This is the accepted manuscript of the following publication:

De Tobel J, Bauwens J, Parmentier G, Franco A, Pauwels N, Verstraete K, Thevissen P. Magnetic resonance imaging for forensic age estimation in living children and young adults: a systematic review. *Ped Radiol.* Accepted on May 8, 2020. https://doi.org/10.1007/s00247-020-04709-x

FULL TITLE PAGE

Title page

Magnetic resonance imaging for forensic age estimation in living children and young adults: a systematic review

Author order	First name	Middle initials	Last name	ORCID	Affiliations
1	Jannick	/	De Tobel	0000-0002-8846-7339	1,2,3
2	Jeroen	/	Bauwens		1
3	Griet	I.L.	Parmentier		1
4	Ademir	/	Franco	0000-0002-1417-2781	2
5	Nele	S.	Pauwels	0000-0002-1862-1927	4
6	Koenraad	L.	Verstraete	0000-0003-3688-1076	1
7	Patrick	W.	Thevissen	0000-0003-0890-6264	2

Affiliations:

1. Diagnostic Sciences – Radiology, Ghent University, Belgium
2. Imaging and Pathology – Forensic Odontology, KU Leuven, Belgium
3. Oral Diseases and Maxillofacial Surgery, Maastricht UMC+, The Netherlands
4. Ghent Knowledge Centre for Health, Ghent University, Belgium

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 64 25

Fax +32 (0)9 332 49 69

jannick.detobel@ugent.be

Abstract

Background The use of magnetic resonance imaging (MRI) in forensic age estimation has been explored extensively during the past decade.

Objective To synthesize the available MRI data for forensic age estimation in living children and young adults, and to provide a comprehensive overview that can guide age estimation practice and future research.

Materials and Methods MEDLINE, Embase and Web of Science were searched. Additionally, cited and citing articles and study registers were searched. Two authors independently selected articles, conducted data extraction, and assessed risk of bias. Study populations including living subjects up to 30 years were considered.

Results Fifty-five studies were included in qualitative analysis and 33 in quantitative analysis. Most studies suffered from bias, including relatively small European (Caucasian) populations, varying MR-approaches and varying staging techniques. Therefore, pooling of the age distribution data was not appropriate.

Reproducibility of staging was remarkably lower in clavicles than in any other anatomical structure. Age estimation performance was in line with the gold standard, which uses radiographs, with mean absolute errors ranging from 0.85 to 2.0 years. The proportion of correctly classified minors ranged from 65% to 91%. Multi-factorial age estimation performed better than based on a single anatomical site.

Conclusion More multi-factorial age estimation studies are necessary, together with studies testing if the MRI data can safely be pooled. The current review results can guide future studies, help medical professionals to decide on the preferred approach for specific cases, and help judicial professionals to interpret the evidential value of age estimation results.

Keywords

magnetic resonance imaging

age estimation

child

adolescent

young adult

Declarations

Funding for this research was entirely provided by the department of Radiology and Nuclear Medicine at Ghent University.

The authors declare that there are no conflicts of interest.

This project was approved by the Ghent University Hospital Ethics Committee (EC 2017/0024, with Belgian registration number B670201730806), as part of an ongoing larger project (EC 2011/0842, B670201112782).

Upon contacting the corresponding author, the original tables with 'study characteristics', 'data extraction', and 'risk of bias assessment' can be provided. Also the tables on which all graphs in the Supplementary Material were based, can be provided by the corresponding author.

Contributions of authors:

- Draft the protocol JDT, GIP, NSP
- Study selection JDT, JB, GIP, AF
- Extract data from studies JDT, JB, GIP, AF
- Carry out the analysis JDT
- Interpret the analysis JDT, GIP, NSP, PWT, KLV
- Draft the final review JDT, GIP
- Disagreement resolution PWT, KLV
- Update the review JDT, JB, GIP, AF, PWT
- Supervision of the process NSP, PWT, KLV

Acknowledgements

We wish to thank Tom Verschoore (Bimetra, Ghent University Hospital), for his advice regarding the review protocol. Furthermore, we want the acknowledge Karoline Elisivdatter Nyhagen for providing us the reference list of her bachelor's thesis at the University of Central Lancashire entitled "Age Estimation in the Living Using

Magnetic Resonance Imaging – A Review of Current Methods Identifying the 18-Years-Old Threshold", which rendered two additional references to screen. We also thank Michiel de Haas (Netherlands Forensic Institute) and Martin Urschler (Ludwig Boltzmann Institute for Clinical Forensic Imaging) for providing us additional papers.

Our sincerest gratitude goes out to the authors who provided additional data for this review: Markus Auf der Mauer [42, 45], Astrid Junge [43], Martin Urschler [40, 44]. Furthermore, special thanks to Peter Roozenbeek (KU Leuven) for his advice and help to create the perfect graphs, and to Steffen Fieuws (I-BioStat, KU Leuven) for statistical support and advice.

Finally, we kindly acknowledge Inès Phlypo and Patrick Davis for their critical review of the manuscript.

MAIN BODY

Introduction

When birth records or other official identification documents reporting the age of an individual are unavailable in criminal, civil and asylum procedures, forensic age estimation can be deemed necessary by the authorities. The estimation usually has to contain a predicted age together with a measure of the uncertainty, and the probability that the examined person has reached a specific legally relevant age threshold. In most countries the age threshold lies between 14 and 22 years of age, representing children and young adults [1]. Furthermore, in sports, age estimation is used to ensure fair play by checking if athletes participate in the correct age category [2].

Established methods for age estimation mainly use radiographs to evaluate teeth, carpal bones and long bones, which are still developing in children and young adults. The 2D radiographic registrations have two major drawbacks. Firstly, they imply an exposure to radiation without a clinical indication, resulting in deontological and ethical issues [3]. In some countries the use of ionizing radiation is prohibited in asylum and civil procedures [4]. Secondly, on plain radiographs, superposition can yield mistakes or impede allocating a developmental status to the anatomical structures of interest [5].

To counter these drawbacks, several research groups have been studying the use of magnetic resonance imaging (MRI) to register the developmental status of the considered anatomical site. Since the details necessary to study development might not be clear in routine clinical MRI, several dedicated MRI protocols have been developed. However, MRI has not found its way to age estimation practice yet, because it remains unclear which is the optimal MRI approach.

The different MRI approaches were reported in pilot studies and cross-sectional reference studies. Compared to reference studies of age estimation based on radiographs of developing teeth or bones, the MRI studies have two shortcomings: (1) they all included a relatively small study population, and (2) few external validation studies (with an independent test sample) on any MRI approach for age estimation have been conducted. Due to these shortcomings, a first attempt to bring forensic age estimation based on MRI into practice resulted in large error rates [6].

To address the small study population, pooling of the MRI data could be considered to increase age estimation performance. However, a review of the MRI studies is indispensable to study if pooling is appropriate. MR-images are highly dependent on the technical parameters of the MRI approach, thus, merging incompatible data would

lead to wrong conclusions. Unfortunately, a review cannot address the lack of external validation studies, but it can provide an overview of the internal validation statistics (within the study population).

To the best of our knowledge, no systematic review has been published on the subject yet. Therefore, the current systematic review was conducted with the following objectives: (1) to synthesize the MRI data for forensic age estimation in living children and young adults, and (2) to provide a comprehensive overview that can guide age estimation practice and future research. The following research questions were studied:

1) How is age estimation on MRI affected by population characteristics and MRI approach?
2) How does the development of different anatomical structures, as registered on MRI, relate to chronological age in living children and young adults?
3) How reproducible is developmental stage allocation based on MRI?
4) What is the performance of age estimation based on development of different anatomical structures as registered on MRI?
    a) Which anatomical structures provide the best MRI information to render a point prediction of age?
    b) Which anatomical structures provide the best MRI information to discern minors from adults?

Materials and methods

Protocol design

The review protocol was drafted according to the Cochrane Guidelines for review protocols (http://training.cochrane.org/) [7], and registered in Prospero, international prospective register of systematic reviews (http://www.crd.york.ac.uk/PROSPERO), with registration number CRD42017061043). This project was approved by the Ghent University Hospital Ethics Committee (EC 2017/0024, with Belgian registration number B670201730806), as part of an ongoing larger project (EC 2011/0842, B670201112782). The reporting of the systematic review complies with the PRISMA statement [8, 9].

Selection of studies

Criteria for considering studies for this review

Types of studies

Cross-sectional observational studies were included. When a pilot study was published, followed by a more recent study including a larger study population, only the final publication was included for the review. When the final publication was not yet published, results of the pilot publication were considered. Furthermore, cohort observational studies were included, but only results of one moment in time were extracted to avoid bias. Case reports and case series were also included, since they might provide information on minimum and/or maximum age per developmental stage. Review articles were excluded. Furthermore, no restrictions were made based on the country of publication, language or publication date.

Types of participants

Study populations including living children, adolescents, and adults up to 30 years old were considered. After the age of 30, age estimation is no longer based on development, but rather on degenerative changes [10, 11]. Moreover, studies which only included deceased individuals were excluded, since MRI is influenced by body temperature [12] and motion artefacts [13].

Types of interventions

MRI of any field strength was included studying hard tissue development related to age. Authors should refer to the staging technique used to assess development. When measurements were made, the way of obtaining them should be described clearly. It was considered inappropriate to compare the age distributions within developmental stages based on MRI with those based on radiographs, since it has been demonstrated that imaging technique specific reference data are required [14-19].

Types of controls

The control for age estimation performance was the chronological age.

Types of outcome measures

The included papers should provide any of these outcome measures:

- Descriptive statistics on age distribution within the different developmental stages of the considered anatomical structures.
- Probabilities of attaining certain threshold ages, diagnostic indices.
- Statistics on the performance of the age estimation model.

Search methods

According to the described eligibility criteria, literature was searched in MEDLINE (via the PubMed interface), Embase (via the embase.com interface), and Web of Science. The search strings are reported in the Supplementary Material. Furthermore, reference lists of included studies were searched for additional suitable papers, and papers citing the included studies were searched using Web of Science and Google Scholar. Finally, grey literature was searched by consulting the following study registers: the United States' ClinicalTrials.gov, EU Clinical Trials Register, the United Kingdoms' ISRCTN registry, German Clinical Trials Register (DRKS). All searches were conducted on September 2, 2018.

Reviewing process and selection of studies

Two authors conducted every step of the reviewing process independently. The first author (JDT) was a reviewer throughout the whole process. Other authors (JB, GP, AF) acted as second reviewers. After a first selection of articles based on title and abstract, the authors considered and compared their selection to achieve a consensus. Of

the retained abstracts, the full text paper was checked independently for eligibility. Discrepancies between reviewers were identified at this stage and resolved by discussion to reach consensus. A record was kept of reasons for excluding studies at each step (either title and abstract, or full text). Reasons for exclusion were checked in the following order:

1. Pilot of other reference.
2. Wrong study design (S): review.
3. Wrong population (P): deceased individuals, insufficient data to differentiate within the group of 1 to 30 years of age.
4. Wrong intervention (I): MRI studying soft tissue
5. Wrong outcome measures (O): no data on age distribution or age estimation performance.

References were managed and duplications removed with Endnote software. Covidence software (Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org) was used for study selection. The process and the results of the literature search and study selection were presented in the PRISMA flowchart (Fig. 1).

Data extraction and management

Study characteristics and outcome data were extracted by two reviewers independently. The study characteristics table summarized data on study population, MRI approach, staging technique, statistical analysis, and observers. The data extraction table summarized data on missing data, correlation between stages and age, age distributions within stages, reproducibility of staging, regression formulas, and age estimation performance.

When multiple records were identified of the same study, they were collated, so that the unit of interest in the review was the study, rather than each record.

Data analysis

The results from data extraction were compared to detect trends relevant to the research questions. Age distributions within stages were summarized into graphs, combining data from multiple studies with similar approaches. Then, it was checked whether meta-analysis of those data would be appropriate.

Quality assessment of studies

Risk of bias assessment of included studies

Risk of bias was assessed by two reviewers independently using a dedicated tool based on the EPOC overview [20] and QUADAS-2 [21]. Questions were phrased in such a way that the preferred approach corresponded with answering “yes”. In case the answer was “no”, the reasons for high risk of bias were elaborated.

Dealing with missing data

When information was missing in a paper, only graphs were reported or clarification was needed, the authors were contacted by e-mail or telephone. When the authors did not provide additional data, but graphs were reported in their paper, data were extracted from the graphs using calibration and the measuring tool in Adobe Photoshop CS2. For missing values due to images not being assessable, the reasons were registered. Moreover, it was evaluated whether missing values depended on age, imaging sequence or research group.

Assessment of heterogeneity

Methodological heterogeneity was checked by comparing biological origin of participants and types of MR sequence. Statistical heterogeneity was taken into account by comparing the different types of statistical analysis that were used.

Results

The essential results are discussed in this section, while the Supplementary Material includes additional considerations, overview tables and graphs.

Selection of studies and data

Figure 1 displays the selection process, whose details are elaborated in the Supplementary Material.

Characteristics and quality of included studies

Study characteristics

Results on age distribution were affected by the study characteristics displayed in Table 1 and Tables 5 and 6 of the Supplementary Material. In those tables, studies are grouped according to anatomical site from head to toe: skull, teeth, chest, upper limb, hip, and lower limb. Note that only one study has been published which integrates information from several anatomical sites into one age estimate (multi-factorial age estimation, as opposed to single site age estimation) [34].

Study populations

Table 1 displays the population characteristics. Most studies included European (Caucasian) populations. In addition, there were limited studies including African, Asian and Latin American populations. Healthy volunteers or athletes were recruited prospectively or patient records were searched retrospectively, excluding pathology. Only one study included patients with possible growth disorders, but their focus was on the agreements between X-ray based bone age and MRI based bone age, rather than on chronological age [19]. Furthermore, the age range of the study populations varied widely, with some studies only including minors, while others included participants from birth to age 30.

Magnetic resonance imaging approaches

The included scanning protocols used scanners with field strengths from 0.2T to 3T (Table 5 of the Supplementary Material). The low field open scanners did not render the highly detailed images necessary for staging and substaging of both the epiphyseal and physeal development, but they allowed assessing individual bone

development of the hand/wrist [36, 39]. Conversely, to study developing teeth and clavicles, 3T appears to be necessary [35, 41, 42, 48-52].

T1 sequences were most frequently used to study bone development, whereas for teeth, T2 sequences were most frequent. The voxel size of those sequences varied widely. Retrospective studies mostly lacked specifics on this, but some reported slice thicknesses ranging from 2 to 4 mm. In-plane resolution was never lower than 1.0 × 1.0 mm^2 and high resolutions were reached in all anatomical sites, with a minimum of 0.188 × 0.188 mm^2 [46]. Unfortunately, the study with the highest resolution [46] did not report the acquisition time. Since 6 minutes 30 seconds could be considered the maximum acceptable acquisition time [55], only the teeth and the iliac crest exceeded this threshold.

Staging techniques and statistical processing

Regarding dental development, the first staging techniques were based on radiographs [56, 57]. However, since the cemento-enamel junction is indiscernible using the reported MRI sequences, these staging techniques were said to be inappropriate for MRI [42]. Consequently, an MRI-specific technique was reported (Table 6 of the Supplementary Material) [42, 49].

Regarding bone development, staging techniques were developed based on radiographs and CT. In contrast to the dental staging techniques, the criteria for staging bone development did not include tissues which are indiscernible on MRI. Therefore, the staging techniques could integrally be applied to MRI (Table 6 of the Supplementary Material). Moreover, they could be grouped when their stages overlapped. The most elaborate staging technique (Table 2) was developed by a German research group and combined stages [58], substages [59] and advanced substaging [60]. When applicable, other staging techniques were transposed to this staging technique to compare studies (Fig. 5 of the Supplementary Material).

In a minority of included papers, regression was used to relate ordinal staging data to age. Most papers only reported descriptive statistics on age per stage in tables. Those statistics were summarized in Fig. 5 of the Supplementary Material and will be elucidated further on. Furthermore, a few papers applied Bayes' rule to nuance the age estimation, which has been stated to be more appropriate than linear regression [47-49, 69, 70]. Finally, advanced machine learning was applied to estimate age in two papers, but no details on the statistical approach were reported [34, 45]. The latter studies, together with four others, applied cross-validation [30, 34, 45, 47-49], while one study tested their results on a validation sample [31].

Risk of bias assessment

Bias was a major concern in almost all included studies (Table 7 of the Supplementary Material). Selection bias was caused by including elite football players, who might be advanced in their development [33], or by including patients in whom developmental disorders could not be ruled out [19, 23, 24, 29]. Furthermore, the small study samples resulted in an uneven distribution among age categories [2, 16, 19, 22, 25, 26, 29, 33, 35, 39-41, 43-46, 53, 61, 62, 69-76], or frequencies per age were not reported [23]. Retrospective studies did not report the biological origin of the population, while some prospective studies included different ethnic groups [2, 35, 44], or only a few individuals of another ethnic group [53]. Moreover, few studies reported the socio-economic status of their study participants. Other sources of bias were elaborated in the Supplementary Material.

Because of the highly biased nature of most studies, it was decided not to conduct meta-analysis on the age distributions per stage. Moreover, it remains unclear if data from an anatomical site can safely be pooled, when the MRI sequences are not identical. To date, only one study compared scanning protocols in the same individuals, but their sample was too small to draw strong conclusions [15].

Quantitative synthesis

Age distributions in relation to development

Statistics were extracted from boxplots for the following references: [2, 69, 70, 79]. Moreover, the following authors provided additional data: Markus Auf der Mauer [43, 46], Jannick De Tobel [42, 47-49], Astrid Junge [44], and Martin Urschler [19, 45].

To provide a clear overview, statistics on age distributions per stage were displayed in box-plots (Figs. 4 and 5 of the Supplementary Material). Note that some boxplots (in early stages) fall entirely below the 18-year-threshold, while others (in late stages) lie entirely above the threshold. Cut-off stages for these absolute statements regarding childhood and adulthood are summarized in Table 3.

Reproducibility of staging

To quantify reproducibility, different statistics were used, with a majority of studies reporting reproducibility statistics > 0.80 (Table 8). However, different studies on clavicular development indicated that staging was less reproducible than at other anatomical sites [35, 48, 83]. Furthermore, for all anatomical sites except the ankle, at

least one study reported considerably lower values than 0.80 [19, 35, 40, 41, 48, 61, 79]. No relation between those lower values and MR-sequence or staging technique seemed apparent.

Age estimation performance

Regression formulas were reported in [25, 30, 36, 72, 82]. Corresponding coefficients of determination ranged from 0.40 [25] to 0.85 [36]. When statistical models were applied to estimate age, two aspects were considered to quantify age estimation performance: (1) the point prediction of age with its uncertainty, and (2) the ability to discern minors from adults.

The first aspect is reflected by the mean absolute error and root mean squared error. Only a few studies reported mean absolute error. For females, mean absolute error reached 2.0 years studying third molars [49]. For males, it reached 1.7 years studying third molars [49], 0.85 years studying the left hand/wrist [45], and 1.14 years combining third molars, both clavicles and the left hand/wrist [34]. Not sex-specific, mean absolute error reached 1.97 years studying both clavicles [48], and 1.79 years studying the left wrist [47]. Moreover, the effect of large differences between chronological and estimated age was quantified by the root mean squared error in three studies: for females root mean squared error was 2.38 years and for males 2.06 years, studying third molars [49], while it was 2.60 years studying both clavicles [48], and 2.24 years studying the left wrist [47]. The latter values were not sex-specific.

The second aspect is reflected by predictive probabilities to be younger/older than 18 and by diagnostic indices (Table 4). For diagnostic indices in the current review, reported statistics were recalculated so specificity would reflect the proportion of correctly classified minors, whereas sensitivity would reflect the proportion of correctly classified adults. Overall, the sensitivity was higher (ranging from 83% to 100%) than the specificity (ranging from 66% to 93%), while the reverse is desirable in forensic age estimation. Still, the reported predictive probabilities to be a minor were very low for the final stages of development, with values under 1% for third molars and clavicles.

Discussion

Characteristics and quality of included studies

Study populations and magnetic resonance imaging approaches

The current systematic review provides an overview of how hard tissue development registered with MRI relates to age. Included studies showed high risk of bias, mainly due to their study population. Since a wide age range was studied, from birth to age 30, large reference populations are indispensable. It has been suggested that at least 10 participants per age category of one year per sex should be included per anatomical structure [1]. Moreover, the age range of the study population affects lower and upper limits of age distributions within developmental stage, as well as the mean age. This phenomenon is called 'age mimicry' and has been a major issue in age estimation for decades [84]. Ideally, a reference study should include participants with an age range starting several years before the studied anatomical structure starts its development, and ending several years after the structure has reached full maturity. For instance, an ideal reference study on third molars' development might include participants between 6 and 28 years old [85]. Unfortunately, these ideally designed studies are scarce even using radiographs, which can easily be done retrospectively. Therefore, it seems self-evident that, in the case of MRI, those ideally designed studies will be rarer still. Only for the clavicles' sternal end did two studies encompass the entire development with lower and upper age margins beyond developmental changes [48, 52]. For other structures, pooling the data of different studies might address this issue, but before this is done, it needs to be ascertained whether or not it is safe to pool data obtained with different MR-sequences. After all, it has been demonstrated that age distributions within stages might differ between sequences for third molars [15] and for the left wrist (Fig. 2) [47]. In the latter study, applying the model derived from one MR-sequence to assessments of the other sequence resulted in a markedly worse age estimation performance [47]. Moreover, different sequences may lead to different staging techniques, impeding the pooling of data [61, 62]. On the other hand, different sequences might provide complementary information, to allow for a more nuanced age estimation [47, 63].

Compared to age estimation studies using radiographs, MRI study populations were relatively small, which could be attributed to the MRI technique. Since developmental stages are based on details, such as bone bridging and apical closure of teeth, routine clinical MRI is mostly not suitable for age estimation. For instance, a thorax MRI will not be suitable to study clavicular development, and neither will a maxillofacial MRI be suitable to assess the apex of third molars. Only larger anatomical structures, such as knee and ankle bones, show sufficient details on

clinical MRI. This also explains why only those structures have been studied for age estimation in retrospective studies [16, 23, 25, 26, 28, 29, 31, 32, 61, 62, 69, 71, 73-76, 82]. Smaller structures require a dedicated scanning protocol, with a dedicated coil and sufficiently high in-plane resolution (Table 5 of the Supplementary Material), and thus, require a prospective study design. Still, such prospective studies have been conducted and it should be investigated whether their data can safely be pooled to create a large reference study.

Ethnic differences between populations have been studied using radiographs. Conclusions vary, with some authors claiming that inter-individual variability within ethnic group is larger than inter-ethnic variability [85-89], and others claiming that socio-economic status is more important than ethnicity [90]. By contrast, differences between ethnic groups have been demonstrated too [91, 92]. Presumably, trends in those studies also apply to MRI, but ethnic differences have only been studied for hand/wrist MRI [2, 44, 72, 79]. Moreover, these studies were only conducted in football players, who might be more advanced in their development than a general population of the same age [93-96]. After all, their advanced development might be part of their talent, i.e. their advanced development might contribute to better performance in sports. Thus, they might be scouted at an earlier age and be more likely to move on to elite sports. The study by Sarkodie et al. (2018) [33] was excluded for quantitative analysis, because it only included elite football players. At the other end of the spectrum, skeletal development in gymnasts might be delayed, allowing more elasticity at a relatively older age [95, 96]. Maybe different standards should be applied to athletes, to take into account their possible advanced or delayed skeletal age.

Staging techniques and statistical processing

MRI-specific staging techniques have been developed [49, 61, 63], but no comparative studies were conducted between staging techniques. Moreover, two studies on clavicle MRI have raised concerns about a possible confusion between stage 1 and stages 4/5 [48, 83]. The authors advise to discard clavicles in those stages for age estimation, and assess development of other structures instead.

Remarkably, only one study [34] has combined the information of three anatomical sites into one age estimate. Other groups have studied different anatomical structures in the same individuals, but did not report how to combine them. From studies using radiographs, it has been demonstrated that linear regression takes on statistical assumptions that do not hold for age estimation [97]. Neither should conditional independence be assumed [98]. Otherwise, artificially narrow uncertainty intervals of the point prediction and artificially high probabilities to be a minor or an adult will cause the judicial evidence to appear stronger than it really is [84, 98, 99].

Quantitative synthesis

Age distributions in relation to development

Bone development has been studied with MRI at most joints of the appendicular skeleton. The only site of the axial skeleton that has been studied was the spheno-occipital synchondrosis. Combined, these anatomical sites cover development from childhood to adulthood. By contrast, dental development has only been studied with MRI in molars, while in children up to age 14, the development of other permanent teeth is essential to estimate age [100, 101].

The graphs (Fig. 5 of the Supplementary Material) revealed some remarkable concerns about how stages relate to age. Firstly, only few anatomical sites and staging techniques provided a steady increase of age with increasing stage, with all participants in the first stage well below the 18-year-threshold and those in the final stage well above it in both sexes. They were Dedouit staging of the distal femur (Fig. 5k) and Vieth staging of the distal femur (Fig. 5n). De Tobel staging of the lower left third molars came close, but the minimum ages of the final stage were still close to 18 (Fig. 5d).

Secondly, the high maximum ages in stage 1 of clavicular development, and the low minimum ages in stages 4 and 5 suggest that those stages might be confused (Fig. 5e), as was pointed out in the original studies[48, 83]. This hinders a logical increase of age with an increase in stages.

Thirdly, although in wrist MRI, Dvorak stage 1 coincides with Schmeling stage 2, Dvorak stage 1 has never been reported above the age of 18 (Fig. 5g), while Schmeling stage 2 has been reported in one male of 18.6 years old (Fig. 5h) [70]. At the other end of the spectrum, in third molar MRI, De Tobel stage 8 coincides with Demirjian stage H. The first has not been reported below the age of 18 (Fig. 5d), while the latter has in males (Fig. 5c) [41, 50].

Fourthly, the influence of the study populations' age ranges is obvious. For instance, Fig. 5e demonstrates that the boxplots of the male participants in Vieth et al. (2014) are situated at the upper ends of other studies' box plots for lower stages, while they are at the lower end of other studies' box plots for higher stages. This can be explained by the narrow age range (5 years) of participants in Vieth et al. (2014). The same applies to Schmidt et al. (2015) in Fig. 5h. Fifthly, the iliac crest does not seem useful for age estimation, since ages within stages all overlap [78]. However, this study suffered from high population bias, with the same narrow age range of participants as Vieth et al. (2014) [51] and Schmidt et al. (2015) [64].

Finally, the introduction of substaging was clearly an attempt to provide more accurate age estimation around the age of 18. They provide a more gradual increase of age with increasing stage than the main stages.

However, there is more to certain staging techniques than the graphs revealed. Some MRI-specific characteristics of skeletal structures have been studied, but their relevance to age estimation remains unclear. The threefold stratification sign was stated to be useful by Timme et al. (2017) [65] while De Tobel et al. (2019) [47] could not confirm its use. Other signs such as the metaphyseal stripe [23], the oreo-sign and the crack-sign [31] still need to be explored in future studies.

Furthermore, considering how stages relate to age, correlation coefficients and coefficients of determination need to be interpreted cautiously, since they depend on the age distribution of the study population. Relatively high coefficients have been reported for single site age estimation based on MRI. Still, they are expected to increase by multi-factorial age estimation, as has been demonstrated for multi-factorial age estimation based on radiographs and computed tomography (CT) [102-107]. Although only one study on multi-factorial age estimation based on MRI has been published [34], all researchers in this field prefer multi-factorial age estimation over single site age estimation [108]. However, no study has been published on how the MRI information of the different sites can be combined appropriately for age estimation. Stern et al. (2017) combined all four third molars, both clavicles and the left hand/wrist [34]. Unfortunately, the statistical approach of their network remains to be elucidated. This combination of third molars, clavicles and hand/wrist complies with international recommendations, but is only partly supported by the current results of the review. Table 3 suggests that in females, combining third molars, the left hand/wrist and the knee might render a more robust model for age estimation. For males, combining third molars, the proximal humerus and the knee might be ideal. However, in practice, a uniform approach for both sexes is desirable.

Reproducibility of staging

Another major concern regarding age estimation based on MRI is the low reproducibility of staging that has been pointed out by some authors (Table 8). An obligatory quality control of centers that perform age estimation is still lacking, resulting in large discrepancies between results from different centers [109]. This already affects the current gold standard of age estimation, using radiographs, and its effect might be even larger using MRI, considering the complexity of interpreting different MR sequences. Therefore, staging development should be

based upon a consensus between experts. These experts should be experienced in age estimation as well as being experienced in interpreting the imaging modality at hand.

To solve this problem, automated approaches have been developed to assess radiographs for age estimation [110, 111]. Since validation studies support the use of these approaches, they are applied in current age estimation practice [112]. Such an automated approach has been developed and optimized for MRI, but still needs to be validated [34, 45, 113-116]. Moreover, should the same automated approaches be used internationally, discrepancies between age estimation performed in different institutes would, presumably, be eliminated [117, 118].

Age estimation performance

Few MRI studies have developed models for age estimation and reported statistical measures of age estimation performance. Remarkably, the same applies to X-ray studies. Studying radiographs of third molars, Thevissen et al. (2010) reported a mean absolute error of 1.13 years [97]. Knowing that their study population included 2513 participants, one might presume that such a mean absolute error value would also be reached by larger MRI studies. Note that this value is almost equal to the one reached by the multi-factorial age estimation MRI study by Stern et al. (2017) [34]. Therefore, the limiting effect of the small study populations in MRI studies might be overcome by the study of multiple anatomical sites with MRI. Furthermore, note that studies applying Bayes' rule to estimate uncertainty of the point prediction are not hampered by 'age mimicry' and counter false assumptions that are made when linear regression is applied [98]. Therefore, interpreting confidence intervals from those studies should be preferred over those obtained from age distribution tables or regression.

Similar to the better (i.e. lower) mean absolute error, the proportion of correctly classified minors is better (i.e. higher) for multi-factorial age estimation than for single site age estimation. This has been demonstrated for MRI [34] as well as for radiographs [119].

Age estimation in practice

To combine the information of different anatomical sites for forensic age estimation two approaches have been put forward. The first approach – called the minimum age principle – is based on descriptive statistics of the age distributions within stages, reported in reference studies [108]. The combined age estimation is an interval (Fig. 3a). For the lower border of the interval, the highest minimum age is retained, since for that anatomical site, no

individuals younger than that age have been reported. For the upper border of the interval, the lowest maximum age is retained, since for that anatomical site, no individuals older than that age have been reported.

The second approach is also based on the age distributions within stages, albeit incorporated in a statistical model [98]. Posterior density curves of age are obtained using a continuation ratio model with Bayesian correction for violation of the conditional independence assumption. The combined age estimation is defined by the combined curve, providing the following statistics: point prediction, 95% prediction interval, and the probability to be an adult (Fig. 3b).

There is no legislation on which approach should be applied or which statistics should be reported. Moreover, the magistrate who decides about a case is free to interpret the findings. For instance, when the age estimation interval of the first approach is close to the threshold of 18, but does not contain it (Fig. 3a), then the magistrate might decide to grant the benefit of the doubt and consider the individual as a minor. Similarly, when the second approach renders a probability to be an adult equal to 0.706 (Fig. 3b), then the magistrate decides if this is sufficient to consider the individual as an adult. Therefore, it is up to the forensic expert who conducts age estimation (e.g. radiologist, odontologist) to be transparent and clear in the report, and to motivate and nuance the findings as much as possible. Moreover, to minimize the effect of inter-observer variability, at least two experts should reach a consensus about the age estimation.

Strengths and weaknesses

This systematic review provides a comprehensive overview of literature that is currently available on age estimation based on MRI. It puts the studies in perspective, allowing medical professionals to decide on which approach seems the most valuable in their casework, and allowing judicial advisors to interpret the evidential value of the age estimation results. According to the PRISMA guidelines, all steps of the review were independently conducted by two reviewers, to avoid errors in the reported data.

However, this review also faced two limitations. Firstly, the search string did not include a part on “development”. Instead only “age estimation” and its variants were used. Therefore, there remains the possibility that studies on development were missed, which may, in turn, have highlighted other MRI-specific signs that might be of interest to age estimation. On the other hand, the encountered studies on development – without a focus on age estimation – were excluded from quantitative analysis, since their data was not sufficiently extensively reported. Secondly, pooling of the data was considered inappropriate, because of discrepancies between the MRI approaches and the

staging techniques. New studies are necessary to compare the age distributions within stages using different MRI approaches in the same population.

Future prospects

Recommendations for new studies

The use of MRI for forensic age estimation has been intensively studied since 2007, because of its major advantage of avoiding ionizing radiation. In its most recent Practical Guide on Age Assessment [117], the European Asylum Support Office states that "radiation-free methods should be applied first and only as a last resort can other methods involving radiation be considered". However, in the European Commission's Science for Policy Report by the Joint Research Centre [120], the authors state that "more studies should be conducted with MRI instead of CT in order to increase the available knowledge base". Consequently, despite the large number of studies discussed in this systematic review, MRI has not found its way into age estimation practice. Thus, the considerations from this review should be taken into account when future studies are designed and when MRI would be taken into practice for age estimation. In particular, the following recommendations can be made:

1) Larger reference populations are desirable. Since the prospective nature of studies impedes a fast expansion of reference data, it would make more sense to try to combine the data of different research groups. However, since small differences exist between MRI approaches and between populations, comparative studies are needed to check if the data can be pooled safely.
2) Multi-factorial age estimation seems to improve age estimation performance, as has been demonstrated using an automated age estimation method. Since most MRI data is based on staging of development, studies are needed in which that staging information is combined using an appropriate statistical approach.
    a) Several research groups have collected MRI data at different anatomical sites, in the same individual, on the same day. Those groups can attempt to combine that information to create age estimation models, taking into account the possible conditional dependence.
    b) It remains unclear if data from different anatomical sites can be combined safely to create age estimation models, when those data were not collected in the same individual. This could be studied, as soon as results from studies complying with the former recommendation (2a) are available.

Incorporating soft tissue development

Since the intervention of interest was MRI, results of the initial search included many studies on brain development and degeneration. However, in literature on age estimation in children, adolescents and young adults, the developing brain is generally not considered. After all, structural changes in the brain are mostly studied in older patients, when degeneration occurs related to age (or disease). However, changes in the developing brain might be useful for age estimation in younger individuals. Another strength of MRI is the possibility of studying dynamic changes in the body, such as diffusion in the brain or blood flow in the heart [121-126].

Therefore, since inter-individual variation remains a challenge in age estimation, adding soft tissue information might allow for a more nuanced age estimation than that based solely on hard tissue information. Moreover, studying functional and anatomical age-related changes in a research context is justifiable because of the lack of ionizing radiation. MRI even enables longitudinal evaluation of the changes over the years in an ethically justifiable way. However, to date, the bridge between hard and soft tissue development remains unexplored.

Conclusion

Single site age estimation using MRI has been studied extensively, providing several reference studies, which all included a relatively small study sample. Although a review might solve the issues of small study samples and disparities in their age distributions by pooling the data, this was currently not appropriate, because of a wide variety in study characteristics. Furthermore, the current review highlighted that age estimation performance was better for multi-factorial age estimation than for single site age estimation. As a next step in the field, more multi-factorial age estimation studies are imminent, since MRI avoids the use of ionizing radiation and, consequently, allows the study of multiple anatomical sites. The current review results can guide those multi-factorial age estimation studies. Moreover, this review can help medical professionals to decide on the preferred approach for specific cases, and it can help judicial professionals to interpret the evidential value of age estimation results.

References

1. Schmeling A, Geserick G, Reisinger W, Olze A (2007) Age estimation. Forensic Sci Int 165:178-181.
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.
3. Thevissen PW, Kvaal SI, Dierickx K, Willems G (2012) Ethics in age estimation of unaccompanied minors. J Forensic Odontostomatol 30 Suppl 1:84-102.
4. Lockemann U, Fuhrmann A, Püschel K, Schmeling A, Geserick G (2004) Arbeitsgemeinschaft für Forensische Altersdiagnostik der Deutschen Gesellschaft für Rechtsmedizin. Rechtsmedizin 14:123-126.
5. Gustafson G, Koch G (1974) Age estimation up to 16 years of age based on dental development. Odontol Revy 25:297-306.
6. Mostad P, Tamsen F (2019) Error rates for unvalidated medical age assessment procedures. Int J Legal Med 133:613-623.
7. The Cochrane Public Health Group (2011) Guide for developing a Cochrane protocol.
8. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj 339:b2700.
9. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Bmj 339:b2535.
10. Cunha E, Baccino E, Martrille L, Ramsthaler F, Prieto J, Schuliar Y, Lynnerup N, Cattaneo C (2009) The problem of aging human remains and living individuals: a review. Forensic Sci Int 193:1-13.
11. Solheim T (1993) A new method for dental age estimation in adults. Forensic Sci Int 59:137-147.
12. Ruder TD, Hatch GM, Siegenthaler L, Ampanozi G, Mathier S, Thali MJ, Weber OM (2012) The influence of body temperature on image contrast in post mortem MRI. Eur J Radiol 81:1366-1370.
13. De Tobel J, van Wijk M, Alberink I, Hillewig E, Phlypo I, van Rijn RR, Thevissen P, Verstraete K, de Haas MB (2019) The influence of motion artifacts on magnetic resonance imaging of the clavicles for age estimation. Int J Legal Med. Accepted on November 17, 2019. doi: 10.1007/s00414-019-02230-x.
14. Vieth V, Kellinghaus M, Schulz R, Pfeiffer H, Schmeling A (2010) Ossification stage of the medial clavicular epiphysis: Comparison of projectional radiography, computed tomography and magnetic resonance imaging. Rechtsmedizin 20:483-488.
15. De Tobel J, Parmentier GIL, Phlypo I, Descamps B, Neyt S, Van De Velde WL, Verstraete KL, Thevissen PW (2019) Magnetic resonance imaging of third molars in forensic age estimation: comparison of the Ghent and Graz protocols focusing on apical closure. Int J Legal Med 133:583-592.
16. Fan F, Zhang K, Peng Z, Cui JH, Hu N, Deng ZH (2016) Forensic age estimation of living persons from the knee: Comparison of MRI with radiographs. Forensic Sci Int 268:145-150.
17. 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.
18. Tangmose S, Jensen KE, Lynnerup N (2013) Comparative study on developmental stages of the clavicle by postmortem MRI and CT imaging. J Forensic Radiol Imaging 1:102-106.
19. Urschler M, Krauskopf A, Widek T, Sorantin E, Ehammer T, Borkenstein M, Yen K, Scheurer E (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.
20. Cochrane Effective Practice and Organisation of Care (EPOC) (2017) EPOC Resources for review authors.
21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529-536.
22. Harcke HT, Synder M, Caro PA, Bowen JR (1992) Growth plate of the normal knee: evaluation with MR imaging. Radiology 183:119-123.
23. Laor T, Chun GF, Dardzinski BJ, Bean JA, Witte DP (2002) Posterior distal femoral and proximal tibial metaphyseal stripes at MR imaging in children and young adults. Radiology 224:669-674.
24. Bollow M, Braun J, Kannenberg J, Biedermann T, Schauer-Petrowskaja C, Paris S, Mutze S, Hamm B (1997) Normal morphology of sacroiliac joints in children: magnetic resonance studies related to age and sex. Skeletal Radiol 26:697-704.

25. Craig JG, Cody DD, Van Holsbeeck M (2004) The distal femoral and proximal tibial growth plates: MR imaging, three-dimensional modeling and estimation of area and volume. Skeletal Radiol 33:337-344.
26. Bray TJ, Vendhan K, Roberts J, Atkinson D, Punwani S, Sen D, Ioannou Y, Hall-Craggs MA (2016) Association of the apparent diffusion coefficient with maturity in adolescent sacroiliac joints. J Magn Reson Imaging 44:556-564.
27. 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.
28. Kercher J, Xerogeanes J, Tannenbaum A, Al-Hakim R, Black JC, Zhao J (2009) Anterior cruciate ligament reconstruction in the skeletally immature: an anatomical study utilizing 3-dimensional magnetic resonance imaging reconstructions. J Pediatr Orthop 29:124-129.
29. Kim HK, Shiraj S, Anton C, Horn PS (2014) The patellofemoral joint: do age and gender affect skeletal maturation of the osseous morphology in children? Pediatr Radiol 44:141-148.
30. 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 277:21-29.
31. Pennock AT, Bomar JD, Manning JD (2018) The Creation and Validation of a Knee Bone Age Atlas Utilizing MRI. J Bone Joint Surg Am 100:e20.
32. Saint-Martin P, Rerolle C, Pucheux J, Dedouit F, Telmon N (2015) Contribution of distal femur MRI to the determination of the 18-year limit in forensic age estimation. Int J Legal Med 129:619-620.
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. 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, Canada, pp 61-69.
35. Tangmose S, Jensen KE, Villa C, Lynnerup N (2014) Forensic age estimation from the clavicle using 1.0T MRI-Preliminary results. Forensic Sci Int 234:7-12.
36. Terada Y, Kono S, Tamada D, Uchiumi T, Kose K, Miyagi R, Yamabe E, Yoshioka H (2013) Skeletal age assessment in children using an open compact MRI system. Magn Reson Med 69:1697-1702.
37. Terada Y, Kono S, Uchiumi T, Kose K, Miyagi R, Yamabe E, Fujinaga Y, Yoshioka H (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.
38. Terada Y, Tamada D, Kose K, Nozaki T, Kaneko Y, Miyagi R, Yoshioka H (2016) Acceleration of skeletal age MR examination using compressed sensing. J Magn Reson Imaging 44:204-211.
39. Tomei E, Sartori A, Nissman D, Al Ansari N, Battisti S, Rubini A, Stagnitti A, Martino M, Marini M, Barbato E, Semelka RC (2014) Value of MRI of the hand and the wrist in evaluation of bone age: Preliminary results. J Magn Reson Imaging 39:1198-1205.
40. Vo A, Beaule PE, Sampaio ML, Rotaru C, Rakhra KS (2015) The femoral head-neck contour varies as a function of physeal development. Bone Joint Res 4:17-22.
41. Baumann P, Widek T, Merkens H, Boldt J, Petrovic A, Urschler M, Kirnbauer B, Jakse N, Scheurer E (2015) Dental age estimation of living persons: Comparison of MRI with OPG. Forensic Sci Int 253:76-80.
42. 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.
43. Jopp E, Schröder I, Maas R, Adam G, Püschel K (2010) Proximal tibial epiphysis in magnetic resonance imaging. Rechtsmedizin 20:464-468.
44. 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.
45. 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.
46. Auf der Mauer M, Saring D, Stanczus B, Herrmann J, Groth M, Jopp-van Well E (2018) A 2-year follow-up MRI study for the evaluation of an age estimation method based on knee bone development. Int J Legal Med 133:205-215.
47. De Tobel J, Hillewig E, de Haas MB, Van Eeckhout B, Fieuws S, Thevissen P, Verstraete K (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 26:2924-2935.

48. De Tobel J, Hillewig E, van Wijk M, Fieuws S, de Haas MB, van Rijn RR, Thevissen PW, Verstraete KL (2019) Staging clavicular development on magnetic resonance imaging: pitfalls and recommendations for age estimation. J Magn Reson Imaging: 2019 Aug 1. doi: 10.1002/jmri.26889. [Epub ahead of print].
49. 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.
50. Guo Y, Olze A, Ottow C, Schmidt S, Schulz R, Heindel W, Pfeiffer H, Vieth V, Schmeling A (2015) Dental age estimation in living individuals using 3.0 T MRI of lower third molars. Int J Legal Med 129:1265-1270.
51. Vieth V, Schulz R, Brinkmeier P, Dvorak J, Schmeling A (2014) Age estimation in U-20 football players using 3.0 tesla MRI of the clavicle. Forensic Sci Int 241c:118-122.
52. Schmidt S, Ottow C, Pfeiffer H, Heindel W, Vieth V, Schmeling A, Schulz R (2017) Magnetic resonance imaging-based evaluation of ossification of the medial clavicular epiphysis in forensic age assessment. Int J Legal Med 131:1665-1673.
53. Hojreh A, Gamper J, Schmook MT, Weber M, Prayer D, Herold CJ, Noebauer-Huhmann IM (2018) Hand MRI and the Greulich-Pyle atlas in skeletal age estimation in adolescents. Skeletal Radiol 47:963-971.
54. 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.
55. Nasel C, Gahleitner A, Breitenseher M, Czerny C, Solar P, Imhof H (1998) Dental MR tomography of the mandible. J Comput Assist Tomogr 22:498-502.
56. Demirjian A, Goldstein H, Tanner JM (1973) A new system of dental age assessment. Hum Biol 45:211-227.
57. Köhler S, Schmelzle R, Loitz C, Puschel K (1994) [Development of wisdom teeth as a criterion of age determination]. Ann Anat 176:339-345.
58. 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.
59. 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.
60. Wittschieber D, Schmidt S, Vieth V, Schulz R, Püschel K, Pfeiffer H, Schmeling A (2014) Subclassification of clavicular substage 3a is useful for diagnosing the age of 17 years. Rechtsmedizin 24:485-488.
61. Dedouit F, Auriol J, Rousseau H, Rouge D, Crubezy E, Telmon N (2012) Age assessment by magnetic resonance imaging of the knee: a preliminary study. Forensic Sci Int 217:232 e231-237.
62. Ekizoglu O, Hocaoglu E, Inci E, Can IO, Aksoy S, Kazimoglu C (2016) Forensic age estimation via 3-T magnetic resonance imaging of ossification of the proximal tibial and distal femoral epiphyses: Use of a T2-weighted fast spin-echo technique. Forensic Sci Int 260:102.e101-107.
63. Vieth V, Schulz R, Heindel W, Pfeiffer H, Buerke B, Schmeling A, Ottow C (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.
64. 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.
65. Timme M, Ottow C, Schulz R, Pfeiffer H, Heindel W, Vieth V, Schmeling A, Schmidt S (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.
66. Tomei E, Semelka RC, Nissman D (2013) Text-atlas of skeletal age determination: MRI of the hand and wrist in children. John Wiley & Sons.
67. Tanner JM, Whitehouse RH, Cameron N, Marshall WA, Healy MJR, Goldstein H (1983) Assessment of skeletal maturity and prediction of adult height (TW2 method). Academic Press Limited, London.
68. Greulich W, Pyle SI (1959) Radiographic atlas of skeletal development of the hand and wrist. Stanford University Press, Stanford, CA.
69. Saint-Martin P, Rerolle C, Dedouit F, Bouilleau L, Rousseau H, Rouge D, Telmon N (2013) Age estimation by magnetic resonance imaging of the distal tibial epiphysis and the calcaneum. Int J Legal Med 127:1023-1030.
70. Serin J, Rerolle C, Pucheux J, Dedouit F, Telmon N, Savall F, Saint-Martin P (2016) Contribution of magnetic resonance imaging of the wrist and hand to forensic age assessment. Int J Legal Med 130:1121-1128.

71. Ekizoglu O, Inci E, Ors S, Kacmaz IE, Basa CD, Can IO, Kranioti EF (2018) Applicability of T1-weighted MRI in the assessment of forensic age based on the epiphyseal closure of the humeral head. Int J Legal Med 133:241-248.
72. 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? Egyptian Journal of Radiology and Nuclear Medicine 49:146-151.
73. Kramer JA, Schmidt S, Jurgens KU, Lentschig M, Schmeling A, Vieth V (2014) The use of magnetic resonance imaging to examine ossification of the proximal tibial epiphysis for forensic age estimation in living individuals. Forensic Sci Med Pathol 10:306-313.
74. Kramer JA, Schmidt S, Jurgens KU, Lentschig M, Schmeling A, Vieth V (2014) Forensic age estimation in living individuals using 3.0 T MRI of the distal femur. Int J Legal Med 128:509-514.
75. Saint-Martin P, Rerolle C, Dedouit F, Rousseau H, Rouge D, Telmon N (2014) Evaluation of an automatic method for forensic age estimation by magnetic resonance imaging of the distal tibial epiphysis--a preliminary study focusing on the 18-year threshold. Int J Legal Med 128:675-683.
76. Ekizoglu O, Hocaoglu E, Can IO, Inci E, Aksoy S, Bilgili MG (2015) Magnetic resonance imaging of distal tibia and calcaneus for forensic age estimation in living individuals. Int J Legal Med 129:825-831.
77. 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.
78. Wittschieber D, Vieth V, Timme M, Dvorak J, Schmeling A (2014) Magnetic resonance imaging of the iliac crest: age estimation in under-20 soccer players. Forensic Sci Med Pathol 10:198-202.
79. 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.
80. Rashid NR, Aliasghar A, Shaker QM (2015) Magnetic resonance imaging of the left wrist: assessment of the bone age in a sample of healthy Iraqi adolescent males. J Fac Med Baghdad 57:22-26.
81. Serinelli S, Panebianco V, Martino M, Battisti S, Rodacki K, Marinelli E, Zaccagna F, Semelka RC, Tomei E (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.
82. Ekizoglu O, Hocaoglu E, Can IO, Inci E, Aksoy S, Sayin I (2016) Spheno-occipital synchondrosis fusion degree as a method to estimate age: A preliminary, magnetic resonance imaging study. Aust J Forensic Sci 48:159-170.
83. Hillewig E, Degroote J, Van der Paelt T, Visscher A, Vandemaele P, Lutin B, D'Hooghe L, Vandriessche V, Piette M, Verstraete K (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.
84. 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 University Press, Cambridge, pp 73-106.
85. Liversidge HM (2008) Timing of human mandibular third molar formation. Ann Hum Biol 35:294-321.
86. Thevissen PW, Alqerban A, Asaumi J, Kahveci F, Kaur J, Kim YK, Pittayapat P, Van VM, Zhang Y, Fieuws S, Willems G (2010) Human dental age estimation using third molar developmental stages: Accuracy of age predictions not using country specific information. Forensic Sci Int 201:106-111.
87. Thevissen PW, Fieuws S, Willems G (2010) Human third molars development: Comparison of 9 country specific populations. Forensic Sci Int 201:102-105.
88. Willems G, Lee SS, Uys A, Bernitz H, Cadenas de Llano-Perula M, Fieuws S, Thevissen P (2017) Age estimation based on Willems method versus new country-specific method in South African black children. Int J Legal Med.
89. Haglund M, Mornstad H (2018) A systematic review and meta-analysis of the fully formed wisdom tooth as a radiological marker of adulthood. Int J Legal Med 133:231-239.
90. Schmeling A, Reisinger W, Loreck D, Vendura K, Markus W, Geserick G (2000) Effects of ethnicity on skeletal maturation: consequences for forensic age estimations. Int J Legal Med 113:253-258.
91. Olze A, van NP, Schmidt S, Wernecke KD, Rosing FW, Geserick G, Schmeling A (2006) Studies on the progress of third-molar mineralisation in a Black African population. Homo 57:209-217.
92. Zhang A, Sayre JW, Vachon L, Liu BJ, Huang HK (2009) Racial differences in growth patterns of children assessed on the basis of bone age. Radiology 250:228-235.
93. 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.

94. Sarkodie B, Ofori E, Pambo P (2013) MRI to determine the chronological age of Ghanaian footballers. S Afr J Sports Med 25:3.
95. Malina RM (2011) Skeletal age and age verification in youth sport. Sports medicine (Auckland, NZ) 41:925-947.
96. Timme M, Steinacker JM, Schmeling A (2017) Age estimation in competitive sports. Int J Legal Med 131:225-233.
97. 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.
98. 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.
99. Konigsberg LW (2015) Multivariate cumulative probit for age estimation using ordinal categorical data. Ann Hum Biol 42:368-378.
100. AlQahtani SJ, Hector MP, Liversidge HM (2010) Brief communication: The London atlas of human tooth development and eruption. Am J Phys Anthropol 142:481-490.
101. Liversidge HM, Smith BH, Maber M (2010) Bias and accuracy of age estimation using developing teeth in 946 children. Am J Phys Anthropol 143:545-554.
102. Bassed RB, Briggs C, Drummer OH (2011) Age estimation using CT imaging of the third molar tooth, the medial clavicular epiphysis, and the spheno-occipital synchondrosis: a multifactorial approach. Forensic Sci Int 212:273.e271-275.
103. Cameriere R, Ferrante L (2008) Age estimation in children by measurement of carpals and epiphyses of radius and ulna and open apices in teeth: A pilot study. Forensic Sci Int 174:60-63.
104. Demirturk Kocasarac H, Sinanoglu A, Noujeim M, Helvacioglu Yigit D, Baydemir C (2016) Radiologic assessment of third molar tooth and spheno-occipital synchondrosis for age estimation: a multiple regression analysis study. Int J Legal Med 130:799-808.
105. Thevissen PW, Kaur J, Willems G (2012) Human age estimation combining third molar and skeletal development. Int J Legal Med 126:285-292.
106. Schmidt S, Schramm D, Ribbecke S, Schulz R, Wittschieber D, Olze A, Vieth V, Ramsthaler HF, Pfischel K, Pfeiffer H, Geserick G, Schmeling A (2016) [Forensic age estimation in juveniles and young adults: Reducing the range of scatter in age diagnosis by combining different methods]. Arch Kriminol 237:25-37.
107. Shi L, Jiang F, Ouyang F, Zhang J, Wang Z, Shen X (2017) DNA methylation markers in combination with skeletal and dental ages to improve age estimation in children. Forensic Sci Int Genet 33:1-9.
108. Schmeling A, Dettmeyer R, Rudolf E, Vieth V, Geserick G (2016) Forensic Age Estimation. Dtsch Arztebl Int 113:44-50.
109. Fournier K (2017) [Age estimation of unaccompanied minors questioned: defining the issue, analysis and recommendations]. Platform for refugee children.
110. Thodberg HH, Kreiborg S, Juul A, Pedersen KD (2009) The BoneXpert method for automated determination of skeletal maturity. IEEE Trans Med Imaging 28:52-66.
111. Thodberg HH, van Rijn RR, Jenni OG, Martin DD (2017) Automated determination of bone age from hand X-rays at the end of puberty and its applicability for age estimation. Int J Legal Med 131.
112. Thodberg HH, Savendahl L (2010) Validation and reference values of automated bone age determination for four ethnicities. Acad Radiol 17:1425-1432.
113. 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 17:421-428.
114. 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 17:220-227.
115. Unterpirker W, Ebner T, Štern D, Urschler M (2015) Automatic third molar localization from 3D MRI using random regression forests. . In: Proceedings of the 19th Conference on Medical Image Understanding and Analysis (MIUA), Lincoln, United Kingdom, pp 195-200.
116. Štern D, Payer C, Lepetit V, Urschler M (2016) Automated age estimation from hand MRI volumes using deep learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp 194-202.
117. European Asylum Support Office (EASO) (2018) Practical Guide on Age Estimation, Second edition. EASO Practical Guides Series, Malta.

118. International Organization for Forensic Odonto-Stomatology (IOFOS) (2018) Recommendations for Quality Assurance: Dental Age Estimation. Leuven
119. Garamendi PM, Landa MI, Ballesteros J, Solano MA (2005) Reliability of the methods applied to assess age minority in living subjects around 18 years old. A survey on a Moroccan origin population. Forensic Sci Int 154:3-12.
120. 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.
121. Das SK, Wang JL, Bing L, Bhetuwal A, Yang HF (2017) Regional Values of Diffusional Kurtosis Estimates in the Healthy Brain during Normal Aging. Clin Neuroradiol 27:283-298.
122. Helpern JA, Adisetiyo V, Falangola MF, Hu C, Di Martino A, Williams K, Castellanos FX, Jensen JH (2011) Preliminary evidence of altered gray and white matter microstructural development in the frontal lobe of adolescents with attention-deficit hyperactivity disorder: A diffusional kurtosis imaging study. J Magn Reson Imaging 33:17-23.
123. Hsu JL, Van Hecke W, Bai CH, Lee CH, Tsai YF, Chiu HC, Jaw FS, Hsu CY, Leu JG, Chen WH, Leemans A (2010) Microstructural white matter changes in normal aging: A diffusion tensor imaging study with higher-order polynomial regression models. NeuroImage 49:32-43.
124. Grady CL, Garrett DD (2014) Understanding variability in the BOLD signal and why it matters for aging. Brain Imaging Behav 8:274-283.
125. Paydar A, Fieremans E, Nwankwo JI, Lazar M, Sheth HD, Adisetiyo V, Helpern JA, Jensen JH, Milla SS (2014) Diffusional kurtosis imaging of the developing brain. AJNR Am J Neuroradiol 35:808-814.
126. Pfefferbaum A, Sullivan EV, Hedehus M, Lim KO, Adalsteinsson E, Moseley M (2000) Age-related decline in brain white matter anisotropy measured with spatially corrected echo-planar diffusion tensor imaging. Magn Reson Med 44:259-268.

Tables

Table 1 Population characteristics of eligible studies. Studies are grouped per anatomical site and ordered per staging technique (see Table 6 of the Supplementary Material).

Anatomical structure	Reference	Year	Study design	Excluded reference because of correspondence	Geographical population	Number of females	Age range (years)	Number of males	Age range (years)
Spheno-occipital synchondrosis	Ekizoglu	2016a	RCS	NA	Turkish	623	7 - 21	455	7 - 21
Molars	Baumann	2015	PCS	NA	Living in Austria	18	13.5 - 23.1	8	13.7 - 21.3
Lower left third molar	Guo	2015	PCS	NA	German	248	12.3 - 25	269	12.1 - 25
Third molars	De Tobel	2017b	PCS	NA	Belgian	26	14.5 - 26.8	26	14.3 - 26.80
Third molars	De Tobel	2017c	PCS	NA	Belgian and Dutch	146	14.10 - 26.80	163	14.10 - 27.00
Clavicle	Hillewig	2013	PCS	Hillewig 2011	Belgian	110	16.0 - 26.9	110	16.1 - 26.9
Clavicle	Tangmose	2014	PCS	NA	Mainly European Caucasian, four from Middle East, Asia or Africa	16	NA	39	NA
Clavicle	Vieth	2014	PCS	NA	German	0	NA	152	18.1 - 23.0
Clavicle	Schmidt	2017	PCS	NA	German	310	12.1 - 25.0	260	12.1 - 24.9
Clavicle	De Tobel	2019c	PCS	Hillewig 2011, 2013	Belgian and Dutch	264	14.1 - 30.5	199	14.1 - 30.1
Manubrium	Martínez Vera	2017	RCS	NA	Austrian	0	NA	130	13 25
Proximal humerus	Ekizoglu	2018	RCS	NA	Turkish hospital	188	12.2 - 30.7	240	12.1 - 30.6
Left distal radius	Dvorak	2007	PCS	NA	Swiss, Malay, Algerian and Argentinan	0	NA	496	14 - 19
Left distal radius	George	2012	PCS	NA	Malaysian Malay	0	NA	150	15 - 19.0
Left distal radius	Bolivar	2015	PCS	NA	Colombian	0	NA	60	12 - 18
Left distal radius	Rashid	2015	PCS	NA	Iraqi	0	NA	179	13 - 18
Left distal radius	Tscholl	2016	PCS	NA	African (Tanzania), Asian (Malaysia), European (Germany), Latin American (Brazil)	487; T 140; M 129; G 117; B 101	13.3 - 19.3	0	NA
Left distal radius	Abdelbary	2018	PCS	NA	Egyptian	0	NA	61	13 - 18
Left distal radius	Sarkodie	2018	PCS	NA	Ghanaian	0	NA	286	13 - 16
Left distal radius	Schmidt	2015	PCS	NA	German	0	NA	152	18.1 - 22.9
Left hand/wrist	Serin	2016	RCS	NA	French hospital	156	9 - 25	107	9 - 25.0
Left distal radius	Timme	2017	PCS	NA	NA	333	12.1 - 24.9	335	12.1 - 24.9
Left wrist	De Tobel	2019b	PCS	NA	Belgian and Dutch	185	14.10 - 26.90	178	14.10 - 27.00
Left hand/wrist	Tomei	2014	PCS	NA	Italian	78	11 - 17	101	11 - 17
Left hand/wrist	Serinelli	2015	PCS	NA	Italian	74	12.00 - 18.8	77	12 - 19.1
Left hand/wrist	Terada	2013	PCS	NA	Japanese	43	4.1 - 16.4	50	4.1 - 16.4
Left hand/wrist	Terada	2014	PCS	NA	Japanese	23	3.4 - 15.7	65	3.4 - 15.7
Left hand/wrist	Terada	2016	PCS	NA	Japanese	24	4.4 - 15.3	35	4.4 - 15.3
Left hand/wrist	Urschler	2016	PCS	NA	Austrian	4	7.57 - 14.1	14	7.92 - 16.8
Left hand/wrist	Hojreh	2018	PCS	Hojreh 2017	European; Iranian, Argentinian, Malian, Philippine excluded for current results	29	12 - 19.8	17	12.8 - 18.5
Left hand/wrist	Urschler	2015	PCS	Stern 2014	Austrian	0	NA	102	13 - 20
Iliac crest	Wittschieber	2014	PCS	NA	German	0	NA	152	18.0 - 22.9
Proximal femur	Vo	2015	PCS	NA	NA	17	8 - 16	26	10 - 18
Sacrum	Bollow	1997	PCS	NA	German hospital	43	8 - 17	71	8 - 17
Sacrum	Bray	2016	RCS	NA	British hospital	36	10.2 - 18.9	19	10.2 - 18.9
Patellofemoral joint	Kim	2014	RCS	NA	NA	51	5 - 22	46	5 - 22
Distal femur	Saint-Martin	2015	RCS	NA	French hospital	0	NA	214	14 - 20
Knee	Dedouit	2012	RCS	NA	French hospital	152	10.1 - 30.9	138	10.3 - 30.3
Knee	Ekizoglu	2016b	RCS	NA	Turkish hospital	198	10 - 30	305	10 - 30
Knee	Harcke	1992	PCS	NA	NA	27	0 - 20	33	0 - 20
Knee	Laor	2002	RCS	NA	American hospital	100	0 - 40	97	0 - 40
Proximal tibia	Jopp	2010	PCS	NA	German	0	NA	41	15.7 - 19.8
Distal femur	Krämer	2014a	RCS	NA	German hospital	124	10.1 - 30.8	166	10.1 - 30.8
Proximal tibia	Krämer	2014b	RCS	NA	German hospital	124	10.1 - 30.8	166	10.1 - 30.8
Knee	Fan	2016	RCS	NA	West China Han	139	11.00 - 29.5	183	11.00 - 29.9
Knee	Ottow	2017	PCS	NA	German	333	12.1 - 25.00	325	12.1 - 25
Knee	Auf der Mauer	2018	PCH	NA	German	0	NA	36	15.3 - 20.7
Knee	Vieth	2018	PCS	NA	German	350	12.1 - 25	344	12.1 - 25
Knee	Pennock	2018	RCS	NA	American hospital	421	2 - 19	438	2 - 19
Knee	Craig	2004	RCS	NA	American hospital	5	3.8 - 15.6	9	3.8 - 15.6
Knee	Kercher	2009	RCS	NA	NA	21	10 - 15	10	10 - 15
Ankle	Saint-Martin	2013	RCS	NA	French hospital	100	8 - 25	80	8 - 25
Distal tibia	Saint-Martin	2014	RCS	NA	French hospital	60	8 - 25	60	8 - 25
Ankle	Ekizoglu	2015	RCS	NA	Turkish hospital	70	8 - 25	97	8 - 25
MFA	Stern	2017	PCS	NA	Austrian	0	NA	103	13 - 24.9

MFA = multi-factorial age estimation; NA = not applicable or not reported; PCH = prospective cohort; PCS = prospective cross-sectional; RCS = retrospective cross-sectional

Table 2 Descriptive criteria for developmental stages of long bones on magnetic resonance imaging.

Main stage	Sub-stage	Advanced substage	
1			Ossification center is invisible (= not yet ossified).
2			Ossification center is visible (= ossified), nonunion 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 over one third until 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 center 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 over one third until 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.
5			Complete union of the epiphysis and metaphysis. Physeal scar is indiscernible.

Table 3 Absolute statements regarding the age threshold of 18 years.

	Minor		**Adult**	
	Anatomical structure	**Stage**	**Anatomical structure**	**Stage**
Females	Spheno-occipital synchondrosis	Bassed stage 1	Lower left third molar	Demirjian stage H
	Lower left third molar *	up to De Tobel stage 2	Lower left third molar *	from De Tobel stage 7 on
	Proximal humerus	up to Kellinghaus stage 3a	Left hand/wrist SE *	Tomei atlas skeletal age 18
	Left hand/wrist SE *	Tomei atlas up to skeletal age 17	Left distal radius SE *	Schmeling stage 5
	Left hand/wrist VIBE	Greulich-Pyle atlas up to skeletal age 16	Distal femur *	Dedouit stage 5
	Left distal radius SE	Dvorak stage 1	Distal femur *	Vieth stage 6
	Left distal radius SE *	up to Kellinghaus stage 3a		
	Left distal radius VIBE	up to Kellinghaus stage 3b		
	Distal femur	up to Kellinghaus stage 2c		
	Distal femur *	up to Dedouit stage 2		
	Distal femur *	up to Vieth stage 2		
	Proximal tibia	up to Kellinghaus stage 2c		
	Proximal tibia	up to Dedouit stage 2		
	Proximal tibia	up to Vieth stage 4		
	Proximal fibula	up to Kellinghaus stage 3c		
	Distal tibia	up to Schmeling stage 2		
	Calcaneum	up to Schmeling stage 3		
Males	Spheno-occipital synchondrosis	Bassed stage 1	Lower left third molar *	from De Tobel stage 7 on
	Lower left third molar	up to Demirjian stage D	Proximal humerus *	Schmeling stage 4
	Lower left third molar *	up to De Tobel stage 2	Left hand/wrist VIBE	Urschler automated skeletal age 19
	Proximal humerus *	up to Kellinghaus stage 3a	Left distal radius SE	Schmeling stage 5
	Left hand/wrist SE	Tomei atlas up to skeletal age 17	Distal femur *	Dedouit stage 5
	Left hand/wrist VIBE	Greulich-Pyle atlas up to skeletal age 17	Distal femur *	Vieth stage 6
	Left hand/wrist VIBE	Urschler up to automated skeletal age 15	Proximal tibia *	Dedouit stage 5
	Left distal radius	Dvorak stage 1	Proximal tibia *	Vieth stage 6
	Distal femur *	up to Dedouit stage 2		
	Distal femur *	Vieth stage 1		
	Proximal tibia *	Dedouit stage 1		
	Proximal tibia *	up to Vieth stage 3		
	Proximal fibula	up to Schmeling stage 2		
	Knee	SKJ up to 5		
	Distal tibia	up to Schmeling stage 2		
	Calcaneum	up to Schmeling stage 2		

SE = T1 spin echo sequence; SKJ = cumulative score of the knee joint; VIBE = T1 gradient echo volumetric interpolated breath-hold examination.

* Anatomical structure and staging technique which allow absolute statements about minority as well as adulthood.

Table 4 Ability to discern minors from adults.

Regarding predictive probabilities, stages or combinations of stages are displayed between brackets and only stages at the end of development were included. Regarding third molars, stages apply to FDI (World Dental Federation) teeth 18, 28, 38, and 48, respectively. For instance, "(6666)" means that all third molars were in stage 6. Regarding clavicles, stages apply to the left and right clavicle, respectively. For instance, "(3,4)" means that the left clavicle was in stage 3, while the right one was in stage 4.

Regarding diagnostic indices, sex-specific results were not reported in all studies. Instead, some studies reported non-sex-specific results, which were displayed in the center of the column.

Anatomical structure	Reference	Year	Predictive probabilities P(Age < 18 years)		Sensitivity		Specificity	
			Females	Males	Females	Males	Females	Males
Third molars	De Tobel	2017	(6666) 0.0491; (7777) 0.0044; (8888) 0.0011	(6666) 0.1117; (7777) 0.0074; (8888) 0.0024	82.6	91.0	65.8	87.2
Clavicles	Hillewig	2013	(3,3) 0.258; (3,4) 0.067; (4,3) 0.070; (4,4) 0.008	(3,3) 0.159; (3,4) 0.026; (4,3) 0.029; (4,4) 0.002	NA		NA	
Clavicles	De Tobel	2019c	(3b, 3c) 0.0059; (3c, 3b) 0.0198; (3c, 3c) 0.0023	(3b, 3c) 0.0053; (3c, 3b) 0.0182; (3c, 3c) 0.0019	86.1		69.4	
Manubrium	Martínez Vera		NA	NA	91.1		82.4	
Left distal radius	Serin	2016	NA	NA	100.0	92.5	89.9	92.5
Left wrist SE	De Tobel	2019b	(4/5) 0.0547	(4/5) 0.0171	88.5		92.8	
Left wrist VIBE	De Tobel	2019b	(4) 0.2570; (5) 0.0840	(4) 0.0547; (5) 0.0248	90.9		87.4	
Distal tibia	Saint-Martin	2013	(4) 0.328	(4) 0.026	NA		NA	
Calcaneum	Saint-Martin	2013	(4) 0.353	(4) 0.064	NA		NA	
Distal tibia and calcaneum	Saint-Martin	2013	NA	NA	97.7	91.7	78.6	90.6
Distal tibia	Saint-Martin	2014	NA	NA	94.3	97.4	71.2	65.5
MFA	Stern	2017	NA	NA	93.2		88.6	

MFA = multi-factorial age estimation; NA = not applicable or not reported.

Figure captions

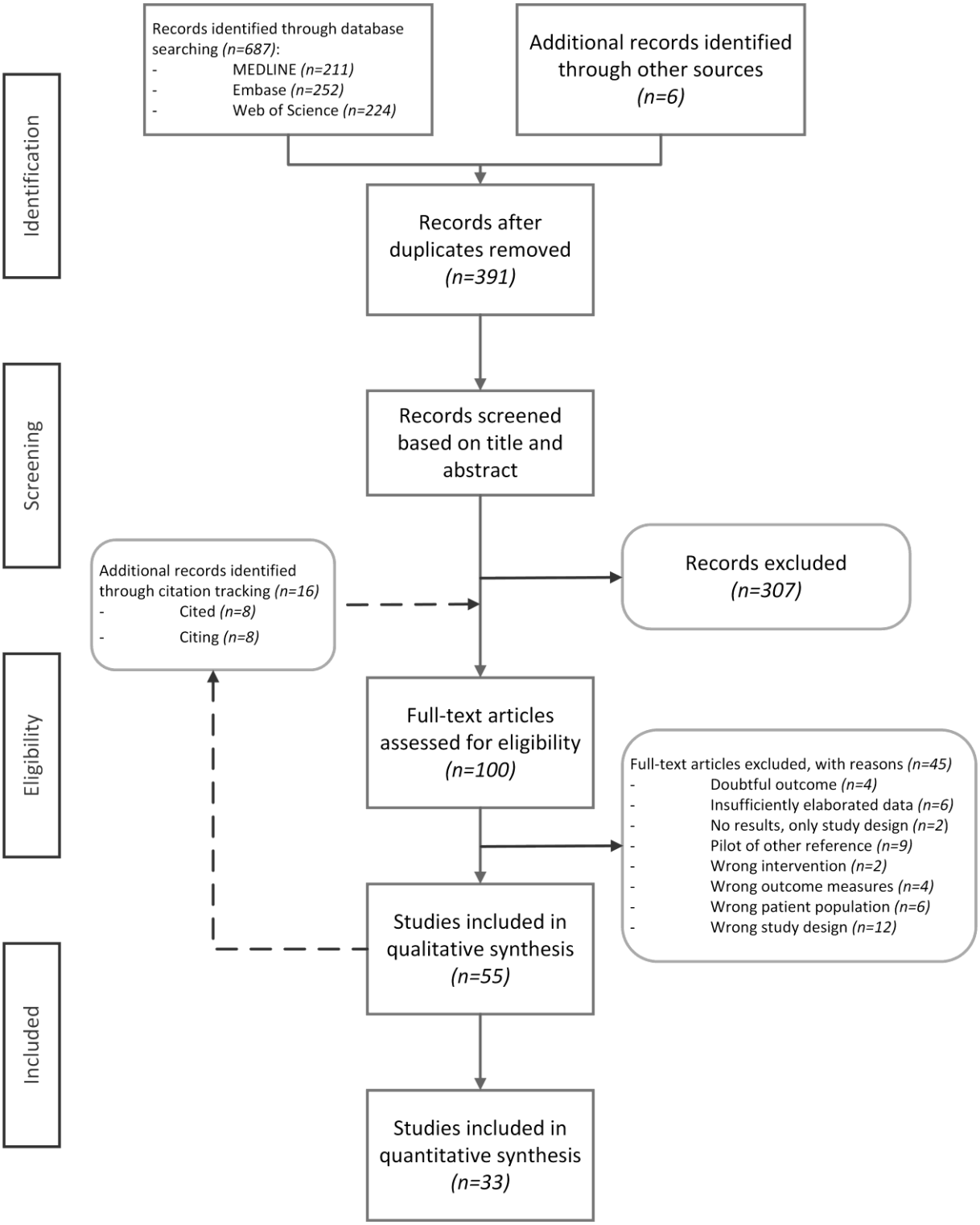
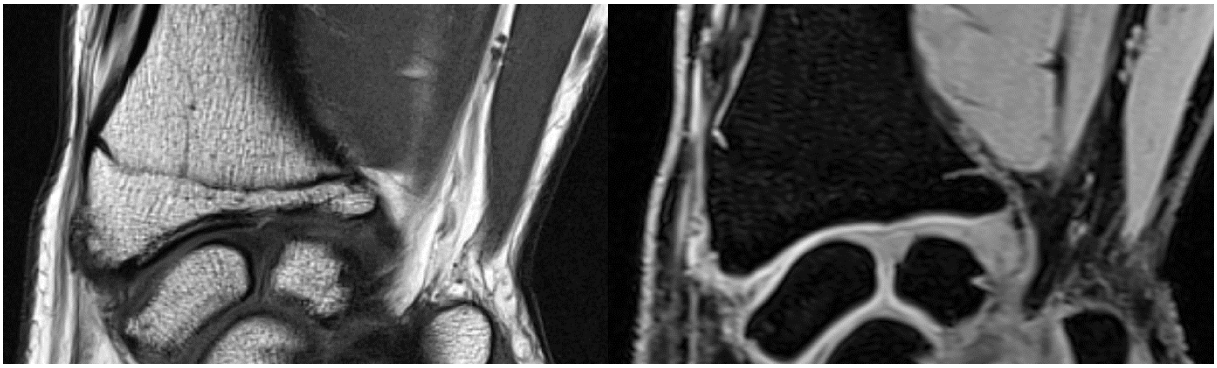


Fig. 1 Flowchart showing the process of literature search and study selection



a b

Fig. 2 Wrist magnetic resonance imaging in a 17.85 year old male. **a** T1 spin echo sequence shows partial bridging of the physeal plate. Stage 3b was allocated. The chemical shift artifact causes a widened appearance of the remaining physeal plate. **b** T1 gradient echo volumetric interpolated breath-hold examination sequence shows more advanced bridging of the physeal plate. Stage 3c was allocated. Fat suppression avoids the chemical shift artifact, causing a more tight delineation of the physeal plate

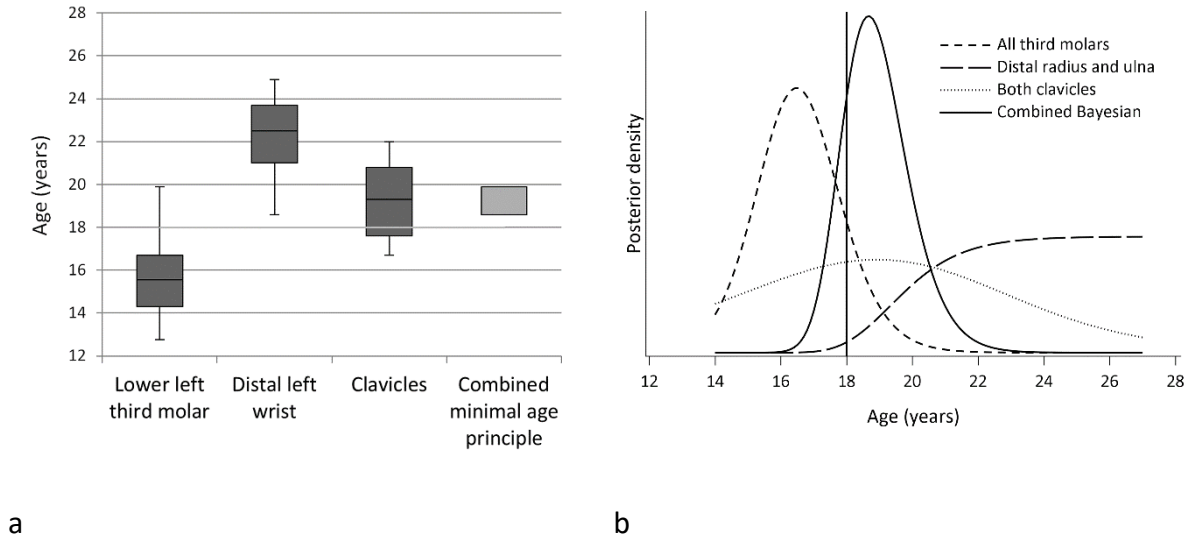


Fig. 3 Male case example of two methods for multi-factorial forensic age estimation in practice. **a** Minimum age principle. Three anatomical sites were assessed. For the third molars and the wrist, only one anatomical structure was considered. For clavicles, both clavicles were assessed and in case of different stages between left and right, the most advanced clavicle was selected. The boxplots show the age distribution for the allocated stage per anatomical site, based on a reference study. The whiskers show the minimum and maximum ages, the box the first and third quartiles, and the central line displays the median. The combined age estimation is an interval: (1) the highest minimum age is retained, since for that anatomical site, no individuals younger than that age have been reported; and (2) the lowest maximum age is retained, since for that anatomical site, no individuals older than that age have been reported. In this male example, the interval was [18.60;19.88]. **b** Continuation ratio model with Bayesian correction for violation of the conditional independence assumption. Three anatomical sites were assessed. For third molars, all four third molars were taken into account. For the wrist, the distal radius and ulna were taken into account. For the clavicles, both of them were taken into account. Thus, the curves per anatomical site already combine the information of the different anatomical structures per site. The curves show the posterior densities of age for the allocated stages to all anatomical structures per anatomical site, based on a reference study. The combined age estimation is defined by the combined curve, providing the following statistics: point prediction, 95% prediction interval, and the probability to be an adult. In this male example these statistics were 19.03 years old, [16.57;22.00], and 0.709, respectively