Leg. Med.* **48**, 9–14. https://doi.org/10.1016/j.jflm.2017.03.005 (2017).

49. El-Bakary, A. A., Hammad, S. M. & Mohammed, F. Dental age estimation in Egyptian children, comparison between two methods. *J. Forensic Leg. Med.* **17**(7), 363–367. https://doi.org/10.1016/j.jflm.2010.05.008 (2010).

50. Mohammed, R. B. Accuracy of four dental age estimation methods in Southern Indian Children. *J. Clin. Diagnost. Res.* Published online (2015). https://doi.org/10.7860/JCDR/2015/10141.5495

51. Singh, H. V., Kalra, N., Tyagi, R. & Khatri, A. Dental age assessment of North Indian origin children using Nolla's method in mandibular second molar. *Egypt. J. Forensic Sci.* **10**(1), 25. https://doi.org/10.1186/s41935-020-00194-3 (2020).

52. Tiwari, P.K., Gupta, M., Verma, A., Pandey, S., Nayak, A. Applicability of the Greulich–Pyle method in assessing the skeletal maturity of children in the Eastern Utter Pradesh (UP) Region: A pilot study. *Cureus*. Published online October 10, 2020. https://doi.org/10.7759/cureus.10880

53. Javadinejad, S., Sekhavati, H. & Ghafari, R. A comparison of the accuracy of four age estimation methods based on panoramic radiography of developing teeth. *J. Dent. Res. Dent. Clin. Dent. Prosp.* **9**(2), 72–78. https://doi.org/10.15171/joddd.2015.015 (2015).

54. Santoro, V. *et al.* Applicability of Greulich and Pyle and Demirijan aging methods to a sample of Italian population. *Forensic Sci. Int.* **221**(1–3), 153.e1-153.e5. https://doi.org/10.1016/j.forsciint.2012.04.014 (2012).

55. Lee, S. S. *et al.* Validity of Demirjian's and modified Demirjian's methods in age estimation for Korean juveniles and adolescents. *Forensic Sci. Int.* **211**(1–3), 41–46. https://doi.org/10.1016/j.forsciint.2011.04.011 (2011).

56. Ambarkova, V., Galić, I., Vodanović, M., Biočina-Lukenda, D. & Brkić, H. Dental age estimation using Demirjian and Willems methods: Cross sectional study on children from the Former Yugoslav Republic of Macedonia. *Forensic Sci. Int.* **234**, 187.e1-187. e7. https://doi.org/10.1016/j.forsciint.2013.10.024 (2014).

57. Yusof, M. Y. P. M., Thevissen, P. W., Fieuws, S. & Willems, G. Dental age estimation in Malay children based on all permanent teeth types. *Int. J. Legal Med.* **128**(2), 329–333. https://doi.org/10.1007/s00414-013-0825-8 (2014).

58. Kumaresan, R., Cugati, N., Chandrasekaran, B. & Karthikeyan, P. Reliability and validity of five radiographic dental-age estimation methods in a population of Malaysian children. *J. Investig. Clin. Dent.* **7**(1), 102–109. https://doi.org/10.1111/jicd.12116 (2016).

59. Baghdadi, Z. D. & Pani, S. C. Accuracy of population-specific Demirjian curves in the estimation of dental age of Saudi children. *Int. J. Pediatr. Dent.* **22**(2), 125–131. https://doi.org/10.1111/j.1365-263X.2011.01179.x (2012).

60. Paz Cortés, M. M., Rojo, R., Alía García, E. & Mourelle Martínez, M. R. Accuracy assessment of dental age estimation with the Willems, Demirjian and Nolla methods in Spanish children: Comparative cross-sectional study. *BMC Pediatr.* **20**(1), 361. https://doi.org/10.1186/s12887-020-02247-x (2020).

61. Melo, M. & Ata-Ali, J. Accuracy of the estimation of dental age in comparison with chronological age in a Spanish sample of 2641 living subjects using the Demirjian and Nolla methods. *Forensic Sci. Int.* **270**, 276.e1-276.e7. https://doi.org/10.1016/j.forsciint.2016.10.001 (2017).

62. Ranasinghe, S. *et al.* Dental age estimation using radiographs: Towards the best method for Sri Lankan children. *Forensic Sci. Int.* **298**, 64–70. https://doi.org/10.1016/j.forsciint.2019.02.053 (2019).

63. Nur, B., Kusgoz, A., Bayram, M., et al. Validity of Demirjian and Nolla methods for dental age estimation for Northeastern Turkish children aged 5–16 years old. *Medicina Oral Patología Oral y Cirugia Bucal*. Published online 2012:e871–e877. https://doi.org/10.4317/medoral.18034

64. Kırzıoğlu, Z. & Ceyhan, D. Accuracy of different dental age estimation methods on Turkish children. *Forensic Sci. Int.* **216**(1–3), 61–67. https://doi.org/10.1016/j.forsciint.2011.08.018 (2012).

65. Noble, S., Scheinost, D. & Constable, R. T. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage* https://doi.org/10.1016/j.neuroimage.2019.116157 (2019).

66. Whiting, P. F. *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* **155**(8), 529–536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009 (2011).

67. IntHout, J., Ioannidis, J. P. A., Rovers, M. M. & Goeman, J. J. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* **6**(7), e010247. https://doi.org/10.1136/bmjopen-2015-010247 (2016).

68. Nyström, M. *et al.* Dental maturity in Finnish children, estimated from the development of seven permanent mandibular teeth. *Acta Odontol. Scand.* **44**(4), 193–198. https://doi.org/10.3109/00016358608997720 (1986).

69. Butti, A. C. *et al.* Haavikko's method to assess dental age in Italian children. *Eur. J. Orthodont.* **31**(2), 150–155. https://doi.org/10.1093/ejo/cjn081 (2009).

70. Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. (Higgins JP, Green S, eds.). Wiley; 2019. https://doi.org/10.1002/9780470712184

71. Ferrante, L. & Cameriere, R. Statistical methods to assess the reliability of measurements in the procedures for forensic age estimation. *Int. J. Legal Med.* **123**(4), 277–283. https://doi.org/10.1007/s00414-009-0349-4 (2009).

72. Nolla, C. The development of permanent teeth. *J. Dent. Child.* **27**, 254–266 (1960).

73. Haavikko, K. Tooth formation age estimated on a few selected teeth. A simple method for clinical use. *Proc. Finnish Dent. Soc. Suomen Hammaslaakariseuran toimituksia.* **70**(1), 15–19 (1974).

74. Greulich, W.W., Pyle, S.I. Radiographic atlas of skeletal development of the hand and wrist. Stanford University Press. Published online 1959:1–10.

75. Smith, B. Standards of human tooth formation and dental age assessment. *Adv. Dent. Anthropol.* Published online 1991:143–168.

76. Willems, G., van Olmen, A., Spiessens, B. & Carels, C. Dental age estimation in Belgian children: Demirjian's technique revisited. *J Forensic Sci.* **46**(4), 893–895 (2001).

Acknowledgements

Author contributions

Concept and design (F.L., C.F.), acquisition of data (M.V., M.C.), data collection (M.V., M.C.), quality assessment (M.V., M.C.), statistical analysis (I.M.), interpretation of data (F.L., M.V., M.C., I.M.), drafting of manuscript (F.L., M.V., M.C., I.M.) and critical revision of manuscript (F.L., M.V., M.C., I.M., F.A., S.E., G.R., C.F.). All authors approved the final manuscript and agreed to be accountable for all aspects of the work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-19944-5.

Correspondence and requests for materials should be addressed to E.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.