Forensic Science International 306 (2020) 110054

Contents lists available at ScienceDirect

Forensic Science International

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

# Multi-factorial age estimation: A Bayesian approach combining dental and skeletal magnetic resonance imaging

Jannick De Tobel[a,b,c,*], Steffen Fieuws[d], Elke Hillewig[a], Inès Phlypo[e], Mayonne van Wijk[f], Michiel Bart de Haas[f], Constantinus Politis[c], Koenraad Luc Verstraete[a], Patrick Werner Thevissen[b]

[a] Department of Diagnostic Sciences – Radiology, Ghent University, Corneel Heymanslaan 10, 9000 Ghent, Belgium
[b] Department of Imaging and Pathology – Forensic Odontology, KU Leuven, Kapucijnenvoer 7 blok a bus 7001, 3000 Leuven, Belgium
[c] Department of Oral and Maxillofacial Surgery, Leuven University Hospitals, Kapucijnenvoer 33, 3000 Leuven, Belgium
[d] KU Leuven – Leuven University & Hasselt University, Department of Public Health and Primary Care, I-BioStat, Leuven, Kapucijnenvoer 35 blok d bus 7001, 3000 Leuven, Belgium
[e] Department of Oral Health Sciences – Special Needs in Dentistry, Ghent University, Corneel Heymanslaan 10, 9000 Ghent, Belgium
[f] Division of Special Services and Expertise, Section of Forensic Anthropology, Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB The Hague, the Netherlands

ARTICLE INFO

*Article history:*
Received 19 May 2019
Received in revised form 19 October 2019
Accepted 12 November 2019
Available online 15 November 2019

*Keywords:*
Age determination by skeleton
Age determination by teeth
Adolescent
Adult
Magnetic resonance imaging

ABSTRACT

*Purpose:* To study age estimation performance of combined magnetic resonance imaging (MRI) data of all four third molars, the left wrist and both clavicles in a reference population of females and males. To study the value of adding anthropometric and sexual maturation data.
*Materials and methods:* Three Tesla MRI of the three anatomical sites was prospectively conducted from March 2012 to May 2017 in 14- to 26-year-old healthy Caucasian volunteers (160 females, 138 males). Development was assessed by allocating stages, anthropometric measurements were taken, and self-reported sexual maturation data were collected. All data was incorporated in a continuation-ratio model to estimate age, applying Bayes' rule to calculate point and interval predictions. Two performance aspects were studied: (1) accuracy and uncertainty of the point prediction, and (2) diagnostic ability to discern minors from adults ($\geq$18 years).
*Results:* Combining information from different anatomical sites decreased the mean absolute error (MAE) compared to incorporating only one site ($P < 0.0001$). By contrast, adding anthropometric and sexual maturation data did not further improve MAE ($P = 0.11$). In females, combining all three anatomical sites rendered a MAE equal to 1.41 years, a mean width of the 95% prediction intervals of 5.91 years, 93% correctly classified adults and 91% correctly classified minors. In males, the corresponding results were 1.36 years, 5.49 years, 94%, and 90%, respectively.
*Conclusion:* All aspects of age estimation improve when multi-factorial MRI data of the three anatomical sites are incorporated. Anthropometric and sexual maturation data do not seem to add relevant information.



## 1. Introduction

Forensic age estimation based on the development at different anatomical sites is recommended by experts [1] and applied by most authorities [2,3]. In adolescents and young adults, combining the developmental information of different anatomical sites improves age estimation performance [1]. In the living, medical imaging can be used to obtain this developmental information. Currently, radiographs and computed tomography (CT) are the commonly applied imaging modalities. Numerous studies on single anatomical sites are available (single site age estimation = SSA), with large reference studies mostly based on retrospective data [4–7].

*Abbreviations:* $\rho_{error}$, Pearson correlation coefficient between the errors of two models; A, anthropometric and sexual maturation data; C, both clavicles' sternal ends; MAE, mean absolute error; MFA, multi-factorial age estimation; PI, prediction interval; RMSE, root mean squared error; SSA, single site age estimation; T, all four third molars; W, wrist=distal radius and ulna.

* Corresponding author at: Department of Diagnostic Sciences – Radiology, Ghent University, Corneel Heymanslaan 10, 9000 Ghent, Belgium.
*E-mail addresses:* jannick.detobel@ugent.be (J. De Tobel), steffen.fieuws@kuleuven.be (S. Fieuws), elke.hillewig@gmail.com (E. Hillewig), ines.phlypo@ugent.be (I. Phlypo), m.van.wijk@nfi.minvenj.nl (M. van Wijk), m.de.haas@nfi.minvenj.nl (M.B. de Haas), constantinus.politis@uzleuven.be (C. Politis), koenraad.verstraete@ugent.be (K.L. Verstraete), patrick.thevissen@kuleuven.be (P.W. Thevissen).

http://dx.doi.org/10.1016/j.forsciint.2019.110054
0379-0738/

Attempts to combine the information from single site studies have been made [8], but controversies about the appropriate statistical approach are remaining [9,10]. By contrast, studying multiple anatomical sites in the same reference population (multi-factorial age estimation = MFA) allows for more relevant conclusions about combined age estimation and the relative contribution of each site. However, studying development of multiple anatomical sites in a living reference population would not be ethical if this requires exposure to ionising radiation without clinical indication. Consequently, MFA studies have been conducted on deceased populations [11], on clinically indicated imaging [10,12–19] or on magnetic resonance imaging (MRI) [20].

MFA studies on deceased populations can be conducted retrospectively as well as prospectively. Mostly, post-mortem full body CT is of high quality because there are no limitations regarding radiation exposure [21]. However, low dose CT is preferred for age estimation in the living. Therefore, (post-mortem) low dose CT reference studies are necessary. Moreover, gathering a sufficiently large reference population of adolescents and young adults might take a long time, since autopsies in this age cohort are rare.

MFA studies applying MRI require a prospective study design. After all, clinical images of the knee and wrist might be of sufficient quality to visualise bridging physeal plates, but clinical maxillofacial and thorax scans will not depict the details of developing teeth [22–26] and clavicles [27–31], respectively. Moreover, different sites of interest for age estimation are unlikely to be scanned at the same time in a clinical context. Consequently, several research groups are collecting MRI data for MFA. The first results were published by Stern et al. (2019) [20]. Their approach was based on a fully automatic assessment of the MRI data and information was combined by deep convolutional neural networks. They demonstrated that MFA combining MRI data of all four third molars (numbered according to the World Dental Federation 18, 28, 38, 48), the left hand and both clavicles outperformed SSA as well as combining only two sites. Unfortunately, their study population only included males. They reported the diagnostic ability of the method to discern minors from adults, as well as mean absolute errors of the point prediction. However, no prediction intervals (PI; i.e. uncertainty intervals of the point prediction) were reported nor was the coverage verified (i.e. the proportion of true ages falling within the 95% PI).

Moreover, few studies have combined developmental information from medical imaging with anthropometric and sexual maturation data. Overall, the studies indicated that anthropometric and sexual maturation data contribute little to age estimation [32,33].

Therefore, the current study aimed to combine the information of human observers, who staged the development of all four third molars, the left wrist and both clavicles for MFA in a reference population of females and males, with specific attention to PIs and coverage. A second aim was to study the value of adding anthropometric and sexual maturation data to MFA.

## 2. Materials and methods

### 2.1. Study population

The Ghent University Hospital ethics committee approved the study. Subsequently, 298 healthy volunteers were prospectively recruited from March 2012 to May 2017 (160 females, 138 males, Table 1). This convenience sample comprised Caucasians of Belgian and Dutch ancestry of middle to high socio-economic status. Exclusion criteria were surgical removal of any third molar, fractures involving the examined metaphysis-physis-epiphysis complex, developmental disorders, chronic diseases or chronic medication intake that might affect development, and contra-indications for MRI. Parts of the study population have been included in earlier publications on SSA [23,24,27,34–37]. All participants gave informed consent, and in minors, the parents also consented. The study design complied with recommendations for reference studies on age estimation [1].

**Table 1**
Number of participants per age per sex.

| Age (y) | Frequency | | |
|---|---|---|---|
| | Female | Male | Total |
| 14 | 11 | 11 | 22 |
| 15 | 11 | 10 | 21 |
| 16 | 10 | 10 | 20 |
| 17 | 11 | 9 | 20 |
| 18 | 13 | 10 | 23 |
| 19 | 15 | 12 | 27 |
| 20 | 20 | 9 | 29 |
| 21 | 14 | 10 | 24 |
| 22 | 12 | 10 | 22 |
| 23 | 11 | 10 | 21 |
| 24 | 11 | 11 | 22 |
| 25 | 11 | 12 | 23 |
| 26 | 10 | 14 | 24 |
| Total | 160 | 138 | 298 |

### 2.2. Anthropometric, socio-economic and sexual maturation data

All data were collected by the first author (JDT). As an oral and maxillofacial surgeon, he has both a medical and a dental degree, which allowed him to assess the complete health status of the participants. He also conducted the MRI (see below).

In a questionnaire, participants were asked about anthropometry, handedness, socio-economic status, medical and dental history. To provide self-reported data on sexual maturation, the questionnaire also included drawings and explanations of the Tanner stages of pubic hair for both sexes [38]. Tanner breast development stages were also included for females, while males assessed their testicular volumes using an orchidometer (intervals of 5 mL). Participants received the questionnaires beforehand, and all aspects were discussed with the researcher at the day of data collection. After said discussion, a clinical dental examination was conducted to check for anomalies.

### 2.3. Image acquisition

Subsequently, 3T MRI (Magnetom Trio Tim, Siemens, Erlangen, Germany) was carried out according to published protocols for third molars [23], the left wrist [35] and both clavicles [28]. Third molars and both clavicles were only scanned with one sequence, i.e. a T2 TSE and a T1 VIBE sequence, respectively. By contrast, the wrist was scanned with a T1 spin echo as well as a T1 VIBE sequence. Nonetheless, reported wrist results were only based on the latter sequence for two reasons:

1) In the original publication [35], it was demonstrated that both sequences are equally suitable.
2) Because of time constraints at the scanner, the spin echo sequence was not performed in the first 30 participants.

Note that the left wrist was scanned in every participant, irrespective of handedness.

### 2.4. Image analysis

The images were pseudonymized and development was studied by allocating stages to all third molars [24], the distalradius

and ulna [35] and both clavicles [27]. Regarding all three anatomical sites, different MRI-specific staging considerations were pointed out in earlier references studies [24,27,35], resulting in the staging techniques depicted in Fig. 1. Criteria for these stages are included in the appendix (Tables A1 and A2). Note that stages of early (stage 1) and late (stages 4 and 5) development of the clavicles were discarded, since they might be confused [27,29]. Moreover, morphological variants of the clavicles' sternal end were excluded for analysis [27]. For each anatomical site, the whole stack of slices was considered. Different observers with different levels of experience assessed the images, resulting in inter- and intra-observer agreements that have been reported in earlier studies [24,27,35]. The staging results of the most experienced observer (JDT) were used for statistical analyses in the current study.

### 2.5. Statistical analysis

Development at each anatomical site and each anthropometric or sexual maturation measure were considered as age indicators. Bayes' rule was used to obtain the posterior distribution of age given these indicators, using an ad hoc procedure [9,39] to obtain an appropriate prediction interval (PI) which corrects for violation of the conditional independence assumption (see below). As in the original approach, a continuation-ratio model was used to incorporate all ordinal age indicators. Moreover, the approach was expanded for the current study to incorporate continuous data too, using a linear regression model for body weight and length. Ten-fold cross-validation handled overfitting.

For each model, two aspects of age estimation performance were studied: (1) point prediction of age, with PIs to quantify uncertainty, and (2) the ability to discern minors (<18 years old) from adults (≥18 years old). Regarding the point prediction, the 5% trimmed mean was used, i.e. the mean in the 95% PI. Uncertainty was quantified by 95% PI, while accuracy and precision were studied by calculating mean absolute error (MAE; error = chronological age − estimated age) and root mean squared error (RMSE). Comparing the RMSE with the RMSE obtained when using the mean observed age as prediction for every participant, a coefficient of determination ($R^2$) was calculated to reflect the percentage of total variance in age explained by the model. Regarding the ability to discern minors from adults, diagnostic indices were defined as follows:

- accuracy = proportion of correctly classified participants,
- sensitivity = proportion of correctly classified adults,
- specificity = proportion of correctly classified minors,
- discrimination slope = the difference between minors and adults in mean predicted probability to be a minor.

Differences in age estimation performance between the models were evaluated by Wilcoxon signed rank tests, comparing MAEs. Furthermore, the conditional independence assumption was tested. In age estimation, two age indicators are conditionally independent if for a given age, the developmental stage of one age indicator provides no information about the developmental stage of the other age indicator. This was tested in two ways. Firstly, the coverage of the 95% PI was studied for the applied models, i.e. for the PIs assuming conditional independence as well as for the PIs corrected for possible violation of the conditional independence. Note that when the conditional independence assumption does not hold, the PIs assuming independence will be too narrow, causing the percentage of true chronological ages falling within the 95% PI being too low. Secondly, the Pearson correlation coefficients ($\rho_{error}$) were calculated between the errors of the different models. If the conditional independence assumption truly holds, this correlation should be close to zero.

MFA was studied by combining different sets of age indicators in the model. The following age indicators comprised those different sets: all four third molars (T), the distal radius and ulna at the left wrist (W), both clavicles' sternal ends (C), anthropometric and sexual maturation data (A). The sets were incorporated into the model separately or using the following combinations: T + W, T + C, W + C, T + W + C, T + W + C + A. SSA based on T was selected as a starting point for comparisons, since third molars go through developmental changes throughout the entire age range of the study population. Conversely, developmental changes of W and C are more pronounced in younger and older age categories, respectively.

The number of included anatomical structures per model varied because of missing data. For all structures, some missing data arose as a result of images not being assessable. Furthermore, third molar data was missing in cases of agenesis, and clavicle data if they were in the discarded stages 1, 4 or 5, or if they were morphological variants. Anthropometric and sexual maturation data were not missing for any participant. Thus, if for instance the performance of T + C was studied and a specific participant did not have T information, the prediction was solely based on C. If there was no information at all within a specific set of age indicators for a specific participant, the following distinction was made between two quantifications: (1) the participant was excluded from the assessment (rendering performance results based only on cases with no missing information), and (2) the prediction for this participant equaled the best available guess, which was chosen to be the mean of the prior distribution of age, i.e. a uniform distribution from 14 to 26 years old (rendering performance results based on the entire study population). It was assumed that the

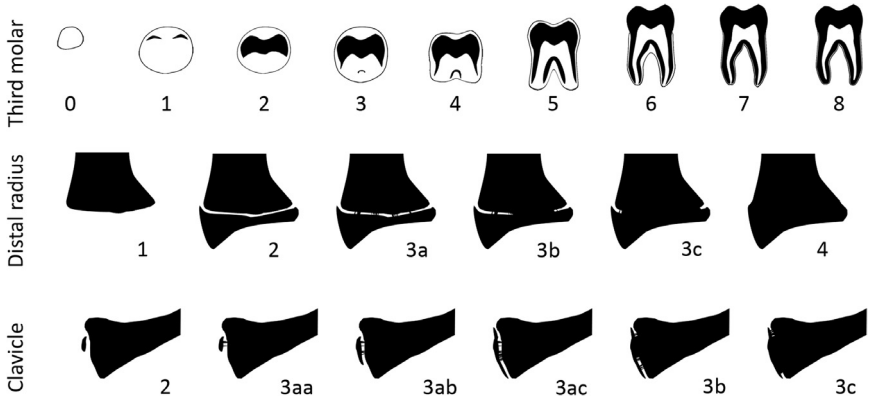


**Fig. 1.** Staging techniques to assess development in three anatomical sites.

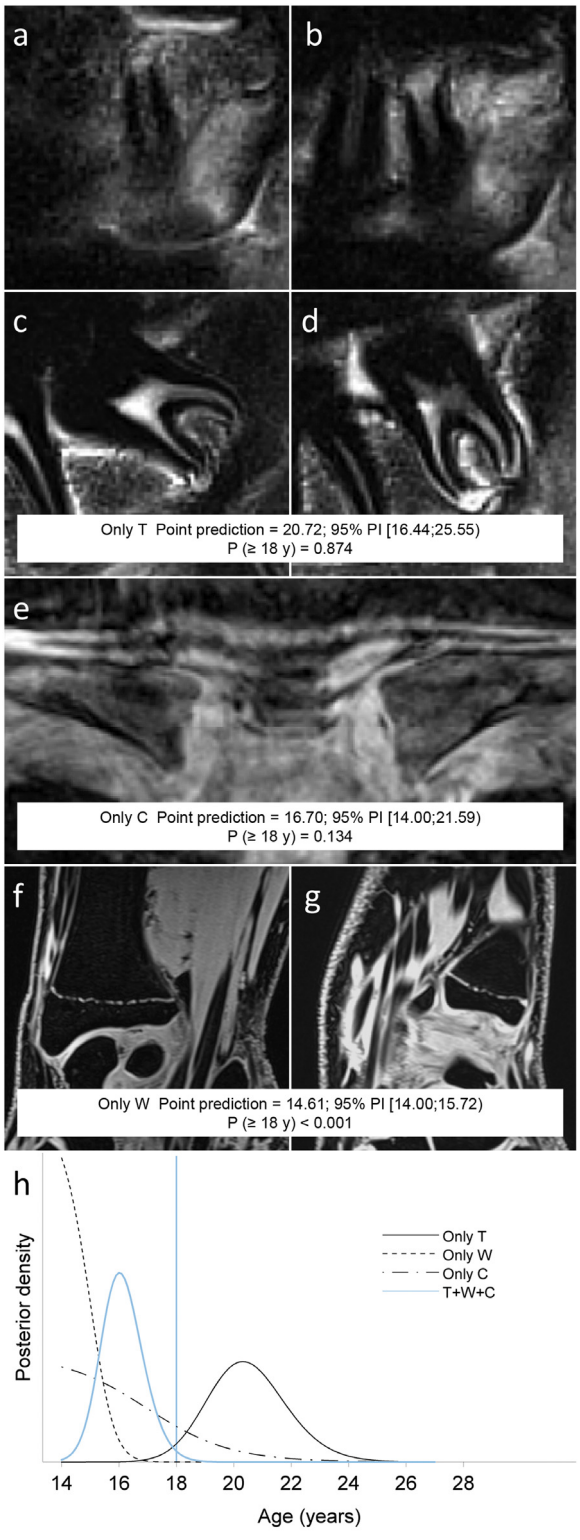


**Fig. 2.** Example of multi-factorial MRI data used for age estimation based on the Bayesian approach. This female participant was 16.05 years old.
**a-d** Third molars 18, 28, 38, and 48, respectively. All third molars are in stage 6, except tooth 28 (**b**) which is in stage 5.
**e** The right clavicle is in stage 1; the left clavicle in stage 2.
**f** The distal radius is in stage 3a.
**g** The distal ulna is in stage 3a.
**h** Posterior density curves for single site age estimation and for multi-factorial age estimation.

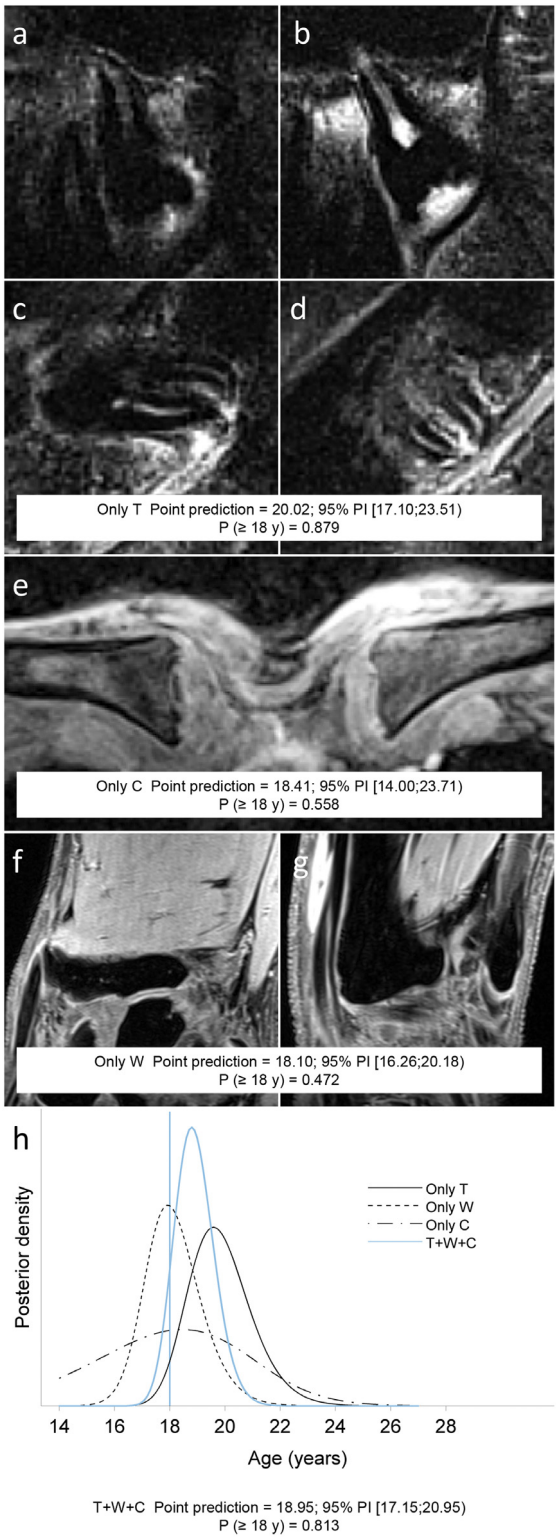


**Fig. 3.** Example of multi-factorial MRI data used for age estimation based on the Bayesian approach. This male participant was 19.41 years old.
**a-d** Third molars 18, 28, 38, and 48, respectively. All third molars are in stage 6, except tooth 38 (**c**) which is in stage 7.
**e** Both clavicles are in stage 3aa.
**f, g** The distal radius and ulna are in stage 3c.
**h** Posterior density curves for single site age estimation and for multi-factorial age estimation.

observed proportion of missing data for the age indicators was representative for the rate of missing data in a practical setting.

Statistical analyses were conducted with SAS 9.4 (SAS Institute, Cary NC, USA), and tests were performed two-sided with the significance level equal to 0.05. Sex-specific results were reported when appropriate.

## 3. Results

To illustrate the Bayesian approach for MFA using MRI, two cases were displayed in Figs. 2 and 3.

### 3.1. Missing data

Table 2 summarizes the number of anatomical structures that were included in the models. Missing clavicle data was more frequent in the younger and older participants, because the earliest and latest stages were discarded. In all other age indicators, missing data showed no predilection for certain age categories.

As a result of the relatively small study sample, the frequency in certain stages was too low to allow for a stable model. Therefore, the following stages were incorporated jointly into the models: radius and ulna stages 2/3a, 3b/3c, 4/5; clavicle stages 3aa/3ab/3ac.

**Table 2**
Number of anatomical structures included in the models for age estimation, per sex.

| Anatomical structure | Frequency | | |
|---|---|---|---|
| | Female | Male | Total |
| Third molar 18 | 135 | 118 | 253 |
| Third molar 28 | 145 | 120 | 265 |
| Third molar 38 | 137 | 106 | 243 |
| Third molar 48 | 137 | 106 | 243 |
| Distal radius | 157 | 136 | 293 |
| Distal ulna | 157 | 136 | 293 |
| Right clavicle | 106 | 78 | 184 |
| Left clavicle | 115 | 82 | 197 |

### 3.2. Which age indicators should be included?

In general, incorporating more age indicators in the model resulted in higher $R^2$ (Table 3). However, adding anthropometric and

**Table 3**
Age estimation performance of different models which include different age indicators.

| Cases without information | Sex | N | | $R^2$ | Accuracy of point prediction | | Uncertainty of point prediction | | Discerning minors from adults | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | RMSE (y) | MAE (y) | Width of 95% PI | | Accuracy | Sensitivity | Specificity | Discrimination slope |
| | | | | | | | Mean | SD | | | | |
| Excluded | Females | 154 | T | 0.317 | 2.378 | 1.920 | 7.82 | 1.88 | 77.3 | 78.8 | 73.2 | 0.477 |
| | | 158 | W | 0.255 | 2.621 | 2.078 | 8.15 | 2.96 | 89.9 | 93.9 | 79.1 | 0.656 |
| | | 121 | C | 0.230 | 2.304 | 1.782 | 8.37 | 1.86 | 86.0 | 92.0 | 70.6 | 0.507 |
| | | 160 | A | 0.118 | 3.103 | 2.464 | 10.42 | 2.63 | 77.5 | 81.2 | 67.4 | 0.362 |
| | | 159 | T+W | 0.418 | 2.027 | 1.566 | 6.47 | 2.00 | 86.2 | 85.3 | 88.4 | 0.641 |
| | | 157 | T+C | 0.397 | 2.110 | 1.677 | 6.90 | 1.81 | 84.7 | 84.2 | 86.0 | 0.582 |
| | | 158 | W+C | 0.342 | 2.296 | 1.825 | 7.23 | 2.50 | 92.4 | 93.9 | 88.4 | 0.682 |
| | | 160 | T+W+C | 0.464 | 1.887 | 1.411 | 5.91 | 1.88 | 92.5 | 93.2 | 90.7 | 0.706 |
| | | 160 | T+W+C+A | 0.478 | 1.836 | 1.353 | 5.76 | 1.77 | 90.6 | 91.5 | 88.4 | 0.718 |
| | Males | 131 | T | 0.467 | 2.117 | 1.706 | 6.79 | 2.33 | 91.6 | 92.4 | 89.7 | 0.715 |
| | | 136 | W | 0.457 | 2.169 | 1.775 | 6.74 | 2.43 | 94.1 | 93.8 | 94.9 | 0.815 |
| | | 94 | C | 0.172 | 2.830 | 2.124 | 8.40 | 1.95 | 78.7 | 85.7 | 58.3 | 0.346 |
| | | 138 | A | 0.155 | 3.361 | 2.672 | 10.17 | 2.99 | 75.4 | 79.6 | 65.0 | 0.369 |
| | | 138 | T+W | 0.526 | 1.861 | 1.500 | 5.96 | 2.22 | 93.5 | 94.9 | 90.0 | 0.807 |
| | | 137 | T+C | 0.513 | 1.905 | 1.518 | 6.36 | 2.33 | 92.0 | 92.8 | 90.0 | 0.724 |
| | | 138 | W+C | 0.473 | 2.069 | 1.628 | 6.20 | 2.16 | 92.8 | 93.9 | 90.0 | 0.793 |
| | | 138 | T+W+C | 0.564 | 1.732 | 1.357 | 5.49 | 1.96 | 92.8 | 93.9 | 90.0 | 0.807 |
| | | 138 | T+W+C+A | 0.566 | 1.724 | 1.342 | 5.45 | 1.94 | 93.5 | 94.9 | 90.0 | 0.811 |
| Mean age imputed | Females | 154 | T | 0.301 | 2.442 | 1.963 | 7.99 | 2.03 | 77.3 | 78.8 | 73.2 | 0.477 |
| | | 158 | W | 0.251 | 2.637 | 2.089 | 8.20 | 2.98 | 89.9 | 93.9 | 79.1 | 0.656 |
| | | 121 | C | 0.119 | 3.088 | 2.420 | 9.34 | 2.36 | 86.0 | 92.0 | 70.6 | 0.507 |
| | | 160 | A | 0.118 | 3.103 | 2.464 | 10.42 | 2.63 | 77.5 | 81.2 | 67.4 | 0.362 |
| | | 159 | T+W | 0.410 | 2.061 | 1.589 | 6.51 | 2.05 | 86.2 | 85.3 | 88.4 | 0.641 |
| | | 157 | T+C | 0.388 | 2.137 | 1.695 | 7.00 | 1.94 | 84.7 | 84.2 | 86.0 | 0.582 |
| | | 158 | W+C | 0.336 | 2.318 | 1.839 | 7.30 | 2.55 | 92.4 | 93.9 | 88.4 | 0.682 |
| | | 160 | T+W+C | 0.464 | 1.887 | 1.411 | 5.91 | 1.88 | 92.5 | 93.2 | 90.7 | 0.706 |
| | | 160 | T+W+C+A | 0.478 | 1.836 | 1.353 | 5.76 | 1.77 | 90.6 | 91.5 | 88.4 | 0.718 |
| | Males | 131 | T | 0.449 | 2.161 | 1.736 | 7.07 | 2.58 | 91.6 | 92.4 | 89.7 | 0.715 |
| | | 136 | W | 0.452 | 2.177 | 1.785 | 6.82 | 2.51 | 94.1 | 93.8 | 94.9 | 0.815 |
| | | 94 | C | 0.084 | 3.613 | 2.904 | 9.66 | 2.45 | 78.7 | 85.7 | 58.3 | 0.346 |
| | | 138 | A | 0.155 | 3.361 | 2.672 | 10.17 | 2.99 | 75.4 | 79.6 | 65.0 | 0.369 |
| | | 138 | T+W | 0.526 | 1.861 | 1.500 | 5.96 | 2.22 | 93.5 | 94.9 | 90.0 | 0.807 |
| | | 137 | T+C | 0.502 | 1.955 | 1.547 | 6.40 | 2.38 | 92.0 | 92.8 | 90.0 | 0.724 |
| | | 138 | W+C | 0.473 | 2.069 | 1.628 | 6.20 | 2.16 | 92.8 | 93.9 | 90.0 | 0.793 |
| | | 138 | T+W+C | 0.564 | 1.732 | 1.357 | 5.49 | 1.96 | 92.8 | 93.9 | 90.0 | 0.807 |
| | | 138 | T+W+C+A | 0.566 | 1.724 | 1.342 | 5.45 | 1.94 | 93.5 | 94.9 | 90.0 | 0.811 |

In cases without information for a specific anatomical site, the case was either excluded (upper panel), or the mean of the prior distribution of age was imputed as point prediction (20.5 years), and the width of the 95% prediction interval equaled that of the prior distribution of age ($0.95 \times 13 = 12.35$ years) (lower panel).
A = anthropometric and sexual maturation data; C = both clavicles; MAE = mean absolute error; N = number of participants; PI = prediction interval; $R^2$ = coefficient of determination; RMSE = root mean squared error; SD = standard deviation; T = all third molars; W = wrist = distal radius and ulna; y = years.
Sensitivity = percentage correctly identified adults; Specificity = percentage correctly identified minors; Discrimination slope = difference between minors and adults in mean predicted probability to be a minor.

**Table 4**
Mean absolute error of age estimation compared in different models which include different age indicators. Females and males are considered jointly. Shaded cells indicate statistically significant results.

| Cases without information | Accuracy of point prediction | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Age indicator(s) | MAE (y) | SD (y) | Age indicator(s) | MAE (y) | SD (y) | N | P-value MAE difference* |
| Excluded | T | 1.83 | 1.32 | C | 1.93 | 1.68 | 206 | 0.9152 |
| | T | 1.83 | 1.35 | W | 1.94 | 1.46 | 282 | 0.4797 |
| | W | 2.01 | 1.48 | C | 1.94 | 1.67 | 213 | 0.7320 |
| | T | 1.82 | 1.34 | T + W | 1.52 | 1.20 | 285 | <0.0001 |
| | T | 1.82 | 1.34 | T + C | 1.59 | 1.23 | 285 | <0.0001 |
| | T | 1.82 | 1.34 | T + W + C | 1.37 | 1.15 | 285 | <0.0001 |
| | T + W | 1.54 | 1.21 | T + W + C | 1.38 | 1.15 | 297 | 0.0041 |
| | T + C | 1.60 | 1.23 | T + W + C | 1.38 | 1.14 | 294 | <0.0001 |
| | W + C | 1.73 | 1.35 | T + W + C | 1.38 | 1.15 | 296 | <0.0001 |
| | T + W + C + A | 1.34 | 1.17 | T + W + C | 1.38 | 1.15 | 297 | 0.1138 |
| Mean age imputed | T | 1.86 | 1.39 | C | 2.64 | 2.05 | 298 | <0.0001 |
| | T | 1.86 | 1.39 | W | 1.95 | 1.46 | 298 | 0.5187 |
| | W | 1.95 | 1.46 | C | 2.64 | 2.05 | 298 | <0.0001 |
| | T | 1.86 | 1.39 | T + W | 1.55 | 1.22 | 298 | <0.0001 |
| | T | 1.86 | 1.39 | T + C | 1.63 | 1.26 | 298 | <0.0001 |
| | T | 1.86 | 1.39 | T + W + C | 1.39 | 1.17 | 298 | <0.0001 |
| | T + W | 1.55 | 1.22 | T + W + C | 1.39 | 1.17 | 298 | 0.0041 |
| | T + C | 1.63 | 1.26 | T + W + C | 1.39 | 1.17 | 298 | <0.0001 |
| | W + C | 1.74 | 1.36 | T + W + C | 1.39 | 1.17 | 296 | <0.0001 |
| | T + W + C + A | 1.35 | 1.17 | T + W + C | 1.39 | 1.17 | 298 | 0.0941 |

In cases without information for a specific anatomical site, the case was either excluded (upper panel), or the mean of the prior distribution of age (20.5 years) was imputed as point prediction (lower panel).
A = anthropometric and sexual maturation data; C = both clavicles; N = number of participants included for comparison; MAE = mean absolute error; SD = standard deviation; T = all third molars; W = wrist = distal radius and ulna; y = years.
* P-value from Wilcoxon signed rank test comparing MAEs.

sexual maturation data to the MRI information of the anatomical sites only yielded a negligible increase in $R^2$ (0.002 to 0.014).

Secondly, the accuracy of the point prediction was considered. When cases without information were excluded, none of the anatomical sites outperformed the other for SSA (Table 4, upper panel). Obviously, when clavicle information lacked, dental or wrist information was indispensable (Table 4, lower panel). Combining the three anatomical sites rendered the most accurate age estimation with MAE = 1.38 years (standard deviation (SD) 1.15), and MAE = 1.39 years (SD 1.17) when imputing mean age for cases without information (Table 4). Moreover, compared with SSA or combining only two anatomical sites, the MAE improvement was not only significant, but also forensically relevant ranging from 0.16 to 0.47 years (i.e. 58–171 days). Still, not all individuals benefit from MFA, compared to SSA, as demonstrated in Fig. 4. Note that developmental changes occur at all three anatomical sites around the age of 18. Therefore, it seems logical that the overall performance of the model benefits from this combination. Finally, adding anthropometric and sexual maturation data to the model did not significantly reduce the MAE (P = 0.1138; P = 0.0941 when imputing mean age for cases without information).

Thirdly, the uncertainty of the point prediction and the ability to discern minors from adults were considered. Combining information narrowed down the 95% PIs and increased the value of diagnostic indices (Table 3). Again, adding anthropometric and sexual maturation data only rendered marginal improvements. Moreover, results for males were better than for females.

Finally, Table A3 provides an overview of probabilities to be an adult, for combinations that might be encountered in practice when the threshold of 18 years is questioned.

### 3.3. Conditional independence

The coverage of the 95% PIs was appropriate, ranging from 94.0–95.8% (Table A4), when they were corrected for the plausible violation of the conditional independence assumption. However, when conditional independence was assumed, they were much too narrow, ranging from 77.2–91.2%. Moreover, the violation of the conditional independence assumption was confirmed by the correlation between the prediction errors when different anatomical sites were incorporated into the models, with $\rho_{error}$ ranging from 0.12 (third molars and clavicles in females) to 0.49 (third molars and wrist in males). Furthermore, the conditional independence assumption was rejected for all anthropometric and sexual maturation data, either among them or between them and each of the anatomical sites, with $\rho_{error}$ ranging from 0.17 (clavicles and pubic hair in females) to 0.76 (testicular volume and pubic hair in males).

## 4. Discussion

The current findings demonstrated that MFA outperforms SSA based on MRI to estimate age in adolescents and young adults of both sexes. With an MAE of 1.41 years in females and 1.36 in males, combining MRI information of all four third molars, the left wrist and both clavicles yielded more accurate age estimations than using less information. Moreover, narrower 95% PIs were obtained, with a width of 5.91 years in females and 5.49 years in males. Furthermore, MFA combining all three anatomical sites reached the highest ability to discern minors from adults, with a specificity of 91% in females and 90% in males, and a sensitivity of 93% in females and 94% in males. Adding anthropometric and sexual maturation data only marginally improved age estimation performance.

The current findings were consistent with those reported by Stern et al. (2019), who performed MFA based on MRI of the same combination of anatomical sites in 322 male Austrian volunteers from 13.0–25.0 years old [20]. They quantified the relative contribution of each anatomical site to age estimation. Third molars showed the most constant contribution throughout the studied age range. The hand/wrist's contribution decreased with increasing age, whereas the clavicles' contribution increased. Similar to our findings, they demonstrated that the combination of all three structures is especially valuable around the age of 18.

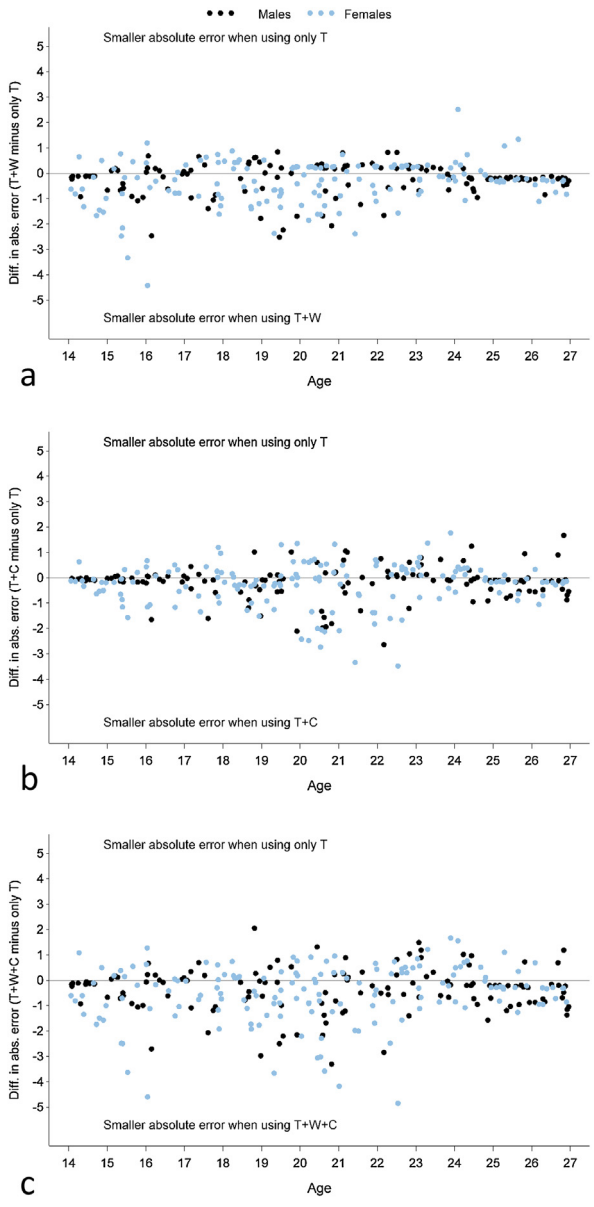


**Fig. 4.** Cross-validated effect of combining anatomical sites on mean absolute error (MAE). On average, MAE is reduced when more information is incorporated into the model, but this does not hold for all participants. **a** Adding wrist information to third molars information mainly renders large corrections of the error in younger participants, whereas the effect of adding wrist information fades away in the older participants. **b** Adding clavicles information to third molars information mainly renders large corrections of the error in participants in their early twenties, whereas the effect of adding clavicles information fades away in the younger participants. **c** Adding wrist and clavicles information to third molars information renders large corrections of the error in participants up to age 23.
C = both clavicles; T = third molars; W = distal radius and ulna.

In their sample, the MAE even reached 1.01 years, and a specificity of 90% corresponded with a sensitivity of 93%. Since they applied a fully automated approach to assess the images and estimate age, they countered intra- and inter-observer variability, which remains a major disadvantage of human observers who stage development [37,40,41]. However, they did not report PIs and it is unclear how conditional dependence is handled within their deep convolutional neural networks approach.

Those PIs are essential to forensic age estimation practice, since they allow the requesting authority to interpret the uncertainty. However, an appropriate statistical approach is required to calculate the PIs [9]. In a first possible approach, the correlation-between the age indicators is modelled explicitly in a multivariate model for the conditional distribution in Bayes' rule [42]. Unfortunately, this approach becomes computationally complex when the number of age indicators increases and when they are of different types. This computational complexity is circumvented in a second approach by factorising the multivariate conditional distribution for the age indicators into univariate distributions, i.e. fitting a separate model for each age indicator. However, the resulting posterior distribution will be too narrow when the conditional independence assumption is violated and an adaption of the posterior distribution is required to obtain appropriate PIs [39]. Correspondingly, several authors rejected the conditional independence assumption, and stated that conditional dependence should be accounted for [9,39,43], which is supported by the current findings. Nevertheless, some authors did assume conditional independence between the different anatomical sites [8,10,44]. For instance, Bleka et al. (2018) reported a mean width of the 95% PI equal to 4.6 years for females and 4.5 years for males, when third molars and hand/wrist information was combined based on radiographs. Unfortunately, they did not verify the coverage.

The major limitation of the current study was its relatively small sample size. This led to the combination of certain (sub)stages to avoid insufficient numbers within stages, which on its turn would hinder analyses. Similarly, we decided to incorporate the wrist information of the VIBE sequence, and not the spin echo sequence, since the spin echo was not available in 10% of the study population. Therefore, when the spin echo wrist information would be considered instead of the VIBE, the added value of wrist information might be underestimated, due to the even smaller number of individuals with wrist information. Thus, to reflect on the full potential of the different age indicators, the original SSA studies [24,27,35] need to be considered, since they included larger study populations, allowing for more nuanced results.

The prospective nature of MFA studies using MRI explains their relatively small sample sizes. Therefore, future studies should investigate if the results of the few MFA studies that have been conducted (or are still ongoing) can safely be combined to increase the sample size. Such studies should bear in mind the need for an appropriate statistical approach to handle conditional dependence and differences in study populations.

In conclusion, MFA based on MRI of all third molars, the left wrist and both clavicles outperforms SSA or any combination of two age indicators. By contrast, adding anthropometric and sexual maturation data does not improve age estimation performance.

## Guarantor

The scientific guarantors of this publication are Koenraad Verstraete, Ghent University, and Patrick Thevissen, KU Leuven.

## Statistics and biometry

One of the authors has significant statistical expertise.

## Informed consent

Written informed consent was obtained from all participants in this study. In case of minors, written informed consent was also obtained from the parents.

## Ethical approval

The study was approved by the Ghent University Hospital Ethics Committee.

## Study subjects or cohorts overlap

Since the study reports the combined approach of the different anatomical sites, parts of the study population have been included in previously published reports. Per anatomical site, those previous reports considered the development of the MRI-protocols [23], the comparison of MRI with radiographs [34], the selection of an appropriate staging technique [24,27,34,35], and age estimation performance [24,27,35]. Moreover, with regard to the practical implementation of MRI for age estimation, two additional studies were conducted. Firstly, we compared our third molars MRI protocol with the one that was developed at the Ludwig Boltzmann Institut für Klinisch-Forensische Bildgebung in Graz, Austria [36]. Secondly, it was studied how motion affects clavicle MRI for age estimation [37].

Results described in this manuscript were orally presented at the American Academy of Forensic Sciences (AAFS) 71st Annual Scientific Meeting in Baltimore, USA, in February 2019. Moreover, they were included in the PhD thesis of the first author, which was defended on March 22nd, 2019 [De Tobel J (2019) Multi-factorial forensic age estimation: Combining magnetic resonance imaging of the third molars, the left wrist and both clavicles. PhD thesis, Ghent University - KU Leuven].

## Methodology

- prospective
- cross sectional study / observational
- performed at one institution

## Funding

Funding for this study was entirely provided by the Department of Diagnostic Sciences – Radiology at Ghent University and the Department of Imaging and Pathology – Forensic Odontology at KU Leuven.

## CRediT authorship contribution statement

**Jannick De Tobel:** Conceptualization, Data curation, Formal-analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing - original draft, Writing - review & editing. **Steffen Fieuws:** Conceptualization, Formal analysis, Methodology, Validation, Writing - review & editing. **Elke Hillewig:** Investigation, Methodology, Writing - review & editing. **Inès Phlypo:** Conceptualization, Investigation, Methodology, Writing - review & editing. **Mayonne van Wijk:** Investigation, Methodology, Writing - review & editing. **Michiel Bart de Haas:** Investigation, Methodology, Writing - review & editing. **Constantinus Politis:** Resources, Supervision, Writing - review & editing. **Koenraad Luc Verstraete:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing - review & editing. **Patrick Werner Thevissen:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - review & editing.

## Declaration of Competing Interest

None.

## Acknowledgements

The authors wish to thank all participants and people who helped with their recruitment. Furthermore, we want to thank Patrick Davis for his critical appraisal of the manuscript.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.forsciint.2019.110054.

## References

[1] A. Schmeling, C. Grundmann, A. Fuhrmann, H.J. Kaatsch, B. Knell, F. Ramsthaler, W. Reisinger, T. Riepert, S. Ritz-Timme, F.W. Rosing, K. Rotzscher, G. Geserick, Criteria for age estimation in living individuals, Int. J. Legal Med. 122 (6) (2008) 457–460.
[2] European Asylum Support Office (EASO), Practical Guide on Age Estimation, second edition, EASO Practical Guides Series, Malta, 2018.
[3] Office of Refugee Resettlement, Children Entering the United States Unaccompanied: Section 1, (2015) https://www.acf.hhs.gov/orr/resource/children-entering-the-united-states-unaccompanied-section-1. (Accessed on October 10, 2018).
[4] M. Haglund, H. Mornstad, A systematic review and meta-analysis of the fully formed wisdom tooth as a radiological marker of adulthood, Int. J. Legal Med. 133 (1) (2018) 231–239.
[5] P.W. Thevissen, Dental Age Estimation: Striving for an Optimal Approach, Leuven University Press, Leuven, 2013.
[6] K. Alshamrani, F. Messina, A.C. Offiah, Is the Greulich and Pyle atlas applicable to all ethnicities? A systematic review and meta-analysis, Eur. Radiol. (2019).
[7] C. Hermetet, P. Saint-Martin, A. Gambier, L. Ribier, B. Sautenet, C. Rerolle, Forensic age estimation using computed tomography of the medial clavicular epiphysis: a systematic review, Int. J. Legal Med. 132 (5) (2018) 1415–1425.
[8] O. Bleka, T. Wisloff, P.S. Dahlberg, V. Rolseth, T. Egeland, Advancing estimation of chronological age by utilizing available evidence based on two radiographical methods, Int. J. Legal Med. 133 (1) (2019) 217–229.
[9] J.L. Boldsen, G.R. Milner, L.W. Konigsberg, J.W. Wood, Transition analysis: a new method for estimating age from skeletons, in: R.D. Hoppa, J.W. Vaupel (Eds.), Paleodemography: Age Distributions from Skeletal Samples, Cambridge University Press, Cambridge, 2002, pp. 73–106.
[10] B. Gelbrich, C. Frerking, S. Weiss, S. Schwerdt, A. Stellzig-Eisenhauer, E. Tausche, G. Gelbrich, Combining wrist age and third molars in forensic age estimation: how to calculate the joint age estimate and its error rate in age diagnostics, Ann. Hum. Biol. 42 (4) (2015) 389–396.
[11] R.B. Bassed, C. Briggs, O.H. Drummer, Age estimation using CT imaging of the third molar tooth, the medial clavicular epiphysis, and the spheno-occipital synchondrosis: a multifactorial approach, Forensic Sci. Int. 212 (1-3) (2011) 273.e1-5.
[12] R. Cameriere, S. De Luca, R. Biagi, M. Cingolani, G. Farronato, L. Ferrante, Accuracy of three age estimation methods in children by measurements of developing teeth and carpals and epiphyses of the ulna and radius, J. Forensic Sci. 57 (5) (2012) 1263–1270.
[13] H. Demirturk Kocasarac, A. Sinanoglu, M. Noujeim, D. Helvacioglu Yigit, C. Baydemir, Radiologic assessment of third molar tooth and spheno-occipital synchondrosis for age estimation: a multiple regression analysis study, Int. J. Legal Med. 130 (3) (2016) 799–808.
[14] L. Kullman, Accuracy of two dental and one skeletal age estimation method in Swedish adolescents, Forensic Sci. Int. 75 (2-3) (1995) 225–236.
[15] M.A. Machado, E. Daruge Junior, M.M. Fernandes, I.F.P. Lima, G.O. Cericato, A. Franco, L.R. Paranhos, Effectiveness of three age estimation methods based on dental and skeletal development in a sample of young Brazilians, Arch. Oral Biol. 85 (2018) 166–171.
[16] V. Pinchi, F. De Luca, M. Focardi, F. Pradella, G. Vitale, F. Ricciardi, G.A. Norelli, Combining dental and skeletal evidence in age classification: pilot study in a sample of Italian sub-adults, Leg. Med. (Tokyo) 20 (2016) 75–79.
[17] L. Shi, F. Jiang, F. Ouyang, J. Zhang, Z. Wang, X. Shen, DNA methylation markers in combination with skeletal and dental ages to improve age estimation in children, Forensic Sci. Int. Genet. 33 (2017) 1–9.
[18] P.W. Thevissen, J. Kaur, G. Willems, Human age estimation combining third molar and skeletal development, Int. J. Legal Med. 126 (2) (2012) 285–292.
[19] A. Kumagai, G. Willems, A. Franco, P. Thevissen, Age estimation combining radiographic information of two dental and four skeletal predictors in children and subadults, Int. J. Legal Med. 132 (6) (2018) 1769–1777.
[20] D. Stern, C. Payer, N. Giuliani, M. Urschler, Automatic age estimation and majority age classification from multi-factorial MRI data, IEEE J. Biomed. Health Inform. 23 (4) (2019) 1392–1403.
[21] D. Gascho, M.J. Thali, T. Niemann, Post-mortem computed tomography: technical principles and recommended parameter settings for high-resolution imaging, Med. Sci. Law 58 (1) (2018) 70–82.
[22] P. Baumann, T. Widek, H. Merkens, J. Boldt, A. Petrovic, M. Urschler, B. Kirnbauer, N. Jakse, E. Scheurer, Dental age estimation of living persons: comparison of MRI with OPG, Forensic Sci. Int. 253 (0) (2015) 76–80.
[23] J. De Tobel, E. Hillewig, S. Bogaert, K. Deblaere, K. Verstraete, Magnetic resonance imaging of third molars: developing a protocol suitable for forensic age estimation, Ann. Hum. Biol. 44 (2) (2017) 130–139.
[24] J. De Tobel, I. Phlypo, S. Fieuws, C. Politis, K.L. Verstraete, P.W. Thevissen, Forensic age estimation based on development of third molars: a staging technique for magnetic resonance imaging, J. Forensic Odontostomatol. 35 (2) (2017) 117–140.

[25] Y. Guo, A. Olze, C. Ottow, S. Schmidt, R. Schulz, W. Heindel, H. Pfeiffer, V. Vieth, A. Schmeling, Dental age estimation in living individuals using 3.0 T MRI of lower third molars, Int. J. Legal Med. 129 (6) (2015) 1265–1270.
[26] T. Widek, P. Genet, H. Merkens, J. Boldt, A. Petrovic, J. Vallis, E. Scheurer, Dental age estimation: the chronology of mineralization and eruption of male third molars with 3T MRI, Forensic Sci. Int. 297 (2019) 228–235.
[27] J. De Tobel, E. Hillewig, M. van Wijk, S. Fieuws, M.B. de Haas, R.R. van Rijn, P.W. Thevissen, K.L. Verstraete, Staging clavicular development on magnetic resonance imaging: pitfalls and recommendations for age estimation, J. Magn. Reson. Imaging (2019) [Epub ahead of print].
[28] E. Hillewig, J. De Tobel, O. Cuche, P. Vandemaele, M. Piette, K. Verstraete, Magnetic resonance imaging of the medial extremity of the clavicle in forensic bone age determination: a new four-minute approach, Eur. Radiol. 21 (4) (2011) 757–767.
[29] E. Hillewig, J. Degroote, T. Van der Paelt, A. Visscher, P. Vandemaele, B. Lutin, L. D'Hooghe, V. Vandriessche, M. Piette, K. Verstraete, Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: towards more sound age estimates, Int. J. Legal Med. 127 (3) (2013) 677–689.
[30] S. Schmidt, C. Ottow, H. Pfeiffer, W. Heindel, V. Vieth, A. Schmeling, R. Schulz, Magnetic resonance imaging-based evaluation of ossification of the medial clavicular epiphysis in forensic age assessment, Int. J. Legal Med. 131 (6) (2017) 1665–1673.
[31] S. Tangmose, K.E. Jensen, C. Villa, N. Lynnerup, Forensic age estimation from the clavicle using 1.0T MRI-preliminary results, Forensic Sci. Int. 234 (2014) 7–12.
[32] P.M. Garamendi, M.I. Landa, J. Ballesteros, M.A. Solano, Reliability of the methods applied to assess age minority in living subjects around 18 years old. A survey on a Moroccan origin population, Forensic Sci. Int. 154 (1) (2005) 3–12.
[33] M. Auf der Mauer, D. Saring, B. Stanczus, J. Herrmann, M. Groth, E. Jopp-van Well, A 2-year follow-up MRI study for the evaluation of an age estimation method based on knee bone development, Int. J. Legal Med. 133 (1) (2018) 205–215.
[34] J. De Tobel, E. Hillewig, K. Verstraete, Forensic age estimation based on magnetic resonance imaging of third molars: converting 2D staging into 3D staging, Ann. Hum. Biol. 44 (2) (2017) 121–129.
[35] J. De Tobel, E. Hillewig, M.B. de Haas, B. Van Eeckhout, S. Fieuws, P. Thevissen, K. Verstraete, Forensic age estimation based on T1 SE and VIBE wrist MRI: do a one-fits-all staging technique and age estimation model apply? Eur. Radiol. 26 (6) (2019) 2924–2935.
[36] J. De Tobel, G.I.L. Parmentier, I. Phlypo, B. Descamps, S. Neyt, W.L. Van De Velde, K.L. Verstraete, P.W. Thevissen, Magnetic resonance imaging of third molars in forensic age estimation: comparison of the Ghent and Graz protocols focusing on apical closure, Int. J. Legal Med. 133 (2) (2019) 583–592.
[37] J. De Tobel, M. van Wijk, I. Alberink, E. Hillewig, I. Phlypo, R.R. van Rijn, P. Thevissen, K. Verstraete, M.B. de Haas, The influence of motion artifacts on magnetic resonance imaging of the clavicles for age estimation, Int. J. Legal Med. (2019) Accepted on November 17, 2019.
[38] J. Adams, N. Shaw, A Practical Guide to Bone Densitometry in Children, National Osteoporosis Society, Camerton, Bath, UK, 2004.
[39] S. Fieuws, G. Willems, S. Larsen-Tangmose, N. Lynnerup, J. Boldsen, P. Thevissen, Obtaining appropriate interval estimates for age when multiple indicators are used: evaluation of an ad-hoc procedure, Int. J. Legal Med. 130 (2) (2016) 489–499.
[40] D. Wittschieber, R. Schulz, V. Vieth, M. Kuppers, T. Bajanowski, F. Ramsthaler, K. Puschel, H. Pfeiffer, S. Schmidt, A. Schmeling, Influence of the examiner's qualification and sources of error during stage determination of the medial clavicular epiphysis by means of computed tomography, Int. J. Legal Med. 128 (1) (2014) 183–191.
[41] G.Y. Levesque, A. Demirjian, The inter-examiner variation in rating dental formation from radiographs, J. Dent. Res. 59 (7) (1980) 1123–1126.
[42] P.W. Thevissen, S. Fieuws, G. Willems, Human dental age estimation using third molar developmental stages: does a Bayesian approach outperform regression models to discriminate between juveniles and adults? Int. J. Legal Med. 124 (1) (2010) 35–42.
[43] M. Seselj, Relationship between dental development and skeletal growth in modern humans and its implications for interpreting ontogeny in fossil hominins, Am. J. Phys. Anthropol. 150 (1) (2013) 38–47.
[44] P. Mostad, F. Tamsen, Error rates for unvalidated medical age assessment procedures, Int. J. Legal Med. 133 (2) (2019) 613–623.