This is the accepted manuscript of the following publication:

Coreelman H, De Tobel J, Widek T, Urschler M, Fieuws S, Thevissen PW, Verstraete KL. Magnetic resonance imaging for multi-factorial age estimation using Bayes' rule: a validation study in two independent samples. Int J Legal Med. 2025 Apr 28. doi: 10.1007/s00414-025-03499-x. Online ahead of print.

FULL TITLE PAGE

Magnetic resonance imaging for multi-factorial age estimation using Bayes' rule: a validation study in two independent samples

Author order	First name	Middle initials	Last name	Academic degree(s) ORCID iD E-mail	Affiliation(s)
1	Heleen	/	Coreelman*	🕮 Master of Science (MSc) in Forensic Science 🕮 Master of Science (MSc) in Biochemistry and Biotechnology 0000-0003-4609-7322 ✉ coreelman.heleen@hotmail.com	a,b,c
2	Jannick	/	De Tobel*	🕮 Doctor of Philosophy (PhD) in Health Sciences 🕮 Medical doctor (MD) 🕮 Doctor of Dental Science (DDS) 🕮 Master of Science (MSc) in Forensic Odontology 0000-0002-8846-7339 ✉ jannick.detobel@ugent.be	a,d
3	Thomas	/	Widek	🕮 Doctor of Philosophy (PhD) in Medical Science 🕮 Master of Science (MSc) in Forensic Radiography 🕮 Diploma in Radiography 0000-0001-8474-1233 ✉ thomas.widek@medunigraz.at	e,f
4	Martin	/	Urschler	🕮 Doctor of Philosophy (PhD) in Computer Science 🕮 Master of Science (MSc) in Computer Science and Electrical Engineering 0000-0001-5792-3971 ✉ martin.urschler@medunigraz.at	g
5	Steffen	/	Fieuws	🕮 Doctor of Philosophy (PhD) 🕮 Master of Science (MSc) in Biostatistics No ORCID iD ✉ steffen.fieuws@kuleuven.be	h
6	Patrick	Werner	Thevissen	🕮 Doctor of Philosophy (PhD) in Biomedical Sciences 🕮 Doctor of Dental Science (DDS) 🕮 Master of Science (MSc) in Forensic Odontology 0000-0003-0890-6264 ✉ patrick.thevissen@kuleuven.be	i

7	Koenraad	Luc	Verstraete	🕮 Doctor of Philosophy (PhD) in Medicine 🕮 Medical doctor (MD) 0000-0003-3688-1076 ✉ koenraad.verstraete@ugent.be	a

Affiliations

a) Department of Diagnostic Sciences – Radiology, Ghent University, Corneel Heymanslaan 10, 9000 Ghent, Belgium
b) Department of Imaging and Pathology – Forensic Biomedical Sciences, KU Leuven, Herestraat 49, 3000 Leuven, Belgium
c) Department of Biology – Biology, KU Leuven – Kulak, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium
d) Clinique dentaire CitySmile, Boulevard de Saint Georges 72, 1205 Genève, Switzerland
e) Diagnostic and Research Institute of Forensic Medicine, Medical University of Graz, Neue Stiftingtalstrasse 6, 8010 Graz
f) BioTechMed-Graz, Mozartgasse 12/II, 8010 Graz, Austria
g) Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Auenbruggerplatz 2/V, 8010 Graz, Austria
h) Interuniversity Institute for Biostatistics and statistical Bioinformatics, KU Leuven – Hasselt University, Kapucijnenvoer 35 blok d bus 7001, 3000 Leuven, Belgium
i) Department of Imaging and Pathology – Forensic Odontology, KU Leuven, Kapucijnenvoer 7 blok a bus 7001, 3000 Leuven, Belgium

* Coreelman H and De Tobel J contributed equally to this manuscript.

Corresponding author and reprint info

Jannick De Tobel

Department of Diagnostic Sciences – Radiology

Ghent University

Corneel Heymanslaan 10, 9000 Ghent, Belgium

Telephone +32 (0)9 332 40 23

Fax +32 (0)9 332 49 69

E-mail jannick.detobel@gmail.com, jannick.detobel@ugent.be

Disclosures

Funding

This work was supported by the Department of Diagnostic Sciences – Radiology at Ghent University and the Department of Imaging and Pathology – Forensic Odontology at KU Leuven.

Conflict of interest

The authors of this manuscript declare no conflict of interest.

Ethical approval

The current study was approved by the Ethics Committee Research of University Hospitals Leuven / KU Leuven (reference number: S67555), and was conducted in accordance with the Declaration of Helsinki.

Informed consent

Written informed consent was obtained from all participants in this study. In case of minors, the legal guardians also consented.

Overlap with previous publications

The Ghent sample was part of the study populations included in the following previous publications: [19-23]. The Graz sample was part of the study populations included in the following previous publications: [14, 24-37].

In those previous publications, the samples were used as reference samples to develop (and internally validate) age estimation methods. By contrast, in the current study, the samples are used as external validation samples for the De Tobel et al. [15] method, since that method was based on a reference population which was different from those in the referred previous studies.

Contributions

Study conception and design were performed by HC and JDT. Data were collected by JDT and TW. Data analysis was performed by HC, JDT, and SF. Data interpretation was performed by HC, JDT, PT and KV. The first draft of the manuscript was written by HC. All authors read and commented on the draft manuscript and approved the final manuscript. Supervision of the process was performed by JDT, MU, PT and KV.

Abstract

Objectives Multi-factorial age estimation (MFA) models have been developed based on Bayes' rule, combining MRI data of the third molars (T), left wrist (W) and/or both clavicles (C). Internal cross-validated performance was reported, but external validation is needed before bringing the approach into practice. This study aimed to validate these MFA models in two independent samples.

Methods In the Ghent sample, W+C MRI was prospectively conducted in 108 healthy Caucasian volunteers (52 males, 56 females) aged 16 to 21 years. In the Graz sample, T+W+C MRI was prospectively conducted in 335 healthy Caucasian males aged 13 to 24 years. Development was staged and checked for intra-observer reliability, and age estimation performances were tested.

Results Staging clavicles was most prone to intra-observer variability. Applying the W+C model to Ghent males rendered a mean absolute error of 1.55 years, a root mean square error of 1.90 years, 70.6% correctly categorised adults and 94.4% correctly categorised minors. In females, the results were 1.49 years, 1.83 years, 92.1% and 66.7%, respectively. Regarding the Graz sample, the W+C results were 1.66 years, 2.08 years, 94.7% and 80.7%, respectively. For the T+W+C model, the results were 1.41 years, 1.80 years, 95.2% and 81.5%, respectively.

Conclusion The T+W+C and W+C models rendered a similar accuracy of the point prediction of age in both validation samples. However, they bared a larger risk of wrongfully categorising a minor as an adult than reported for internal validation, stressing the importance of the prediction interval for age estimation in practice.

Keywords

Age Estimation by Skeleton

Age Estimation by Teeth

Adolescent

Adult

Magnetic Resonance Imaging

Abbreviations

CT	Computed tomography
GE	Gradient echo
GRE	Gradient-recalled echo
MAE	Mean absolute error
MFA	Multi-factorial age estimation
MRI	Magnetic resonance imaging
PI	Prediction interval
RMSE	Root mean square error
T+W+C	Third molars, left wrist and both clavicles
VIBE	Volumetric interpolated breath-hold examination
W+C	Left wrist and both clavicles

Introduction

Studying dental and skeletal development is crucial in forensic age estimation of living adolescents and young adults, particularly with regard to legal applications and sports [1-4]. Moreover, dental and skeletal age determination are also implemented in clinical medicine considering growth-related and endocrine disorders [5]. Unfortunately, the use of medical imaging for forensic age estimation purposes remains a controversial subject and is often labelled as (1) unethical and (2) imprecise and inaccurate [6-9].

Firstly, to study developing anatomical structures, conventional radiography and/or computed tomography (CT) are the currently used imaging modalities [3, 10]. Nevertheless, the exposure of an individual to ionising radiation without a medical indication brings about ethical issues [11]. Indeed, risk assessment is an important consideration in forensic age estimation procedures, as the diagnostic value needs to outweigh possible harmful effects [12]. Moreover, the use of ionising radiation is prohibited for civil procedures and asylum cases in several countries [13]. Hence, medical imaging modalities that do not expose the examinee to ionising radiation have been proposed, with magnetic resonance imaging (MRI) as the most promising alternative [10, 14].

Secondly, due to inter-individual variability in developmental changes of different anatomical structures, single site age estimation procedures result in uncertainty and possibly an ethically unacceptable error, i.e. wrongly categorising a minor as an adult [1, 10]. To increase the accuracy of the point prediction of age, to reduce its uncertainty and to decrease the chance of incorrect categorisation, forensic age estimation based on multiple anatomical sites (multi-factorial age estimation; MFA) is preferred [1, 3, 10]. Based on MRI, two research groups independently demonstrated that age estimation around the age of 18 years should combine data of the third molars, left hand/wrist and both clavicles [14, 15].

Several statistical approaches can be implemented to estimate the examinee's chronological age from multivariate data, with application of Bayes' rule being the most appropriate [16-18]. Using such an approach, De Tobel et al. [15] developed MFA MRI models for females and males combining developmental stages of the third molars, left wrist and/or both clavicles, and age estimation performance was internally validated using cross-validation. Yet, to date, no external validation on the MFA MRI models has been performed. Hence, the current study aimed to externally validate MFA MRI models published by De Tobel et al [15].

Material and Methods

Study population

The current study was approved by the Ethics Committee Research of University Hospitals Leuven / KU Leuven (reference S67555), and was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants in this study. In case of minors, the legal guardians also consented.

Development was assessed in two independent samples to test the age estimation performance of the De Tobel et al. models. As a first sample, healthy Belgian participants were prospectively recruited between December 2007 and August 2011. MRI was conducted in 108 individuals, comprising 52 males and 56 females, with ages 16 to 21 years old (Table 1). This sample will be referred to as the 'Ghent sample', and was part of the study populations included in earlier research [19-23]. As a second sample, healthy Austrian participants were prospectively recruited between May 2011 and November 2018. MRI was conducted in 335 males with ages 13 to 24 years old (Table 1). This sample will be referred to as the 'Graz sample', and was part of the study populations included in earlier research [14, 24-37].

Both samples included Caucasians of middle to high socio-economic status [23], with a known verified age. Exclusion criteria were surgical removal of three or four third molars (only applies to the Graz sample), fractures of the region of interest, undernourishment, growth-related and endocrine disorders, chronic diseases or chronic medication usage that could disturb normal development, and general MRI contra-indications.

Image acquisition

In the Ghent sample, the left wrist and sternal extremity of both clavicles were visualised by conducting 3T MRI using a Siemens scanner (Magnetom Trio Tim, Siemens, Erlangen, Germany), according to published protocols for the wrist [19] and clavicles [22]. In the Graz sample, the third molars, left hand/wrist and sternal extremity of both clavicles were visualised by conducting 3T MRI using Siemens scanners (Magnetom Trio Tim/Prisma and Skyra, Siemens, Erlangen, Germany), according to similar published protocols for the third molars [33], hand/wrist [32] and clavicles [31]. Details about the applied MRI sequences were included in Table S1 of the Supplementary Material.

Image analysis

All images were exported in DICOM format and pseudonymised. Assessment data of an experienced observer (JDT; a resident in oral and maxillofacial surgery with 10 years of experience in assessing MR images for age estimation) were used for all analyses. The Ghent sample was included in published

studies that reported on inter- and intra-observer agreements [19, 20]. For the Graz sample, fifty randomly selected Graz cases per anatomical site were assessed twice to investigate intra-observer agreements. The Ghent sample images were displayed on a Barco MDCC-6230 monitor (Barco, Kortrijk, Belgium) with a resolution of 3280 x 2048 pixels and evaluated using KPacs viewing software (Version 1.6.0). The Graz sample images were displayed on an iMac monitor (Apple Inc., Cupertino, California, USA) with a resolution of 2560 x 1440 pixels and evaluated using OsiriX viewing software (OsiriX 4.1, https://www.osirixviewer.com) embedded in an aycan software environment (aycan Medical Systems, Wuerzburg, Germany).

Stage allocation was performed according to published staging techniques for the third molars [38] (Table S2), left wrist [19] and both clavicles [20] (Table S3). In essence, third molar development was assessed using the De Tobel [38] staging technique, wrist development was assessed using the Schmeling [39] and Kellinghaus [40] staging technique, and clavicle development was assessed using the Schmeling [39], Kellinghaus [40] and Wittschieber [41] staging technique. Moreover, morphological variants of the sternal extremity of the clavicle were documented. All anatomical sites were analysed independently to avoid bias. Stage allocation was conducted by considering the whole stack of slices, and in case of doubt the lower stage was allocated.

Statistical analysis

Data were analysed with SPSS Statistics 27.0 (IBM SPSS Statistics for Windows, Armonk, NY). Reproducibility of staging was quantified using linear weighted Cohen's kappa statistics.

The current samples were used to externally validate two different MFA MRI models published by De Tobel et al [15]. Since the Ghent sample included wrist and clavicle MRI (W+C), it was used to validate the W+C combination model. By contrast, the Graz sample included third molar, wrist and clavicle MRI (T+W+C). Thus, it was used to validate the T+W+C and W+C combination models. To develop the original MFA models, radius and ulna stages 2/3a, 3b/3c, 4/5, and clavicle stages 2a/2b/2c, 3aa/3ab/3ac were merged [15]. Additionally, clavicle stages 1 and 4/5 were discarded because they could be confused [15]. Study participants with one or more excluded clavicles were still included in the analyses, however, the affected clavicle(s) were treated as missing value(s). Moreover, morphological variants of the clavicles' sternal extremity were excluded, since they might develop in an abnormal manner compared to clavicles with a straight, or slightly indented or rounded morphology of their sternal extremity [15, 42].

To estimate the chronological age of each examinee, point predictions and 95% prediction intervals were obtained using estimates from the MFA MRI model published by De Tobel et al [15]. This model employs Bayes' rule in combination with a continuation-ratio model for the ordinal stages of the age

indicators (T, W , and C), which serve as the conditional distribution of the age indicator given age [16]. Assuming a uniform prior distribution between 14.0 and 27.0 years, the MFA MRI model provides both point estimates and interval predictions for all possible combinations of T, W, and C data, including cases where one or more age indicators are missing. In the current study, these predictions were retrieved for each subject included in the validation sets. Consequently, the statistical analysis was strictly limited to the external validation of the MFA MRI model and did not involve model development or optimization. Age estimation performance was then studied by considering two major aspects: (1) a point estimate of age along with its 95% prediction interval (PI), and (2) the ability to discriminate between minors (i.e. <18 years of age) and adults (i.e. ≥18 years of age).

The accuracy of the point prediction of age was quantified by the mean absolute error (MAE) and root mean square error (RMSE), whereas the model's ability to discriminate between minors and adults was represented by diagnostic indices:

- Accuracy; which is the fraction of correctly categorised examinees.
- Sensitivity; defined as the fraction of correctly categorised adults.
- Specificity; defined as the fraction of correctly categorised minors.

Finally, the discrimination slope was calculated for discerning minors from adults, which represents the difference between minors and adults in mean predicted probability to be a minor.

Results

Reproducibility of staging

Intra-observer agreements for stage allocation in the Ghent [19; 20] and Graz samples were summarised in Table 2. Note that the results for intra-observer agreements for wrist stage allocation in the Ghent sample are missing. Instead, inter-observer agreements were reported, which are expected to be lower than intra-observer agreements. Most disagreements were attributed to one-stage differences (Tables S4-S6). Moreover, interchanged stages of early and late clavicle development were encountered in both the Ghent and Graz sample (Table S6).

Validation of age estimation performance

The results on age estimation performance of the MFA MRI models based on internal ten-fold cross-validation [15] and external validation were summarised in Table 3.

Considering the accuracy of the point prediction of age, the MAE and RMSE in the external samples were mostly similar to those reported for internal validation. However, considering the ability to discern minors from adults, a strikingly low specificity of 66.7% was obtained in Ghent females (W+C). In Graz males, the specificity was also relatively low, equalling to 80.7% (W+C) and 81.5% (T+W+C). By contrast, sensitivity was mostly high (92.1 to 95.2%), except for a striking 70.6% in Ghent males (W+C). Comparing the discrimination slopes in the different validation sets allows for a more intuitive and comprehensive interpretation of the classification performance. In Graz males, slopes around 0.70 were obtained, compared to slopes around 0.80 for internal cross-validation. In Ghent males and females, lower slopes were obtained equalling 0.62 and 0.47 respectively. For all diagnostic indices, 95% confidence intervals were very large (range 20.7 to 45.7) in the Ghent set, due to the relatively small sample size. Therefore, conclusions should be drawn with caution regarding this validation set.

Overall, the best validation results were obtained using the T+W+C model, with results that were similar to the internal validation results, except for a lower specificity (81.5% versus 90% in internal validation).

To illustrate the practical application of the MFA models in the different samples, case examples are displayed in Figs. 1-3. Furthermore, the extremes in age estimation performance are illustrated in Fig. 4, representing a scatter plot of the 5% cases with the largest prediction error of each sample per combination model. Overall, overestimation occurred more often in these cases than underestimation. Note a trend towards lower absolute prediction errors for the T+W+C model compared to the W+C model. Moreover, in the original Ghent males (T+W+C), the largest absolute prediction errors only occurred in adults, while the trend was shifted towards younger ages in both validation samples.

Discussion

The use of MRI for forensic age estimation of adolescents and young adults has already been studied extensively [43], but research is still needed before its practical implementation could be possible. A first important gap in the literature is the lack of validation studies. After all, the implementation of methods that have not been adequately validated could bare a huge risk for wrongfully categorising a minor as an adult or vice versa [44]. Although the publication of large scale reference studies is indispensable, often only descriptive statistics of age are reported as a function of developmental stage. These descriptive statistics are still insufficient to render an applicable age estimation method. To gain insight into the age estimation performance of a method, its validated quantification (MAE, RMSE, diagnostic indices for minor/adult classification) is necessary. A second gap in literature is the lack of MFA studies. Most age estimation reference studies are retrospective single site studies, even those based on MRI. Yet, MRI allows to study development for age estimation prospectively and at several anatomical sites, because it avoids radiation exposure. To date, only two research groups reported on the performance of MFA based on MRI. Štern et al [14] used a deep convolutional neural network to integrate MRI data of the third molars, hand-wrist bones and clavicles for fully automatic MFA. This overcomes the major limitation of subjectivity in age estimation practice. However, it remains unclear how they accounted for conditional dependence of development at various anatomical sites, and they did not report how to obtain prediction intervals. Conversely, De Tobel et al [15] developed MFA MRI models based on Bayes' rule combining observer-based data of the third molars, left wrist and/or both clavicles. Statistical analysis using a Bayesian framework is preferred in forensic age estimation practice because it reduces age mimicry (i.e. the effect of the reference sample's age distribution on the outcomes) and systematic bias (i.e. systematically estimating young individuals as older and vice versa) [23, 45-48].

Reproducibility of staging

In the literature, intra-observer (weighted) Cohen's kappa values ranging from 0.85 to 0.89, 0.84 to 0.99 and 0.66 to 0.99 have been reported for stage allocation of the third molars, wrist and clavicles, respectively [21, 43]. Overall, our kappa values were lower, which might be explained by the differences in MRI protocols, making the observer less accustomed to the MR sequences used for the Graz sample. This further underlines the importance of externally validating developed age estimation models on independent datasets. Note that our kappa values for stage allocation of both clavicles correspond to those reported for clavicle CT (0.61 to 0.96) [21, 49], which might reflect the difficulty of assessing the clavicles as skeletal structures, rather than reflecting that MRI is a more difficult imaging modality to assess.

Validation of age estimation performance

The major advantage of the Ghent sample is that the MRI protocol to visualise the left wrist and both clavicles is identical to that reported by De Tobel et al [15]. Therefore, possible MRI-induced differences (i.e. field inhomogeneities, gradient field differences, coil, and MR sequence dissimilarities) are excluded.

Applying the original W+C model to the external validation sample yielded a rather low sensitivity in males (70.6%) and specificity in females (66.7%). These findings are exactly what one can expect from a W only model, because the wrist bones tend to reach the fused stage earlier in females than in males. Apparently, in this sample, the clavicle could not sufficiently correct the outcome, stressing the following two considerations. Firstly, the model discarded stages 1 and 4/5, while they hold the potential to increase performance drastically, but only when they are not confused. Secondly, third molar information lacked in this sample, while both De Tobel et al [15] and Štern et al [14] independently demonstrated that integrating information of the third molars, left hand/wrist and both clavicles renders the best age estimation results in adolescents and young adults. Still, among these three anatomical sites, third molar information is the most frequently unavailable due to agenesis and/or extraction [50, 51]. Thus, the W+C results reported here provide insight into how the absence of third molar information may impact age estimation in practice.

Third molars were included in the Graz sample. Adding the third molar information mainly affected the accuracy of the point prediction of age, rather than the diagnostic indices. This corresponds with the original findings, where diagnostic indices remained unaffected by adding third molar information in males. Moreover, applying the original W+C and T+W+C model to the Graz validation sample yielded a rather low specificity in males (80.7% and 81.5%, respectively). Therefore, in practice, the prediction interval should be taken into account to decide on the minor versus adult state.

A first limitation of the Graz sample was that it did not include females. Therefore, it was not possible to verify whether adding information of all third molars increases specificity in females, as was observed in the original sample (W+C 88.4% versus T+W+C 90.7%).

A second limitation was the different MRI protocol. Regarding the third molars, De Tobel et al [48] demonstrated that the Ghent and Graz MRI protocols rendered different staging results, with a lower stage often being allocated based on the Ghent protocol than based on the Graz protocol. Conversely, regarding the left wrist and both clavicles, there are no studies that compared the different MRI protocols in the same study population. Since differences in MRI protocols might result in different age distributions per stage, age estimation performance might be affected. Still, the differences between

the protocols are subtle, and by taking the prediction interval into account, rather than just the point prediction of age, the effect of protocol differences can be expected to decrease in practice.

Future prospects

To date, it remains unclear if MRI data from different research groups can safely be merged when the applied MRI protocols differ. Therefore, validation studies are necessary to first investigate whether the results of external validation are similar to those published for internal cross-validation. After all, obtaining similar results after external validation should mean that the between-population differences and differences in the applied MRI protocols are inferior to the within-population differences (i.e. inter-individual dissimilarities). Consequently, the data could be integrated into the original model to increase its sample size, but without affecting the width of the PIs. Thus, our current study only represents a first step towards merging MRI data for age estimation.

Moreover, it might be interesting to investigate whether the fully automatic method published by Štern et al [14] performs equally well in the De Tobel et al [15] study sample. Still, the deep learning algorithm cannot simply be applied to the De Tobel et al [15] sample, because the Graz algorithm was designed to assess MR images obtained with the Graz protocol. Thus, the Graz software needs adaption before it could be applied to the De Tobel et al [15] sample. This might be the focus of follow-up research.

Finally, the current study only considered the point prediction of age to discern minors from adults. In practice, however, the lower bound of the prediction interval is decisive for this distinction [10], ensuring that the benefit of the doubt is granted, and thus minimising the risk of wrongfully classifying a minor as an adult. Note that currently, the minimum age concept is regarded as the gold standard for integrating information from multiple anatomical sites into a multi-factorial age estimation [3, 52]. The relationship and differences between applying Bayes' rule versus the minimum age concept are the focus of an ongoing study by our research group, and were therefore not addressed in the current study.

Conclusion

The models published by De Tobel et al [15] for MFA of adolescents and young adults based on MRI rendered a similar accuracy of the point prediction of age in the two independent external validation samples as in the internal cross-validation. However, certain diagnostic indices for minor versus adult classification were lower than those from internal cross-validation, stressing the importance of taking the point prediction of age and its related prediction interval into account for age estimation in practice.

References

1. De Tobel J, de Haas MB, van Wijk M, Verstraete KL, Thevissen PW. (2018) De toekomst van forensische leeftijdsschatting bij levende adolescenten en jongvolwassenen: magnetische resonantie beeldvorming en automatisering. In: Aps JKM, Boxum SC, De Bruyne MAA, Jacobs R, van der Meer WJ, Nienhuijs MEL, eds. Het tandheelkundig Jaar 2018. Bohn Stafleu van Loghum Houten. pp. 1-19.
2. 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. doi: bjsm.2006.031021 [pii];10.1136/bjsm.2006.031021 [doi]
3. Schmeling A, Dettmeyer R, Rudolf E, Vieth V, Geserick G (2016) Forensic Age Estimation. Dtsch Arztebl Int 113: 44-50. doi: 10.3238/arztebl.2016.0044
4. Schmeling A, Geserick G, Reisinger W, Olze A (2007) Age estimation. Forensic Sci Int 165: 178-81.
5. Martin DD, Wit JM, Hochberg Z et al (2011) The use of bone age in clinical practice - part 1. Horm Res Paediatr 76: 1-9. doi: 10.1159/000329372
6. Axelsson I (2019) Bone maturation cannot be used to estimate chronological age in asylum-seeking adolescents. Acta Paediatr 108: 590-2. doi: 10.1111/apa.14645
7. Milani S, Benso L (2019) Why we can't determine reliably the age of a subject on the basis of his maturation degree. J Forensic Leg Med 61: 97-101. doi: 10.1016/j.jflm.2018.12.002
8. Mishori R (2019) The Use of Age Assessment in the Context of Child Migration: Imprecise, Inaccurate, Inconclusive and Endangers Children's Rights. Children (Basel) 6. doi: 10.3390/children6070085
9. Ording Muller LS, Offiah A, Adamsbaum C et al (2019) Bone age for chronological age determination - statement of the European Society of Paediatric Radiology musculoskeletal task force group. Pediatr Radiol 49: 979-82. doi: 10.1007/s00247-019-04379-4
10. De Tobel J, Ottow C, Widek T et al (2020) Dental and skeletal imaging in forensic age estimation: disparities in current approaches and the continuing search for optimization. Semin Musculoskelet Radiol 24: 510-22. doi: 10.1055/s-0040-1701495
11. Thevissen PW, Kvaal SI, Dierickx K, Willems G (2012) Ethics in age estimation of unaccompanied minors. J Forensic Odontostomatol 30 Suppl 1: 84-102.
12. Ramsthaler F, Proschek P, Betz W, Verhoff MA (2009) How reliable are the risk estimates for X-ray examinations in forensic age estimations? A safety update. Int J Legal Med 123: 199-204. doi: 10.1007/s00414-009-0322-2
13. Lockemann U, Fuhrmann A, Püschel K, Schmeling A, Geserick G (2004) Arbeitsgemeinschaft für Forensische Altersdiagnostik der Deutschen Gesellschaft für Rechtsmedizin. Rechtsmedizin 14: 123-6. doi: 10.1007/s00194-004-0243-9
14. Štern D, Payer C, Giuliani N, Urschler M (2019) Automatic Age Estimation and Majority Age Classification From Multi-Factorial MRI Data. IEEE J Biomed Health Inform 23: 1392-403. doi: 10.1109/jbhi.2018.2869606
15. De Tobel J, Fieuws S, Hillewig E et al (2020) Multi-factorial age estimation: A Bayesian approach combining dental and skeletal magnetic resonance imaging. Forensic Sci Int 306: 110054. doi: 10.1016/j.forsciint.2019.110054
16. 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: 489-99. doi: 10.1007/s00414-015-1200-8
17. Konigsberg LW, Frankenberg SR, Liversidge HM (2016) Optimal trait scoring for age estimation. Am J Phys Anthropol 159: 557-76. doi: 10.1002/ajpa.22914
18. Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH (2008) Estimation and evidence in forensic anthropology: age-at-death. J Forensic Sci 53: 541-57. doi: 10.1111/j.1556-4029.2008.00710.x

19. De Tobel J, Hillewig E, de Haas MB et al (2019) 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 29: 2924-35. doi: 10.1007/s00330-018-5944-7
20. De Tobel J, Hillewig E, van Wijk M et al (2020) Staging Clavicular Development on MRI: Pitfalls and Suggestions for Age Estimation. J Magn Reson Imaging 51: 377-88. doi: 10.1002/jmri.26889
21. De Tobel J, van Wijk M, Alberink I et al (2020) The influence of motion artefacts on magnetic resonance imaging of the clavicles for age estimation. Int J Legal Med 134: 753-68. doi: 10.1007/s00414-019-02230-x
22. 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: 757-67. doi: 10.1007/s00330-010-1978-1 [doi]
23. 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: 677-89. doi: 10.1007/s00414-012-0798-z
24. Ebner T, Štern D, Donner R, Bischof H, Urschler M (2014) Towards automatic bone age estimation from MRI: localization of 3D anatomical landmarks. Med Image Comput Comput Assist Interv 17: 421-8.
25. Martinez Vera NP, Holler J, Widek T, Neumayer B, Ehammer T, Urschler M (2017) Forensic age estimation by morphometric analysis of the manubrium from 3D MR images. Forensic Sci Int 277: 21-9. doi: 10.1016/j.forsciint.2017.05.005
26. Neumayer B, Lesch A, Thaler F et al (2020) The four-minute approach revisited: accelerating MRI-based multi-factorial age estimation. Int J Legal Med 134: 1475-85. doi: 10.1007/s00414-019-02231-w
27. Neumayer B, Schloegl M, Payer C et al (2018) Reducing acquisition time for MRI-based forensic age estimation. Sci Rep 8: 2063. doi: 10.1038/s41598-018-20475-1
28. Štern D, Ebner T, Bischof H, Grassegger S, Ehammer T, Urschler M (2014) Fully automatic bone age estimation from left hand MR images. Med Image Comput Comput Assist Interv 17: 220-7.
29. Štern D, Payer C, Urschler M (2019) Automated age estimation from MRI volumes of the hand. Med Image Anal 58: 101538. doi: 10.1016/j.media.2019.101538
30. 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: 358-67. doi: 10.3109/03014460.2015.1043945
31. Widek T, De Tobel J, Ehammer T, Genet P (2023) Forensic age estimation in males by MRI based on the medial epiphysis of the clavicle. Int J Legal Med 137: 679-89. doi: 10.1007/s00414-022-02924-9
32. Widek T, Genet P, Ehammer T, Schwark T, Urschler M, Scheurer E (2021) Bone age estimation with the Greulich-Pyle atlas using 3T MR images of hand and wrist. Forensic Sci Int 319: 110654. doi: 10.1016/j.forsciint.2020.110654
33. Widek T, Genet P, Merkens H et al (2019) Dental age estimation: The chronology of mineralization and eruption of male third molars with 3T MRI. Forensic Sci Int 297: 228-35. doi: 10.1016/j.forsciint.2019.02.019
34. Štern D, Ebner T, Bischof H, Urschler M. (2014) Determination of legal majority age from 3D magnetic resonance images of the radius bone. . IEEE 11th International Symposium on Biomedical Imaging (ISBI). pp. 1119-22.
35. Štern D, Urschler M. (2016) From individual hand bone age estimates to fully automated age estimation via learning-based information fusion. . IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE. pp. 150-4.
36. Štern D, Kainz P, Payer C, Urschler M. (2017) Multi-Factorial Age Estimation from Skeletal and Dental MRI Volumes. Springer Lecture Notes in Computer Science, 10541. pp. 61-9.
37. Štern D, Payer C, Lepetit V, Urschler M. (2016) Automated age estimation from hand MRI volumes using deep learning. Springer Lecture Notes in Computer Science, 9901. pp. 194-202.

38. 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: 117-40.
39. Schmeling A, Schulz R, Reisinger W, Muhler 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.
40. 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: 321-5. doi: 10.1007/s00414-010-0448-2
41. 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: 485-8. doi: 10.1007/s00194-014-0990-1
42. Schmidt S, Ottow C, Pfeiffer H et al (2017) Magnetic resonance imaging-based evaluation of ossification of the medial clavicular epiphysis in forensic age assessment. Int J Legal Med 131: 1665-73. doi: 10.1007/s00414-017-1676-5
43. De Tobel J, Bauwens J, Parmentier GIL et al (2020) Magnetic resonance imaging for forensic age estimation in living children and young adults: a systematic review. Pediatr Radiol 50: 1691-708. doi: 10.1007/s00247-020-04709-x
44. Mostad P, Tamsen F (2019) Error rates for unvalidated medical age assessment procedures. Int J Legal Med 133: 613-23. doi: 10.1007/s00414-018-1916-3
45. Braga J, Heuze Y, Chabadel O, Sonan NK, Gueramy A (2005) Non-adult dental age assessment: correspondence analysis and linear regression versus Bayesian predictions. Int J Legal Med 119: 260-74. doi: 10.1007/s00414-004-0494-8
46. Saint-Martin P, Rerolle C, Dedouit F et al (2013) Age estimation by magnetic resonance imaging of the distal tibial epiphysis and the calcaneum. Int J Legal Med 127: 1023-30. doi: 10.1007/s00414-013-0844-5
47. Serinelli S, Panetta V, Pasqualetti P, Marchetti D (2011) Accuracy of three age determination X-ray methods on the left hand-wrist: a systematic review and meta-analysis. Leg Med (Tokyo) 13: 120-33. doi: 10.1016/j.legalmed.2011.01.004
48. 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. doi: 10.1007/s00414-009-0329-8 [doi]
49. Hermetet C, Saint-Martin P, Gambier A, Ribier L, Sautenet B, Rerolle C (2018) Forensic age estimation using computed tomography of the medial clavicular epiphysis: a systematic review. Int J Legal Med 132: 1415-25. doi: 10.1007/s00414-018-1847-z
50. Levesque GY, Demirijian A, Tanguay R (1981) Sexual dimorphism in the development, emergence, and agenesis of the mandibular third molar. J Dent Res 60: 1735-41.
51. Vandeplas C, Vranckx M, Hekner D, Politis C, Jacobs R (2020) Does Retaining Third Molars Result in the Development of Pathology Over Time? A Systematic Review. J Oral Maxillofac Surg 78: 1892-908. doi: 10.1016/j.joms.2020.06.014
52. Schumacher G, Schmeling A, Rudolf E. (2018) Medical age assessment of juvenile migrants: An analysis of age marker-based assessment criteria. Joint Research Centre (JRC) Science for Policy Report, European Union Luxembourg.
53. 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

Figure captions

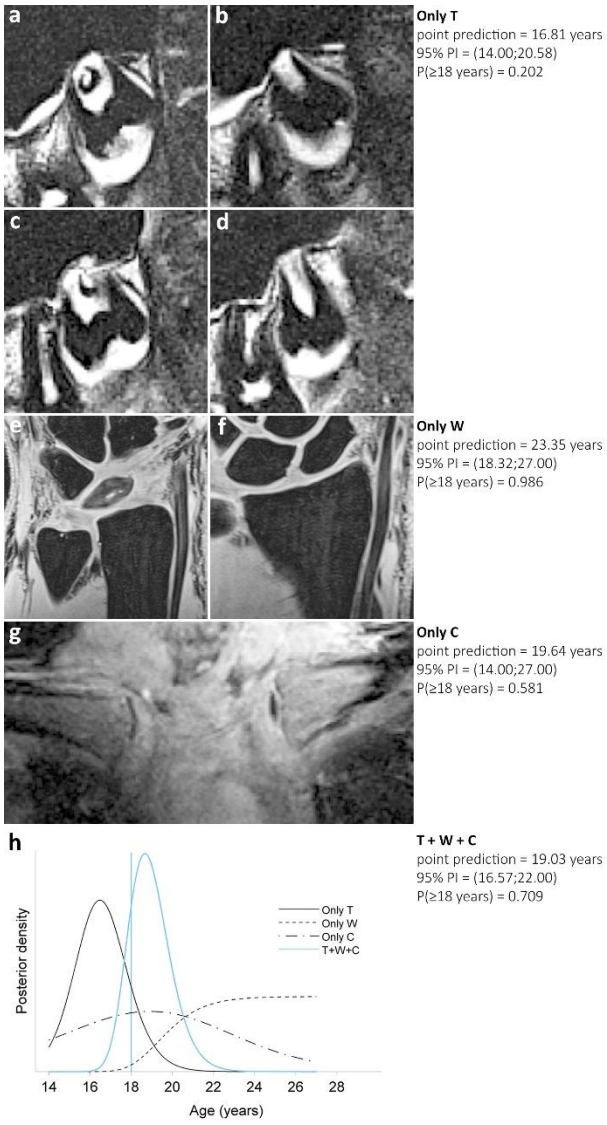

Fig. 1 | Case example from the original De Tobel et al. reference sample, for whom age estimation is based on a model that integrates MRI data of the third molars (T), left wrist (W) and both clavicles (C), applying Bayes' rule. The male participant was 19.56 years old. PI = prediction interval. **(a,b)** Third molar 18, with its buccal roots **(a)** and its palatal root **(b)**. Stage 4 is allocated to this third molar. **(c,d)** Third molar 28, with its buccal roots **(c)** and its palatal root **(d)**. Stage 4 is allocated to this third molar. The lower third molars were agenetic. **(e)** Stage 4 is allocated to the left distal ulna. **(f)** Stage 4 is allocated to the left distal radius. **(g)** The sternal extremity of the right clavicle is unsuitable for staging, due to poor image quality. Stage 3ab is allocated to the sternal extremity of the left clavicle. **(h)** Posterior distributions of age for both single site age estimation and multi-factorial age estimation. Figure is reproduced with permission from De Tobel [53].

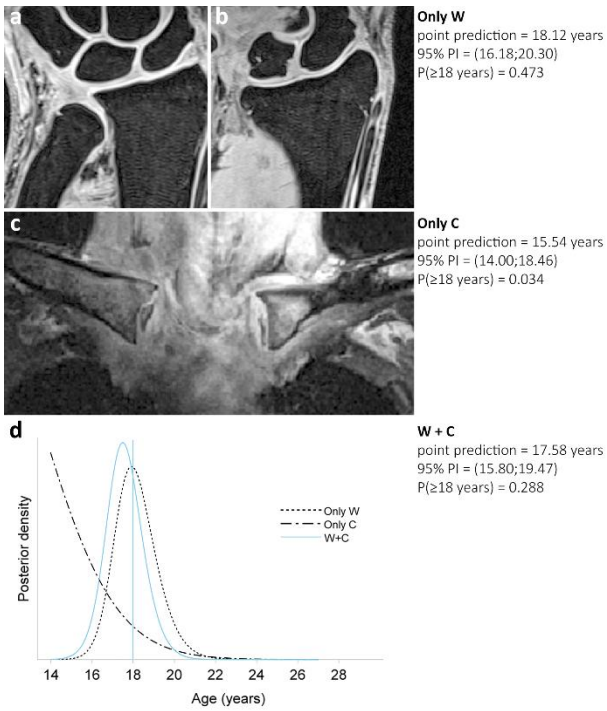

Fig. 2 | Case example from the Ghent validation sample. Age estimation is based on a model that integrates MRI data of the left wrist (W) and both clavicles (C). The male participant was 18.23 years old. PI = prediction interval. **(a)** Stage 3c is allocated to the left distal ulna. **(b)** Stage 3c is allocated to the left distal radius. **(c)** Stage 2b is allocated to the sternal extremity of both the right and left clavicle. **(d)** Posterior distributions of age for both single site age estimation and multi-factorial age estimation. Note that the combined result grants the benefit of the doubt to this participant who only recently became an adult.

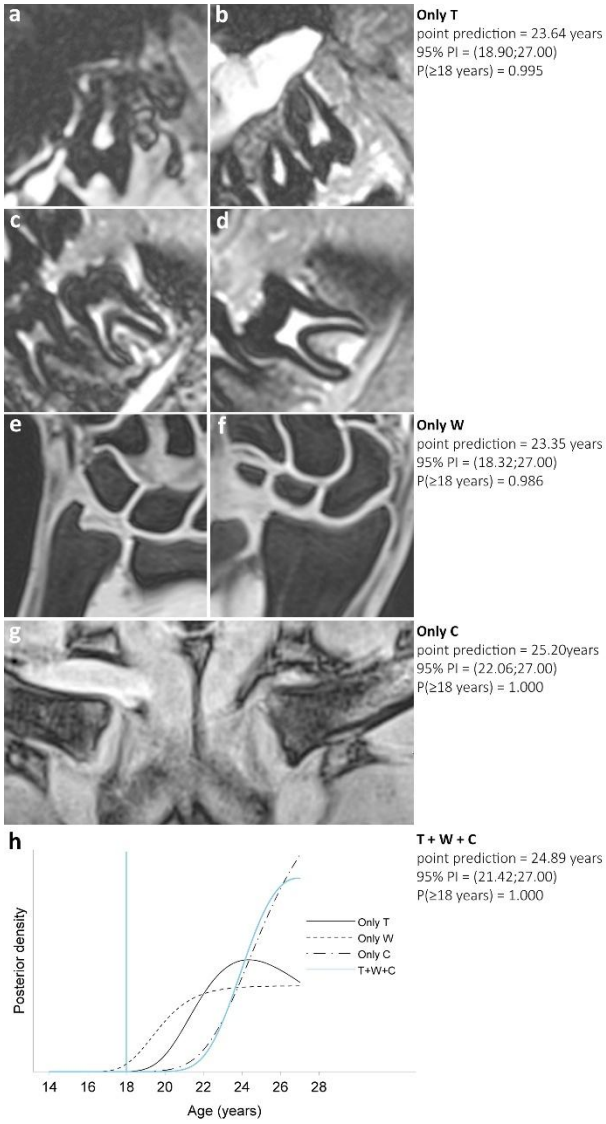

Fig. 3 | Case example from the Graz validation sample. Age estimation is based on a model that integrates MRI data of the third molars (T), left wrist (W) and both clavicles (C). The male participant was 21.14 years old. PI = prediction interval. **(a)** Stage 6 is allocated to third molar 18. **(b)** Stage 8 is allocated to third molar 28. **(c)** Stage 8 is allocated to third molar 38. **(d)** Stage 8 is allocated to third molar 48. **(e)** Stage 5 is allocated to the left distal ulna. **(f)** Stage 5 is allocated to the left distal radius. **(g)** Stage 3c is allocated to the sternal extremity of both the right and left clavicle. **(h)** Posterior distributions of age for both single site age estimation and multi-factorial age estimation.

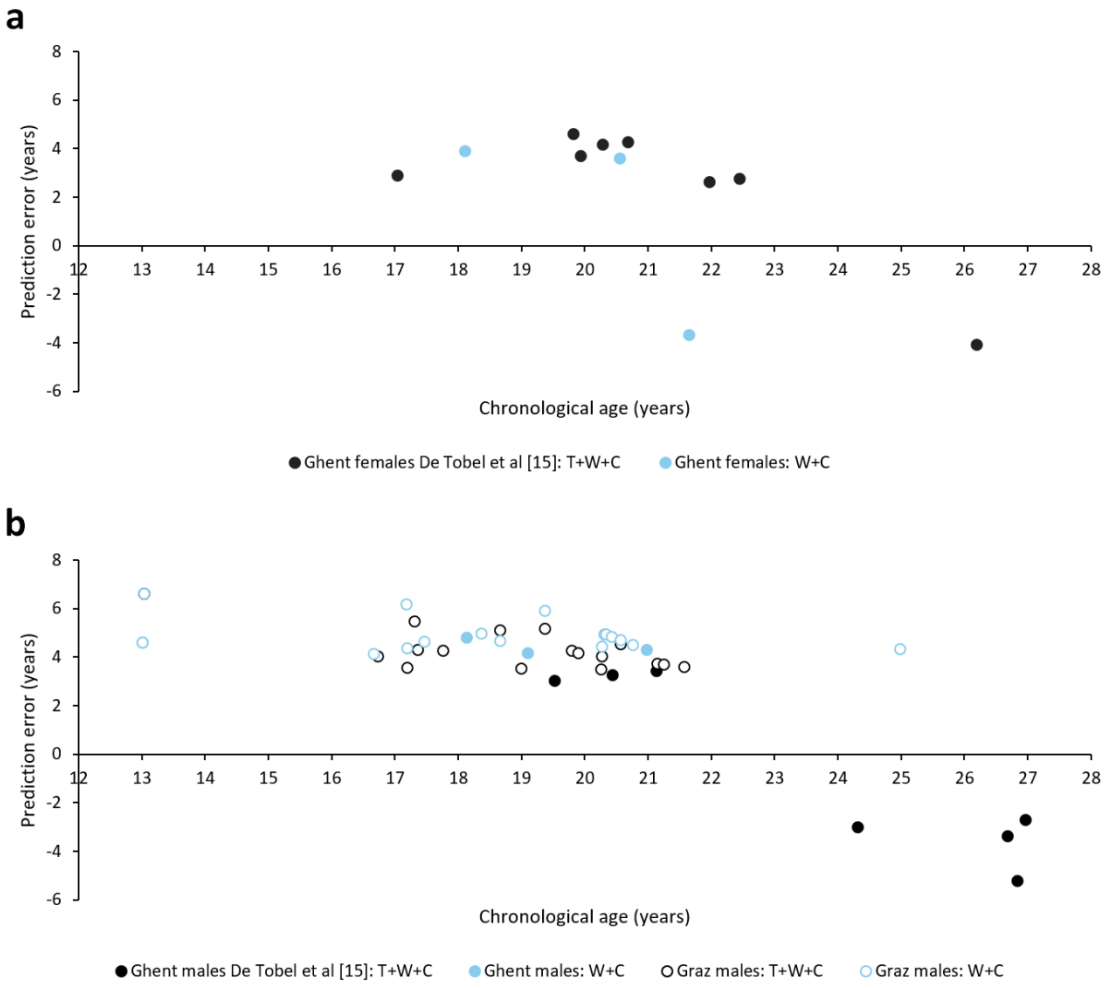

Fig. 4 | Scatter plot showing the 5% cases with the largest prediction error, per sample per multi-factorial forensic age estimation MRI combination model. A positive error indicates an overestimation of the chronological age, while a negative error indicates an underestimation. Cases for the original sample and validation samples are represented. Scatter plots corresponding to females **(a)** and males **(b)** are depicted in different plots. Results obtained using different samples and combination models are indicated with different grey scales. C = both clavicles; T = third molars; W = left wrist.

Tables

Table 1 | Distribution of the study population by sex across the age categories.

AGE (y) *	FREQUENCY GHENT SAMPLE			FREQUENCY GRAZ SAMPLE
	MALE	FEMALE	BOTH SEXES	MALES
13	-	-	-	27
14	-	-	-	19
15	-	-	-	27
16	10	9	19	35
17	8	9	17	38
18	10	10	20	26
19	9	9	18	30
20	10	10	20	35
21	5	9	14	24
22	-	-	-	28
23	-	-	-	23
24	-	-	-	23
TOTAL	52	56	108	335

* Each age category was considered from the birthday until the day before the next birthday, e.g. 13 means from 13.00 to 13.99 years.

Table 2 | Reproducibility of stage allocation per anatomical site, investigating intra-observer agreements.

ANATOMICAL SITE	ANATOMICAL STRUCTURE	INTRA-OBSERVER GHENT		INTRA-OBSERVER GRAZ	
		WEIGHTED KAPPA (95% CI)	N	WEIGHTED KAPPA (95% CI)	N
T	18	-	-	0.785 (0.681;0.888)	43
	28	-	-	0.832 (0.703;0.962)	47
	38	-	-	0.634 (0.500;0.769)	38
	48	-	-	0.827 (0.710;0.944)	41
W	RADIUS	0.672 (0.563; 0.782) *	99 [19]	0.795 (0.684;0.907)	50
	ULNA	0.740 (0.646; 0.835) *	99 [19]	0.937 (0.871;1.003)	50
C	RIGHT	0.730 (0.571;0.889)	41 [20]	0.631 (0.448;0.814)	44
	LEFT	0.623 (0.438;0.808)	42 [20]	0.537 (0.326;0.749)	42

* The result is the inter-observer agreement, since no result for intra-observer agreement is available. The inter-observer values are expected to be lower than the intra-observer agreement.
C = both clavicles; CI = confidence interval; N = number of participants included; T = all third molars; W = left wrist (radius and ulna).

Table 3 | Internal and external validation results on age estimation performance of the De Tobel et al [15] MRI combination models.

ANATOMICAL SITES	SEX	VALIDATION SAMPLE	N	ACCURACY OF THE POINT PREDICTION OF AGE		DISCERNING MINORS FROM ADULTS			
				MAE (y)	RMSE (y)	ACCURACY (%)	SENSITIVITY (%)	SPECIFICITY (%)	DISCRIMI-NATION SLOPE
W+C	Males	Internal [15] *	138	1.68	2.07	92.8	93.9	90.0	0.793
		External Ghent	52	1.55 [1.24;1.86]	1.90 [1.52;2.27]	78.8 [65.3;88.9]	70.6 [52.5;84.9]	94.4 [72.7;99.9]	0.615 [0.452;0.777]
		External Graz	329	1.66 [1.52;1.80]	2.08 [1.92;2.24]	88.8 [84.8;92.0]	94.7 [90.5;97.4]	80.7 [73.2;86.9]	0.700 [0.639;0.761]
	Females	Internal [15] *	158	1.83	2.30	92.4	93.9	88.4	0.682
		External Ghent	56	1.49 [1.20;1.77]	1.83 [1.48;2.18]	83.9 [71.7;92.4]	92.1 [78.6;98.3]	66.7 [41.0;86.7]	0.471 [0.314;0.628]
T+W+C	Males	Internal [15] *	138	1.36	1.73	92.8	93.9	90.0	0.807
		External Graz	335	1.41 [1.29;1.53]	1.80 [1.67;1.94]	89.3 [85.4;92.4]	95.2 [91.2;97.8]	81.5 [74.2;87.4]	0.733 [0.675;0.791]

* Results for internal validation were reported in the referenced paper, in which also the model development was reported; they did not result from the current study.

C = both clavicles; MAE = mean absolute error; N = number of participants included; RMSE = root mean square error; T = all third molars; W = left wrist (radius and ulna).

Accuracy = fraction of correctly categorised examinees; sensitivity = fraction of correctly categorised adults (≥18 years of age); specificity = fraction of correctly categorised minors (<18 years of age); discrimination slope = difference between minors and adults in mean predicted probability to be a minor.

95% confidence intervals are given between brackets. These were not reported in the reference paper [15].

SUPPLEMENTARY MATERIAL

Material and Methods

Image acquisition

Table S1 Overview of the applied MRI sequence parameters, per protocol per anatomical site.

PROTOCOL	ANATOMICAL SITE	MRI SEQUENCE TYPE	SEQUENCE PARAMETER DETAILS					
			TR (ms)	TE (ms)	FA (°)	FS	Voxel (mm^3)	TA (min)
GHENT	Wrist	T1 GE VIBE	12.70	4.82	10	Yes	0.4 × 0.4 × 0.4	5:57
	Clavicles	T1 GE VIBE	10	2.92	20	Yes	0.7 × 0.7 × 0.9	4:02
GRAZ	Teeth	PD 3D TSE	172-254	10-12	150	Yes	0.6 × 0.6 × 1.0 acquired 0.3 × 0.3 × 1.0 reconstructed	± 10:00
		T2 3D CISS	5.41	2.33	30		0.6 × 0.6 × 1.0	9:34
	Hand/wrist	T1 3D GRE VIBE	14	4.01	15	Yes	0.9 × 0.9 × 0.9 acquired 0.45 × 0.45 × 0.9 reconstructed	3:46
		T2 3D GRE DESS	14.28	5.18	30	Yes	0.8 × 0.8 × 0.8 acquired 0.4 × 0.4 × 0.8 reconstructed	2:04
	Clavicles	T1 3D GRE VIBE	9.77	3.72	12	Yes	0.9 × 0.9 × 0.9	5:29
		T2 2D TSE	2905	65	150		1.0 × 1.0 × 2.0	5:57

CISS = constructive interference in steady state; DESS = dual echo steady state; FA = flip angle; FS = fat suppression; GE = gradient echo; GRE = gradient-recalled echo; MRI = magnetic resonance imaging; PD = proton density-weighted; TA = acquisition time; TE = echo time; TR = repetition time; TSE = turbo spin echo; T1 = T_1-weighted; T2 = T_2-weighted; VIBE = volumetric interpolated breath-hold examination.

Image analysis

Table S2 Description of the developmental stages for third molars on MRI [39].

STAGE	CRITERIA
0	The crypt of the third molar is suspected without any calcification.
1	A beginning of calcification is seen at the superior level of the crypt in the form of an inverted cone or cones. There is no fusion of these calcified points.
2	a) Fusion of the calcified points forms one or several cusps which unite to give a regularly outlined occlusal surface.
	b) The outline of the pulp chamber has a flat or curved shape at the occlusal border.
	c) Initial formation of the radicular bifurcation is seen in the form of a hypo-intense calcified point.
3	a) The pulp chamber has a trapezoidal shape. The outline of the pulp horns is pointy and shaped like an umbrella top.
	b) Further downshaping of the crown and/or beginning of root formation is seen in the form of a spicule. The spicule is shorter than MR crown height.
	c) The calcified region of the bifurcation has developed further into a hypo-intense semi-lunar shape.
4	a) MR root length reaches at least one MR crown height.
	b) The calcified region of the bifurcation still has a semi-lunar shape or has developed further down.
5	a) MR root length reaches at least one and a half MR crown height.
	b) The calcified region of the bifurcation has developed further down from its semi-lunar shape to give the roots a more definite and distinct outline with funnel shaped endings. The funnel shape persists for some millimetres (i.e. it is not limited to a few pixels on the image)
6	a) The walls of the distal root canal are parallel and its apical end is still partially open.
	b) The walls at the apex of the root canal show relatively thin dentin.
	c) Remnants of the dental follicle are seen in the form of a hyper-intense area surrounding the apex.
7	a) The walls of the distal root canal are convergent and its apical end is still partially open.
	b) The walls at the apex of the root canal show relatively thin dentin.
	c) Remnants of the dental follicle are seen in the form of a hyper-intense area surrounding the apex.
8	a) The apical end of the distal root canal is completely closed.
	b) The walls at the apex of the root canal show relatively thick dentin.

Table S3 Description of the developmental (profound) (sub)stages for long bones [40-42].

STAGE	CRITERIA
1	Ossification centre is invisible (i.e. not yet ossified).
2	Ossification centre is visible (i.e. 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 (i.e. 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	The lengthwise measurement of the epiphysis is one third or less compared to the widthwise measurement of the metaphyseal ending.
3ab	The lengthwise measurement of the epiphysis is between one third and two thirds compared to the widthwise measurement of the metaphyseal ending.
3ac	The lengthwise measurement of the epiphysis is over two thirds compared to 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 (i.e. physeal plate is completely ossified). Physeal scar is still visible.
5	Complete union of the epiphysis and metaphysis. Physeal scar is indiscernible.

Results

Reproducibility of staging

Table S4 Cross tabulation of frequencies of allocated stages of the third molars by the observer in both staging sessions of the Graz sample. Third molar 18, 28, 38 and 48 are considered jointly since the results of all third molars were similar.

THIRD MOLARS

STAGE SESSION 1	STAGE SESSION 2									TOTAL
	0	1	2	3	4	5	6	7	8	
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
2	0	0	5	0	0	0	0	0	0	5
3	0	0	8	23	3	0	0	0	0	34
4	0	0	0	4	27	1	0	0	0	32
5	0	0	0	0	2	10	2	0	0	14
6	0	0	0	0	1	1	19	1	3	25
7	0	0	0	0	0	0	7	9	3	19
8	0	0	0	0	0	0	4	8	28	40
TOTAL	0	0	13	27	33	12	32	18	34	

Shaded cells indicate intra-observer agreements.

Table S5 Cross tabulation of frequencies of allocated (sub)stages of the left distal radius and ulna by the observer in both staging sessions of the Graz sample.

LEFT RADIUS

STAGE SESSION 1	STAGE SESSION 2							TOTAL
	1	2	3a	3b	3c	4	5	
1	0	0	0	0	0	0	0	0
2	0	1	3	0	0	0	0	4
3a	0	0	10	0	0	0	0	10
3b	0	0	0	0	0	0	0	0
3c	0	0	0	0	2	0	1	3
4	0	0	0	0	0	1	0	1
5	0	0	0	0	4	2	26	32
TOTAL	0	1	13	0	6	3	27	

LEFT ULNA

STAGE SESSION 1	STAGE SESSION 2							TOTAL
	1	2	3a	3b	3c	4	5	
1	0	0	0	0	0	0	0	0
2	0	5	1	0	0	0	0	6
3a	0	0	8	0	0	0	0	8
3b	0	0	0	0	0	0	0	0
3c	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	1	2	33	36
TOTAL	0	5	9	0	1	2	33	

Shaded cells indicate intra-observer agreements.

Table S6 Cross tabulation of frequencies of allocated (profound) (sub) stages of the sternal extremity of the clavicles by the observer in both staging sessions of the Graz sample. Right and left clavicle are considered jointly since the results of both sides were similar.

CLAVICLES													
		STAGE SESSION 2											TOTAL
		1	2a	2b	2c	3aa	3ab	3ac	3b	3c	4	5	
STAGE SESSION 1	1	6	1	0	0	0	0	0	0	2	2	0	11
	2a	2	2	0	0	2	0	0	0	0	1	1	8
	2b	0	1	1	0	0	0	0	0	0	0	0	2
	2c	0	0	0	0	0	0	0	0	0	0	0	0
	3aa	0	1	0	0	6	3	0	0	2	0	0	12
	3ab	0	0	0	0	1	9	0	1	0	0	0	11
	3ac	0	0	0	0	0	1	8	0	0	0	0	9
	3b	0	0	0	0	0	1	1	3	0	0	0	5
	3c	0	0	0	0	0	2	0	0	9	3	1	15
	4	0	0	0	0	0	0	0	0	0	10	2	12
	5	0	0	0	0	0	0	0	0	1	0	0	1
TOTAL		8	5	1	0	9	16	9	4	14	16	4	

Shaded cells indicate intra-observer agreements.