

RESEARCH ARTICLE

# Observer agreement of imaging measurements used for evaluation of dentofacial deformity in juvenile idiopathic arthritis

<sup>1</sup>Johannes Fischer, <sup>2</sup>Josefine Halbig, <sup>3,4</sup>Thomas Augdal, <sup>5,6</sup>Oskar Angenete, <sup>7</sup>Peter Bangsgaard Stoustrup, <sup>7</sup>Kasper Dahl Kristensen, <sup>1,8</sup>Marit Slättelid Skeie, <sup>9</sup>Karin Tylleskär, <sup>1,10</sup>Annika Rosén, <sup>1,11</sup>Xieqi Shi and <sup>12</sup>Karen Rosendahl

<sup>1</sup>Department of Clinical Dentistry, University of Bergen, Bergen, Norway; <sup>2</sup>Public Dental Service Competence Centre of Northern-Norway (TkNN), Tromsø, Norway; <sup>3</sup>Department of Radiology, University Hospital of North Norway, Tromsø, Norway; <sup>4</sup>UiT the Arctic University of North Norway, Tromsø, Norway; <sup>5</sup>Department of Radiology and Nuclear Medicine, St. Olav University Hospital, Olav Kyrres gate, Trondheim, Norway; <sup>6</sup>Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Faculty of Medicine, Trondheim, Norway; <sup>7</sup>Section of Orthodontics, Faculty of Health Sciences, Aarhus University, Aarhus, Denmark; <sup>8</sup>Center for Oral Health Services and Research, Mid-Norway (TkMidt), Trondheim, Norway; <sup>9</sup>The Children's Clinic at Haukeland University Hospital, Bergen, Norway; <sup>10</sup>Department of Oral and Maxillofacial Surgery, Haukeland University Hospital, Bergen, Norway; <sup>11</sup>Department of Oral and Maxillofacial Radiology, Faculty of Odontology, University of Malmö, Malmö, Sweden; <sup>12</sup>UiT the Arctic University of North Norway, Tromsø, Norway

**Objectives:** To examine the precision of imaging measures commonly used to assess mandibular morphology in children and adolescents with juvenile idiopathic arthritis (JIA). Secondly, to compare cone-beam computed tomography (CBCT) and magnetic resonance imaging (MRI) in the measurement of condylar height.

**Methods:** Those included were children diagnosed with JIA during 2015–18 who had had an MRI, a CBCT of the temporomandibular joints (TMJs) and a lateral cephalogram (ceph) of the head within one month of each other. Agreement within and between observers and methods was examined using Bland-Altman mean-difference plots and 95% limits of agreement (LOA). A 95% LOA within 15% of the sample mean was considered acceptable. Minimal detectable change (MDC) within and between observers was estimated.

**Results:** 90 patients (33 males) were included, with a mean age of 12.8 years. For MRI, intra- and interobserver 95% LOA were relatively narrow for total mandibular length: 9.6% of the sample mean. For CBCT, condylar height, both intra- and interobserver 95% LOA were wide: 16.0 and 28.4% of the sample mean, respectively. For ceph, both intra- and interobserver 95% LOA were narrow for the SNA-angle and gonion angle: 5.9 and 8% of the sample mean, and 6.2 and 6.8%, respectively.

**Conclusions:** We have identified a set of precise measurements for facial morphology assessments in JIA, including one MRI-based (total mandibular length), one CBCT-based (condylar height), and three ceph-based. Condylar height was higher for MRI than for CBCT; however, the measurement was too imprecise for clinical use. MDC was also determined for a series of measurements.

*Dentomaxillofacial Radiology* (2022) **51**, 20210478. doi: [10.1259/dmfr.20210478](https://doi.org/10.1259/dmfr.20210478)

**Cite this article as:** Fischer J, Halbig J, Augdal T, Angenete O, Stoustrup PB, Dahl Kristensen K, et al. Observer agreement of imaging measurements used for evaluation of dentofacial deformity in juvenile idiopathic arthritis. *Dentomaxillofac Radiol* (2022) 10.1259/dmfr.20210478.

**Keywords:** Observer agreement; Temporomandibular joint; Magnetic resonance imaging; Cone-beam computed tomography; Cephalometry

## Introduction

Juvenile idiopathic arthritis (JIA) is an autoimmune, heterogeneous condition that includes different forms of chronic arthritis of unknown origin and affects around 1–2 in 1000 under the age of 16 years.<sup>1,2</sup> The disease is characterised by synovial inflammation, with a potential risk of developing progressive joint destruction and severe functional disability.<sup>1</sup> The temporo-mandibular joint (TMJ) is more frequently involved than previously believed (in up to 78% of cases), of which a high proportion appears to be clinically silent.<sup>3–6</sup> TMJ arthritis is associated with all JIA subtypes, and active inflammation is often difficult to detect clinically.<sup>7,8</sup> Moreover, around one-third of JIA patients with TMJ arthritis with JIA onset during growth will develop mandibular growth disturbances before skeletal maturity.<sup>9</sup>

Monitoring mandibular growth during childhood and puberty in children with JIA and TMJ arthritis has traditionally been performed using cephalometric tracings (cephs) and their superimpositions.<sup>10</sup> However, evaluation of growth using two-dimensional radiographs is flawed by distortion and overlapping of 3D structures, varying magnification and issues with positioning. Thus, over the past 20 years, the method has been replaced by cone-beam CT (CBCT) in many centres, at the cost of higher radiation doses.<sup>11</sup> A recent paper by Maspero and colleagues comparing measurements of mandibular body length and growth by CBCT and reconstructed lateral cephalograms, the authors found that although the direct measurements differed between the two methods, mandibular growth assessment was almost identical. They concluded that two-dimensional radiographs remain the preferred method in evaluating mandibular body growth.<sup>12</sup> However, knowledge on the precision or accuracy of cephs or CBCT using Fryback and Thornbury's widely cited principles for radiological research is lacking.<sup>13</sup> They suggest a hierarchical model, where level one addresses technical image quality, level two addresses diagnostic accuracy, sensitivity, and specificity, and so on until level six, which addresses the examination's impact on social costs and benefits. The demonstration of efficacy at each lower level in this hierarchy is logically necessary but not sufficient to assure efficacy at higher levels.<sup>13</sup> Their statements have fuelled our efforts to examine the precision of imaging markers used in children.<sup>14</sup> For CBCT, validation of the measurements has mainly been performed on specimens using small datasets.<sup>15,16</sup> Systematic reviews on accuracy of measurements and reliability of landmark identification with CT techniques in the maxillofacial area concluded, that there is just a limited number of studies and that most studies had methodological limitations and were of moderate quality.<sup>17,18</sup> According to recently published guidelines for imaging of TMJs in patients with JIA, magnetic resonance imaging (MRI) of the TMJs is advised for the assessment of inflammatory change.<sup>19,20</sup> The potential of MRI for the evaluation

of growth disturbances secondary to TMJ involvement has been addressed, using T1W 3D sequences to construct oblique sections through the mandible on which measurements are based.<sup>21,22</sup> By adding a T1W 3D sequence of the whole head to the TMJ protocol, we aimed to examine the precision of three MRI-derived measurements for mandibular morphology. Moreover, we examined the precision of one measurement derived from CBCT (small field of view) and commonly used ceph-measurements in a large cohort of children and adolescents with JIA. Finally, a comparison between CBCT and MRI for the measurement of condylar height was performed. The overall aim was to identify the most precise measurements for the assessment of mandibular morphology, for further use in the monitoring of mandibular growth.

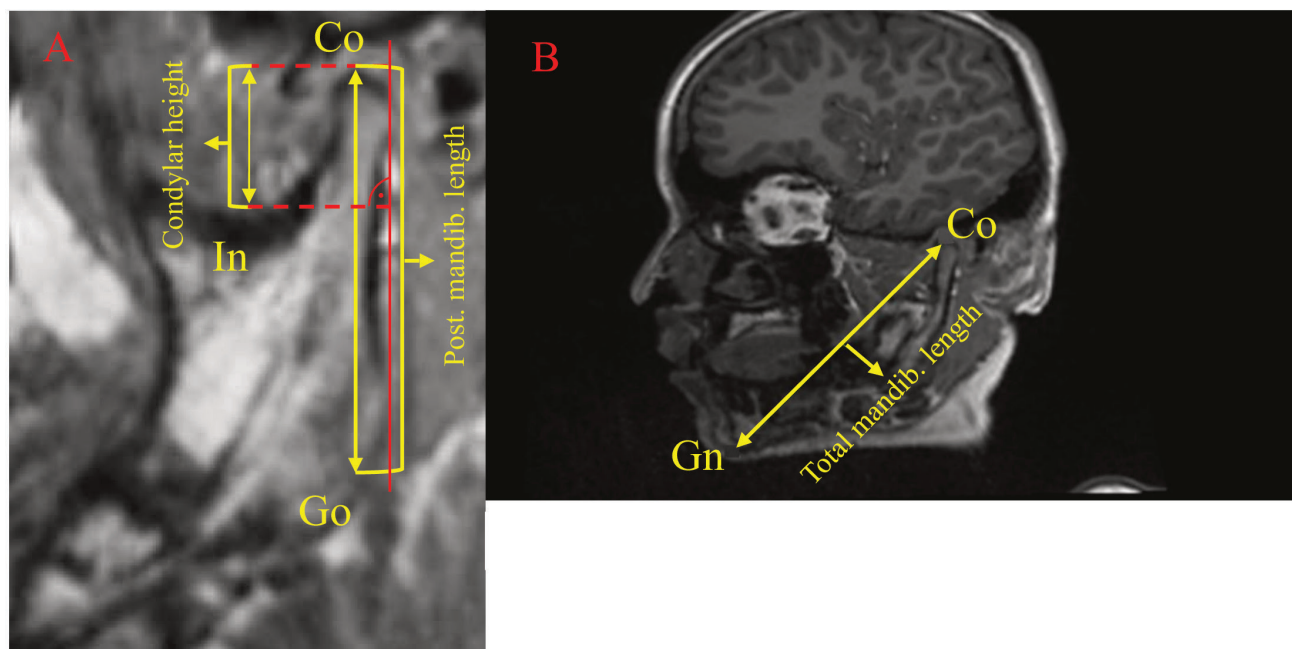
## Methods and materials

The present study is part of a longitudinal multicentre study addressing children and adolescents diagnosed with JIA, which was performed during 2015–2020 (NCT03904459 in [www.clinicaltrials.gov](http://www.clinicaltrials.gov)).<sup>1</sup> The study was approved by the Regional Ethics Committee (REK number 2012/542), and written informed consent was obtained from each participant and/or a caregiver according to the national guidelines. This particular study includes a subset of 90 patients who had undergone an examination of MRI, a CBCT of the TMJs and a lateral cephalogram of the head, taken within one month of each other between March 2015 and May 2018. The patients were identified on the basis of clinical, demographic information to reflect the whole range of disease duration, JIA subgroup and severity, in order to robustly test the variables in question.

### *Magnetic Resonance Imaging (MRI)*

All MRI examinations were performed on a 3 Tesla system (Skyra, Siemens Healthcare, Erlangen, Germany), using a 64-channel head coil (or a 32-channel in 30 patients). For the present study, a sagittal T1W magnetisation prepared rapid gradient echo (ultra-fast gradient-echo 3D) sequence (TR/TE/FA/SL = 2000/2.26/8/1) was used. Following several calibration-meetings and discussions, the images were assessed independently by two consultant radiologists using a high-resolution PACS-screen, twice (at an interval of 4 weeks) by KR and once by TAA (30 and 14 years of experience in paediatric imaging, respectively), without any other information available. The following three measurements were made: a) posterior mandibular length (Go-Co) and condylar height Co-In, b) total mandibular length Co-Gn ([Figure 1](#)).

To measure the total mandibular length, gnathion was used as origo when the multiplanar reconstruction



**Figure 1** Constructions and both linear measurements of posterior mandibular length measured from the gonion to the top of condyle (Co-Go) and condylar height measured from the most caudal point of incisura mandibulae to the top of the condyle (Co-In). **Figure 1B**: Construction of total mandibular base length measured between the gnathion and the top of the condyle (Co-Gn). Co: Condyle; Go: Gonion; Gn: Gnathion; In: Incisura; Post: Posterior; Mandib: Mandibular.

(MPR) volume was reoriented to include the top of the condyle. Co-Gn was measured, and the reorientation was repeated for the contralateral side. The method for measuring the posterior mandibular length and condyle height included reorienting the volume to a ramus-corrected oblique sagittal view, and determining the ramus tangent, gonion, top of the condyle and the lowest (caudal) point of the incisura mandibulae, to then measure the Co-Gn and Co-In (Figure 1).

#### Cone-beam computed tomography (CBCT)

The CBCT examinations were performed on one of three CBCT machines, with kVp / mAs / field of view (mm) / voxel dimension (isotropic, mm) settings: 3D Accuitomo 170 (Morita Mfg Corp, Kyoto, Japan) 85/175/40\*40\*40/0.08; Promax 3D (Planmeca Oy, Helsinki, Finland) 90/13.6/200\*200\*60/0.40; or Scanora 3D (Soredex, Tuusula, Finland) 90/45/60\*60\*60/0.13. The participants were positioned in the Frankfort horizontal plane, with their teeth in maximal intercuspal position.

The images were exported together with the accompanying image viewers included in the three respective CBCT systems – Planmeca Romexis Viewer (Planmeca Oy, Helsinki, Finland); OnDemand3DApp Project Viewer Limited (version 1.0.10.4304, CyberMed, Daejeon, Republic of Korea); and iDixel One Volume Viewer (J. Morita MFG. Corp., Kyoto, Japan) – and were assessed independently by two consultant radiologists, twice (at an interval >4 weeks) by TAA and once by OWA (both with 13 years of experience). Prior

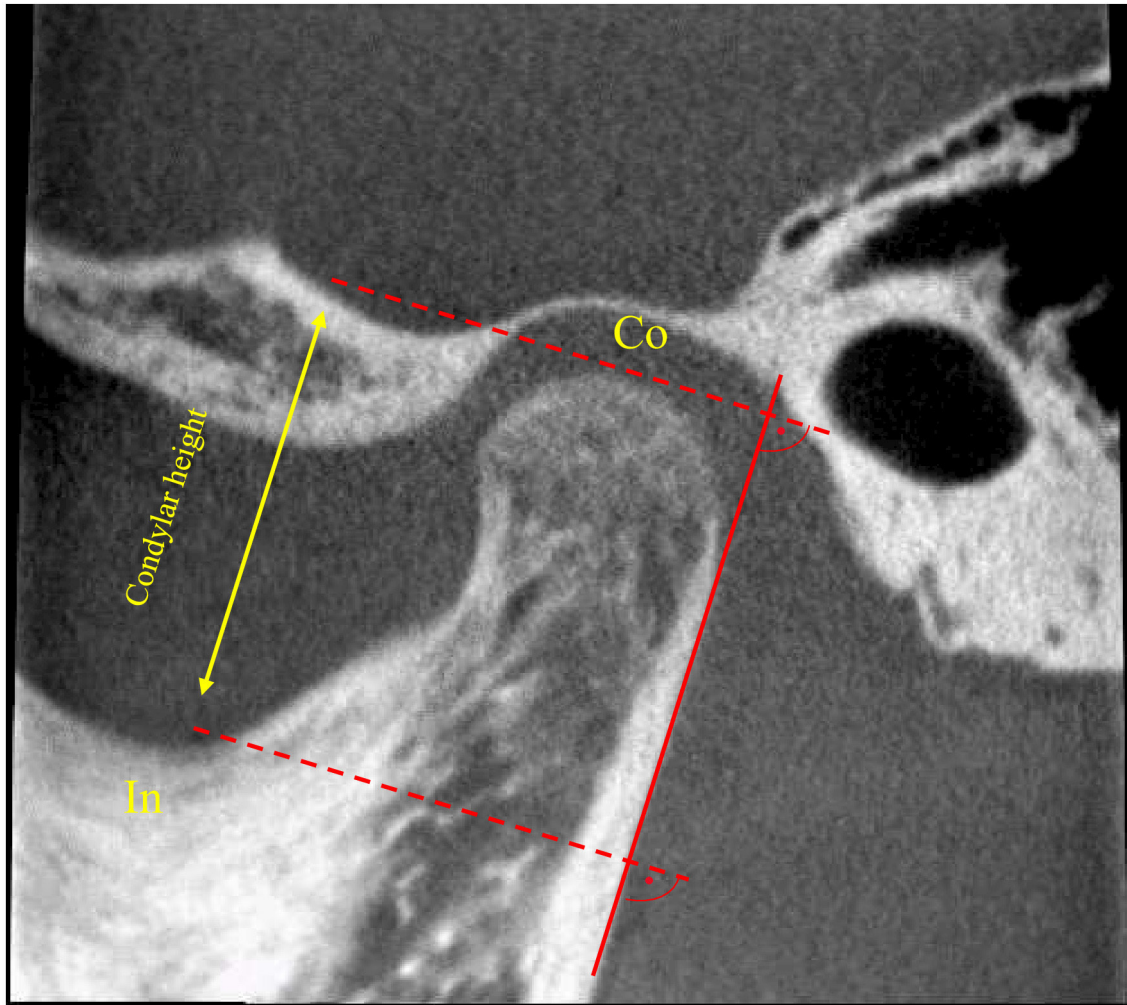
to scoring, meticulous calibration was performed. Condylar height was measured in the same manner as for MRI; however, due to the limited field of view, we approximated the ramus corrected sagittal view to include the coronoid process and the ramus tangent along the posterior border at the lowermost point of the condyle or ramus included in the field of view (Figure 2).

#### Cephalogram (ceph)

Lateral cephs were performed using one of three different pan/ceph systems: Orthophos xg 5 (Sirona Dental Systems, Bensheim, Germany), with the following settings: 73 kVp, 15 mA, exposure times 9.7 and 9.4 sec for adolescents and children, respectively, and a magnification factor of 1.1 with a 16-bit pixel depth for all images; Promax Type Version 3.8.1.0 (Planmeca, Helsinki, Finland), kV: 62–70, mAs: 7–10 mA, 6.7 sec; and Soredex Cranex D (Helsinki, Finland) 70 kV, 10 mA, exposure time 5.8 sec. The radiographs were taken under standardised conditions with a natural head position (Frankfort horizontal line parallel to the floor) and teeth in maximal intercuspal position.

#### Calibration of ceph-measurements

Two observers (JH and JF) underwent four calibration exercises (two on five cases and two on 30 cases – not included in the study) under the guidance of an expert (KDK), where nine measurements based on 16 anatomical landmarks were calculated (Supplementary Material S1) and (Figure 3). At the completion of the



**Figure 2** CBCT measurement of condylar height (Co-In). Condyle; In: Incisura

calibration phase, the bias level between observers was acceptable (Supplementary Material S2).

#### Statistical analysis

The normality of the data was confirmed using Q-Q plots and the Shapiro Wilks test. Agreement within and between observers and methods was analysed using Bland-Altman mean-difference plots.<sup>23,24</sup> The mean difference was reported as a measure of constant bias, whilst the 95% limits of agreement (95% LOA = mean difference  $\pm 1.96 \times$  standard deviation) of the differences ( $SD_{diff}$ ) was reported as a measure of intra- and interobserver variation. 95% LOA was expressed in the actual units of the measurements and as a percentage relative to the mean measurement, since there was a clear dependency of the measurement variation on the mean values. The limit for clinically acceptable agreement was informally set at a 95% LOA of 15%.<sup>25,26</sup>

Absolute reliability was also determined by standard error of measurement (SEM) and minimal detectable change at a 95% confidence interval ( $MDC_{95}$ ).<sup>27,28</sup>

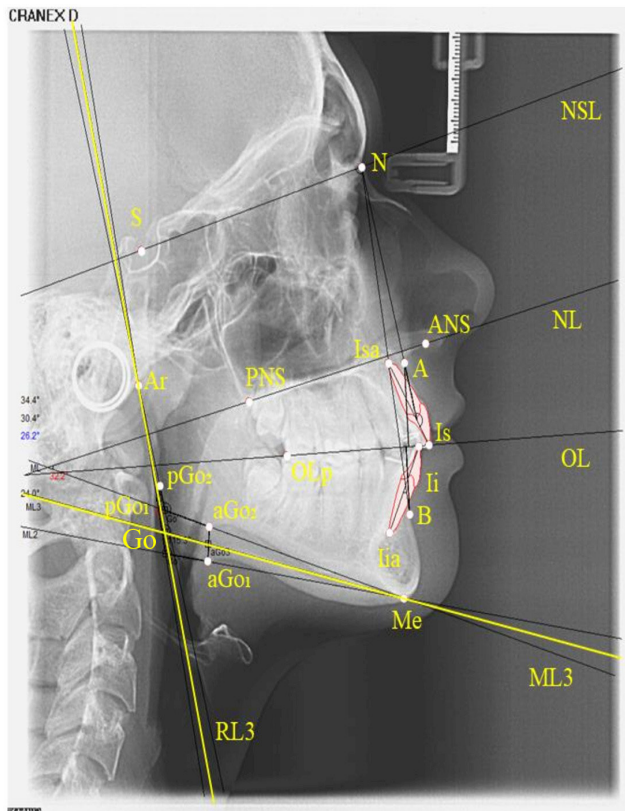
A one-way between-groups analysis of variance was conducted to explore the impact of CBCT machine type on the intraobserver variation of measurements of condylar length, right side.

All statistical analysis was performed using IBM SPSS version 26 (IBM, Chicago, IL). The level of statistical significance was set at 5% ( $p \leq .05$ ).

The NorJIA study was approved by the Regional Ethics Committee; REK nr 2012/542. Informed consents were given by the children if  $\geq 16$  years, and by the parents if the child were  $< 16$  years. Data were collected and stored according to the General Data Protection Regulation (GDPR).

#### Results

90 JIA patients (33 boys and 57 girls) were included, with a mean age of 12.8 years (range 4.9–16.3 years). Among the 90 cases, 39 (43.3%) were oligoarticular, 26 (28.9%) were polyarticular, 10 (11.1%) were enthesitis-related, 3



**Figure 3** Geometrical redesigning of Go: Go was defined as the intersection of RL3 and ML3. RL3 was the average of two lines drawn from the point Ar to the posterior border of the left and right ramus (pGo<sub>1</sub> and pGo<sub>2</sub>, respectively). Similarly, ML3 was the average of two lines drawn from the point Me to the lower border of the left and right mandible (aGo<sub>1</sub> and aGo<sub>2</sub>, respectively). Point A; ANS: Anterior Nasal Spine; Ar: Articulare; B: Point B; Me: Menton; N: Nasion; OLp: Occlusal line, posterior point; pGo1+2: Posterior gonion (posterior point on ramus); PNS: Posterior Nasal Spine. S: Sella; aGo1+2: Anterior gonion (lower border of mandible). Ii: Tip of the crown of the left/right first inferior incisor. IiA: Apex of the left/ right first inferior incisor. Is: Tip of the crown of the left /right first superior incisor. Isa: Apex of the left/right first superior incisor. ML3: Mandibular line; RL3: Ramus line; Go: Gonion angle.

(3.3%) were systemic, 3 (3.3%) were psoriasis-related, and 9 (10%) were undifferentiated types of JIA. The median duration of the disease was 5.0 years (IQR 6.2, 0.4–14.4). Two JIA patients had a poor head posture and/or a lack of maximal intercuspitation at the image acquisition.

#### MRI measurements

The intra- and interobserver 95% limits of agreement (LOA) were relatively narrow for total mandibular length (9.6% of the sample mean, right side) (Table 1 and Supplementary Material S3). For the posterior mandibular height, both intra- and interobserver 95% LOA were wide: 17.2 and 17.3% of the sample mean, respectively. The variation within and between observers was even higher for the condylar height, with 95% LOA of 55.4 and 34.8%, respectively. The intra- and

interobserver mean differences (bias) were low, ranging from 0.1 to 0.9 mm (2.0 and 18.4%) (Table 1). For individual subjects, a change in overall distance of at least 5.4 mm for mandibular length, 4.8 mm for posterior mandibular height and 5.4 mm for condylar height right side would have to be observed to confirm that a true change, beyond measurement error, has occurred (Table 1).

#### CBCT measurements

For the condylar height, both intra- and interobserver 95% LOA were wide: 16.0% and 28.4% of the sample mean, right side, respectively (Table 2). The intraobserver mean difference (bias) was zero (Supplementary Material S4), whilst the interobserver mean difference was 0.1 mm and 0.4 mm for right and left condyle, respectively (Table 3). MDC varied between 1.4 mm and 3.1 mm (Table 2).

#### Ceph measurements

For the SNA angle, both intra- and interobserver 95% LOA were narrow: 5.9 and 8% of the sample mean, respectively (Supplementary Material S5). The variation within and between observers was even lower for the SNB angle, with 95% LOA of 4.5 and 6%, respectively. The variation within and between observers was narrow for the gonion angle (RL3/ML3) (Supplementary Material S5), with 95% LOA of 6.2 and 6.8%, respectively. For the mandibular plane angle demonstrated by ML3/NSL, only the intraobserver 95% LOA was narrow with 13.4% (Supplementary Material S5). The remaining ceph-based measurements showed wide limits of agreement. The MDC varied between 1.4° and 7.0° (Table 3).

#### Condylar height; comparison between CBCT and MRI

Mean condylar height as measured by CBCT was 17.3 mm (SD 3.6), as compared to 19.3 mm (SD 3.6) by MRI (95% LOA = -1.3 to 5.3, right TMJ) in 52 patients who had both CBCT and MRI examinations at either baseline or at two years follow-up (Table 4).

A one-way between-groups analysis showed that the mean difference between observer one's first and second measurement was 0.01 mm for the CBCT machine in Bergen, vs 0.001 mm for Trondheim and 0.09 for Tromsø ( $p = 0.875$ ), implying that the 95% limits of agreement did not differ significantly across different CBCTs.

## Discussion

In their hierarchal model of efficacy in diagnostic imaging, Fryback and Thornbury pointed out that each level of efficacy is necessary but not sufficient to assure efficacy at higher levels such as diagnostic thinking.<sup>13</sup> Our study addresses the lower end of this hierarchy, namely the precision of measurements for mandibular morphology. We found little or no constant bias, but varying agreement within and between observers for

**Table 1** Intra- and interobserver agreement and MDC<sub>95</sub> of MRI-based measurements (mm) of total mandibular length, posterior mandibular length and condylar height in 78\* children and adolescents with JIA

	<b>OBS1 (Intraobserver) (first vs second measurements)</b>						<b>OBS2 (Interobserver) (OBS one vs OBS 2)</b>					
	first mean (SD)	second mean (SD)	Mean diff. (SD)	95% LOA	first LOA %	MDC <sub>95</sub> individ.	Mean (SD)	Mean diff. (SD)	95% LOA	LOA %	MDC <sub>95</sub> individ.	
Total mandibular length Rt	112.6 (8.4)	111.7 (9.0)	0.9 (2.7)	-4.5-6.3	9.6	5.4	112.6 (7.6)	-0.2 (2.7)	-4.9-5.4	9.6	5.4	
Total mandibular length Lt	112.9 (7.9)	112.4 (8.8)	0.5 (3.5)	-6.5-7.5	12.4	7.0	112.3 (7.8)	0.6 (2.9)	-5.6-6.0	10.3	5.8	
Posterior mandibular length Rt	55.3 (7.0)	55.8 (6.9)	0.5 (2.4)	-4.3-5.2	17.2	4.8	55.4 (6.3)	-0.1 (2.4)	-4.3-5.3	17.3	4.8	
Posterior mandibular length Lt	55.4 (6.4)	54.8 (6.5)	0.6 (2.3)	-4.0-5.2	16.6	4.6	55.4 (6.1)	0.0 (2.9)	-5.6-6.0	20.9	5.8	
Condylar height Rt	19.5 (3.5)	19.0 (3.4)	0.5 (2.7)	-4.9-5.9	55.4	5.4	19.8 (3.4)	-0.3 (1.7)	-2.8-4.0	34.3	3.4	
Condylar height Lt	19.0 (3.4)	19.1 (3.2)	-0.1 (2.4)	-4.6-5.0	50.5	4.8	19.8 (3.4)	-0.8 (2.2)	-3.5-5.3	44.4	4.4	

Diff: Difference. LOA: Limit of agreement. Lt: Left side. MDC<sub>95</sub>individ: Individual minimal detectable change. MRI: Magnetic resonance imaging. OBS: Observer. Rt: Right side. SD: Standard deviation.

\* 78\* 78 of the 90 children and adolescents included had MRIs available for these analyses.

a set of MRI, CBCT and ceph-based measurements commonly used for the assessment of mandibular morphology in JIA. The measurements with the highest test/retest agreement, were the ceph-based SNA, SNB and RL3/ML3, and the MRI-based total mandibular length, with LOA within 15% of the sample means.

Using SEM data, we were able to calculate the MDC at an individual level, providing a more clinically useful means of interpreting agreement. The MDC indicates the lower boundary for detectable change, whilst the MDC around the mean difference provides the LOA.<sup>29</sup> Thus, based on MDC from this study, we are 95% confident that differences lower than 4.6 to 7 mm for the MRI-measurements of both total mandibular and posterior

mandibular lengths are attributable to measurement error. In the evaluating of an intervention, one might argue that these MDC values, lying within 11% of the measured means but within 17% of the 95% LOA, are acceptable. This may well be the case, illustrating the difference between MDC and minimally clinically important differences. Thus, the a priori set limit for clinically acceptable agreement, e.g., an LOA of 15%, is adjustable and obviously depends on different clinical scenarios.

Measurement of the condylar height by CBCT had a suboptimal interobserver agreement, but acceptable intraobserver agreement and MDC, whilst the MRI-based measurement was too imprecise for clinical use,

**Table 2** Intra- and interobserver agreement and MDC<sub>95</sub> of CBCT-based measurements (mm) of condylar height in 73\* children and adolescents with JIA

	<b>OBS1 (Intraobserver) (first vs second measurements)</b>						<b>OBS2 (Interobserver) (OBS one vs OBS 2)</b>				
	first mean (SD)	second mean (SD)	Mean diff. (SD)	95% LOA	first LOA %	MDC <sub>95</sub> individ.	Mean (SD)	Mean diff. (SD)	95% LOA	LOA %	MDC <sub>95</sub> individ.
Condylar height Rt	17.5 (3.6)	17.5 (3.4)	0.0 (0.7)	-1.4-1.4	16.0	1.4	17.6 (3.7)	-0.1 (1.3)	-2.4-2.6	28.4	2.8
Condylar height Lt	18.1 (4.1)	17.9 (3.8)	0.2 (1.1)	-2.0-2.4	24.3	2.2	18.2 (3.7)	-0.4 (1.6)	-2.7-3.5	34.1	3.0

CBCT: Cone-beam computed tomography. Diff: Difference. LOA: 95% Level of agreement. Lt, Left side; MDC<sub>95</sub>individ, Individual minimal detectable change; OBS: Observer. Rt: Right side. SD: Standard deviation.

\* 73 of the 90 children and adolescents with JIA had CBCT examinations available for these analyses because the field of view (FOV) did not cover the structure of incisura mandibulae in 17 participants.

**Table 3** Intra- and interobserver agreement and MDC<sub>95</sub> of cephalometric measurements in 88\* children and adolescents with JIA

Angle (°), distance (mm)	OBS1 (Intraobserver) (first vs second measurements)						OBS2 (Interobserver measures) (OBS one vs OBS 2)				
	first mean (SD)	second mean (SD)	Mean diff. (SD)	95% LOA	first LOA %	MDC <sub>95</sub> individ.	Mean (SD)	Mean diff. (SD)	95% LOA	LOA %	MDC <sub>95</sub> individ.
SNA °	81.1 (3.6)	81.5 (3.7)	-0.4 (1.2)	-2.8–2.0	5.9	2.2	81.3 (3.2)	-0.2 (1.7)	-3.5–3.0	8	3.3
SNB °	78.2 (3.4)	78.6 (3.6)	-0.4 (0.9)	-2.2–1.3	4.5	1.4	78.2 (3.1)	0 (1.2)	-2.4–2.3	6	2.2
ANB °	2.9 (2.6)	2.8 (2.7)	0.1 (1.0)	-1.8–1.9	127.6	1.9	3.1 (2.4)	-0.2 (1.0)	-2.3–1.7	125.8	1.9
ML3-NSL °	32 (6.3)	31.4 (6.3)	0.6 (1.1)	-1.6–2.7	13.4	2.2	32.3 (6.3)	-0.3 (1.3)	-3.0–2.2	16.1	2.5
ML3-NL °	24.3 (5.6)	24.5 (5.8)	-0.2 (1.3)	-2.8–2.5	21.8	2.5	24.5 (5.9)	-0.2 (1.9)	-3.9–3.7	31.0	3.6
Wits-appraisal (mm)	-2.1 (3.1)	-2.3 (3.0)	0.2 (1.2)	-2.1–2.5	-219	2.2	-1.8 (3.2)	-0.3 (1.4)	-3.0–2.4	-300	2.8
ILsNA°	23.0 (6.8)	22.5 (7.3)	0.5 (2.1)	-3.6–4.6	35.7	4.2	21.5 (6.7)	1.5 (3.5)	-5.3–8.3	63.3	7.0
ILiNB°	26.5 (7.1)	26.8 (7.3)	-0.3 (1.9)	-4.0–3.4	27.9	3.6	26.9 (7.4)	-0.4 (2.7)	-5.8–5.0	40.1	5.4
RL3ML3°	123.4 (6.8)	123.1 (6.9)	0.3 (2.0)	-3.5–4.1	6.2	3.9	124 (6.7)	-0.6 (2.2)	-4.9–3.7	6.8	4.4

ANB, Nasion-AB; Diff, Difference; ILiNB, Inclination inferior incisors to NB line; ILsNA, Inclination superior incisors to NA line; LOA, Level of agreement; MDC<sub>95</sub>individ, Individual minimal detectable change; ML3-NL, Angle of mandibular line and palatal plane; ML3/NSL, Angle between mandibular line and cranial base; OBS, Observer; RL3ML3, Angle between posterior mandibular ramus line and the mandibular line; SD, Standard deviation; SNA, Sella-Nasion-A angle; SNB, Sella-Nasion-B angle.

\*of the 90 children had cephalometric examinations available for these analyses. two had poor head posture and/or a lack of maximal intercuspitation at the image acquisition.

with wide variation, both for the same and between observers. Of note is that the MRI-measurement of condylar height was higher than that obtained by CBCT, which is the opposite of what was reported in a study of eight cadaver skulls, comparing MRI, CBCT and radiographs.<sup>21</sup> This may in part be due to slightly different planes, measurement points and image qualities, such as slice thickness. For example, Markic et al used a temporomandibular surface coil, which provided higher resolution images than the head coil used in our 3T MRI scanner.<sup>21</sup> Further, Markic et al reported

that measurements of condylar height with MRI were comparable to those of CBCT in terms of intra- and interobserver agreement.<sup>21</sup> However, their 95% LOA were around 2mm for CBCT and 4mm for MRI for the same observer. Given a mean condylar height of around 18mm similar to a recent study, the 95% LOA for the CBCT-based condylar height can be estimated at 10% of the sample mean, as compared to 22% for MRI, suggesting that CBCT, but not MRI, has an acceptable precision for clinical use.<sup>21</sup> This compares well with our results, although our 95% LOA for MRI was even higher,

**Table 4** Comparison of mean condylar height (Co-In) between MRI and CBCT

Condylar height	OBS1 (Intraobserver) (first vs second measurements)			OBS2 (Interobserver) (OBS 2 MRI vs OBS 2 CBCT)					
	MRI n = 52	CBCT n = 52	Mean difference (95% LOA)	MRI n = 49	CBCT n = 49	Mean difference (95% LOA)	Mean (SD)	Mean (SD)	Mean difference (95% LOA)
Co-In Right (mm)	19.3 (3.6)	17.3 (3.6)	2.0 (-1.3–5.3)	18.9 (3.4)	17.3 (3.5)	1.6 (-7.4–10.6)	19.6 (3.6)	17.5 (3.8)	2.1 (-6.3–10.5)
Co-In Left (mm)	19.2 (3.3)	17.8 (3.6)	1.4 (-6.0–8.8)	19.2 (3.0)	17.5 (3.7)	1.7 (-5.7–9.1)	19.9 (3.6)	18.3 (4.1)	1.6 (-7.8–11.0)

CBCT: Cone-beam computed tomography. LOA: Level of agreement. MDC<sub>95</sub>individ: Individual minimal detectable change. MRI: Magnetic resonance imaging. OBS: Observer. SD: Standard deviation.

up to 55%. In another comparative study including 18 adults, with a mean age of 37.8 years, CBCT performed better than MRI with regard to intra- and interobserver variation for a set of direct measurements and angles, using the Mimics Research program.<sup>22</sup> The interobserver intraclass correlation (ICC) for mandibular body length was excellent (ICC = 0.95) for CBCT and only moderate (ICC = 0.74) for MRI. However, there was no information on the absolute reliability, such as the limits of agreement or MDC, thus, preventing a direct comparison with our results.<sup>22</sup>

High interobserver variations were found for condylar height based on both CBCT and MRI, and for MRI-measurements of posterior mandibular length and condylar height. The wide variation might be explained by the fact that we oriented the CBCT and MR volumes to reconstruct the multiplanar views prior to all measurements – *i.e.*, we identified the landmarks during each of the reading sessions. Previous 2D and 3D analyses have shown that condylar height represents one of the most critical measurements in assessing dentofacial growth deviation.<sup>30</sup> In their radiographic study, Kjellberg and colleagues found significantly shorter relative condylar height in 35 children and adolescents with JIA (aged 7–16 years) compared to their healthy peers; however, their results were based on condylar ratio and not on linear measurements.<sup>31</sup> In their reliability and validity study of 23 3D measurements, Stoustrup and colleagues identified, and highly recommended seven measures for the study of dentofacial growth in JIA.<sup>32</sup> In addition, they recommended several additional measures, including condylar height.

Studies have shown that identifying landmarks introduces errors that contribute to measurement inaccuracy.<sup>33</sup> Our results contrast with those of Ludlow *et al* in a study of 20 patients, which show that landmark identification with CBCT-MPR was accomplished with less variability than conventional ceph, implying that MPR-based measurements are more precise than measurements based on cephalograms.<sup>34</sup>

Similarly, some authors suggest that measuring directly on the 3D surface-rendered CBCT images introduces higher variability of certain landmarks – *e.g.*, in the mediolateral direction, probably related to the inadequate definition of the landmarks in the third dimension.<sup>35</sup> Baumrind, Broch and colleagues argued that cephalometric landmarks-based measurements such as edges are easier to localise, whereas landmarks placed on curves showed a higher measurement error.<sup>36,37</sup> Taken together, the body of studies published on the reliability of both ceph and CBCT is heterogeneous with respect to design and statistical analysis used; thus, the results are difficult to compare (Supplementary Material S6).

Several of the measurements obtained from conventional cephalograms, *i.e.*, the SNA, SNB, gonion angle (RL3/ML3), showed high precision and small MDC, which is a finding that has also been reported for these measurements obtained from MPRs derived from

volumetric CBCT scans.<sup>22,38</sup> Conversely, poor precision was found for the remainder of the measurements – for example, an intraobserver LOA as high as 127.6% of the measured mean for the ANB angle. The poor precision of ANB can, in part, be explained by its small value, but more crucial; point A is more challenging to locate than point B.<sup>39,40</sup> Moreover, numerous studies have shown that dental landmarks tend to have poorer validity than skeletal landmarks.<sup>41,42</sup> Kamoen *et al* addressed the high variability of landmark identification, and the cumulative effect of errors in a study of 50 cephs.<sup>43</sup> In our study, the clinical acceptance of the Wits appraisal is insufficient. This measurement is determined by perpendiculars from points A and B to the occlusal plane, and any change in the occlusal plane enhances the measurement error.<sup>44</sup> The clinical implication and the use of these variables to detect actual treatment effects can be questioned, but the literature reveals that the clinical significance is usually regarded as a difference of less than one or two measuring units.<sup>45</sup> Clinical relevance becomes more evident using the Bland-Altman approach when reporting on differences between observers and methods, rather than on relative reliability, such as the ICC or paired t-test.<sup>42,46</sup> Hitherto, few cephalometric studies and TMJ-imaging studies have tested the precision of CBCT and MRI-measurements applying the Bland-Altman mean-difference plots and 95% LOA (Supplementary Material S6). Thus, further studies addressing the accuracy of commonly used measures for morphological assessment of the mandibular complex in children and adolescents with JIA.

Our study has several limitations. Firstly, there is the subjective nature of identifying the landmarks with inherent biases in the reader's past experiences and understanding of the images. We endeavoured to overcome this by hosting several calibration sessions between all readers prior to scoring and analysis. Our study does not address the clinical validity of the measurements; however, this was not our intention, which was instead to primarily examine whether adding a 3D anatomical sequence to the routine MRI protocol for TMJ-imaging might provide precise measurements of dentofacial deformity, with the view to then assessing these for clinical validity.<sup>13</sup> Similarly, we intended to test the precision of one CBCT-based measurement derived from a routine examination with a small field of view, as well as commonly used cephalometric-based measurements. The strengths of this study include the reasonably high numbers, the thorough calibration process, and the multireader aspect of our data analysis.

## Conclusion

We have identified a set of precise radiological measurements for the assessment of dentofacial deformity in JIA. The measurements include one MRI-based, one CBCT-based and three ceph-based, in the hope that these can be helpful for studies that assess clinical



validity and long-term patient outcomes. MRI-based measurement of condylar height was higher than that obtained by CBCT; however, the measurement was too imprecise for clinical use. Moreover, we have determined the MDC for a set of measurements.

## Acknowledgment

This study is part of the multicentre NorJIA Study (The Norwegian JIA Study – Temporomandibular Involvement, Oral Health, Uveitis, Bone Health and Quality of Life in Children with Juvenile Idiopathic Arthritis (JIA)). NorJIA is a collaboration among universities (University of Bergen, Norwegian University of Science and Technology, The Arctic University of Norway), university hospitals (Haukeland University Hospital, St. Olav's Hospital, University Hospital of North Norway), and oral health centres (Oral Health Centre of Expertise in Western Norway-Vestland, Centre for Oral Health Services and Research, Mid-Norway, Public Dental Health Service Competence Centre of Northern Norway) in Bergen, Trondheim and Tromsø

represented by Karen Rosendahl MD PhD (PI), Marit Slättelid Skeie DDS PhD, Marite Rygg MD PhD, Ellen Nordal MD PhD, Anne N. Åstrøm DDS PhD, Karin Tylleskär MD, Annika Rosén DDS PhD, Elisabeth Grut Gil DDS, Johannes Maria Fischer DDS, Xieqi Shi DDS PhD, Oskar Angenete MD, Lena Cetrelli DDS, Gunnar Lyngstad DDS, Marie Sager DDS, Astrid J Feuerheim PhD, Anette Lundestad MD, Thomas Augdal MD, Paula Frid DDS, Veronika Rypdal MD, Josefine Halbig DDS, Athanasia Bletsas DDS PhD, Marit Midtbø DDS PhD, Larissa von Wangenheim Marti DDS and Mats Säll DDS. We are indebted to radiographers Marianne Lothe Vollan and Erik Haro and the study nurses Tone Kvinnsland Amdal, Susanne Irene Tobiesen Eidset, Line Rapp Simonsen, Marte Grimsmo Teige, Brita Lena Hansen, and Lisbeth Aune. Our gratitude also goes to all the participants to be included in the study. We would also like to thank biostatistician Stein Atle Lie PhD, for helpful comments on the article's statistical issues and Tore Larheim DDS PhD and Keijo Luukko DDS PhD, regarding consensus-based recommendations of anatomical landmarks.

## REFERENCES

- Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al. International league of associations for rheumatology classification of juvenile idiopathic arthritis: second revision, edmonton, 2001. *J Rheumatol* 2004; **31**: 390–92.
- Prakken B, Albani S, Martini A. Juvenile idiopathic arthritis. [Internet]. *Lancet* 2011; **377**: 2138–49. [https://doi.org/10.1016/S0140-6736\(11\)60244-4](https://doi.org/10.1016/S0140-6736(11)60244-4)
- Arvidsson LZ, Smith H-J, Flatø B, Larheim TA. Temporomandibular joint findings in adults with long-standing juvenile idiopathic arthritis: CT and MR imaging assessment. *Radiology* 2010; **256**: 191–200. <https://doi.org/10.1148/radiol.10091810>
- Weiss PF, Arabshahi B, Johnson A, Bilaniuk LT, Zarnow D, Cahill AM, et al. High prevalence of temporomandibular joint arthritis at disease onset in children with juvenile idiopathic arthritis, as detected by magnetic resonance imaging but not by ultrasound. *Arthritis Rheum* 2008; **58**: 1189–96. <https://doi.org/10.1002/art.23401>
- Glerup M, Stoustrup P, Matzen LH, Rypdal V, Nordal E, Frid P, et al. Longterm outcomes of temporomandibular joints in juvenile idiopathic arthritis: 17 years of followup of a nordic juvenile idiopathic arthritis cohort. *J Rheumatol* 2020; **47**: 730–38. <https://doi.org/10.3899/jrheum.190231>
- Billiau AD, Hu Y, Verdonck A, Carels C, Wouters C. Temporomandibular joint arthritis in juvenile idiopathic arthritis: prevalence, clinical and radiological signs, and relation to dentofacial morphology. *J Rheumatol* 2007; **34**: 1925–33.
- Cannizzaro E, Schroeder S, Müller LM, Kellenberger CJ, Saurenmann RK. Temporomandibular joint involvement in children with juvenile idiopathic arthritis. *J Rheumatol* 2011; **38**: 510–15. <https://doi.org/10.3899/jrheum.100325>
- Keller H, Müller LM, Markic G, Schraner T, Kellenberger CJ, Saurenmann RK. Is early TMJ involvement in children with juvenile idiopathic arthritis clinically detectable? clinical examination of the TMJ in comparison with contrast enhanced MRI in patients with juvenile idiopathic arthritis. *Pediatr Rheumatol Online J* 2015; **13**. <https://doi.org/10.1186/s12969-015-0056-2>
- Stoustrup P, Glerup M, Bilgrau AE, Küseler A, Verna C, Christensen AE, et al. Cumulative incidence of orofacial manifestations in early juvenile idiopathic arthritis: a regional, three-year cohort study. *Arthritis Care Res (Hoboken)* 2020; **72**: 907–16. <https://doi.org/10.1002/acr.23899>
- Fjeld M, Arvidsson L, Smith H-J, Flatø B, Ogaard B, Larheim T. Relationship between disease course in the temporomandibular joints and mandibular growth rotation in patients with juvenile idiopathic arthritis followed from childhood to adulthood. *Pediatr Rheumatol Online J* 2010; **8**. <https://doi.org/10.1186/1546-0096-8-13>
- Al-Okshi A, Lindh C, Salé H, Gunnarsson M, Rohlin M. Effective dose of cone beam CT (CBCT) of the facial skeleton: A systematic review. *Br J Radiol* 2015; **88**(1045). <https://doi.org/10.1259/bjr.20140658>
- Maspero C, Farronato M, Bellincioni F, Cavagnetto D, Abate A. Assessing mandibular body changes in growing subjects: a comparison of CBCT and reconstructed lateral cephalogram measurements. *Sci Rep* 2020; **10**(1). <https://doi.org/10.1038/s41598-020-68562-6>
- Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991; **11**: 88–94. <https://doi.org/10.1177/0272989X9101100203>
- deSouza NM, Achten E, Alberich-Bayarri A, Bamberg F, Boellaard R, Clément O, et al. Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the eiball\* subcommittee of the european society of radiology (esr). *Insights Imaging* 2019; **10**(1). <https://doi.org/10.1186/s13244-019-0764-0>
- Gribel BF, Gribel MN, Frazão DC, McNamara JA Jr, Manzi FR. Accuracy and reliability of craniometric measurements on lateral cephalometry and 3D measurements on CBCT scans. *Angle Orthod* 2011; **81**: 26–35. <https://doi.org/10.2319/032210-166.1>
- García-Sanz V, Bellot-Arcis C, Hernández V, Serrano-Sánchez P, Guarinos J, Paredes-Gallardo V. Accuracy and reliability of cone-beam computed tomography for linear and volumetric mandibular condyle measurements. A human cadaver study. *Sci Rep* 2017; **7**(1). <https://doi.org/10.1038/s41598-017-12100-4>

17. Rossini G, Cavallini C, Cassetta M, Barbato E. 3D cephalometric analysis obtained from computed tomography. review of the literature. *Ann Stomatol (Roma)* 2011; **2**: 31–39.
18. Sam A, Currie K, Oh H, Flores-Mir C, Lagravère-Vich M. Reliability of different three-dimensional cephalometric landmarks in cone-beam computed tomography: A systematic review. [Internet]. *Angle Orthod* 2019; **89**: 317–32. <https://doi.org/10.2319/042018-302.1>
19. Stoustrup P, Resnick CM, Pedersen TK, Abramowicz S, Michelotti A, Küsel A, et al. Standardizing terminology and assessment for orofacial conditions in juvenile idiopathic arthritis: international, multidisciplinary consensus-based recommendations. *J Rheumatol* 2019; **46**: 518–22. <https://doi.org/10.3899/jrheum.180785>
20. Hemke R, Herregods N, Jaremko JL, Åström G, Avenarius D, Becce F, et al. Imaging assessment of children presenting with suspected or known juvenile idiopathic arthritis: ESSR-ESPR points to consider. *Eur Radiol* 2020; **30**: 5237–49. <https://doi.org/10.1007/s00330-020-06807-8>
21. Markic G, Müller L, Patcas R, Roos M, Lochbühler N, Peltomäki T, et al. Assessing the length of the mandibular ramus and the condylar process: A comparison of OPG, CBCT, CT, MRI, and lateral cephalometric measurements. *Eur J Orthod* 2015; **37**: 13–21. <https://doi.org/10.1093/ejo/cju008>
22. Maspero C, Abate A, Bellincioni F, Cavagnetto D, Lanteri V, Costa A, et al. Comparison of a tridimensional cephalometric analysis performed on 3T-MRI compared with CBCT: a pilot study in adults. *Prog Orthod* 2019; **20**(1). <https://doi.org/10.1186/s40510-019-0293-x>
23. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 1990; **20**: 337–40. [https://doi.org/10.1016/0010-4825\(90\)90013-f](https://doi.org/10.1016/0010-4825(90)90013-f)
24. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol* 2003; **22**: 85–93. <https://doi.org/10.1002/uog.122>
25. Oehme NHB, Roelants M, Saervold Bruserud I, Madsen A, Eide GE, Bjercknes R, et al. Reference data for testicular volume measured with ultrasound and pubic hair in norwegian boys are comparable with northern european populations. *Acta Paediatr* 2020; **109**: 1612–19. <https://doi.org/10.1111/apa.15159>
26. Bruserud IS, Roelants M, Oehme NHB, Madsen A, Eide GE, Bjercknes R, et al. References for ultrasound staging of breast maturation, tanner breast staging, pubic hair, and menarche in norwegian girls. *J Clin Endocrinol Metab* 2020; **105**: 1599–1607. <https://doi.org/10.1210/clinem/dgaa107>
27. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; **8**: 135–60. <https://doi.org/10.1177/0962282029900800204>
28. Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: A practical guide. 2011; 1–338.
29. Ferri-Morales A, Alegre LM, Basco A, Aguado X. Test-retest relative and absolute reliability of knee extensor strength measures and minimal detectable change. *IES* 2014; **22**: 17–26. <https://doi.org/10.3233/IES-130522>
30. Stoustrup P, Küsel A, Kristensen KD, Herlin T, Pedersen TK. Orthopaedic splint treatment can reduce mandibular asymmetry caused by unilateral temporomandibular involvement in juvenile idiopathic arthritis. *Eur J Orthod* 2013; **35**: 191–98. <https://doi.org/10.1093/ejo/cjr116>
31. Kjellberg H, Ekkestubbe A, Kiliaridis S, Thilander B. Condylar height on panoramic radiographs. A methodologic study with A clinical application. *Acta Odontol Scand* 1994; **52**: 43–50. <https://doi.org/10.3109/00016359409096375>
32. Stoustrup P, Iversen CK, Kristensen KD, Resnick CM, Verna C, Nørholt SE, et al. Assessment of dentofacial growth deviation in juvenile idiopathic arthritis: reliability and validity of three-dimensional morphometric measures. *PLoS ONE* 2018; **13**(3). <https://doi.org/10.1371/journal.pone.0194177>
33. Lou L, Lagravère MO, Compton S, Major PW, Flores-Mir C. Accuracy of measurements and reliability of landmark identification with computed tomography (CT) techniques in the maxillofacial area: a systematic review. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2007; **104**: 402–11. <https://doi.org/10.1016/j.tripleo.2006.07.015>
34. Ludlow JB, Gubler M, Cevidanes L, Mol A. Precision of cephalometric landmark identification: cone-beam computed tomography vs conventional cephalometric views. *Am J Orthod Dentofacial Orthop* 2009; **136**: 312.. <https://doi.org/10.1016/j.ajodo.2008.12.018>
35. van Vlijmen OJC, Maal T, Bergé SJ, Bronkhorst EM, Katsaros C, Kuijpers-Jagtman AM. A comparison between 2D and 3D cephalometry on CBCT scans of human skulls. *Int J Oral Maxillofac Surg* 2010; **39**: 156–60. <https://doi.org/10.1016/j.ijom.2009.11.017>
36. Baumrind S, Frantz RC. The reliability of head film measurements. 1. landmark identification. *Am J Orthod* 1971; **60**: 111–27. [https://doi.org/10.1016/0002-9416\(71\)90028-5](https://doi.org/10.1016/0002-9416(71)90028-5)
37. Broch J, Slagsvold O, Røsler M. Error in landmark identification in lateral radiographic headplates. *Eur J Orthod* 1981; **3**: 9–13. <https://doi.org/10.1093/ejo/3.1.9>
38. Oz U, Orhan K, Abe N. Comparison of linear and angular measurements using two-dimensional conventional methods and three-dimensional cone beam CT images reconstructed from a volumetric rendering program in vivo. *Dentomaxillofac Radiol* 2011; **40**: 492–500. <https://doi.org/10.1259/dmfr/15644321>
39. Houston WJ, Maher RE, McElroy D, Sherriff M. Sources of error in measurements from cephalometric radiographs. *Eur J Orthod* 1986; **8**: 149–51. <https://doi.org/10.1093/ejo/8.3.149>
40. Freeman RS. Adjusting A-N-B angles to reflect the effect of maxillary position. *Angle Orthod* 1981; **51**: 162–71. [https://doi.org/10.1043/0003-3219\(1981\)051<0162:AAARTT>2.0.CO;2](https://doi.org/10.1043/0003-3219(1981)051<0162:AAARTT>2.0.CO;2)
41. Tng TTH, Chan TCK, Hägg U, Cooke MS. Validity of cephalometric landmarks. an experimental study on human skulls. *Eur J Orthod* 1994; **16**: 110–20. <https://doi.org/10.1093/ejo/16.2.110>
42. Damstra J, Huddleston Slater JJR, Fourie Z, Ren Y. Reliability and the smallest detectable differences of lateral cephalometric measurements. *Am J Orthod Dentofacial Orthop* 2010; **138**: 546.. <https://doi.org/10.1016/j.ajodo.2010.05.013>
43. Kamoen A, Dermaut L, Verbeeck R. The clinical significance of error measurement in the interpretation of treatment results. *Eur J Orthod* 2001; **23**: 569–78. <https://doi.org/10.1093/ejo/23.5.569>
44. Haynes S, Chau MNY. The reproducibility and repeatability of the wits analysis. *Am J Orthod Dentofacial Orthop* 1995; **107**: 640–47. [https://doi.org/10.1016/s0889-5406\(95\)70108-7](https://doi.org/10.1016/s0889-5406(95)70108-7)
45. Richardson A. A comparison of traditional and computerized methods of cephalometric analysis. *Eur J Orthod* 1981; **3**: 15–20. <https://doi.org/10.1093/ejo/3.1.15>
46. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003; **326**: 219. <https://doi.org/10.1136/bmj.326.7382.219>