European Radiology (2019) 29:2924–2935
https://doi.org/10.1007/s00330-018-5944-7

FORENSIC MEDICINE

Forensic age estimation based on T1 SE and VIBE wrist MRI: do a one-fits-all staging technique and age estimation model apply?

Jannick De Tobel[1,2,3] · Elke Hillewig[1] · Michiel Bart de Haas[4] · Bram Van Eeckhout[1] · Steffen Fieuws[5] · Patrick Werner Thevissen[2] · Koenraad Luc Verstraete[1]

Received: 9 September 2018 / Revised: 31 October 2018 / Accepted: 3 December 2018 / Published online: 7 January 2019

Abstract

Objectives Providing recommendations for wrist MRI in age estimation by determining (1) which anatomical structures to include in the statistical model, (2) which MRI sequence to conduct, and (3) which staging technique to apply.

Methods Radius and ulna were prospectively studied on 3 T MRI in 363 healthy Caucasian participants (185 females, 178 males) between 14 and 26 years old, using T1 spin echo (SE) and T1 gradient echo VIBE. Bone development was assessed applying a 5-stage staging technique with several amelioration attempts to optimise staging. A Bayesian model rendered point predictions of age and diagnostic indices to discern minors from adults.

Results All approaches rendered similar results, with none of them outperforming the others. A single bone assessment of radius or ulna sufficed. SE and VIBE sequences were both suitable, but needed sequence-specific age estimation. A one-fits-all 5-stage staging technique—with substages in stage 3—was suitable and did not benefit from profound substaging. Age estimation based on SE radius resulted in a mean absolute error of 1.79 years, a specificity (correctly identified minors) of 93%, and a discrimination slope of 0.640.

Conclusion Radius and ulna perform similarly to estimate age, and so do SE and VIBE. A one-fits-all staging technique can be applied.

Key Points

- *Radius and ulna perform similarly to estimate age.*
- *SE and VIBE perform similarly, but age estimation should be based on the corresponding sequence-specific reference data.*
- *A one-fits-all 5-stage staging technique with substages 3a, 3b, and 3c can be applied to both bones and both sequences.*

Keywords Age determination by skeleton · Wrist · Adolescent · Adult · Magnetic resonance imaging

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-018-5944-7) contains supplementary material, which is available to authorized users.

✉ Jannick De Tobel
jannick.detobel@ugent.be

[1] Department of Radiology and Nuclear Medicine, Ghent University, Corneel Heymanslaan 10, 9000 Ghent, Belgium

[2] Department of Imaging and Pathology – Forensic Odontology, KU Leuven, Leuven, Belgium

[3] Department of Head, Neck and Maxillofacial Surgery, Ghent University Hospital, Ghent, Belgium

[4] Department of Forensic Anthropology, Netherlands Forensic Institute, Den Haag, The Netherlands

[5] Department Public Health and Primary Care, KU Leuven – University of Leuven & Universiteit Hasselt, I-BioStat, Leuven, Belgium

Abbreviations

CI	Confidence interval
END	End stage
MAE	Mean absolute error
SAR	Specific absorption rate
SE	T1-weighted spin echo MR-sequence
TFS	Threefold stratification sign
VIBE	T1-weighted gradient echo volumetric interpolated breath-hold examination MR-sequence

Introduction

Established methods for forensic age estimation in the living use radiographs and computed tomography (CT) to study the development of anatomical structures. However, the use of ionising radiation limits their forensic applicability in several

countries [1]. Consequently, magnetic resonance imaging (MRI) has been studied for age estimation using different approaches [2–6]. Therefore, evidence-based recommendations are necessary to implement MRI in age estimation practice. Three aspects can be considered for this implementation: anatomical structures to include, MRI sequence types, and staging technique. The current study aims to provide recommendations for each of these aspects.

Anatomical structures

Age estimation focusing on the age threshold of 18 years should include analysing wrist development [7–9]. Using MRI, 11 developing hand/wrist features were used by Urschler et al (2015) [10]. By contrast, Serin et al (2016) demonstrated that age estimation results based on combined developmental information of distal radius, distal ulna, and proximal first metacarpal bone did not outperform those based on the distal radius alone [11]. Consequently, other authors only analysed the distal radius [6, 11–17]. Because of these varying approaches in literature and since the distal ulna and radius are the last regions to mature in the hand/wrist, the current study aims to analyse both of them, to verify which structure should be included for an optimal age estimation.

MRI sequence types

Two MRI sequence types have been reported to study hand/wrist development: T1-weighted spin echo images (SE) without fat suppression [6, 11, 13–19] and T1-weighted gradient echo VIBE images (VIBE) with fat suppression [5, 10, 18, 19]. Since the relative value of both sequence types remains unclear, they were both analysed in the current study.

Staging techniques for wrist development

Different staging techniques for wrist development have been applied to MRI. The most elaborate staging technique for long bone development was established for clavicle CT and includes five stages with several substages and profound substaging (Table 1) [16, 20–22]. However, those substages might not be consecutive, questioning their relevance for age estimation [6, 23, 24]. Additionally, on MRI, a threefold stratification sign (TFS) could be of interest to subdivide stages 3c and 4, but its relevance to age estimation needs to be explored [6, 16]. Hence, the current study aims to provide a comprehensive staging technique, keeping the balance between a sufficient number of stages and sufficient reproducibility [25].

Materials and methods

Study population

Healthy Caucasian volunteers were prospectively included between January 2008 and May 2017. Part of this sample has been used in previous studies [2, 3, 26–28]. The current study population included for analysis comprised 363 individuals (185 females, 178 males) between 14 and 26 years old (Table 2). Ethical approval and informed consent were obtained.

Image acquisition

Three Tesla MRI of the left wrist was conducted using a Siemens scanner (Magnetom Trio Tim, Siemens) with two MRI sequences:

- SE (TR 534 ms, TE 16 ms, flip angle 150°, no fat suppression, voxel size $0.20 \times 0.20 \times 2.0$ mm^3, 2:43 min);
- VIBE (TR 12.7 ms, TE 4.82 ms, flip angle 10°, fat suppression, voxel size $0.40 \times 0.40 \times 0.40$ mm^3, 5:57 min).

Participants were positioned supine, with their left wrist in a dedicated 8-channel coil (Invivo).

Image analysis

MRI datasets were anonymised and blinded for the age of the participants. Three observers assessed different parts of the datasets independently (Fig. 1). Before starting to stage images for the current study, randomly selected cases stratified on age categories of 1 year were discussed at an initial training session. Then, the most experienced observer (JDT) assessed the whole study sample. To study intra-observer agreement, he assessed part of the sample ($N = 98$) a second time after 9 months. Two other less experienced observers (MdH and BVE) assessed the whole study sample complementarily. To study inter-observer agreement, their staging results were jointly considered and compared with the staging results of the first observer.

Images were displayed on radiological monitors with resolutions of 3280×2048 pixels (MDCC-6130, Barco) or 1280×1024 pixels (MFCD 1219, Barco). The most elaborate staging technique for long bone development was applied (Table 1, Figs. 2 and 3) [16, 20–22]. Stages were allocated scrolling through the whole stack of slices. When deciding between stages 2 and 3 (including substages), stage 3 was allocated as soon as one bone bridge was visible on any image. When deciding between stages 4 and 5, stage 4 was allocated when the physeal scar or its remnants were observed on at least one image.

Table 1 Descriptive criteria for developmental stages of long bones on CT or MRI. Most elaborate staging reported in literature so far

1	Ossification centre is invisible (= not yet ossified).
2	Ossification centre is visible (= ossified), non-union of the epiphysis and metaphysis.
2a	The lengthwise epiphyseal measurement is one third or less compared to the widthwise measurement of the metaphyseal ending.
2b	The lengthwise epiphyseal measurement is between one third and two thirds compared to the widthwise measurement of the metaphyseal ending.
2c	The lengthwise epiphyseal measurement is over two thirds compared to the widthwise measurement of the metaphyseal ending.
3	Physeal plate is partially ossified (= bone trabeculae cross the physeal plate from ossification centre to metaphysis).
3a	The epiphyseal-metaphyseal fusion completes one third or less of the former gap between epiphysis and metaphysis.
3aa	Lengthwise measurement of the epiphysis is one third or lower compared with the widthwise measurement of the metaphyseal ending.
3ab	Lengthwise measurement of the epiphysis is between one third and two thirds compared with the widthwise measurement of the metaphyseal ending.
3ac	Lengthwise measurement of the epiphysis is over two thirds compared with the widthwise measurement of the metaphyseal ending.
3b	The epiphyseal-metaphyseal fusion completes between one third and two thirds of the former gap between epiphysis and metaphysis.
3c	The epiphyseal-metaphyseal fusion completes over two thirds of the former gap between epiphysis and metaphysis.
4	Complete union of the epiphysis and metaphysis (= physeal plate is completely ossified). Physeal scar is still visible.
4a	Threefold stratification sign is visible.
4b	No threefold stratification sign.
5	Complete union of the epiphysis and metaphysis. Physeal scar is indiscernible.

Two types of amelioration attempts for the staging technique were applied. Firstly, the influence of merging (sub) stages was tested. One possibility to merge stages was to discard the substages in stage 3. A second possibility was to jointly consider stages 4 and 5 as one stage.

Secondly, the influence of splitting up (sub) stages was tested. To start with, two possibilities to split up stage 3c were explored. A first possibility was by registering when an end stage of stage 3c was allocated (stage 3c END), i.e. when only one or two consecutive slices were suggestive for a small physeal remnant. A second possibility was by registering the presence of the TFS (Fig. 2), which was also registered when stage 4 was allocated. Note that the TFS was only assigned when a clear parallel tram-track-like sign was distinguishable from other trabecular patterns (Fig. 4).

Furthermore, three confounding factors for staging were documented mimicking a lateral remnant of the physeal plate: blood vessels, persisting lateral fissures, or surrounding soft tissue (supplementary Fig. A1). Finally, if images were not assessable, the reason was registered.

Table 2 Number of participants per age per sex. Age categories of 1 year were considered from the birthday until the day before the next birthday. For instance, the age of 14-year-old participants could range from 14.00 to 14.99

Age (years)	Frequency		
	Female	Male	Total
14	9	11	20
15	11	10	21
16	18	17	35
17	18	17	35
18	18	19	37
19	20	20	40
20	22	15	37
21	21	12	33
22	9	10	19
23	9	10	19
24	9	11	20
25	11	11	22
26	10	15	25
Total	185	178	363

Statistical analysis

SPSS Statistics 25.0 (IBM SPSS Statistics for Windows) and SAS 9.4 (SAS Institute) were used for statistical analyses. Descriptive statistics were calculated, and statistical tests were performed two-sided with the significance level equal to 0.05. Unless otherwise stated, reported results were based on the first observer's assessments. Females and males were considered together, since similar results were obtained when sex-specific results were calculated.

Weighted kappa statistics and tests for symmetry (extension of the McNemar test) were used to evaluate the inter-observer reliability of staging and the agreement between staging obtained with SE and VIBE.

A linear model with three factors (sex, stage, and sequence) was used to compare the mean age of participants with their radius or ulna in particular (sub) stages and to verify if the difference in mean age depended on sequence. Stages obtained with both sequences were considered as independent samples for this analysis, although the same subject appeared twice. In the model, the variance of age was allowed to differ between the stages. Independent *t* tests were used to compare age between two stages. A one-way ANOVA was used to compare age between more than two stages.

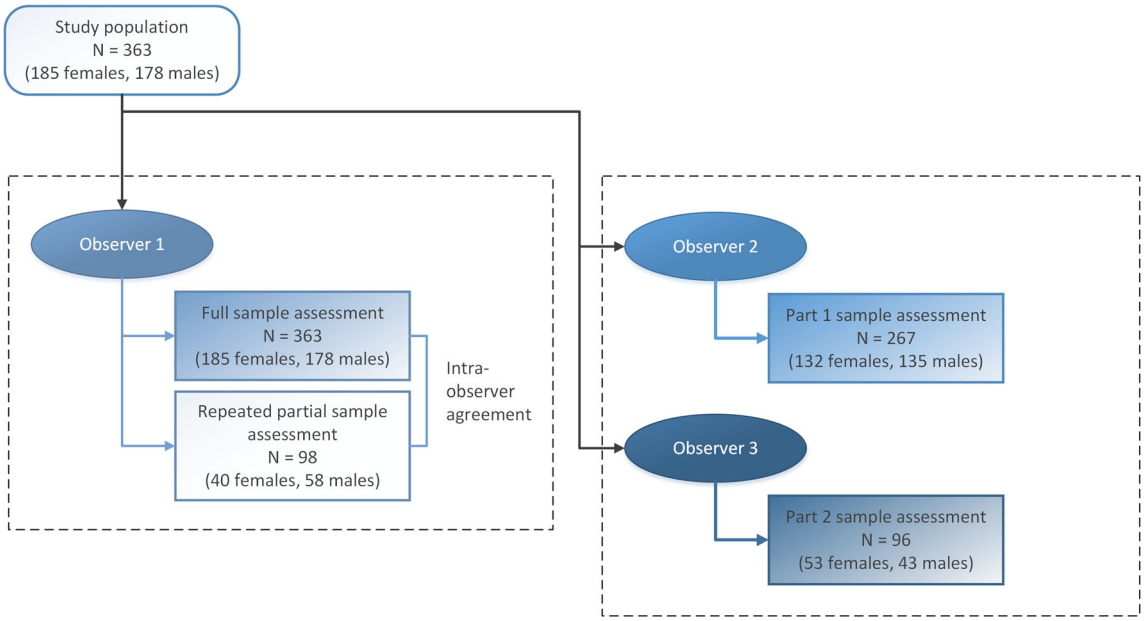

Fig. 1 Schematic overview of the MRI assessments by three observers. Observer 1 assessed part of the sample twice to study intra-observer agreement. Observers 2 and 3 assessed the whole study sample complementarily. Therefore, they were jointly considered to study inter-observer agreement

Age estimation was based on the application of Bayes' rule, with a continuation ratio model for the distribution of the stage [29, 30]. Tenfold cross-validation was used to handle overfitting. Two aspects of age estimation were studied to quantify and compare performance: (1) point prediction for age and (2) the ability to discern minors from adults.

Regarding the first aspect, the modus is the 'most likely' age estimate. However, it is known not to minimise the error [31]. Therefore, the mean was used as point prediction, more specifically the 5% trimmed mean, i.e. the mean in the 95% prediction interval. The mean absolute error (MAE) was evaluated, as measure of accuracy. The Wilcoxon signed rank test was used to compare the MAE.

Regarding the second aspect, diagnostic indices for the ability to discern minors from adults were calculated as accuracy (proportion of correctly classified participants), sensitivity (proportion of correctly classified adults), and specificity (proportion of correctly classified minors), as well as the discrimination slope (the difference between minors and adults in the predicted probability of being an adult).

Besides studying the different amelioration attempts, sequence-specificity of the age estimation model was tested by applying the VIBE model to the SE data and vice versa.

Results

Assessing the images was impossible in three cases due to motion artefacts or a poor signal-to-noise ratio. By contrast, confounding factors did not hinder staging, although they

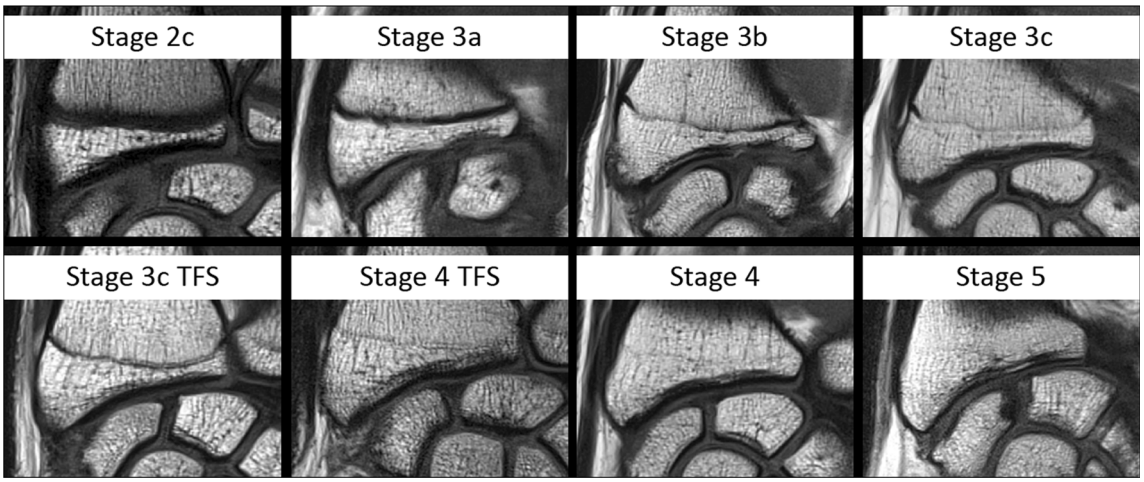

Fig. 2 Representative examples of radii in consecutive developmental stages on SE. TFS, threefold stratification sign

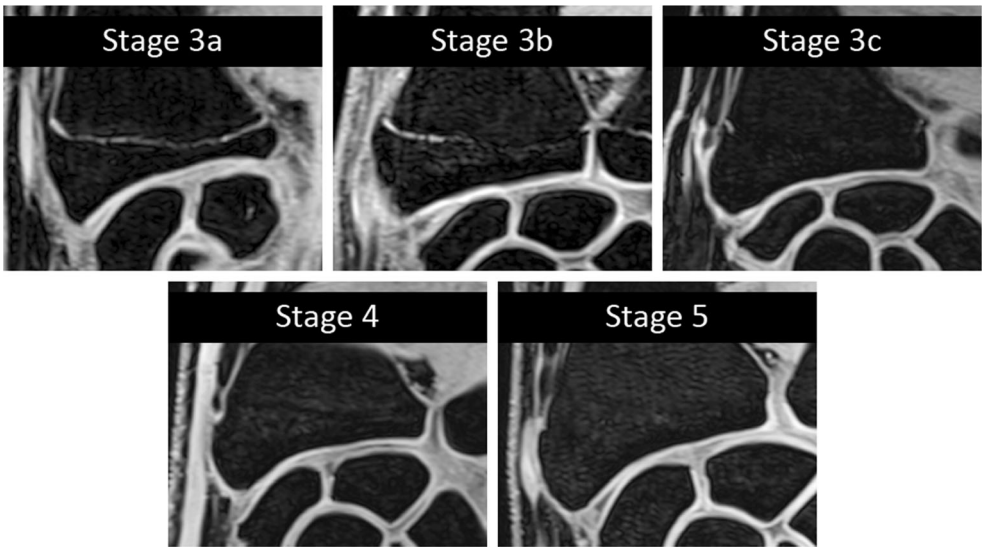

Fig. 3 Representative examples of radii in consecutive developmental stages on VIBE

were often seen. Their frequencies were included in supplementary Table A1.

Stages 2c, 3ac, 3b, 3c, 4, and 5 were allocated, with stage 4 rarely being allocated on VIBE (6.1% = 88 times), while stage 5 was rarely allocated on SE (4.3% = 62 times) (Fig. 4). For comparison with literature, descriptive statistics per stage were included in supplementary Tables A2 to A5.

Reproducibility of staging

No systematic inter- and intra-observer differences were encountered, and all kappa values were ≥ 0.70 (Table 3). One-stage differences occurred frequently. A three-stage difference was encountered twice between observers and a four-stage difference was encountered three times between observers. Within observers, the largest difference was a two-stage difference.

Anatomical structures

Table 4 summarises age estimation performance of the different tested approaches. Regarding the accuracy of point prediction, no statistically significant differences could be demonstrated between radius and ulna separately. Although combined information of radius and ulna statistically significantly outperformed separate information per structure, this was clinically insignificant, since the MAE differences ranged from − 0.09 to 0.16.

Regarding the ability to discern minors from adults, the highest specificity (93%) was obtained for the radius. By contrast, the highest discrimination slopes were obtained with combined information of radius and ulna.

MRI sequence types

The observers judged staging on VIBE to be easier than on SE because of the high contrast with adjacent bone (Fig. 3), and the lack of a chemical shift artefact, which caused blurring on the SE images (Fig. 2). Conversely, the physeal scar was hard to discern on VIBE (Fig. 4).

Consequently, staging differed systematically between SE and VIBE, with lower stages being allocated on SE than VIBE ($p < 0.0001$ in the symmetry tests for radius as well as for ulna). Weighted kappas and their 95% confidence interval

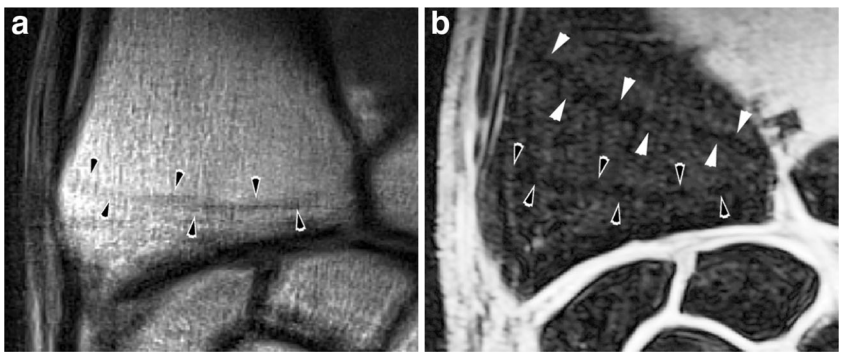

Fig. 4 **a** SE radius in stage 4. The physeal scar is clearly visible (black/white arrowheads). **b** Corresponding VIBE. A possible physeal scar is hardly discernible (black/white arrowheads). A similar trabecular pattern is visible more proximally (white arrowheads). Therefore, stage 5 was allocated

Table 3 Reproducibility of staging radius and ulna development per MRI sequence

Anatomical structure	Sequence	Inter-observer 1–2			Inter-observer 1–3			Intra-observer		
		p value	Weighted kappa (95% CI)	Number of cases	*p* value	Weighted kappa (95% CI)	Number of cases	*p* value	Weighted kappa (95% CI)	Number of cases
Radius	SE	0.371	0.796 (0.744; 0.848)	266	1.000	0.770 (0.679; 0.861)	95	1.000	0.856 (0.781; 0.931)	98
	VIBE	0.268	0.699 (0.628; 0.770)	265	0.317	0.696 (0.593; 0.798)	94	1.000	0.843 (0.764; 0.923)	98
Ulna	SE	0.447	0.794 (0.747; 0.842)	264	1.000	0.836 (0.764; 0.909)	95	0.999	0.822 (0.742; 0.901)	98
	VIBE	0.949	0.785 (0.720; 0.849)	264	0.412	0.774 (0.694; 0.854)	94	0.998	0.836 (0.761; 0.910)	98

CI confidence interval, *p value p* value for test of symmetry (extension of the McNemar test)

(CI) equalled 0.550 (0.504; 0.595) and 0.595 (0.551; 0.638) for radius and ulna, respectively, confirming the differences between both sequence types. One-stage differences occurred frequently, whereas two-stage differences were encountered only twice for radius and twice for ulna. Hence, applying the sequence-specific model rendered better point prediction accuracy and diagnostic indices (Table 4).

Regarding the accuracy of point prediction, MAE did not differ significantly between SE and VIBE within one anatomical structure. Neither did the diagnostic indices for discerning minors from adults differ systematically between sequence types (Table 4).

Staging techniques for wrist development

Two aspects of the amelioration attempts need clarification. First, note that the TFS was only visible in SE, since it is part of the physeal scar. Neither in stage 3c nor in stage 4 did the TFS improve age estimation performance. Second, it was noticed that radii in SE stage 4 showed a dispersion in VIBE stages from 3c to 5. Thus, one amelioration attempt was to add VIBE information to SE information in SE stage 4. To illustrate the Bayesian approach, Fig. 5 displays posterior density curves of age for SE radius with added VIBE information in males.

Regarding point prediction, none of the amelioration attempts remarkably outperformed the original staging technique (Table 4). Although statistically significant differences were obtained between suggested substages for mean age (Table 5) as well as for MAE (Table 4), the differences in MAE were too small to be of clinical significance (range −0.02 to 0.24 years). Note that merging stages rendered the least accurate point predictions.

Similarly, regarding the ability to discern minors from adults, none of the amelioration attempts remarkably outperformed the original staging technique. Merging SE radius stages 4 and 5 considerably decreased specificity (Table 4), suggesting that these stages should be considered separately. By contrast, merging SE radius stage 3 rendered the highest specificity, but caused a major decrease in sensitivity (Table 4), suggesting that substages in stage 3 should also be considered separately. This avoids an artificially high number of wrongly identified adults that would only marginally increase the number of correctly identified minors.

For comparison with literature, probabilities to be an adult per stage per sex were included in supplementary Tables A6 to A8. Remarkably, looking at these probabilities for the different amelioration attempts, stage 3c END still seems valuable for the radius, although this is not reflected by the diagnostic indices in Table 4. For instance, the probability to be an adult was 0.937 for males with SE radius stage 3c END. When the 3c END radii were excluded from stage 3c, that probability dropped from 0.737 to 0.664.

Table 4 Age estimation performance based on radius and ulna development, per MRI sequence and attempts to ameliorate the staging technique

Anatomical structure	Sequence	Amelioration attempt	Accuracy of point prediction					Discerning minors from adults			
			Considering all cases	Considering the subset of cases within the relevant stages				Considering all cases			
			MAE (years)	N	MAE reference[#]	MAE amelioration	*p* value MAE difference[‡]	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Discrimination slope
Radius + Ulna	SE	None	1.72		NA			91.2 (87.8; 93.9)	94.0 (90.4; 96.6)	84.7 (76.6; 90.8)	0.725
	VIBE	None	1.78		NA			92.3 (89.0; 94.8)	95.6 (92.3; 97.8)	84.7 (76.6; 90.8)	0.721
Radius	SE	None	1.79		NA		NA	89.8 (86.2; 92.7)	88.5 (83.9; 92.2)	92.8 (86.3; 96.8)	0.640
	VIBE	None	1.90		NA		NA	89.8 (86.2; 92.7)	90.9 (86.6; 94.1)	87.4 (79.7; 92.9)	0.584
	SE	Merge stage 3	1.95	150	1.17	1.56	< 0.001	86.2 (82.2; 89.6)	81.0 (75.6; 85.6)	98.2 (93.6; 99.8)	0.511
	VIBE	Merge stage 3	2.03	143	1.34	1.66	< 0.001	87.9 (84.1; 91.1)	84.9 (79.9; 89.1)	94.6 (88.6; 98.0)	0.492
	SE	Merge stages 4 and 5	1.93	172	2.25	2.26	0.183[†]	86.1 (82.2; 89.5)	94.4 (90.8; 96.9)	67.6 (58.0; 76.1)	0.571
	VIBE	Merge stages 4 and 5	2.02	187	2.36	2.35	0.687[†]	85.2 (81.1; 88.7)	85.1 (80.1; 89.3)	85.5 (77.5; 91.5)	0.494
	SE	Stage 3c END	1.78	80	1.40	1.35	0.707	92.0 (88.7; 94.6)	92.5 (88.5; 95.4)	91.0 (84.1; 95.6)	0.683
	VIBE	Stage 3c END	1.87	74	1.66	1.53	0.042	89.3 (85.6; 92.2)	93.3 (89.4; 96.0)	80.2 (71.5; 87.1)	0.615
	SE	Stage 3c TFS	1.80	80	1.40	1.42	0.584	89.0 (85.3; 92.0)	88.5 (83.9; 92.2)	90.1 (83.0; 94.9)	0.647
	SE	Stage 4 TFS	1.80	205	2.27	2.29	< 0.001	89.8 (86.2; 92.7)	88.5 (83.9; 92.2)	92.8 (86.3; 96.8)	0.640
	SE	Add VIBE info to SE stage 4	1.77	205	2.27	2.23	0.162	89.8 (86.2; 92.7)	88.5 (83.9; 92.2)	92.8 (86.3; 96.8)	0.640
Ulna	SE	None	1.88		NA			91.7 (88.4; 94.4)	94.0 (90.4; 96.6)	86.5 (78.7; 92.2)	0.652
	VIBE	None	1.82		NA			92.3 (89.0; 94.8)	95.6 (92.3; 97.8)	84.7 (76.6; 90.8)	0.676
	SE	Merge stage 3	1.98	109	0.96	1.29	< 0.001	89.5 (85.9; 92.5)	90.9 (86.6; 94.1)	86.5 (78.7; 92.2)	0.574
	VIBE	Merge stage 3	2.03	124	1.03	1.40	< 0.001	85.4 (81.3; 88.9)	82.9 (77.7; 87.4)	91.0 (84.1; 95.6)	0.465
	SE	Merge stages 4 and 5	1.93	228	2.33	2.38	0.082	88.0 (84.2; 91.2)	90.7 (86.4; 94.0)	82.0 (73.6; 88.6)	0.584
	VIBE	Merge stages 4 and 5	2.05	232	2.35	2.35	0.127	88.5 (84.8; 91.7)	92.7 (88.8; 95.6)	79.1 (70.3; 86.3)	0.526
	SE	Stage 3c END			NA[§]						
	VIBE	Stage 3c END	1.83	48	0.97	1.00	0.705	92.3 (89.0; 94.8)	95.6 (92.3; 97.8)	84.7 (76.6; 90.8)	0.672
	SE	Stage 3c TFS			NA[§]						
	SE	Stage 4 TFS	1.84	224	2.36	2.29	0.103	91.7 (88.4; 94.4)	94.0 (90.4; 96.6)	86.5 (78.7; 92.2)	0.652
	SE	Add VIBE info to SE stage 4	1.78	224	2.36	2.18	0.031	92.6 (89.4; 95.0)	93.7 (89.9; 96.3)	90.1 (83.0; 94.9)	0.682
Radius	SE	VIBE model*	1.98	355	1.81	1.97	0.006	89.6 (85.9; 92.6)	88.4 (83.8; 92.1)	92.3 (85.4; 96.6)	0.536
	VIBE	SE model°	1.43	165	1.39	1.43	0.166	81.2 (74.4; 86.9)	61.7 (48.2; 73.9)	92.4 (85.5; 96.7)	0.414

END end stage, *CI* confidence interval, *N* total number of participants in relevant stages, *NA* not applicable, *MAE* mean absolute error, *TFS* threefold stratification sign

[#] Reference for comparison is the original equivalent (radius or ulna; SE or VIBE). [‡] *p* value from Wilcoxon signed rank test comparing MAE. [†] Based on the second observer's assessment. [§] Too few cases in one category. *Restricted to SE stages 3ac, 3b, 3c, and 4 (too few subjects in stage 5). Applying the SE-sequence-specific model in these cases resulted in MAE = 1.81. ° Restricted to VIBE stages 3ac, 3b, 3c, and 4 (too few subjects in stage 5). Applying the VIBE-sequence-specific model in these cases resulted in MAE = 1.29

Discussion

Reproducibility of staging

Inter-observer agreements with Cohen's kappa ranging from 0.23 [32] to 0.97 [6] have been reported in literature on wrist MRI for age estimation, with kappa > 0.80 in most studies [6, 11, 13, 16, 17, 32, 33]. The inter-observer agreements in the current study ranged from 0.71 to 0.81, which might be due to the difference in experience [34] and reflects the difficulty of the interpretation.

Anatomical structures

Combined assessment of radius and ulna did not outperform separate assessments, neither regarding the accuracy of the point prediction nor regarding the ability to discern minors from adults. This corresponds with findings of Serin et al (2016) [11]. Conversely, Urschler et al (2015) included hand and wrist bones on VIBE [10], suggesting that hand bones could add age information to the wrist model. However, no comparing studies have been conducted so far between hand/wrist and wrist alone.

MRI sequence types

Both wrist SE and VIBE have been reported in literature on age estimation. The subjective finding that the physeal plate appears more clear-cut on VIBE corresponds with the findings of Urschler et al (2016) [18] and Hojreh et al (2018) [19] and is caused by fat suppression. Conversely, SE clearly depicts the physeal scar. Together with the blurry appearance of the physeal plate, the latter can be explained by the chemical shift artefact [35], which causes an enlarged appearance of the physeal cartilage. Fat suppression in VIBE avoids chemical shift artefacts. These differences result in different age estimation models. Therefore, MRI sequence-specific reference data should be used. This has also been demonstrated for age estimation based on third molars' MRI [36].

Combining information of different sequences is common in a clinical context, but has only once been suggested in forensic age estimation by Vieth et al [37]. Their combined staging technique allowed discerning minors from adults based on the distal femoral epiphysis. By contrast, the current study did not render better results by adding VIBE information to SE.

Moreover, the high field MRI might not even be necessary to study hand/wrist development. Several authors have successfully applied open 0.2–0.3 T scanners to study the hand/wrist for age estimation [38–43]. However, only one of those studies included participants older than 18 [39]. An open scanner is more appropriate for children, who might be scared of lying in the small bore of a large scanner. By contrast, 3 T MRI seemed appropriate in the current study, since it focussed on subadults and young adults and was part of a larger project that also included MRI of the third molars [2] and clavicles.

Staging techniques for wrist development

The current study confirmed the applicability of the 5-stage staging technique to assess radius and ulna development, including substages to stage 3. Merging stages decreased age estimation performance, while splitting up stages did not ameliorate age estimation performance.

Splitting up clavicular stage 3a into 3aa, 3ab, and 3ac has been suggested on clavicle CT [22]. However, in the current study, only stage 3ac was seen. Furthermore, two possibilities to split up stage 3c were encountered. A first possibility was by registering END, which might valuable to correctly identify minors. In judicial context, this is the major concern. However, only the probabilities to be adult suggested any value of this substage. A larger study sample will be needed to verify the value of stage 3c END. A second possibility to split up stage 3c was by registering the TFS on SE [16]. The TFS can be explained by the laminar structure of the maturing metaphysis-physis-epiphysis complex [44]. It has been hypothesised that this sign occurs gradually before the physeal scar condenses into a single line.

Accordingly, it has been suggested to split up stage 4, with or without a TFS [6]. However, current results suggest that the TFS is of no value for age estimation. Even our 25- and 26-year-olds demonstrated no trend regarding the presence (or disappearance) of the TFS, which contrasts with Timme et al (2017), who only included subjects until age 24. Another possibility to split up stage 4 on SE radius was to add VIBE information. This allowed a further differentiation of age within that group, but did not outperform the original approach. Unfortunately, no other studies have combined SE and VIBE information of the wrist to estimate age.

Recommendations and future prospects

The current findings and the available literature demonstrated that no anatomical structure should be preferred over another and that both SE and VIBE are equally suitable. A one-fits-all 5-stage staging technique with substages 3a, 3b, and 3c can be applied based on the corresponding sequence-specific reference data.

Moreover, compared to literature, the methods in the current study seem to perform equally well regarding the distinction between minors and adults, but perform worse regarding the point prediction of age. Few MRI studies have applied a model to estimate age and have reported MAE and diagnostic indices. For the hand/wrist, a MAE of 0.85 in males has been reported, applying an automated age estimation method [10]. For the third molars, a MAE of 1.7 years in males and 2.0 years

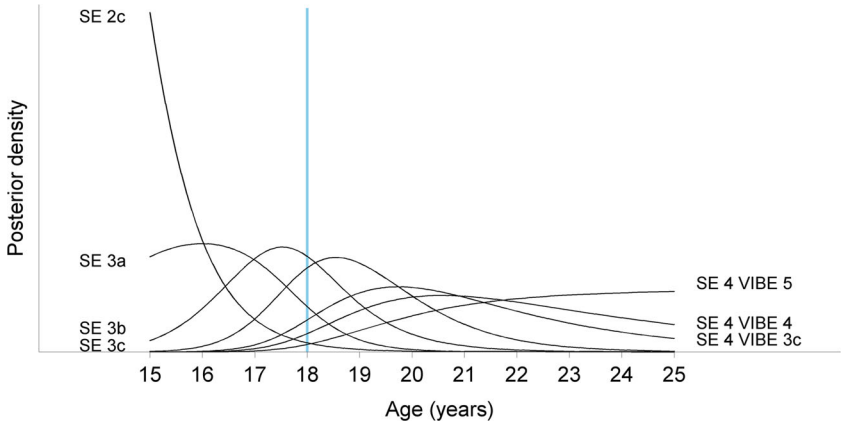

Fig. 5 Posterior density of age per radius stage in males. SE is considered and when stage 4 is allocated, additional VIBE staging allows for a differentiation within that group. All distributions of age are right-skewed. Around the age of 18 years, most individuals have a radius in stage 3b or 3c. Per curve, the point prediction of age corresponds with the trimmed mean, and the probability to be adult is represented by the area under the curve to the right of the 18-year threshold (blue vertical line)

in females has been reported, with corresponding specificities of 96% and 73%, respectively [2]. Combined MRI of the third molars, both clavicles, and the left hand/wrist rendered a MAE of 1.14 and 89% correctly classified minors [5].

The search for the optimal distinction between minors and adults is characteristic for forensic age estimation, which should always strive for the highest possible proportion of correctly classified minors and report appropriate confidence

Table 5 Age difference between substages based on different amelioration attempts

Anatomical structure	Sequence	Amelioration attempt	Considered subset of cases for comparison of mean age	Lower substage		Middle substage		Higher substage		*p* value age difference‡
				N	Mean age (SD)	*N*	Mean age (SD)	*N*	Mean age (SD)	
Radius	SE	Merge stage 3	Stages 3a, 3b, and 3c	62	16.1 (1.25)	8	16.8 (1.03)	80	18.4 (1.83)	< 0.001
	VIBE	Merge stage 3	Stages 3a, 3b, and 3c	62	15.8 (1.24)	7	16.9 (1.13)	74	18.3 (2.27)	< 0.001
	SE	Merge stages 4 and 5	Stages 4 and 5	165	22.4 (2.63)			7	23.0 (2.72)	0.528†
	VIBE	Merge stages 4 and 5	Stages 4 and 5	27	23.1 (2.66)			160	22.1 (2.75)	0.112†
	SE	Stage 3c END	Stage 3c without or with END	63	17.9 (1.59)			17	20.0 (1.76)	< 0.001
	VIBE	Stage 3c END	Stage 3c without or with END	53	17.9 (2.22)			21	19.1 (2.25)	0.057
	SE	Stage 3c TFS	Stage 3c without or with TFS	50	18.4 (1.89)			30	18.3 (1.75)	0.674
	SE	Stage 4 TFS	Stage 4 with or without TFS	94	22.4 (2.83)			111	22.5 (2.46)	0.875
	SE	Add VIBE info to SE stage 4	VIBE stages 3c, 4, and 5	14	21.3 (2.38)	18	20.7 (2.16)	173	22.7 (2.61)	0.002
Ulna	SE	Merge stage 3	Stages 3a, 3b, and 3c	67	16.2 (1.30)	12	16.7 (1.03)	30	18.3 (1.66)	< 0.001
	VIBE	Merge stage 3	Stages 3a, 3b, and 3c	66	16.0 (1.35)	12	16.5 (1.30)	48	17.8 (1.58)	< 0.001
	SE	Merge stages 4 and 5	Stages 4 and 5	224	21.7 (2.83)			20	23.4 (2.13)	0.010
	VIBE	Merge stages 4 and 5	Stages 4 and 5	27	20.2 (1.77)			210	22.2 (2.77)	< 0.001*
	SE	Stage 3c END	Stage 3c without or with END	29	18.1 (1.56)			1	21.6 (NA)	NA§
	VIBE	Stage 3c END	Stage 3c without or with END	40	17.7 (1.60)			8	18.2 (1.52)	0.385
	SE	Stage 3c TFS	Stage 3c without or with TFS	28	18.2 (1.72)			2	18.5 (0.20)	NA§
	SE	Stage 4 TFS	Stage 4 with or without TFS	43	21.3 (2.57)			181	21.9 (2.88)	0.211
	SE	Add VIBE info to SE stage 4	VIBE stages 3c, 4, and 5	15	18.8 (1.73)	24	20.2 (1.79)	185	22.2 (2.79)	< 0.001

END end stage, *N* number of participants in subgroup, *NA* not applicable, *SD* standard deviation, *TFS* threefold stratification sign

‡ *p* value from independent *t* test comparing age between two stages, or from one-way ANOVA comparing age between three stages. † Based on the second observer's assessment. § Too few cases in one category. * Based on the other observer $p = 0.299$

intervals around the point prediction of age [29, 30]. Traditionally, age estimation put too much focus on the point prediction, which increases the risk of over- or underestimation of the age. After all, age estimation will always be prone to inter-individual variation, impeding a true age determination. Fortunately, more recent literature and ongoing research focus more on the uncertainty intervals and probabilities to have reached the 18-year-threshold [31, 45]. It should be stressed that they are part of the age estimation results and should always be reported [46].

This ensures an appropriate use of medical imaging for forensic age estimation. Conversely, many authors have stated that age estimation based on for instance the Greulich and Pyle atlas [47] or the Demirjian method [48] should not be applied in age estimation, since they were developed to detect developmental disorders. However, when reference studies have validated certain methods of age estimation, their use is scientifically justified. If their use is ethically justified remains debatable [49, 50], but when forensic age estimation is compared with other forensic clinical activities, it seems similar: the medical expertise is used to answer a question asked by an authority, with respect for the individual's integrity but in a societal context.

Study limitations

Finally, two limitations of the current study need to be taken into account. First, the suitability of stages 1 to 2b and substages of 3a could not be studied, with age 14 as our lower cut-off. By contrast, the sample of participants up to age 26 did allow studying stages 4 and 5, which is more important considering the 18-year threshold. Second, the relatively small study sample hindered to draw strong conclusions, especially regarding the amelioration attempts of the staging technique. Profound substaging sometimes resulted in too few cases to allow any analysis.

Conclusion

To estimate age in adolescents and subadults based on wrist MRI, it suffices to apply a one-fits-all 5-stage staging technique—optimised with substages in stage 3—to either radius or ulna on T1 SE or VIBE. The age estimate and the probability to have reached a threshold age need to be based on sequence-specific reference data.

Acknowledgements The authors wish to express their gratitude to all participants and everyone who helped with the recruitment. We thank Maarten Peleman and Dries Ovaere for installing the viewing software on the department's computers. Finally, we wish to acknowledge Inès Phlypo for her indispensable critical appraisal of the manuscript.

Results described in this manuscript were presented at the 20th Meeting of the Study Group on Forensic Age Diagnostics (Arbeitsgemeinschaft für Forensische Altersdiagnostik, AGFAD) in Berlin, Germany, on March 17, 2017.

Funding Funding for this research was entirely provided by the Department of Radiology and Nuclear Medicine at Ghent University and the Department of Imaging and Pathology - Forensic Odontology at KU Leuven.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Koenraad Verstraete, Ghent University.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

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 Parts of the study population have been previously reported in [2, 3, 26–28, 36]. In those studies, the development of their clavicles and third molars was studied for age estimation.

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

References

1. Schmeling A, Geserick G, Reisinger W, Olze A (2007) Age estimation. Forensic Sci Int 165:178–181
2. De Tobel J, Phlypo I, Fieuws S, Politis C, Verstraete KL, Thevissen PW (2017) Forensic age estimation based on development of third molars: a staging technique for magnetic resonance imaging. J Forensic Odontostomatol 35:117–140
3. Hillewig E, Degroote J, Van der Paelt T et al (2013) Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: towards more sound age estimates. Int J Legal Med 127:677–689
4. Ottow C, Schulz R, Pfeiffer H, Heindel W, Schmeling A, Vieth V (2017) Forensic age estimation by magnetic resonance imaging of the knee: the definite relevance in bony fusion of the distal femoral- and the proximal tibial epiphyses using closest-to-bone T1 TSE sequence. Eur Radiol 27:5041–5048
5. Štern D, Kainz P, Payer C, Urschler M (2017) Multi-factorial age estimation from skeletal and dental MRI Volumes. In: International workshop on machine learning in medical imaging. Springer, Quebec City, pp 61–69

6. Timme M, Ottow C, Schulz R et al (2017) Magnetic resonance imaging of the distal radial epiphysis: a new criterion of maturity for determining whether the age of 18 has been completed? Int J Legal Med 131:579–584
7. Schmeling A, Dettmeyer R, Rudolf E, Vieth V, Geserick G (2016) Forensic age estimation. Dtsch Arztebl Int 113:44–50
8. Thevissen PW (2013) Dental age estimation in sub-adults: striving for an optimal approach. Leuven University Press, Leuven
9. Cunha E, Baccino E, Martrille L et al (2009) The problem of aging human remains and living individuals: a review. Forensic Sci Int 193:1–13
10. Urschler M, Grassegger S, Štern D (2015) What automated age estimation of hand and wrist MRI data tells us about skeletal maturation in male adolescents. Ann Hum Biol 42:358–367
11. Serin J, Rérolle C, Pucheux J et al (2016) Contribution of magnetic resonance imaging of the wrist and hand to forensic age assessment. Int J Legal Med 130:1121–1128
12. Dvorak J (2009) Detecting over-age players using wrist MRI: science partnering with sport to ensure fair play. Br J Sports Med 43: 884–885
13. Dvorak J, George J, Junge A, Hodler J (2007) Age determination by magnetic resonance imaging of the wrist in adolescent male football players. Br J Sports Med 41:45–52
14. Dvorak J, George J, Junge A, Hodler J (2007) Application of MRI of the wrist for age determination in international U-17 soccer competitions. Br J Sports Med 41:497–500
15. George J, Nagendran J, Azmi K (2012) Comparison study of growth plate fusion using MRI versus plain radiographs as used in age determination for exclusion of overaged football players. Br J Sports Med 46:273–278
16. Schmidt S, Vieth V, Timme M, Dvorak J, Schmeling A (2015) Examination of ossification of the distal radial epiphysis using magnetic resonance imaging. New insights for age estimation in young footballers in FIFA tournaments. Sci Justice 55:139–144
17. Tscholl PM, Junge A, Dvorak J, Zubler V (2016) MRI of the wrist is not recommended for age determination in female football players of U-16/U-17 competitions. Scand J Med Sci Sports 26: 324–328
18. Urschler M, Krauskopf A, Widek T et al (2016) Applicability of Greulich-Pyle and Tanner-Whitehouse grading methods to MRI when assessing hand bone age in forensic age estimation: a pilot study. Forensic Sci Int 266:281–288
19. Hojreh A, Gamper J, Schmook MT et al (2018) Hand MRI and the Greulich-Pyle atlas in skeletal age estimation in adolescents. Skelet Radiol 47:963–971. https://doi.org/10.1007/s00256-017-2867-3
20. Kellinghaus M, Schulz R, Vieth V, Schmidt S, Pfeiffer H, Schmeling A (2010) Enhanced possibilities to make statements on the ossification status of the medial clavicular epiphysis using an amplified staging scheme in evaluating thin-slice CT scans. Int J Legal Med 124:321–325
21. Schmeling A, Schulz R, Reisinger W, Mühler M, Wernecke KD, Geserick G (2004) Studies on the time frame for ossification of the medial clavicular epiphyseal cartilage in conventional radiography. Int J Legal Med 118:5–8
22. Wittschieber D, Schmidt S, Vieth V et al (2014) Subclassification of clavicular substage 3a is useful for diagnosing the age of 17 years. Rechtsmedizin 24:485–488
23. Ekizoglu O, Hocaoglu E, Inci E, Can IO, Aksoy S, Sayin I (2015) Estimation of forensic age using substages of ossification of the medial clavicle in living individuals. Int J Legal Med 129:1259–1264
24. Wittschieber D, Schulz R, Vieth V et al (2014) The value of substages and thin slices for the assessment of the medial clavicular epiphysis: a prospective multi-center CT study. Forensic Sci Med Pathol 10:163–169
25. Thevissen PW, Fieuws S, Willems G (2013) Third molar development: evaluation of nine tooth development registration techniques for age estimations. J Forensic Sci 58:393–397
26. De Tobel J, Hillewig E, Bogaert S, Deblaere K, Verstraete K (2017) Magnetic resonance imaging of third molars: developing a protocol suitable for forensic age estimation. Ann Hum Biol 44:130–139
27. De Tobel J, Hillewig E, Verstraete K (2017) Forensic age estimation based on magnetic resonance imaging of third molars: converting 2D staging into 3D staging. Ann Hum Biol 44:121–129
28. Hillewig E, De Tobel J, Cuche O, Vandemaele P, Piette M, Verstraete K (2011) Magnetic resonance imaging of the medial extremity of the clavicle in forensic bone age determination: a new four-minute approach. Eur Radiol 21:757–767
29. Boldsen JL, Milner GR, Konigsberg LW, Wood JW (2002) Transition analysis: a new method for estimating age from skeletons. In: Hoppa RD, Vaupel JW (eds) Paleodemography: Age distributions from skeletal samples. (Cambridge Studies in Biological and Evolutionary Anthropology). Cambridge University Press, Cambridge, pp 73–106
30. Fieuws S, Willems G, Larsen-Tangmose S, Lynnerup N, Boldsen J, Thevissen P (2016) Obtaining appropriate interval estimates for age when multiple indicators are used: evaluation of an ad-hoc procedure. Int J Legal Med 130:489–499
31. Thevissen PW, Fieuws S, Willems G (2010) 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 124:35–42
32. Bolívar J, Sandoval Ó, Osorio J, Dib G, Gallo J (2015) Relationship of chronological age and sexual maturity with skeletal maturity by magnetic resonance imaging of the distal radial epiphysis in adolescent football players. Apunts Medicina de l'Esport 50:129–137
33. Sarkodie BD, Botwe BO, Pambo P, Brakohiapa EK, Mayeden RN (2018) MRI age verification of U-17 footballers: the Ghana study. J Forensic Radiol Imaging 12:21–24
34. Wittschieber D, Schulz R, Vieth V et al (2014) 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 128:183–191
35. McGibbon CA, Bencardino J, Palmer WE (2003) Subchondral bone and cartilage thickness from MRI: effects of chemical-shift artifact. MAGMA 16:1–9
36. De Tobel J, Parmentier GIL, Phlypo I et al (2018) 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. https://doi.org/10.1007/s00414-018-1905-6
37. Vieth V, Schulz R, Heindel W et al (2018) Forensic age assessment by 3.0T MRI of the knee: proposal of a new MRI classification of ossification stages. Eur Radiol 28:3255–3262. https://doi.org/10.1007/s00330-017-5281-2
38. Abdelbary MH, Abdelkawi MM, Nasr MA (2018) Age determination by MR imaging of the wrist in Egyptian male football players. How far is it reliable? The Egyptian Journal of Radiology and Nuclear Medicine 49:146–151
39. Serinelli S, Panebianco V, Martino M et al (2015) Accuracy of MRI skeletal age estimation for subjects 12-19. Potential use for subjects of unknown age. Int J Legal Med 129:609–617
40. Terada Y, Kono S, Tamada D et al (2013) Skeletal age assessment in children using an open compact MRI system. Magn Reson Med 69: 1697–1702
41. Terada Y, Kono S, Uchiumi T et al (2014) Improved reliability in skeletal age assessment using a pediatric hand MR scanner with a 0.3T permanent magnet. Magn Reson Med Sci 13:215–219
42. Terada Y, Tamada D, Kose K et al (2016) Acceleration of skeletal age MR examination using compressed sensing. J Magn Reson Imaging 44:204–211

43. Tomei E, Sartori A, Nissman D et al (2014) Value of MRI of the hand and the wrist in evaluation of bone age: preliminary results. J Magn Reson Imaging 39:1198–1205
44. Jaramillo D, Connolly SA, Mulkern RV, Shapiro F (1998) Developing epiphysis: MR imaging characteristics and histologic correlation in the newborn lamb. Radiology 207:637–645
45. Bleka Ø, Wisloff T, Dahlberg PS, Rolseth V, Egeland T (2018) Advancing estimation of chronological age by utilizing available evidence based on two radiographical methods. Int J Legal Med. https://doi.org/10.1007/s00414-018-1848-y
46. International Organization for Forensic Odonto-Stomatology (IOFOS) (2018) Recommendations for quality assurance. Dental Age Estimation, Leuven
47. Greulich W, Pyle SI (1959) Radiographic atlas of skeletal development of the hand and wrist, 2nd edn. Stanford University Press, Stanford
48. Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. Hum Biol 45:211–227
49. Aynsley-Green A, Cole T, Crawley H, Lessof N, Boag L, Wallace R (2012) Medical, statistical, ethical and human rights considerations in the assessment of age in children and young people subject to immigration control. Br Med Bull 102:17–42
50. Thevissen PW, Kvaal SI, Dierickx K, Willems G (2012) Ethics in age estimation of unaccompanied minors. J Forensic Odontostomatol 30(Suppl 1):84–102