

This is the accepted manuscript of the following publication:

De Tobel J, van Wijk M, Alberink I, Hillewig E, Phlypo I, van Rijn RR, Thevissen PW, Verstraete KL, de Haas MB. The influence of motion artefacts on magnetic resonance imaging of the clavicles for age estimation. *Int J Legal Med*. Accepted on November 17, 2019.

FULL TITLE PAGE

The influence of motion artefacts on magnetic resonance imaging of the clavicles for age estimation

Author order	First name	Middle initials	Last name	Academic degrees ORCID	Affiliations
1	Jannick	/	De Tobel*	- Doctor of Philosophy in Health Sciences (PhD) - Medical doctor (MD) - Doctor of Dental Science (DDS) - Master of Forensic Odontology (MSc) 0000-0002-8846-7339	1,2,3,4
2	Mayonne	/	van Wijk*	- Master of Science (MSc) in Forensic Science - Bachelor of Science (BSc) in Biomedical Science 0000-0002-5285-8899	5
3	Ivo		Alberink	- Master of Science (MSc) - Doctor of Philosophy in Mathematical Statistics (PhD)	5
4	Elke	/	Hillewig	- Master of Science (MSc) in Criminology - Bachelor in Medical Imaging 0000-0003-0584-4897	1
5	Inès	/	Phlypo	- Doctor of Dental Science (DDS) 0000-0001-8749-4493	6
6	Rick	R.	van Rijn	- Doctor of Philosophy in Medicine (PhD) - Medical doctor (MD) 0000-0002-8876-6515	7
7	Patrick	Werner	Thevissen	- Doctor of Philosophy in Biomedical Sciences (PhD) - Doctor of Dental Science (DDS) - Master of Forensic Odontology (MSc) 0000-0003-0890-6264	2
8	Koenraad	Luc	Verstraete	- Doctor of Philosophy in Medicine (PhD) - Medical doctor (MD) 0000-0003-3688-1076	1

9	Michiel	Bart	de Haas	/	5
0000-0001-7946-5676					

Affiliations:

1. Department of Diagnostic Sciences – Radiology, Ghent University, Belgium
2. Department of Imaging and Pathology – Forensic Odontology, KU Leuven, Belgium
3. Department of Oral Diseases and Maxillofacial Surgery, Maastricht UMC+, The Netherlands
4. Department of Oral and Maxillofacial Surgery, Leuven University Hospitals, Belgium
5. Division of Special Services and Expertise, Section of Forensic Anthropology, Netherlands Forensic Institute, The Hague, The Netherlands
6. Department of Oral Health Sciences – Special Needs in Dentistry, Ghent University, Belgium
7. Amsterdam UMC, University of Amsterdam, Department of Radiology and Nuclear Medicine, Amsterdam, The Netherlands

* De Tobel J and van Wijk M equally contributed to this paper.

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

jannick.detobel@ugent.be

Acknowledgements

We wish to thank all participants and everybody who helped with recruitment. Special thanks to Patrick Davis for his critical appraisal of the manuscript. This study was funded by the Dutch Immigration and Naturalisation Service (IND).

Funding

This study has received funding by the Dutch Immigration and Naturalisation Service (IND).

Compliance with ethical standards

This project was approved by the Ghent University Hospital Ethics Committee. Written informed consent was obtained from all volunteers, and in case of minors, from their parents.

Conflict of interest

The authors declare no conflict of interest.

Related work

Results described in this manuscript were orally presented at the 6th International Society of Forensic Radiology and Imaging (ISFRI) Congress in Odense, Denmark, on May 12, 2017.

Part of the study population has been included in previous papers to study the development of their third molars and left wrist:

- De Tobel J, Hillewig E, de Haas MB et al (2018) 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.* 2019; 26:2924-2935.
- De Tobel J, Parmentier G, Phlypo I, Descamps B, Neyt S, Van De Velde W, Politis C, Verstraete K, Thevissen P (2018) Magnetic resonance imaging of third molars in forensic age estimation: comparison of the Ghent and Graz protocols focusing on apical closure. *Int J Legal Med.* 2018; 133:583-592.
- De Tobel J, Phlypo I, Fieuws S, Politis C, Verstraete K, Thevissen P (2017) Forensic age estimation based on development of third molars: a staging technique for magnetic resonance imaging. *J Forensic Odontostomatol* 35: 117-140.
- 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.
- 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.

Part of the study population has been included in previous papers to study the development of their clavicles, without a specific focus on motion artefacts:

- 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 suggestions for age estimation. *J Magn Reson Imaging*. doi: 10.1002/jmri.26889. [Epub ahead of print].
- Hillewig E, Degroote J, Van der Paelt T et al (2013) Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: towards more sound age estimates. *Int J Legal Med* 127: 677-689.
- 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.

Abstract

Purpose: To determine how motion affects stage allocation to the clavicle's sternal end on MRI.

Materials and methods: Eighteen volunteers (9 females, 9 males) between 14 and 30 years old were prospectively scanned with 3T MRI. One resting state scan was followed by five intentional motion scans. Additionally, a control group of 72 resting state scans were selected from previous research. Firstly, six observers allocated developmental stages to the clavicles independently. Secondly, they re-assessed the images, allocating developmental statuses (immature, mature). Finally, the resting state scans of the 18 volunteers were assessed in consensus to decide on the 'correct' stage/status. Results were compared between groups (control, prospective resting state, prospective motion), and between staging techniques (stages/statuses).

Results: Inter-observer agreement was low (Krippendorff α 0.23-0.67). The proportion of correctly allocated stages (64%) was lower than correctly allocated statuses (83%). Overall, intentional motion resulted in fewer assessable images and less images of sufficient evidential value. The proportion of correctly allocated stages did not differ between resting state (64%) and motion scans (65%), while correctly allocated statuses were more prevalent in resting state scans (83% versus 77%). Remarkably, motion scans did not render a systematically higher or lower stage/status, compared to the consensus.

Conclusion: Intentional motion impedes clavicle MRI for age estimation. Still, in case of obvious disturbances, the forensic expert will consider the MRI unsuitable as evidence. Thus, the development of the clavicle as such and the staging technique seem to play a more important role in allocating a faulty stage for age estimation.

Keywords

Age determination by skeleton

Clavicle

Magnetic resonance imaging

Motion artefact

Abbreviations

CI	confidence interval
ICC	intra-class correlation coefficient
MRI	magnetic resonance imaging
TE	echo time
TR	repetition time

MAIN TEXT

Introduction

Magnetic resonance imaging (MRI) has been studied by several research groups to visualise developing structures for forensic age estimation in living adolescents and subadults. Developing structures for which MRI protocols have been reported in forensic literature include long bones, the iliac crest and teeth [1-8]. Most protocols will include a method of fixation or stabilisation of the region of interest, since MRI acquisition requires several minutes and is prone to motion artefacts. For example, the wrist can be fixed with a flexible cloth belt [9] or stabilised in the coil using pillows [10] or a vacuum bed [11]. The head can be fixed by an individualised bite plate [2]. This minimises motion artefacts that might impede an adequate assessment of the developmental stage of the considered anatomical structure.

However, when clavicles are studied for age estimation, absolute fixation is impossible and breathing artefacts are unavoidable. Hillewig et al. (2011) reported the use of a specially shaped vacuum pillow and rubber device to position the coil during the scan of the sternal ends of the clavicles. Moreover, they scanned in prone position to decrease breathing artefacts [4]. Authors of more recent publications agreed that scanning in prone position resulted in less breathing artefacts [12-14].

Since absolute fixation is impossible when the clavicles are scanned, more extensive movements than breathing movements are possible. If participants would be aware of the capability of motion to disturb the age diagnostics, they might move intentionally. One can assume that mainly older subjects would try to frustrate the examination to try to appear younger. Intentional motion has been noted in practice during X-ray examinations of clavicles or hand/wrist by the authors (Fig. 1). Therefore, the aim of the current study was to determine the influence of motion on stage allocation to the sternal end of both clavicles based on MRI. It was hypothesised that intentional motion artefacts would impede stage allocation, whereas mere breathing artefacts would not.

Materials and methods

Study population and staging technique

The study was approved by the ethics committee of Ghent University Hospital. From every participant, written informed consent was obtained. In case of minors, the parents' informed consent was also obtained.

Healthy Belgian and Dutch Caucasian volunteers between 14 and 30 years old were prospectively included to participate in motion scans (Table 1). The participants were part of a sample of a larger ongoing study at Ghent University [15, 16]. They were selected based on the expected developmental stage of their clavicles, aiming for a balanced sample that would cover all stages of development. The staging technique by Schmeling (2004), modified by Kellinghaus (2010) was used to assess the images [17, 18]. Because of the low number of participants with clavicles in stage 2, no substaging was included for that stage. Moreover, since Hillewig et al. (2013) demonstrated that the physal scar could not be distinguished on clavicle MRI [19], stages based on the presence or absence of the physal scar were not included (Table 2).

In a total of 18 participants, one motionless resting state scan and five motion scans were conducted. Moreover, for every motion scan that would be assessed, a resting state matched control scan was included (Fig. 2). Matching was done based on the developmental stages of the clavicles, allocated in consensus. Since five motion scans were conducted per participant and every motion participant already had one resting state scan, an additional 72 resting state scans were needed. Therefore, 55 participants were retrospectively selected from the study sample of Hillewig et al. (2013) [19], complemented with 17 participants from the sample of the larger ongoing study at Ghent University [14, 15]. Regarding the scans from Hillewig et al. (2013), only cases were selected in which at least three out of four observers had agreed on the stage. Regarding the scans from the larger ongoing study, cases were selected based on preliminary staging results.

Image acquisition and motion protocol

Three Tesla MRI (Magnetom Trio Tim, Siemens, Erlangen, Germany) was performed in prone position, using a loop-shaped surface coil, according to the T1 gradient-echo VIBE protocol by Hillewig et al. (2011) (TR/TE 10/2.9; voxel size $0.7 \times 0.7 \times 0.9 \text{ mm}^3$; matrix 536×576 pixels; FOV 400 mm; flip angle 20° ; water excitation filter; acquisition time 4 minutes 2 seconds). Prospective scans were conducted between June and November 2015, after a small pilot of scanning three independent volunteers to decide on the study design. Retrospectively selected control scans had been conducted between April 2009 and May 2015.

For the motion study, six scans in two sessions of three consecutive scans were conducted in every participant (Fig. 2). For all of them, the first scan was a resting state scan without intentional motion. The applied scanning sequence filled k-space in a linear way. The position of the sternal ends of the clavicles was in the centre of the images, corresponding with k-space being filled at half the acquisition time. Thus, the influence of motion on image quality was less at the beginning and at the end of the acquisition time, while it peaked at half the acquisition time. Therefore, the participant was instructed to move voluntarily at 2 minutes 1 second during the motion scans (Fig. 3).

Participants moved in five distinctive ways (Fig. 2), four of which were standardised movements designed by the authors:

- cough twice,
- move both shoulders twice anteriorly,
- move both shoulders twice posteriorly,
- move both shoulders twice cranially.

One movement (further on referred to as 'freestyle') was invented by the participant in advance, after a short explanation about the purpose of the movements and without prior knowledge about the standardised movements. It was explained that by moving, a participant would intend to distort the image in order to impede assessment. On the other hand, the movement should be quite discrete, since it should remain unnoticed by the scanning personnel.

In order to eliminate the influence of fatigue on the latter scans, the order of the motion scans was randomised, so that a specific order would only occur once.

Image analysis

All images were pseudonymised and every scan was allocated a random code, in order to randomly distribute motion scans and resting state scans. Six observers (JDT, MvW, EH, RvR, KV, MdH) assessed the images. They represented a multidisciplinary team, all of whom were experienced in age estimation and had followed the rise of MRI in this field closely. Their level of experience and the monitors used for the assessments are shown in Table 3. Training and calibration occurred by assessing two training cases with known motion artefacts in consensus. These cases were not included for analysis. Moreover, the first 20 – randomly selected – assessments occurred a second time in the total dataset. Those first 20 assessments were considered as further training. Consequently, they were excluded for analysis.

Firstly, each observer individually assessed both clavicles simultaneously. Four variables were considered. For every one of them, the observer took the whole sequence of slices into account.

- 1) It was recorded if motion could be noticed on one or more images of the sequence.
- 2) A stage of epiphyseal development of the clavicles (Table 2) was allocated. No exact measurements were done whilst assessing the MR-images. Instead, the degree of fusion was estimated by the observers. During staging, the benefit of the doubt was always granted. This means that in cases of reasonable doubt, the younger stage was allocated. If the evaluation proved to be impossible, a reason was reported (anatomical impediment, morphological variant, insufficient technical quality, motion artefacts, other). This staging technique corresponds with the current practice in Belgium, albeit currently still applied to radiographs [20, 21].
- 3) Then, the observer judged if the images could be used as forensic evidence in a court of law.
- 4) Finally, observers described the ease of the assessment, i.e. to allocate a stage or to decide that the images were not assessable and to judge the evidential value (easy, not easy/not difficult, difficult).

To eliminate the learning effect of assessing the images, the first 20 cases were randomly repeated in the second half of the dataset. These cases were disregarded in the analyses.

Secondly, the six observers re-assessed all scans independently from the previous assessment, collecting data on the above described four variables, but with an alternative to the staging technique. This time, only two statuses were considered: immature or mature (Table 2). This status technique corresponds with the current practice in the Netherlands, albeit currently still applied to radiographs [22].

Finally, five out of the six original observers (JDT, MvW, EH, KV, MdH) assessed the resting state scan of the 18 prospectively included participants in consensus, joined by an additional observer (PT) (Table 3).

Statistical analysis

For analysis, the data were transferred from Microsoft Access 2010 to Microsoft Excel 2010, SPSS Statistics 22.0 (IBM SPSS Statistics for Windows, Armonk, NY, USA) and Matlab R217a (Mathworks Inc., Natick, MA, USA). Descriptive statistics were calculated and statistical tests were performed two-sided and evaluated at the 0.05 significance level.

Influences other than motion

In order to be able to study the influence of motion artefacts, other influences had to be taken into account first. To study the influences other than motion, the proportions of the four study variables were compared between resting state scans of the control group and those of the prospective group.

Reproducibility of the process was studied using the Krippendorff α coefficient, a statistical measure of the agreement achieved when coding a set of units. The coefficient is well fit for nominal data with a considerable amount of missing data, which was the case here. In these tests, stages were considered as nominal instead of ordinal data, because of the relatively large number of participants in stages 1 and 4, while those stages might be confused. The SPSS software of Hayes and Krippendorff was used, including the ability to bootstrap the results [23]. Rules of thumb describe that one may rely on data with $\alpha \geq 0.80$, consider data with $0.80 > \alpha \geq 0.67$ fit to draw tentative conclusions, and to discard data with $\alpha < 0.67$. Bootstrap intervals were used to construct 95% confidence intervals (CI) for Krippendorff α .

To detect possible population bias, the inter-observer agreement of the image assessment was compared between the 18 prospectively included participants' resting state scans and the 72 controls. Since this was done for each observer separately, matched pairs were considered in a Wilcoxon rank-sum test.

To study the influence of the imaging technique, the stages allocated by the different observers were compared with the consensus stages. When the allocated stage corresponded with the consensus stage, this was called a 'correctly allocated stage'. When all observations would randomly differ from the consensus, this would reflect the difficulty in interpreting the MRI. This also allowed studying observer bias and checking for the need to weigh the assessments of different observers or even the need to exclude an observer (if his/her allocated stages would differ systematically from the consensus and the other observers). Status allocation was studied in the same way. Thus, the influence of staging technique was reflected by differences between allocating stages versus statuses.

Influence of motion

To study the influence of motion, the proportions of the four study variables were compared between prospective resting state scans and motion scans. A final step was to compare the proportion of correctly allocated stages and statuses based on:

- resting state scans,
- motion scans,
- scans in which the observer indicated that the images could be used as evidence in a court of law.

A Wilcoxon matched-pairs signed-rank test was used for these analyses.

Results

All reported results were based on the evaluation of the right clavicle. Results for the left clavicle were similar. Morphological variants impeded staging in 0.7% of the assessed clavicles (8/1080 right, 8/1080 left). Those clavicles were excluded from analysis. Table 4 summarises the results of the studied variables in the different groups.

Influences other than motion

Population bias

The upper and middle parts of Table 4 display the results of resting state scans in the control and prospective populations, respectively. A population bias was noted, with scans in the prospective population showing more motion artefacts, and consequently being more difficult to assess. This needs to be taken into account when interpreting all other results.

Reproducibility and imaging technique, observer bias and staging technique

Inter-observer agreement was low, with α ranging from 0.23 to 0.67. The middle part of Table 4 summarises resting state scan results of stage/status allocation in the prospective population. The proportion of correctly allocated stages was low (64%). Results per observer indicated that this was not due to one or more observers performing worse than others. By contrast, the proportion of correctly allocated statuses was higher (83%). However, allocating statuses instead of stages did not result in more assessable images or more images of sufficient evidential value. Overall, the process of status allocation was considered easier than stage allocation.

Influence of motion

To illustrate the motion artefacts, Figs. 4 and 6 show one corresponding slice of the six conducted scans from two participants. Correspondingly, Figs. 5 and 7 illustrate the resulting variation in allocated stages and statuses.

The lower part of Table 4 summarises motion scan results. Motion artefacts were more frequent than in resting state scans, indicating that intentional motion did affect the images. Hence, motion resulted in fewer assessable images and less images of sufficient evidential value. The proportion of correctly allocated stages did not differ significantly between resting state scans (64%) and motion scans (65%). By contrast, correctly allocated statuses were more prevalent in resting state scans (83% versus 77%). Surprisingly, when the observer was confident to take the images to court as evidence, the proportions of correctly allocated stages (58%; 121/210) was lower than in images of low evidential value (68%; 69/101). Conversely, correctly allocated statuses in these cases were 80% (209/261) and 73% (69/94) respectively.

Overall, the process of assessing motion scans was considered easier than assessing resting state scans. After all, in case of clear image distortion by motion, it was easy to decide to discard the images. By contrast, when the artefacts were more subtle, it was harder to discern intentional motion from physiological motion artefacts,.

Furthermore, Tables 5 and 6 summarise all observers' allocated stages/statuses based on motion scans, compared to consensus stage/status. Motion scans did not render a systematically higher or lower stage/status, compared to the consensus stages/statuses. Neither did the ability to assess correlate with stage/status. Intentional motion lead to wrongfully allocating a younger stage (stage 1 to 3) to a fully mature clavicle in 35% (24/68) of assessable cases. For status allocation, this was in 56% (41/73) of assessable cases. Conversely, wrongfully deciding on a fully mature clavicle was less frequent with 19% (32/165) for stages and 11% (23/202) for statuses.

Discussion

Influence of motion

The current study demonstrated that at a certain degree of motion, assessment becomes impossible and the images rendered unsuitable as evidence. Deciding this is relatively easy and does not depend on staging technique. When this is obvious during the scanning session, a rescan can be done. This corresponds with the way radiographs are obtained in current age estimation. If a radiograph proves to be unsuitable, a new radiograph is made [22].

Motion artefacts are known to distort MR-images of the clavicles for age estimation in the living [12, 19, 24]. It has been reported that the magnitude of the disturbing effect is directly proportional to the magnetic field strength [25], so artefacts become more important, as line by line, higher field scanners are replacing lower field scanners all around the world. Table 7 displays the frequencies of untenable MRIs that have been reported in age estimation literature. However, these studies only included participants who volunteered to be scanned. One can assume that they tried their best to lie still during the image acquisition. Conversely, in age estimation practice, the examinee might try to distort the image by voluntary motion. This might result either in untenable images or in a faulty assessment of the clavicle's development. Nevertheless, the observer needs to decide whether or not the images are of sufficient quality to be of judicial evidence.

However, motion artefacts might be caused by breathing or by the pulsations of the blood vessels near the clavicles, even in resting state scans. Those cannot be distinguished from artefacts caused by small intentional movements. Neither can small unintentional movements be excluded, such as swallowing or occasional contraction of muscles in the arm. Therefore, the current study aimed to quantify the effect of intentional motion on stage allocation. Results demonstrated that motion artefacts bear the risk of wrongfully assessing the clavicle's development, regardless of the developmental stage. No trend could be demonstrated regarding the direction of this wrong assessment, i.e. no systematic over- or under-staging. Remarkably, development was not significantly more correctly assessed in resting state scans than in motion scans.

Therefore, age estimation based on clavicle MRI might benefit from further reducing motion artefacts. One way to do this, is to decrease acquisition time. Using an open compact 0.3T MRI to apply a 3D coherent gradient-echo sequence, Terada et al. (2015) demonstrated that compressed sensing with an acceleration factor of 3 rendered the optimal protocol to study hand/wrist development for age estimation [26]. This resulted in an acquisition time of 55 seconds instead of 2 minutes 44 seconds. Similarly, Neumayer et al. (2018) used a 3D FLASH VIBE sequence on 3T MRI to study hand/wrist [27].

They undersampled their data to study acceleration factors between 4 and 20. They suggested that an acceleration factor up to 7.49 was feasible for age estimation by a radiologist. This corresponded with an acquisition time of 15 sec instead of 3 min 46 sec. Automatic age estimation [28] was possible within the range of reported minimum standard errors, based on images with an acceleration factor up to 19.58. Both publications suggest that a major reduction in acquisition time would be possible to study hand/wrist for age estimation. However, it remains unclear if this could also apply to MRI of other anatomical regions – such as clavicles – to estimate age.

In pulmonary MRI, motion artefacts can be avoided by using either a coached breathing procedure or an active breathing control device in combination with segmented k-space acquisition [29]. That way, data acquisition only occurs while breathing is suspended. However, this might not be feasible for age estimation practice, since it respectively requires good compliance of the participant and an external device, which adds to the cost of the procedure. Moreover, intentional motion would still be possible with this procedure.

Influences other than motion

Staging technique

The influence of motion artefacts on stage and status allocation proved to be similar. Rather than depending on motion artefacts, correctly assessing the clavicle's development seemed to depend on the staging technique. After all, status allocation was more frequently correct than stage allocation. This is in line with status allocation being more frequently considered as an easy process, as opposed to stage allocation. Thus, the interpretation of the clavicle MRI is more difficult when one needs to allocate a stage than when one only needs to decide if the clavicle is mature. By contrast, Hillewig et al. (2013) demonstrated that stage 1 can be mistaken for a stage 4 and vice versa [19], a problem which was confirmed in the current study (Fig. 5) and is not countered by status allocation. De Tobel et al. (2019) even suggested discarding clavicles for age estimation if they are both in stages 1, 4, or 5 [14]. Instead, the authors suggest assessing other developing anatomical structures for age estimation.

Reproducibility and imaging technique

The difficult interpretation of the clavicle MRI is reflected by the low reproducibility of the process. Surprisingly, status allocation did not render higher inter- and intra-observer agreements than stage allocation. By contrast, Tangmose et al. (2014) reported higher values for reproducibility statistics when allocating statuses (Table 7). Note that the low values of Krippendorff's α cannot directly be compared with literature, since other statistics were reported. Still, the reproducibility statistics for clavicle assessments displayed in Table 7 are lower than those for other long bones, with kappa > 0.80

in most studies [30]. Since the clavicles' depiction on CT is more clear-cut, relatively higher reproducibility statistics have been reported than with MRI [31]. However, great disparity exists among reported reproducibility statistics. This might indicate that the clavicle as a bone is hard to assess, rather than the MRI as an imaging technique being hard to assess.

Morphological variants

A final disturbing influence lies in the wide range of anatomical variants of the clavicle's sternal end. Several authors have suggested that variants should be excluded for age estimation, since their development is unsure [13, 17, 32-36]. Recently, that suggestion has been proven by studying age estimation performance, when morphological variants were staged. Applying their original model for age estimation – which excluded morphological variants – to the variants, De Tobel et al. (2019) found significantly higher mean absolute errors of the point prediction of age [14]. Still, in the current study population, the prevalence of variants (0.7%) was lower than the range that has been reported in literature (2.1% [14] to 21.3% [13]).

Strengths, weaknesses, and future prospects

With 18 participants, the population in which intentional motion was studied was small. Moreover, a population bias was noted, which might obfuscate the conclusions. This might have been caused by the selection of control scans from Hillewig et al. (2013). Since cases were only selected when three out of four observers had agreed on the stage, those scans were more likely less affected by artefacts. Still, intentional motion showed the potential to distort the images without them being considered of insufficient quality to hold as judicial evidence. Consequently, the risk of a wrong assessment of the clavicle's development was evident. Since intentional motion cannot be derived from the image, the only way to detect intentional motion – and to prevent a wrong assessment – is to monitor the examinee during the scan. This could simply be done by someone keeping an eye on the examinee at all times during the scan, or by videotaping the examination. Moreover, the MR scanning procedure might be overwhelming to the examinee – even more than a CT scan – with the impressive device, being slid in the bore, the need to lie still for several minutes, and the noise during the scan. This might cause restlessness and unintentional motion during the scan. Thus, to avoid this restlessness, the process needs to be explained clearly, and a mock scanner can support the explanation with a demonstration [15]. By and large, this preparation is part of obtaining informed consent.

With six observers, this study included a wide variety of observers' backgrounds, their experience with MRI, and their experience with age estimation. Some observers were more cautious and decided to discard the images as evidence more often than other observers. This reflects the wide variety of practitioners who are performing age estimation currently [37]. As long as there is no internationally

defined and accepted approach to estimate age, the inter-observer variety of assessments will add dispersion to the already large inter-individual variety of development. Therefore, in practice, assessments of the clavicle should always be based on a consensus between at least two observers [14]. This is not limited to clavicle MRI, but applies to all imaging techniques and all anatomical structures. Moreover, age estimation should always be based on the assessment of different anatomical structures, i.e. be multi-factorial [38, 39]. If discrepancies are noted between the structures, experts need to beware of possibly distorted clavicle images by subtle motion artefacts. Fortunately, a promising step towards a more uniform age estimation process has been made by Stern et al. (2019), introducing fully automated multi-factorial age estimation [40]. In a next step, the value of this approach needs to be explored in validation studies. Moreover, motion artefacts will undoubtedly have occurred in their sample and it remains unclear how machine learning handled them.

Conclusion

Intentional motion impedes clavicle MRI for age estimation. Still, in case of obvious disturbances, the forensic expert will consider the MRI unsuitable as evidence. Thus, the development of the clavicle as such and the staging technique seem to play a more important role in allocating a faulty stage for age estimation. Although status allocation might be a safer approach than stage allocation, validation studies are needed to verify this statement.

References

1. Baumann P, Widek T, Merkens H et al (2015) Dental age estimation of living persons: Comparison of MRI with OPG. *Forensic Sci Int* 253: 76-80. doi: <http://dx.doi.org/10.1016/j.forsciint.2015.06.001>
2. 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: 130-9. doi: 10.1080/03014460.2016.1202321
3. 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](https://doi.org/10.1136/bjsm.2006.031021) [pii];10.1136/bjsm.2006.031021 [doi]
4. 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]
5. Jopp E, Schröder I, Maas R, Adam G, Püschel K (2010) Proximal tibial epiphysis in magnetic resonance imaging. *Rechtsmedizin* 20: 464-8. doi: 10.1007/s00194-010-0705-1
6. Ottow C, Krämer JA, Olze A et al (2014) Magnetresonanztomographiestudie zur Altersschätzung von unbegleiteten minderjährigen Flüchtlingen. *Rechtsmedizin* 25: 12-20. doi: 10.1007/s00194-014-0991-0
7. 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
8. 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. doi: 10.1007/s12024-014-9548-5
9. Terada Y, Kono S, Tamada D et al (2013) Skeletal age assessment in children using an open compact MRI system. *Magn Reson Med* 69: 1697-702. doi: 10.1002/mrm.24439
10. Tomei E, Sartori A, Nissman D et al (2014) Value of MRI of the hand and the wrist in evaluation of bone age: Preliminary results. *J Magn Reson Imaging* 39: 1198-205. doi: 10.1002/jmri.24286
11. Urschler M, Krauskopf A, Widek T et al (2016) Applicability of Greulich-Pyle and Tanner-Whitehouse grading methods to MRI when assessing hand bone age in forensic age estimation: A pilot study. *Forensic Sci Int* 266: 281-8. doi: 10.1016/j.forsciint.2016.06.016
12. 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. doi: 10.1016/j.forsciint.2013.10.027
13. 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: 118-22. doi: 10.1016/j.forsciint.2014.05.008
14. De Tobel J, Hillewig E, van Wijk M et al (2019) Staging clavicular development on magnetic resonance imaging: pitfalls and recommendations for age estimation. *J Magn Reson Imaging*: [Epub ahead of print]. doi: 10.1002/jmri.26889
15. 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
16. De Tobel J, Fieuws S, Hillewig E et al (2019) Multi-factorial age estimation: a Bayesian approach combining dental and skeletal magnetic resonance imaging *Forensic Sci Int* [Revision submitted after peer review].
17. 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
18. 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.
19. 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
20. Thevissen P, Willems G. (2013) [The Triple Test: The K.U.Leuven-protocol for age estimation of unaccompanied minor refugees]. In: Aps JKM, Brand HS, Duyck J, van Es RJJ, Jacobs R, Vissink A, eds. *Het Tandheelkundig Jaar 2013*. Bohn Stafleu van Loghum Houten. pp. 175-90.

21. Thevissen PW (2013) Dental age estimation: striving for an optimal approach. PhD thesis, Leuven University Press
22. Netherlands Forensic Institute. (2014) [Protocol Age Assessment].
23. Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. *Communication methods and measures* 1: 77-89.
24. 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
25. van Gelderen P, de Zwart JA, Starewicz P, Hinks RS, Duyn JH (2007) Real-time shimming to compensate for respiration-induced B0 fluctuations. *Magn Reson Med* 57: 362-8. doi: 10.1002/mrm.21136
26. Terada Y, Tamada D, Kose K et al (2016) Acceleration of skeletal age MR examination using compressed sensing. *J Magn Reson Imaging* 44: 204-11. doi: 10.1002/jmri.25140
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. 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
29. Arnold JF, Morchel P, Glaser E, Pracht ED, Jakob PM (2007) Lung MRI using an MR-compatible active breathing control (MR-ABC). *Magn Reson Med* 58: 1092-8. doi: 10.1002/mrm.21424
30. 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* 26: 2924-35. doi: 10.1007/s00330-018-5944-7
31. 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
32. Ekizoglu O, Hocaoglu E, Inci E, Can IO, Aksoy S, Sayin I (2015) Estimation of forensic age using substages of ossification of the medial clavicle in living individuals. *Int J Legal Med* 129: 1259-64. doi: 10.1007/s00414-015-1234-y
33. Ekizoglu O, Hocaoglu E, Inci E et al (2015) Forensic age estimation by the Schmeling method: computed tomography analysis of the medial clavicular epiphysis. *Int J Legal Med* 129: 203-10. doi: 10.1007/s00414-014-1121-y
34. 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
35. Wittschieber D, Schulz R, Vieth V et al (2014) The value of sub-stages and thin slices for the assessment of the medial clavicular epiphysis: a prospective multi-center CT study. *Forensic Sci Med Pathol* 10: 163-9. doi: 10.1007/s12024-013-9511-x
36. Rudolf E, Kramer J, Schmidt S, Vieth V, Winkler I, Schmeling A (2019) Anatomic shape variants of extremitas sternalis clavicularae as collected from sternoclavicular thin-slice CT-studies of 2820 male borderline-adults. *Int J Legal Med* 133: 1517-28. doi: 10.1007/s00414-019-02065-6
37. European Asylum Support Office (EASO). (2018) Practical Guide on Age Estimation, Second edition. EASO Practical Guides Series Malta.
38. 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
39. De Tobel J, Bauwens J, Parmentier GIL et al (2019) The use of magnetic resonance imaging in forensic age estimation of living children and subadults systematically reviewed. *Pediatr Radiol* [Revision submitted after peer review].
40. Stern 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
41. 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: 102-6. doi: <http://dx.doi.org/10.1016/j.jofri.2013.05.008>
42. Schmidt S, Henke CA, Wittschieber D et al (2016) Optimising magnetic resonance imaging-based evaluation of the ossification of the medial clavicular epiphysis: a multi-centre study. *Int J Legal Med* 130: 1615-21. doi: 10.1007/s00414-016-1442-0

43. Schulze D, Rother U, Fuhrmann A, Richel S, Faulmann G, Heiland M (2006) Correlation of age and ossification of the medial clavicular epiphysis using computed tomography. *Forensic Sci Int* 158: 184-9. doi: 10.1016/j.forsciint.2005.05.033
44. Bassed RB, Drummer OH, Briggs C, Valenzuela A (2011) Age estimation and the medial clavicular epiphysis: analysis of the age of majority in an Australian population using computed tomography. *Forensic Sci Med Pathol* 7: 148-54. doi: 10.1007/s12024-010-9200-y
45. Milenkovic P, Djuric M, Milovanovic P, Djukic K, Zivkovic V, Nikolic S (2014) The role of CT analyses of the sternal end of the clavicle and the first costal cartilage in age estimation. *Int J Legal Med* 128: 825-39. doi: 10.1007/s00414-014-1026-9
46. Wittschieber D, Schulz R, Vieth V et al (2014) Influence of the examiner's qualification and sources of error during stage determination of the medial clavicular epiphysis by means of computed tomography. *Int J Legal Med* 128: 183-91. doi: 10.1007/s00414-013-0932-6
47. El Morsi DA, Abo El-Atta HM, ElMaadawy M, Tawfik AM, Batouty NM (2015) Age Estimation from Ossification of the Medial Clavicular Epiphysis by Computed Tomography. *Int J Morphol* 33.
48. Franklin D, Flavel A (2015) CT evaluation of timing for ossification of the medial clavicular epiphysis in a contemporary Western Australian population. *Int J Legal Med* 129: 583-94. doi: 10.1007/s00414-014-1116-8
49. Houpert T, Rérolle C, Savall F, Telmon N, Saint-Martin P (2016) Is a CT-scan of the medial clavicle epiphysis a good exam to attest to the 18-year threshold in forensic age estimation? *Forensic Sci Int* 260: 103.e1-.e3. doi: <http://dx.doi.org/10.1016/j.forsciint.2015.12.007>
50. Gurses MS, Inanir NT, Gokalp G, Fedakar R, Tobcu E, Ocakoglu G (2016) Evaluation of age estimation in forensic medicine by examination of medial clavicular ossification from thin-slice computed tomography images. *Int J Legal Med* 130: 1343-52. doi: 10.1007/s00414-016-1408-2
51. Gurses MS, Inanir NT, Soylu E, Gokalp G, Kir E, Fedakar R (2017) Evaluation of the ossification of the medial clavicle according to the Kellinghaus substage system in identifying the 18-year-old age limit in the estimation of forensic age-is it necessary? *Int J Legal Med* 131: 585-92. doi: 10.1007/s00414-016-1515-0
52. Ramadan SU, Gurses MS, Inanir NT, Hacifazlioglu C, Fedakar R, Hizli S (2017) Evaluation of the medial clavicular epiphysis according to the Schmeling and Kellinghaus method in living individuals: A retrospective CT study. *Leg Med (Tokyo)* 25: 16-22.

Tables

Table 1 Age and sex distribution of the motion study population.

Age	Frequency	
	Female	Male
14	0	2
15	1	1
16	0	1
17	0	1
20	1	0
21	3	0
23	1	0
24	1	0
25	1	1
26	0	2
27	0	1
30	1	0
Total	9	9

Table 2 Applied statuses and stages. Staging criteria for the sternal end of the clavicle were based on Schmeling et al. (2004) and Kellinghaus et al. (2010), modified for MRI. Stages 1 to 3 were combined into one status, “Immature”, whereas stage 4 corresponded with status “Mature”.

Status	Stage	Stage criteria
Immature	1	Ossification centre is invisible (= not yet ossified).
	2	Ossification centre is visible (= ossified), non-union of the epiphysis and metaphysis.
	3	Growth plate is partially ossified (= bone trabeculae cross the growth plate from ossification centre to metaphysis).
	3a	The epiphyseal-metaphyseal fusion completes one third or less of the former gap between epiphysis and metaphysis.
	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.
	Mature	4

Table 3 Observers' background and monitor properties.

Observer	Background		Monitor		
	Profession	Level of experience	Brand	Type	Resolution (pixels)
EH	Researcher at Radiology	9 years of interpreting MR-images as researcher on age estimation	Barco	MDCC-6130	3280 x 2048
JDT	Resident Oral and Maxillofacial Surgery	7 years of interpreting MR-images as researcher on age estimation	Barco	MDCC-6130	3280 x 2048
KV	Professor Radiology and Medical Imaging	more than 25 years of experience as musculoskeletal radiologist; involved in age estimation for more than 25 years	Toshiba	173HT02-T01	1920 x 1080
MdH	Physical Anthropologist	7 years of interpreting X-ray images as researcher on age estimation	Barco	MDCC-6130	2048 x 1640
MvW	Forensic Anthropologist	5 years of interpreting X-ray images as researcher on age estimation	Barco	MDCC-6130	2048 x 1640
PT	Professor Forensic Odontology	12 years of interpreting X-ray images as researcher on age estimation	Barco	MDCC-6130	3280 x 2048
RR	Professor Forensic and Paediatric Radiology	14 years of experience as clinical paediatric radiologist	Barco	MDCC-6130	2048 x 1640

Table 4 Results of stage and status allocation, per population group and type of scan. The table contains six similar blocks of results. Visual comparisons can be made horizontally, comparing staging techniques (stage allocation versus status allocation) within a population group and within the type of scan. Vertical comparisons are between population groups and types of scans (e.g. control group versus prospective group, or resting state scan versus motion scan), within a staging technique. Moreover, vertical comparisons were quantified using statistical tests.

Population	Type of scan	Variable	Stage allocation						Status allocation							
			Average		Range		P-value	Average		Range		P-value				
Control	Resting state	Motion artefacts	51%	(219/432)	17%	(12/72) -	85%	(61/72)	0.002	52%	(223/432)	32%	(23/72) -	90%	(65/72)	< 0.001
		Assessable	83%	(358/432)	67%	(48/72) -	92%	(66/72)	0.007	84%	(363/432)	65%	(47/72) -	92%	(66/72)	0.132
		Inter-observer agreement	0.53			[0.46 -	0.60]		0.39				[0.16 -	0.61]		
		Sufficient evidence in court	72%	(310/432)	54%	(39/72) -	89%	(64/72)	0.013	69%	(299/432)	50%	(36/72) -	85%	(61/72)	0.048
		Easy assessment	31%	(136/432)	4%	(3/72) -	69%	(50/72)	0.003	48%	(208/432)	17%	(12/72) -	100%	(72/72)	0.331
Prospective	Resting state	Motion artefacts	74%	(80/108)	44%	(8/18) -	100%	(18/18)		81%	(87/108)	44%	(8/18) -	100%	(18/18)	
		Assessable	71%	(77/108)	50%	(9/18) -	94%	(17/18)		75%	(81/108)	61%	(11/18) -	94%	(17/18)	
		Inter-observer agreement	0.23			[0.12 -	0.34]		0.67				[0.49 -	0.84]		
		Correctly allocated	64%	(49/77)	44%	(4/9) -	77%	(10/13)		83%	(67/81)	69%	(9/13) -	92%	(11/12)	
		Sufficient evidence in court	55%	(59/108)	28%	(5/18) -	83%	(15/18)		56%	(61/108)	28%	(5/18) -	94%	(17/18)	
		Easy assessment	17%	(18/108)	0%	(0/18) -	33%	(6/18)		40%	(43/108)	11%	(2/18) -	100%	(18/18)	
Prospective	Motion	Motion artefacts	89%	(483/540)	77%	(69/90) -	99%	(89/90)	0.001	84%	(452/540)	59%	(53/90) -	100%	(90/90)	0.433
		Assessable	43%	(233/540)	23%	(21/90) -	62%	(56/90)	< 0.001	51%	(275/540)	37%	(33/90) -	72%	(65/90)	< 0.001
		Inter-observer agreement	0.58			[0.50 -	0.65]		0.39				[0.12 -	0.61]		
		Correctly allocated	65%	(151/233)	48%	(21/44) -	80%	(40/50)	0.550	77%	(211/275)	53%	(26/49) -	90%	(38/42)	0.070
		Sufficient evidence in court	28%	(151/540)	10%	(9/90) -	46%	(41/90)	< 0.001	37%	(202/540)	24%	(22/90) -	57%	(51/90)	0.005
		Easy assessment	36%	(193/540)	10%	(9/90) -	62%	(65/90)	0.010	52%	(282/540)	30%	(27/90) -	100%	(90/90)	0.021

P-value (Control) = P-value of Wilcoxon rank-sum test comparing results between the resting scans in the control and prospective populations.

P-value (Motion) = P-value of Wilcoxon signed-rank test comparing results between resting state and motion scans in the prospective population.

Range = Range of frequencies encountered by different observers. For inter-observer agreement, the 95% bootstrap interval is displayed instead.

Table 5 Cross tabulation of allocated stages based on motion scans by all observers, per consensus stage. Shaded cells indicate perfect correspondence between the observer and the consensus.

Consensus stage	Individual stage allocation (all motion scans and all observers combined)									
	1		2		3		4		Not assessable	
1	23%	(27/120)	1%	(1/120)	4%	(5/120)	6%	(7/120)	67%	(80/120)
2	-	-	-	-	-	-	-	-	-	-
3	6%	(16/270)	1%	(4/270)	30%	(80/270)	9%	(25/270)	54%	(145/270)
4	7%	(11/150)	1%	(1/150)	8%	(12/150)	29%	(44/150)	55%	(82/150)

Table 6 Cross tabulation of allocated statuses based on motion scans by all observers, per consensus status. Shaded cells indicate perfect correspondence between the observer and the consensus.

Consensus status	Individual status allocation (all motion scans and all observers combined)					
	Immature		Mature		Not assessable	
Immature	46%	(179/390)	6%	(23/390)	48%	(188/390)
Mature	27%	(41/150)	21%	(32/150)	51%	(77/150)

Table 7 Frequencies of motion artefacts and reproducibility of assessing clavicular development based on 3D imaging modalities. Note that different statistics have been reported to quantify reproducibility of staging, and that the number of stages varies between studies.

Imaging modality	Subjects	Reference	Untenable images due to motion artefacts	Number of applied stages	Intra-observer agreement			Inter-observer agreement		
					Statistic		N	Statistic		N
MRI	Living	Hillewig et al. 2011 [4]	1.0%	4	-		-	ICC	0.743 - 0.787	121
		Hillewig et al. 2013 [19]	2.8%	4	Cohen's kappa	0.75	20	Fleiss' kappa	0.74 - 0.76	220
		Tangmose et al. 2014 [12]	-	4	kappa	(prone) 0.675	32	kappa	0.077 - 0.125	49 - 53
			-	2 (statuses)	kappa	(prone) 0.788	32	kappa	0.040 - 0.446	55
		Vieth et al. 2014 [13]	5.9%	9	-		-	-		-
		Schmidt et al. 2017 [24]	-	9	weighted kappa	0.991	160	weighted kappa	0.987	160
		De Tobel et al. 2019 [14]	2.5%, 4.0%, 19.1%, 21.5% [#]	11	weighted kappa	0.82	186	weighted kappa	0.60 - 0.64	387 - 471
MRI	Deceased	Tangmose et al. 2013 [41]		4	Cohen's kappa	0.685	94	-		-
		Tangmose et al. 2014 [12]		4	Cohen's kappa	0.662	47	kappa	0.410 - 0.447	45 - 47
		Schmidt et al. 2016 [42]		9	weighted kappa	0.992	80	weighted kappa	0.986	80
CT	Living	Schulze et al. 2006 [43]	-	4	-		-	Cohen's kappa	0.63 - 0.80	100
		Bassed et al. 2011 [44]	-	5	kappa	0.866	20	kappa	0.734	20
		Milenkovic et al. 2014 [45]	-	5	Cohen's kappa	0.817 - 0.832	308	Cohen's kappa	0.678 - 0.708	308
		Wittschieber et al. 2014 [46]	-	5	-		-	Cohen's kappa	0.494 - 0.674	710
			-	9	-		-	Cohen's kappa	0.470 - 0.605	710
		Ekizoglu et al. 2015a [33]	-	5	Cohen's kappa	0.768	503	Cohen's kappa	0.826	503
			-	5	weighted kappa	0.928	503	weighted kappa	0.946	503
		Ekizoglu et al. 2015b [32]	-	6	Cohen's kappa	0.916	193	Cohen's kappa	0.868	193
	-	5	kappa	0.854		kappa	0.753			

Figures

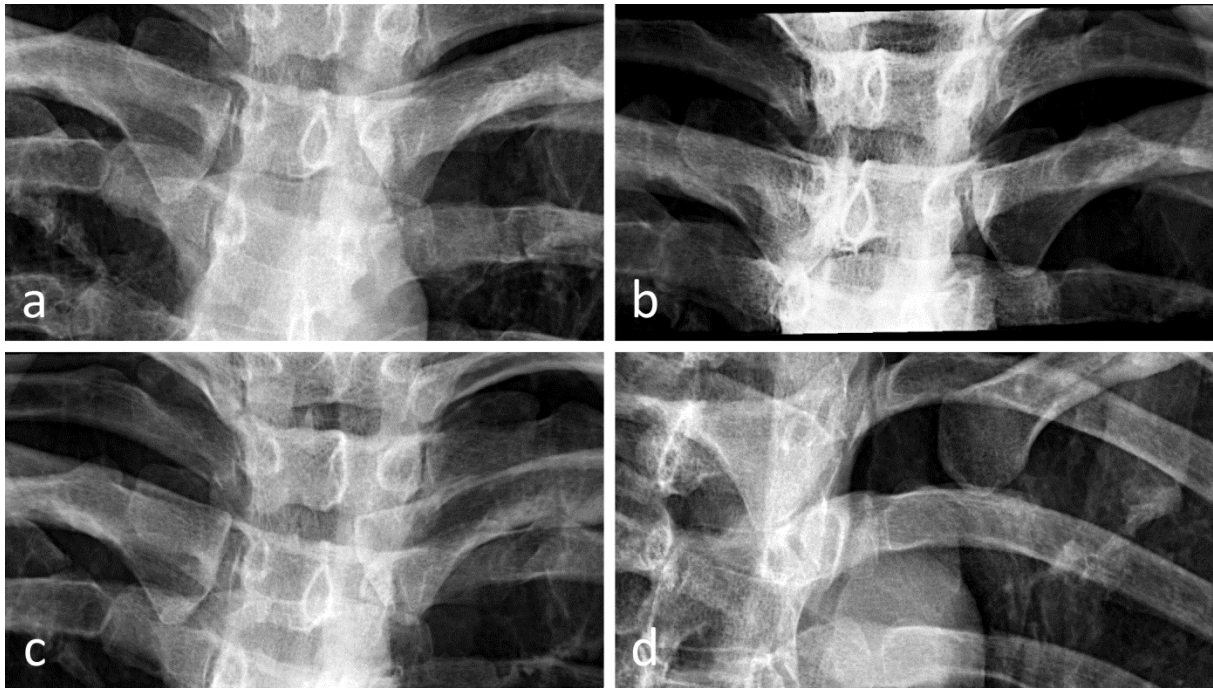


Fig. 1 Corresponding radiographs of both clavicles. All radiographs were produced postero-anteriorly (PA), but they were mirrored here to correspond with the standard view of imaging. **a** Standard PA resting state. The right clavicle can be assessed and appears fully mature. The left clavicle cannot be assessed, due to superpositioning. **b** Left anterior oblique 10° resting state, intended to better visualise the left clavicle. The left clavicle can be assessed and appears fully mature. **c** Standard PA while moving both shoulders caudally. The right clavicle can be assessed, while the left one is cannot, due to superposition. **d** Left anterior oblique 10° while moving the left shoulder from 10° to 45°. None of both clavicles can be assessed. Because of the overly turned left shoulder, a (remnant of the) growth plate and epiphysis would be projected over the metaphysis.

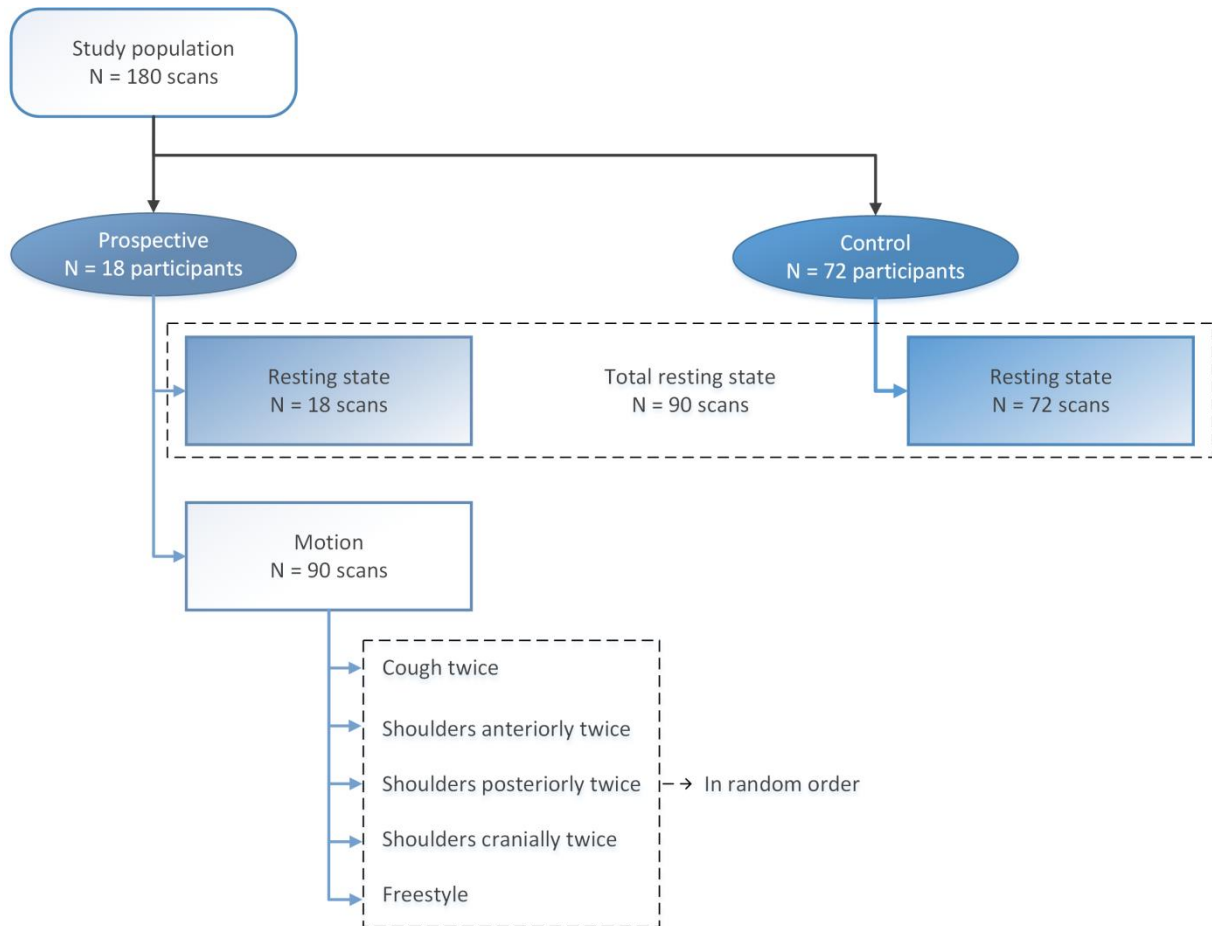


Fig. 2 Schematic overview of the study population. All 180 scans were assessed by six observers.

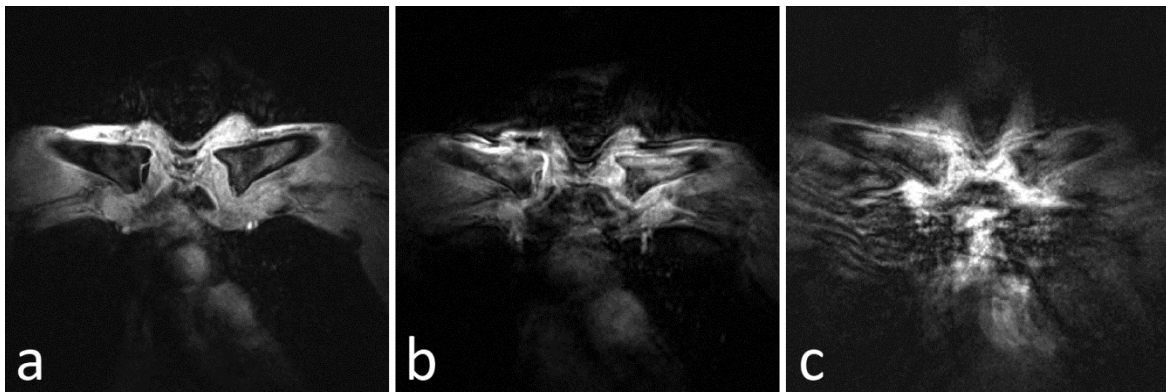


Fig. 3 One corresponding slice of three scans from the same 18-year-old male volunteer, who was only included in the small pilot study. **a** Resting state scan. Note the ossification centre of the right clavicle, without bone bridges, and a remnant of the ossification centre of the left clavicle, at its caudal end (the ossification centre appeared more clearly on other slices of the scan). **b** Freestyle motion, as instructed after 2 minutes 1 second of scanning. In this case, the participant chose to move the shoulders posteriorly. The image is clearly distorted by motion artefacts, albeit more subtle than in **c**. **c** Freestyle motion repetitively during the whole scan, resulting in major motion artefacts.

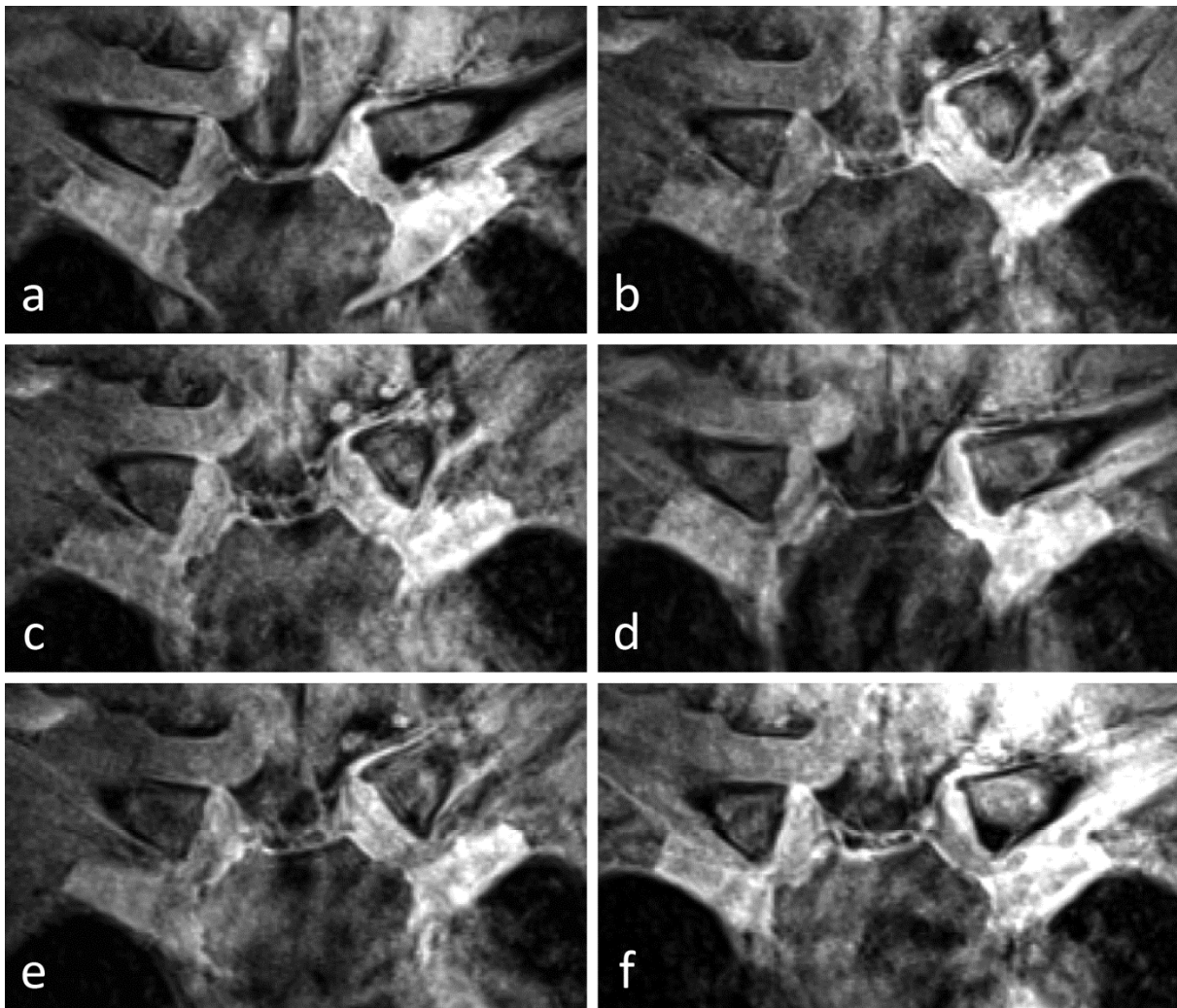


Fig. 4 One corresponding slice of the six conducted scans from the same 15-year-old female participant. Allocated stages and statuses are displayed in **Fig. 5**. **a** Resting state scan. **b** Cough twice. **c** Move both shoulders twice anteriorly. **d** Move both shoulders twice posteriorly. **e** Move both shoulders twice cranially. **f** Freestyle motion, for which the participant chose to move the shoulders posteriorly, corresponding with the movement in **d**.

In this case, scans were conducted in the following order: a, d, f; c, b, e. Note that because of the break in between the third and fourth sequence, the position of the participant changed.

The resting state scan (**a**) rendered a clean-cut depiction of the sternal end of both clavicles, while motion (**b-e**) created double lines which could be interpreted as a bridging growth plate. Only the freestyle motion scan (**f**) looked very similar to the resting state scan.

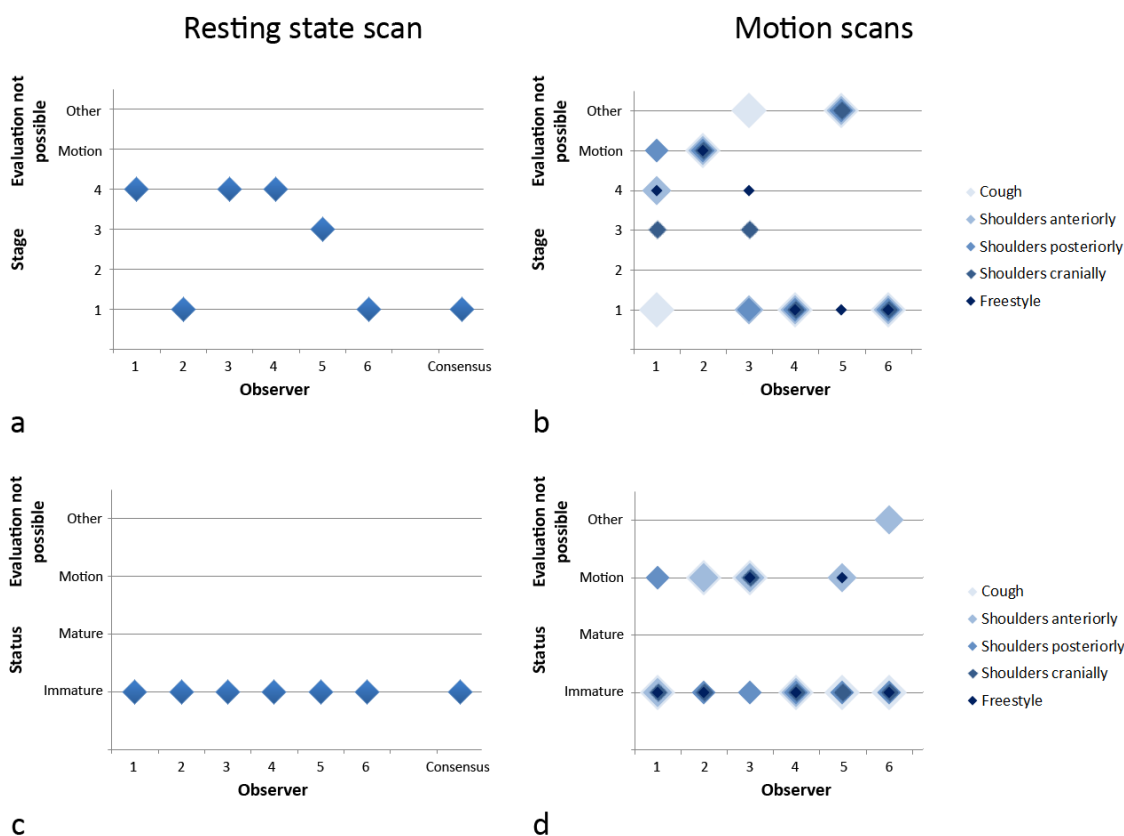


Fig. 5 Allocated stages (**a, b**) and statuses (**c, d**) to the right clavicle of the same participant as shown in **Fig. 4**, based on the different scans, per observer. **a** Although consensus stage 1 was allocated to the resting state scan, individual observers allocated stages 1, 3 or 4. **b** This dispersion was also seen based on the motion scans. Note that one observer judged all motion scans to be unsuitable for stage allocation, while others allocated a stage to all of them. **c** Allocating a status, all observers agreed on an immature clavicle. **d** This agreement was also seen based on the motion scans. Note that the observer who had judged all motion scans to be unsuitable to allocate a stage, did find those scans suitable for status allocation.

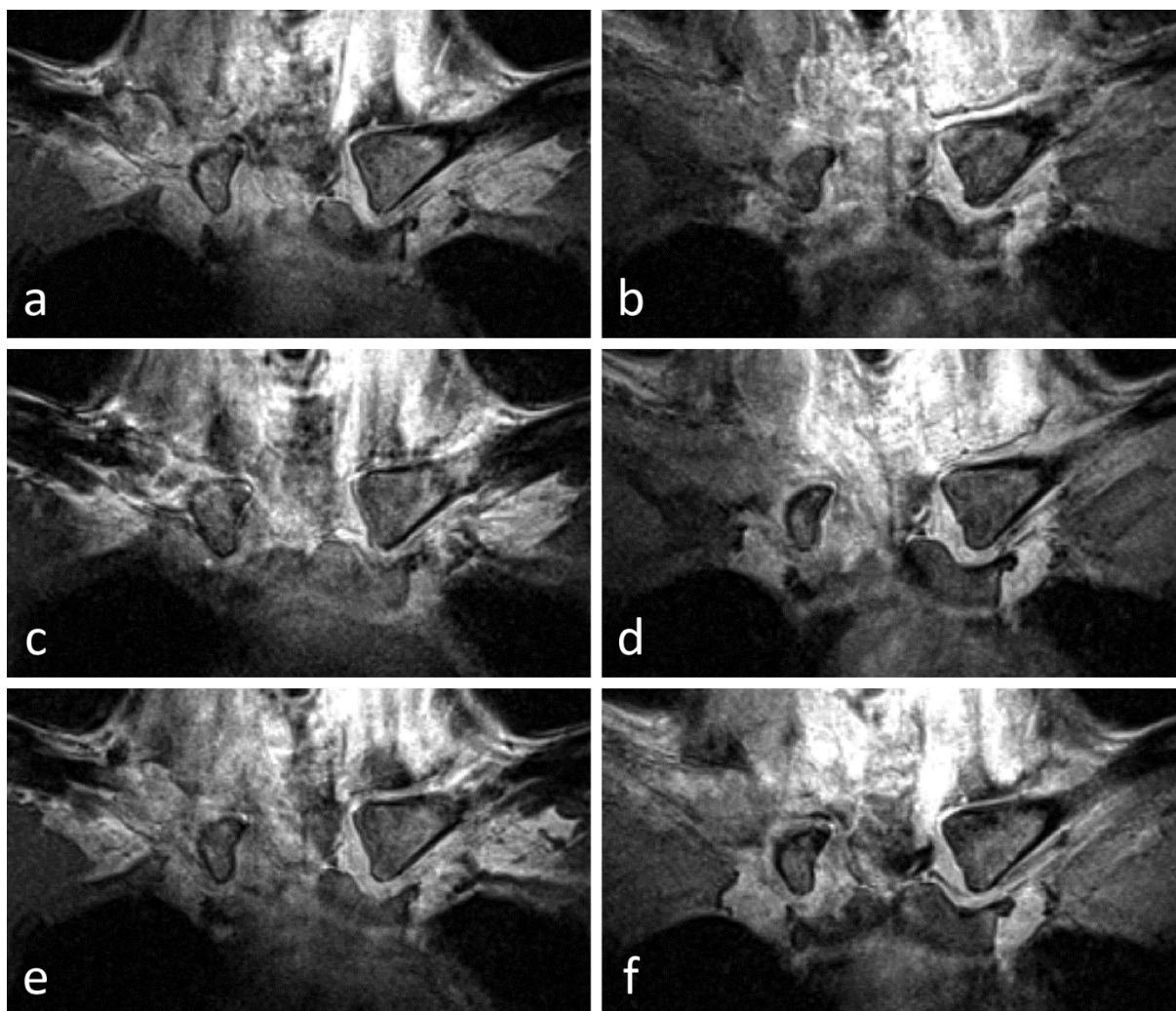


Fig. 6 One corresponding slice of the six conducted scans from the same 30-year-old female participant. Allocated stages and statuses are displayed in **Fig. 7**. **a** Resting state scan. **b** Cough twice. **c** Move both shoulders twice anteriorly. **d** Move both shoulders twice posteriorly. **e** Move both shoulders twice cranially. **f** Freestyle motion, for which the participant chose to move the chin towards the chest.

In this case, scans were conducted in the following order: a, e, c, f; d, b. The break in between the third and fourth sequence hardly changed the position of the participant.

The resting state scan (**a**) rendered a clean-cut depiction of the sternal end of both clavicles. Note that especially the moving both shoulders anteriorly (**c**) caused artefacts mimicking ossification centres. Nonetheless, five of the six observers indicated that motion impeded stage allocation to this motion scan **c** (see **Fig. 7b**). Therefore, motion did not result in a faulty stage allocation.

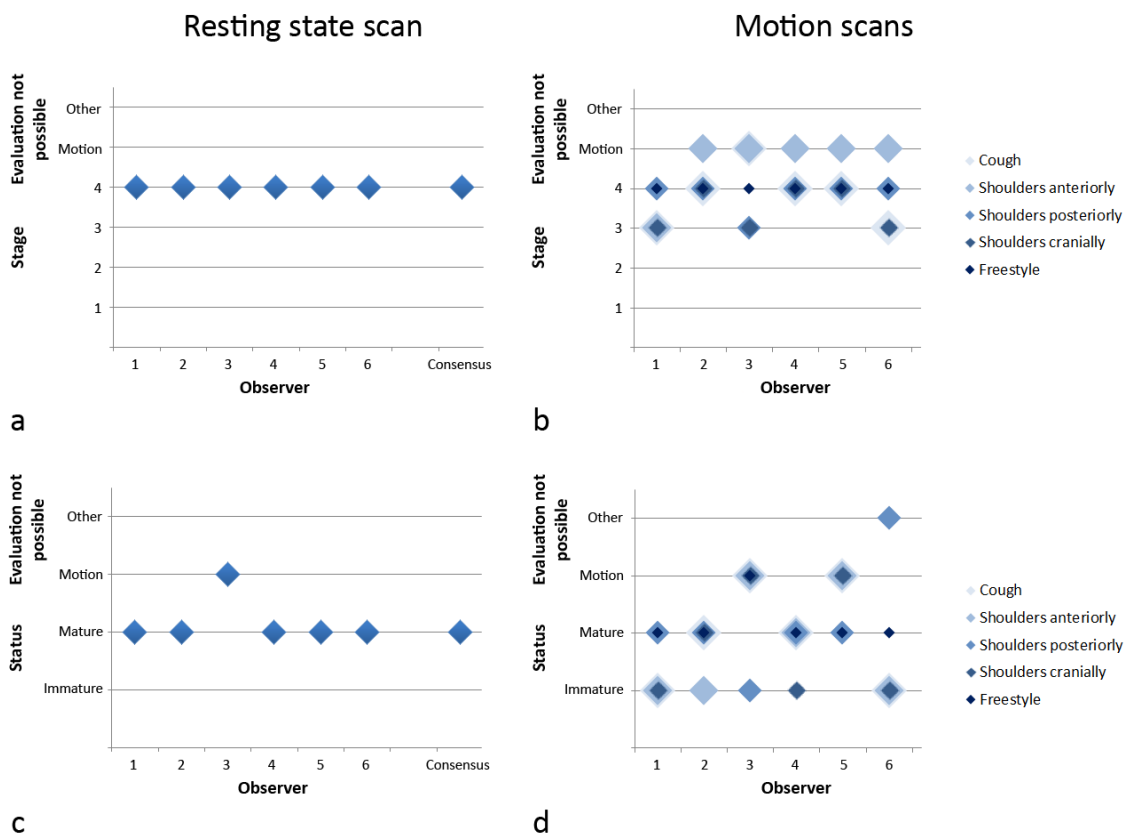


Fig. 7 Allocated stages (**a, b**) and statuses (**c, d**) to the right clavicle of the same participant as shown in **Fig. 6**, based on the different scans, per observer. **a** All observers agreed on a clavicle in the final stage. **b** Motion caused dispersion, with artefacts being interpreted as bridging growth plates. Note that all but one observer judged the motion scan of moving the shoulders anteriorly to be unsuitable for stage allocation (see **Fig 5c**). **c** Strikingly, allocating statuses, one observer judged the resting state scan to be unsuitable due to motion artefacts, while that same observer did allocate a stage. **d** The dispersion caused by motion was even larger when allocating statuses than when allocating stages. Motion was sometimes interpreted as an immature status.