

**Reliability of three-dimensional anterior cranial base superimposition methods for assessment of overall hard tissue changes:  
A systematic review**

**Cecilia Ponce-Garcia<sup>a</sup>; Manuel Lagravere-Vich<sup>b</sup>; Lucia Helena Soares Cevidanes<sup>c</sup>; Antonio Carlos de Olivera Ruellas<sup>d</sup>; Jason Carey<sup>e</sup>; Carlos Flores-Mir<sup>f</sup>**

**ABSTRACT**

**Objective:** The purpose of this systematic review was to synthesize the available literature concerning the reliability of three-dimensional superimposition methods when assessing changes in craniofacial hard tissues.

**Materials and Methods:** Four electronic databases were searched. Two authors independently reviewed potentially relevant articles for eligibility. Clinical trials, cohort, case-control, and cross-sectional studies that evaluated the reliability of three-dimensional superimposition methods on the anterior cranial base were included.

**Results:** Six studies fulfilled the inclusion criteria. Four studies used the voxel-based registration method, one used the landmark-based method and one used the surface-based method. Regarding reliability, the voxel-based studies showed on average a difference of 0.5 mm or less between images. The optimized analysis using a six-point correction algorithm in the landmark-based method showed 1.24 mm magnitude of error between images.

**Conclusions:** Although reliability appears to be adequate, the small sample size and high risk of bias among studies make available evidence still insufficient to draw strong conclusions. (*Angle Orthod.* 2018;88:233–245.)

**KEY WORDS:** Superimposition; Cone-beam computed tomography; Cranial base

**INTRODUCTION**

Superimposition of cephalometric headfilms taken at defined intervals is used by researchers and clinicians to

help in orthodontic diagnosis and treatment planning and to obtain a general view of growth changes and treatment outcomes in the dentofacial complex.<sup>1–3</sup> Conventional lateral cephalometric radiographs have proven to be an invaluable part of initial and final orthodontic records to quantify and determine craniofacial growth changes and effects of orthodontic treatment.<sup>1,4</sup> However, two-dimensional (2D) cephalometric radiographs suffer from a number of inherent flaws, such as errors generated because of inadequate patient head position, alignment of the imaging device, inherent geometric distortions, and differential magnification created by projection distance and beam divergence.<sup>1,5–9</sup>

During the past decade, craniofacial three-dimensional (3D) digital records have become increasingly popular among orthodontists as the specialty progressed toward a 3D virtual representation of the patient for diagnosis, treatment planning, and surgical simulation. The advanced imaging capabilities of cone-beam computed tomography (CBCT) are depicted through 3D cephalometric analysis, temporomandibular joint visualization, and 3D evaluation of dental anomalies, to name only a few.<sup>4,10</sup> A single scan

<sup>a</sup> Postgraduate Student, Division of Orthodontics, School of Dentistry, University of Alberta, Edmonton, Alberta, Canada.

<sup>b</sup> Assistant Professor, Division of Orthodontics, School of Dentistry, University of Alberta, Edmonton, Alberta, Canada.

<sup>c</sup> Assistant Professor, Department of Orthodontics, School of Dentistry, University of Michigan, Ann Arbor, Michigan, USA.

<sup>d</sup> Assistant Professor, Department of Orthodontics, School of Dentistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

<sup>e</sup> Professor and Associate Dean, Department of Mechanical Engineering, University of Alberta, Edmonton, Alberta, Canada.

<sup>f</sup> Department Head, Division of Orthodontics, School of Dentistry, University of Alberta, Edmonton, Alberta, Canada.

Corresponding author: Cecilia Ponce-Garcia, School of Dentistry, 11405 87Avenue, Edmonton, Alberta T6G 1C9, Canada  
(e-mail: poncegar@ualberta.ca)

Accepted: October 2017. Submitted: July 2017.

Published Online: November 30, 2017

© 2018 by The EH Angle Education and Research Foundation, Inc.

provides an overlap-free 3D visualization of different components of the skull, enables volumetric measurements to be made, and allows a detailed assessment of the maxillofacial structures in variable thickness of the axial, coronal, and sagittal slices, providing real measurements with no magnification.<sup>9,11</sup>

Recently, similar to 2D cephalometric tracings, CBCT images can be superimposed, allowing a 3D evaluation of growth changes, treatment effects, and stability over a certain time interval through registration points, angles, shapes, and volumes.<sup>12-14</sup> One of the main challenges of 3D superimposition of serial images is to understand that linear/angular measurements in 2D and 3D images are not directly comparable because of differences in size, shape, and relative spatial location of the skeletal, dental, and soft tissue between the two imaging systems.<sup>4,15</sup>

The following three general methods of 3D cephalometric superimposition have been published and used for clinical diagnosis and assessment of orthodontic treatment outcomes: (1) voxel based,<sup>3,4,16-19</sup> (2) landmark based,<sup>7,20</sup> and (3) surface based.<sup>4,21</sup>

A review addressing the 3D CBCT superimposition methods was published in 2015.<sup>4</sup> Although it discussed the three main techniques, it focused mainly on their clinical applications, benefits, and limitations. It did not consider the measuring capabilities of any of those methods. No systematic review has been specifically conducted to investigate the reliability of these 3D superimposition methods when assessing changes in craniofacial hard tissues. Without an in-depth understanding of the measurement properties of each method, indiscriminate use should be questioned, as treatment decisions/assessments may not have been based on sound superimposition evidence.

The purpose of this systematic review was to synthesize the available literature concerning the reliability of 3D superimposition methods to evaluate craniofacial hard tissues changes.

## MATERIALS AND METHODS

This systematic review followed, whenever applicable, the Preferred Reporting Items for Systematic Reviews and Meta-Analysis checklist.<sup>22</sup>

### Protocol and Registration

The study protocol was not registered in advance.

### Eligibility Criteria

The following selection criteria were applied for the review:

- Study design: clinical trials, cohort, case-control, and cross-sectional studies that evaluated the reliability,

repeatability, or reproducibility of 3D superimposition methods on the anterior cranial base were included. No restrictions were applied regarding language or year of publication.

- Exclusion criteria: review articles, meeting abstracts, book chapters, case reports, editorial letters, and personal opinions were excluded from the review.

### Information Sources and Search Strategy

A systematic search of four electronic databases (Embase, Medline via OVID, Web of Science, and SCOPUS) was performed. All searches were inclusive until December 2016. The search strategy was designed with the assistance of a health science senior librarian. Appropriate truncation and word combinations were selected and adjusted for each database search. Keywords used in the search and combination of terms per database can be found in Table 1.

### Study Selection

The relevant articles were selected through a two-phase process. In phase 1, two authors (CPG and MLV) independently reviewed the titles and abstracts of all references. In phase 2, full texts of potentially relevant abstracts were retrieved, reviewed, and screened by the same two reviewers according to the same selection criteria to confirm final selection while considering the full manuscript. Any disagreement was settled by means of discussion until a mutual consensus was reached.

### Data Collection Process and Data Items

Data were extracted from each of the selected studies using a developed standardized data collection form based on the Cochrane Consumers and Communication Review.<sup>23</sup> One reviewer (CPG) collected the required information from the selected articles. The second reviewer (MLV) cross-checked the gathered data and confirmed its accuracy. Once again, any disagreement in either phase was resolved by consensus.

### Risk of Bias in Individual Studies

The Consensus-Based Standards for the Selection of Health Measurement Instruments checklist, a standardized tool for assessing the methodological quality of studies that evaluates measurement properties, was used for quality assessment of included studies.<sup>24</sup> Disagreements between the reviewers in relation to quality assessment were resolved by means of discussion, and the third reviewer (CFM) made a

**Table 1.** Database Search Strategy

Database	Search Strategy
WEB OF SCIENCE (1945–August 2016)	(#1 TOPIC: cone beam computed tomography OR #2 TOPIC: CBCT) AND ((#3 TOPIC: three-dimensional cephalometric superimposition OR #4 TOPIC: three dimensional cephalometric superimposition OR #5 TOPIC: 3-dimensional cephalometric superimposition OR #6 TOPIC: 3D cephalometric superimposition) OR (#7 TOPIC: three-dimensional superimposition OR #8 TOPIC: three dimensional superimposition OR #9 TOPIC: 3-dimensional superimposition OR #10 TOPIC: 3D superimposition) OR (#11 TOPIC: craniofacial three-dimensional superimposition OR #12 TOPIC: craniofacial three dimensional superimposition OR #13 TOPIC: craniofacial 3-dimensional superimposition OR #14 TOPIC: craniofacial 3D superimposition) OR (#15 TOPIC: three-dimensional superimposition methods OR #16 TOPIC: three dimensional superimposition methods OR #17 TOPIC: 3-dimensional superimposition methods OR #18 TOPIC: 3D superimposition methods) OR (#19 TOPIC: three-dimensional cephalometric superimposition methods OR #20 TOPIC: three dimensional cephalometric superimposition methods OR #21 TOPIC: 3-dimensional cephalometric superimposition methods OR #22 TOPIC: 3D cephalometric superimposition methods))
EMBASE (1974–August 2016)	(cone beam computer tomography.mp. OR craniofacial three dimensional imaging.mp. OR CBCT) AND (3D cephalometric superimposition.mp. OR three-dimensional cephalometric superimposition.mp. OR three dimensional cephalometric superimposition.mp. OR three-dimensional superimposition.mp. OR three dimensional superimposition.mp. 3D superimposition OR craniofacial superimposition.mp.)
MEDLINE (1966–August 2016)	(cone beam computer tomography.mp. OR craniofacial three dimensional imaging.mp. OR CBCT) AND (3D cephalometric superimposition.mp. OR three-dimensional cephalometric superimposition.mp. OR three dimensional cephalometric superimposition.mp. OR three-dimensional superimposition.mp. OR three dimensional superimposition.mp. 3D superimposition OR craniofacial superimposition.mp.)
SCOPUS (until August 2016)	((TITLE-ABS-KEY (cone beam computed tomography)) OR (TITLE-ABS-KEY (cbct))) AND (((TITLE-ABS-KEY (3d cephalometric superimposition)) OR (TITLE-ABS-KEY (three dimensional cephalometric superimposition)) OR (TITLE-ABS-KEY (three-dimensional cephalometric superimposition))) OR ((TITLE-ABS-KEY (3d superimposition)) OR (TITLE-ABS-KEY (three dimensional superimposition)) OR (TITLE-ABS-KEY (three-dimensional superimposition))) OR ((TITLE-ABS-KEY (three-dimensional superimposition methods)) OR (TITLE-ABS-KEY (three-dimensional cephalometric superimposition methods)) OR (TITLE-ABS-KEY (three dimensional cephalometric superimposition methods)) OR (TITLE-ABS-KEY (3d cephalometric superimposition methods)) OR (TITLE-ABS-KEY ( 3d superimposition methods)) OR (TITLE-ABS-KEY (three dimensional superimposition methods))) OR ((TITLE-ABS-KEY (craniofacial three dimensional superimposition)) OR (TITLE-ABS-KEY (craniofacial three-dimensional superimposition)) OR (TITLE-ABS-KEY (craniofacial 3d superimposition))))
Limitation: human subjects	

final decision if consensus was not reached by the first 2 reviewers.

**Synthesis of Studies**

As a result of the nature of the question and the available data, a meta-analysis was not possible. Included studies assessed reliability of measurements from different craniofacial anatomical regions.

**RESULTS**

**Study Selection**

A flow chart of the selection process of articles included in this study is outlined in Figure 1. A total of 254 manuscripts were selected for a phase 1 assessment. Thereafter a total of 219 studies were excluded following abstract/title assessment. Only 35 references and 1 additional study found through a manual search (from the reference list) were subsequently selected and received a full-text reading (phase 2). From the total full-text articles retrieved and reviewed, 30 studies were later excluded; there-

fore, only 6 studies fulfilled the criteria to be included in this review.

**Study Characteristics**

All of the selected studies for the qualitative synthesis were published in English. Sample size ranged from 3 to 18 patients and included adults undergoing orthognathic surgery and children undergoing either orthopedic treatment with miniplates for Class III correction or rapid maxillary expansion. To assess treatment or growth changes, all of the studies used pre- and postimages to apply the 3D superimposition method. A large field of view was used in all of the studies, and voxel size ranged from 0.25 to 0.8. An abbreviated summary of the descriptive characteristics of the included articles is provided in Table 2.

**Risk of Bias Within Studies**

The methodological quality scores based on the Consensus-Based Standards for the Selection of Health Measurement Instruments checklist when assessing reliability were evaluated to have a high risk of

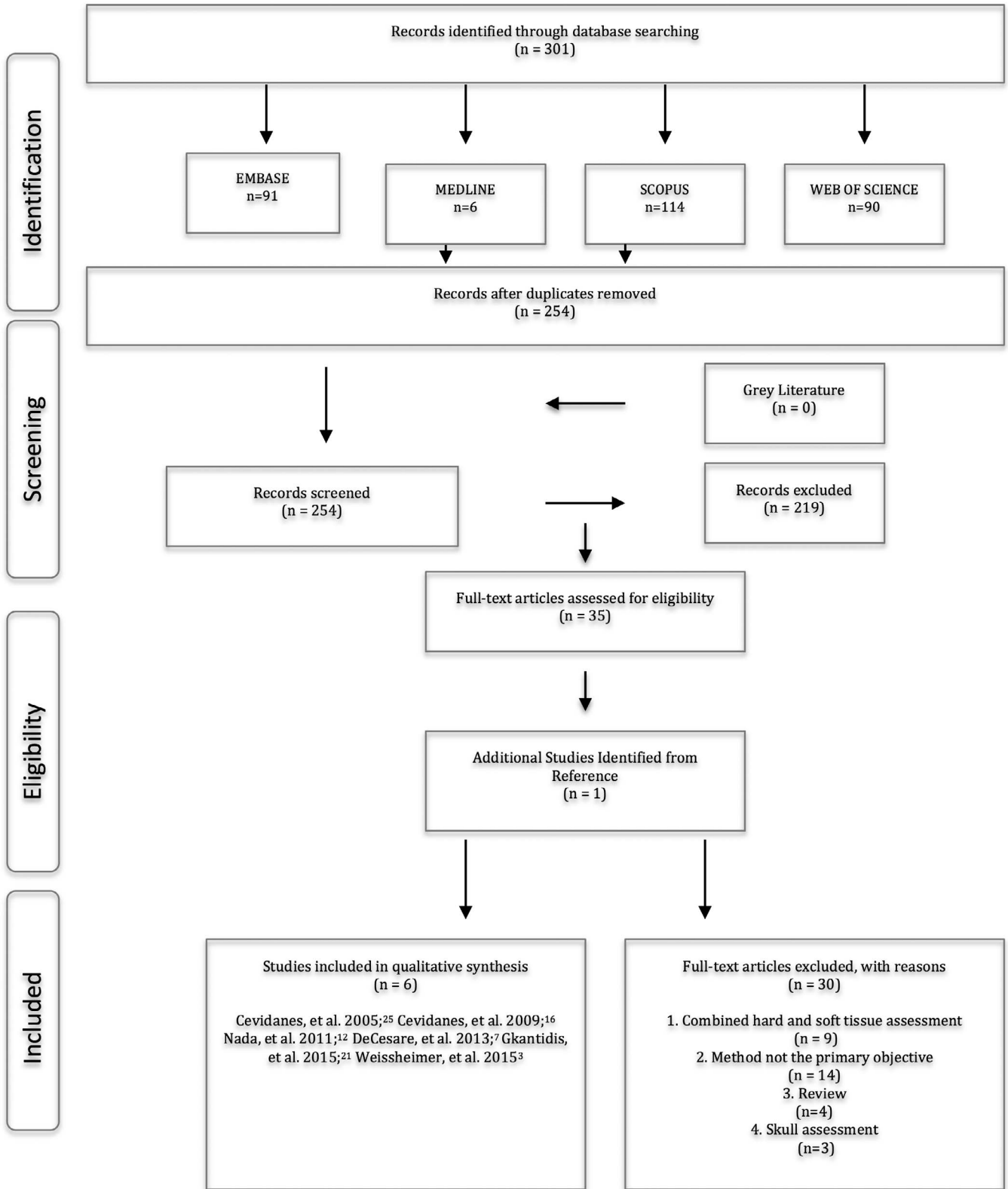


Figure 1. Flow diagram of the data search using the PRISMA guidelines.

bias (poor methodological quality) in all of the included studies. Similarly, when evaluating measurement error and validity, five studies scored poor methodological quality except for one study on each measurement property that obtained fair methodological quality. The critical appraisal details about each of the items and the evaluation criteria are described in Tables 3 to 6.

### Results of Individual Studies

For better interpretation, the results were separated according to the superimposition techniques: voxel based, landmark based, and surface based. Pre- and posttreatment images were registered on the anterior cranial base surface on all studies,<sup>3,7,12,16,21,25</sup> and one was also superimposed on the left zygomatic arch.<sup>12</sup>

#### Voxel-Based Method

Four studies tested this method. A first study carried out by Cevidane et al.<sup>25</sup> assessed interobserver reproducibility in a subset of 10 CBCT scans (before and after treatment) of five patients undergoing orthognathic surgery using three observers. They showed the similarity between the 3D color-coded maps and that pre- to postsurgery surface distance measurements differed among the three observers by no more than 0.26 mm (maximal error measured as displacement at the mandibular rami surface). The average inward displacement for all surfaces (mandibular rami, posterior border of the mandibular ramus, and condyles) was smaller than the image spatial resolution of 0.6 mm. The one-sample *t*-test *P* values were statistically significant at all surfaces, despite the small values of displacements that were observed.

A second study<sup>12</sup> assessed the voxel-based method on pairs of CBCT scans of 16 adult patients. The mean absolute distances between the two 3D images were calculated in four different regions (cranial base, forehead, and right and left zygomatic arches). The results showed small interobserver variability when the 3D model construction and superimposition procedure was repeated by a second observer. Mean differences between superimpositions performed by the first and second observer were 0.01 mm (95% confidence interval [CI], 0.03–0.05) for the forehead region, –0.07 mm (95%CI, 0.13 to –0.003) for the right zygomatic arch, and –0.01 mm (95%CI, –0.09 to 0.07) for the left zygomatic arch. The correlation coefficient between the repeated superimpositions (intraobserver repeatability) ranged from 0.53 to 0.94.

A third study<sup>3</sup> that used a sample of CBCT images of 18 patients also assessed the voxel-based method. A total of 10 patients were used as a reference standard, reorienting the spatial position of the pretreatment

CBCT volume and then superimposing on the original image. The other eight patients (four nongrowing and four growing) had pre- and posttreatment superimposed images. The results showed that the surface distance error was less than 0.25 mm for the sample that tested the superimpositions of CBCT images with a 1-year interval for growing patients treated with rapid maxillary expansion. Similarly, the adult sample, which underwent orthognathic surgery, revealed discrepancies in the anterior cranial base between the registered surface models that were less than 0.5 mm for most regions.

A fourth study<sup>16</sup> assessed the voxel-based method in growing patients. Three observers were trained for analysis of CBCT images using two images not included in the study. After calibration, each observer examined pre- and posttreatment CBCT scans of three growing patients. The interexaminer range of the measurements across anatomic regions was equal to or less than 0.5 mm.

#### Landmark-Based Registration

DeCesare et al.<sup>7</sup> assessed the six-landmark superimposition method when defining the coordinate system using data from 10 growing patients. The error reported was the absolute value of the difference in distances between the points calculated for the first image and the second image. The results showed high intertest reproducibility and great consistency between trials. The average error seen in the distances between the first image and the second image was 1.24 mm.

#### Surface-Based Registration

Gkantidis et al.<sup>21</sup> tested five surface-based 3D superimposition techniques in a sample of eight nongrowing orthodontic patients treated with rapid maxillary expansion (three-point registration, anterior cranial base, anterior cranial base + foramen magnum, both zygomatic arches, one zygomatic arch). The results showed that all of the techniques differed from each other ( $P < .005$ ), except for the anterior cranial base and both zygomatic arches superimpositions ( $P = .43$ ) using CT scans. The anterior cranial base + foramen magnum was the most accurate technique ( $P = .07$ ). The reproducibility and precision of all the techniques were acceptable because there were no significant differences between the repeated measurements and among examiners on the measured structural changes (anterior cranial base + foramen magnum: examiner 1, 0.11 mm [95%CI, 0.09–0.17]; examiner 2, 0.07 mm [95%CI, 0.04–0.09]; examiner 3, 0.09 mm [95%CI, 0.04–0.14]).

**Table 2.** Summary of Descriptive Characteristics of Finally Selected Studies<sup>a</sup>

Author/Year	Aim of Study	Participant Demographics and Characteristics		Method Details
		Age	Sample Size/Study Population/ Time Between Scan	Imaging System Used to Acquired Scan
Cevitanes et al. 2005 <sup>25</sup>	Evaluate the registration of 3D models from CBCT images taken before and after orthognathic surgery for the assessment of mandibular anatomy and position	20.6 ± 5.2 years	10 patients/undergoing maxillary surgery at the University of North Carolina Dentofacial Deformities Program/scans taken before and 1 week after orthognathic surgery	NewTom 9000
Cevitanes et al. 2009 <sup>16</sup>	Evaluate a new method for superimposition of 3D models of growing patients	11.4 years	Three patients/undergoing orthopedic treatment with miniplates/scans were taken before and after treatment (about 1 year follow-up)	i-CAT 3D Imaging System
Nada et al. 2011 <sup>12</sup>	Evaluate accuracy and reproducibility of a semiautomated voxel based image registration technique for the superimposition of 3D CBCT models on two different regions, the anterior cranial base and the zygomatic arches	26 ± 9 years	16 adult patients/underwent two surgical interventions at Radboud University Nijmegen Medical Centre/scans were taken prior to treatment and before the second orthognathic surgery—average 18 (±4.6) months later	i-CAT 3D Imaging System
DeCesare et al. 2013 <sup>7</sup>	Minimize errors that occur when using a four vs six-landmark superimpositioning method in the cranial base to define the coordinate system	Not mentioned	10 patients/participated in a maxillary expansion clinical trial/scans were taken 12 months apart	NewTom 3G
Gkantidis et al. 2015 <sup>21</sup>	Test the applicability, accuracy, precision, and reproducibility of various 3D superimposition techniques for radiographic data, transformed to triangulated surface data	16.2 (range: 15.1, 22.9) years	Eight young adult patients/treated with rapid maxillary expansion performed by a mini-implant supported device/scans were acquired just before placement of the appliance and at the end of the activation period at a median of 15 days later	Philips Brilliance 16 CT Scanner
Weissheimer et al. 2015 <sup>3</sup>	The aim of this study was to validate a method for fast 3D superimposition of CBCT in growing patients and adults (surgical cases)	11.4 ± 1 years (pretreatment scans) 26.3 ± 5.7 years (nongrowing adults) 9.5 ± 1.8 years (growing patients)	18 patients total. 10 patients, saved as a reoriented volume, and then superimposed on the original image. Four nongrowing and four growing/scans were taken pre- and 1 year postorthognathic surgery for adult sample and post-rapid palatal expansion for growing sample	i-CAT 3D Imaging System

**Table 2.** Extended

Field of View and Voxel Size	Software Used	Method Details			Main Results
		Superimposition Method	Time to Complete Superimposition	Statistics Analysis	Outcomes ( <i>P</i> value) or (OR; 95%CI)
FOV: 23 × 23 cm Voxel size: 0.58 × 0.58 × 0.6 mm	Insight SNAP: regional semiautomatic segmentation MIRIT: computing fully automated registration VALMET: comparison of 3D models	Voxel based	Not mentioned	Interobserver reliability	Mandibular rami surface: Inward (0.26 SD 0.12) Outward (0.24 SD 0.10) Posterior border of the mandibular ramus: Inward (0.15 SD 0.07) Outward (0.11 SD 0.04) Condyles: Inward (0.05 SD 0.02) Outward (0.13 SD 0.08) <i>P</i> value < .001
FOV: 22 × 16 cm Voxel size: 0.5 mm	ITK-SNAP: creating 3D surface models Imagine: computing the rigid registration CFM: overlaying 3D surface models that are registered in the same coordinate system	Voxel based	Not mentioned	Not mentioned	Interexaminer range of measurement across anatomic regions was equal or less than 0.5 mm, which are clinically insignificant
FOV: 22 × 16 cm Voxel size: 0.4 mm	MAXILIM	Voxel based	30–40 minutes	Intraobserver and interobserver reliability using the Pearson correlation coefficient for the mean distances at four anatomical regions following the first and second superimpositions. Paired-sample <i>t</i> -test to compare the means of corresponding measurements following registration on the anterior cranial base and the left zygomatic arch. The significance level set at 5%	The correlation coefficients registered on the anterior cranial base ranged between 0.53 and 0.94 for the mean distances at the four regions and between 0.24 and 0.71 for the mean distances at the four anatomic regions registered on the zygomatic arches The <i>P</i> values ranged between .001 and .025 and were statistically significant for the four regions
FOV: 12" Voxel size: 0.25 mm	AVIZO	Point based	Not mentioned	Intrareliability values were determined using intraclass correlation coefficient for all four landmarks, repeating the process three times for each image	When analyzing real patient data, it was found that the six-point correction algorithm reduced errors between images and increased intrapoint reliability. This method demonstrated greater reliability and reproducibility than the previous four-point correction algorithm
FOV: 21 × 21 × 12 cm Voxel size: 0.8 mm	OSIRIX	Surface based	25 minutes	Differences in the measured variables were evaluated using permutational multivariate analysis of variance (MANOVA) with factorial mixed effects models. In all cases, a two-sided significance test was carried out at an alpha level of 0.05. The level of significance used for the study was set at 0.05. Bonferroni correction was applied for pairwise a posteriori multiple comparison tests	Pairwise a posteriori tests between superimposition techniques showed that all techniques differed from each other ( <i>P</i> < .005). The AC + F technique was the most accurate (D < 0.17 mm), as expected, followed by AC and BZ superimpositions that presented similar level of accuracy (D < 0.5 mm)
FOV: Large Voxel size: 0.25 mm	OnDemand3D: automatic voxel-based rigid registration ITK-SNAP: automatic segmentation STL to SGI Inventor 2.0: convert files from STL to IV CMF app: provided closest point color maps between registered 3D surface models	Voxel based	10–15 seconds	Not mentioned	The quantification of the superimposition errors by color-coded surface distances revealed that distances in the anterior cranial base between registered surface models were less than 0.5 mm for most regions for both growing patients and adults

**Table 2.** Continued.

<sup>a</sup> OR indicates odds ratio; CI, confidence interval; SD, standard deviation; CBCT, cone-beam computed tomography; 3D, three-dimensional. ITK Software, National Library of Medicine Insight Segmentation and Registration Toolkit; ITK-SNAP Software, open software interactive tool for semi-automatic segmentation of multi-modality biomedical images; MIRIT Software, multi-modality image registration by maximization of mutual information; VALMET Software, tool for validation and comparison of object segmentation; AVIZO Software, image analysis software package designed to work with 3D images—whether data from confocal or CT, or 3D rendered images; MAXILIM Software, Medicim-Medical Image Computing; CFM Software, developed at Maurice Muller Institute, Bern, Switzerland - overlays 3D surface models that are registered in the same coordinate system; FOV, Field of View; STL, stereolithography file; SGI, IV, open inventor file; CMF Software, Cranio-Maxillo Facial application software, developed at the M. E. Muller Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland.

**Synthesis of Results**

All of the included studies reported adequate reliability of all 3D superimposition methods. Nevertheless, the quality of the studies was consistently poor. Hence it is unknown how the poor quality of evidence influenced the results.

**Risk of Bias Across Studies**

The main methodological limitations across the studies were related to small sample size, different age groups, treatment type, flaws in the study design, and the lack of a detailed description of statistical analysis.

**Table 3.** COSMIN Risk of Bias Assessment—Reliability<sup>a</sup>

Box B. Reliability: Relative Measures (Including Test-Retest Reliability, Interrater Reliability, and Intrarater Reliability)	Cevidanes et al. <sup>25</sup>				Cevidanes et al. <sup>16</sup>				Nada et al. <sup>12</sup>				DeCesare et al. <sup>7</sup>				Gkantidis et al. <sup>21</sup>				Weissheimer et al. <sup>3</sup>			
	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P
<b>Design requirements</b>																								
Was the percentage of missing items given?			x				x				x				x				x				x	
Was there a description of how missing items were handled?			x				x				x				x				x				x	
Was the sample size included in the analysis adequate?				x				x			x					x				x				x
Were at least two measurements available?	x					x				x				x				x						x
Were the administrations independent?	x					x				x				x				x						x
Was the time interval stated?			x				x			x					x			x						x
Were patients stable in the interim period on the construct to be measured?			x				x				x				x				x					x
Was the time interval appropriate?			x				x			x					x			x						x
Were the test conditions similar for both measurements (e.g., type of administration, environment, instructions)?			x				x			x					x				x					x
Were there any important flaws in the design or methods of the study?				x				x			x					x			x					x
<b>Statistical methods</b>																								
For continuous scores: Was an intra class correlation coefficient calculated?				x				x	x							x	x							x
For dichotomous/nominal/ordinal scores: Was kappa calculated?				x				x			x					x				x				x
For ordinal scores: Was a weighted kappa calculated?				x				x			x					x				x				x
For ordinal scores: Was the weighting scheme described (e.g., linear, quadratic)?				x				x			x					x	x							x
Score	Poor				Poor				Poor				Poor				Poor							

<sup>a</sup> COSMIN box with four-point scale for methodological quality: E, excellent; G, good; F, fair; P, poor. A methodological quality score per box was obtained by taking the lowest rating of any item in a box (“worst score counts”). A poor score on any item was thus considered to represent a fatal flaw. In the scoring system, items 1 and 2 (on the number of missing items and how missing items are handled) were scored less strictly than the other items as this information is often not reported in articles. In all boxes, a small sample size was considered poor methodological quality.<sup>24</sup> COSMIN indicates Consensus-Based Standards for the Selection of Health Measurement Instruments.



**Table 4.** COSMIN Risk of Bias Assessment–Measurement Error<sup>a</sup>

Box C. Measurement Error: Absolute Measures	Cevidanes et al. <sup>25</sup>				Cevidanes et al. <sup>16</sup>				Nada et al. <sup>12</sup>				DeCesare et al. <sup>7</sup>				Gkantidis et al. <sup>21</sup>				Weissheimer et al. <sup>3</sup>			
	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P
<b>Design requirements</b>																								
Was the percentage of missing items given?			x				x				x				x				x				x	
Was there a description of how missing items were handled?			x				x				x				x				x				x	
Was the sample size included in the analysis adequate?				x				x			x					x				x			x	
Were at least two measurements available?	x					x				x				x				x					x	
Were the administrations independent?	x					x				x				x				x						x
Was the time interval stated?			x				x			x					x			x					x	
Were patients stable in the interim period on the construct to be measured?			x				x				x				x				x				x	
Was the time interval appropriate?	x					x				x				x				x					x	
Were the test conditions similar for both measurements (e.g., type of administration, environment, instructions)?			x				x				x				x				x				x	
Were there any important flaws in the design or methods of the study?				x				x			x					x			x					x
For CTT: Was the standard error of measurement, smallest detectable change, or limits of agreement calculated?			x				x				x				x				x				x	
Score	Poor				Poor				Fair				Poor				Poor							

<sup>a</sup> COSMIN box with four-point scale for methodological quality: E, excellent; G, good; F, fair; P, poor. A methodological quality score per box was obtained by taking the lowest rating of any item in a box (“worst score counts”). A poor score on any item was thus considered to represent a fatal flaw. In the scoring system, items 1 and 2 (on the number of missing items and how missing items are handled) were scored less strictly than the other items as this information is often not reported in articles. In all boxes, a small sample size was considered poor methodological quality.<sup>24</sup> COSMIN indicates Consensus-Based Standards for the Selection of Health Measurement Instruments. CTT, classical test theory.

**Additional Analysis**

All of the articles used different anatomical regions to assess reliability and measurement error; this made the application of a meta-analysis questionable.

**DISCUSSION**

**Summary of Evidence**

In this systematic review, the available evidence concerning the reliability of the 3D superimposition methods when assessing the changes in the craniofacial hard tissues was investigated. Although all of the included studies for all three methods in this review reported acceptable reliability, the quality of evidence was low. Therefore, any reported conclusions are not to be supported with a high level of certainty.

CBCT is currently a well-established diagnostic tool for the 3D assessment of growth and/or treatment changes on craniofacial structures. However, it is

important to understand that challenges remain because 3D superimposition is much more complicated than 2D superimposition. The difficulties assessing the reliability of 3D superimpositions are not only a result of registration issues but also a result of the choice of regions to test the reproducibility of the superimposition, with landmark locations on various anatomic surfaces in the three planes of space.<sup>26</sup>

To be suitable for routine application in medical image processing, a superimposition method should be able to register precisely and aid understanding of the changes a result of growth or treatment relative to the structures of reference. The image analysis procedures include 3D construction, registration, superimposition, and quantification of changes.

In the voxel-based registration method, all of the steps involved are automated, which may allow image analysis procedures independent of observer errors. The application of this method has been widely

**Table 5.** COSMIN Risk of Bias Assessment–Validity<sup>a</sup>

Box H. Criterion Validity	Cevidanes et al. <sup>25</sup>				Cevidanes et al. <sup>16</sup>				Nada et al. <sup>12</sup>				DeCesare et al. <sup>7</sup>				Gkantidis et al. <sup>21</sup>				Weissheimer et al. <sup>3</sup>			
	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P
<b>Design requirements</b>																								
Was the percentage of missing items given?			x				x				x				x				x				x	
Was there a description of how missing items were handled?		x				x				x				x				x				x		
Was the sample size included in the analysis adequate?				x				x	x							x				x				x
Can the criterion used or employed be considered as a reasonable “gold standard”?				x				x				x				x				x				x
Were there any important flaws in the design or methods of the study?			x				x				x				x				x				x	
<b>Statistical methods</b>																								
For continuous scores: Were correlations, or the area under the receiver-operating curve calculated?			x				x				x				x				x				x	
For dichotomous scores: Were sensitivity and specificity determined?			x				x				x				x				x				x	
Score	Poor				Poor				Poor				Poor				Poor				Fair			

<sup>a</sup> COSMIN box with four-point scale for methodological quality: E, excellent; G, good; F, fair; P, poor. A methodological quality score per box was obtained by taking the lowest rating of any item in a box (“worst score counts”). A poor score on any item was thus considered to represent a fatal flaw. In the scoring system, items 1 and 2 (on the number of missing items and how missing items are handled) were scored less strictly than the other items as this information is often not reported in articles. In all boxes, a small sample size was considered poor methodological quality.<sup>24</sup> COSMIN indicates Consensus-Based Standards for the Selection of Health Measurement Instruments.

described in the literature to assess changes after orthognathic surgery and orthopedic treatment.<sup>2,17,27–29</sup> Cevidanes et al.<sup>25</sup> introduced this method into dentistry. This first study used this method to assess mandibular anatomy and position before and after maxillary advancement. They applied distance measurement to quantify mandibular rotation and displacement. The results showed that the interobserver errors had a range of 0.26 mm. Similarly, Nada et al.<sup>12</sup> used the voxel-based image registration method to test the reliability and measurement error of CBCT superimposition on the anterior cranial base in adult patients who underwent combined surgical orthodontic treatment. The authors reported small differences within 0.5 mm, which were considered to be clinically insignificant. It was also mentioned that the registration of the superimposed scans on the zygomatic arch could be contemplated as an alternative to the anterior cranial base when using smaller Field of View scans in nongrowing patients. However, it is important to be aware that the regions used for quantification of error in this study were closer to the region of reference, thus reducing the magnitude of error. It is known that the further the region of interest is relative to the superimposition structures, the larger the theoretical

error of measurement. Weissheimer et al.<sup>3</sup> also tested the voxel-based method, and their results revealed that distances in the anterior cranial base between registered surface scans were <0.5 mm for most regions, indicating reliable superimposition. Nevertheless, the statistical analysis was not reported and therefore their conclusion should be taken warily.

When comparing the reliability of the voxel-based and the surface-based methods, Almkhhtar et al.<sup>30</sup> reported no significant difference between the two methods using pre- and post-CBCT images of orthognathic surgical patients, although voxel-based registration was associated with less variability. The higher variability in the surface-based method could have been a result of the extra step involving 3D model rendering that this registration required to generate a 3D surface mesh model, which may have introduced a possible source of error. This was also reported by Kang et al.<sup>31</sup> when comparing four different software packages to produce the 3D surface meshes. They found that all four software programs generated reasonably similar meshing accuracies for clinical use. However, there were statistically significant differences at all anatomic regions between them, revealing that there was an inherent range of error in

**Table 6.** COSMIN Risk of Bias Assessment–Generalizability<sup>a</sup>

Generalizability Box	Cevidanes et al. <sup>25</sup>				Cevidanes et al. <sup>16</sup>				Nada et al. <sup>12</sup>				DeCesare et al. <sup>7</sup>				Gkantidis et al. <sup>21</sup>				Weissheimer et al. <sup>3</sup>			
	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P	E	G	F	P
Was the sample in which the HR-PRO instrument was evaluated adequately described? In terms of:																								
Median or mean age (with standard deviation or range)?				x				x	x							x	x							x
Distribution of sex?								x				x				x	x	x						x
Important disease characteristics (e.g., severity, status, duration) and description of treatment?		x				x			x							x	x	x						x
Setting(s) in which the study was conducted (e.g., general population, primary care or hospital/rehabilitation care)?		x						x	x							x				x				x
Countries in which the study was conducted?				x				x	x							x				x				x
Language in which the HR-PRO instrument was evaluated?				x				x				x				x				x				x
Was the method used to select patients adequately described (e.g., convenience, consecutive, or random)?				x				x				x				x				x				x
Was the percentage of missing responses (response rate) acceptable?				x				x				x				x				x				x

<sup>a</sup> COSMIN box with four-point scale for methodological quality: E, excellent; G, good; F, fair; P, poor. Generalizability box is used to extract data on the characteristics of the study population and sampling procedure, with no scoring system used.<sup>24</sup> COSMIN indicates Consensus-Based Standards for the Selection of Health Measurement Instruments. HR-PRO, health-related patient-reported outcome.

the CT image-based meshing process and highlighted that precautions should be taken in selecting the appropriate software and/or anatomic regions to avoid potential error in specific clinical applications.

When using the landmark-based method, the main drawback is that it requires landmark registration, which can increase the risk of observer-dependent errors. However, as previous studies have reported, images can offer consistent and reproducible data if protocols for operator training and calibration are followed.<sup>32,33</sup> This method uses a reference point for 3D cephalometric analysis with CBCT.<sup>34,35</sup> Lagravère et al.<sup>6</sup> evaluated the potential errors associated with the superimposition of serial CBCT images. They used reference planes based on cranial base landmarks using a sensitivity analysis. DeCesare et al.<sup>7</sup> later optimized this analysis using a six-point correction algorithm. This optimized method added two extra landmarks, foramen ovale right and left, which were shown to decrease the envelope of error when determining the coordinate system and increase the intrapoint reliability when comparing images.

Assessing reproducibility is similarly relevant for a 3D superimposition method to be used in research and clinical settings. Cevidanes et al.<sup>16</sup> assessed

reproducibility using the voxel-based method for superimposition in growing patients, although in a small sample. They analyzed before and after treatment CBCT scans of only three growing patients who had orthopedic treatment with miniplates as a treatment to correct a Class III malocclusion. The changes with growth and treatment were measured on the 3D models constructed by three examiners. They reported an interobserver range of measurements across anatomic regions equal or less than 0.5 mm, concluding that these variations were clinically insignificant; therefore, the technique provided a reproducible 3D assessment of growing patients. Comparable reproducibility results were reported by Nada et al.<sup>12</sup> using the voxel-based method and Gkantidis et al.<sup>21</sup> using the surface-based method when repeating superimpositions on the anterior cranial base. Although the last study used CT scans instead of CBCT, the authors claimed that the validity of the proposed superimposition method would not change substantially when applied to CBCT images. Although CBCT images have lower segmentation accuracy to some extent when compared to CT, anatomical landmarks and models are generated in a reliable and clinically applicable way.<sup>36</sup>

## CONCLUSIONS

- Findings from most of the studies included in the current review suggest that all three methods for 3D superimposition provide an acceptable level of reliability when assessing changes in craniofacial hard tissues.
- However, due to a low methodological quality of the identified evidence, the overall results should be considered cautiously.
- In addition, although the 3D superimposition methods are more convenient for craniofacial assessment than conventional 2D methods, to date no studies have used a gold standard to determine the real accuracy of any of these methods.

## ACKNOWLEDGMENTS

The authors acknowledge the assistance of health sciences librarians at the University of Alberta. The authors declare no conflicts of interest and no funding.

## REFERENCES

1. Jacobson A, Sadowsky L. Superimposition of Cephalometric Radiographs. In: Jacobson A, Jacobson R. *Radiographic Cephalometry*. Chicago, IL: Quintessence; 2006:145–151.
2. Cevidanes LH, Styner MA, Proffit WR. Image analysis and superimposition of 3-dimensional cone-beam computed tomography models. *Am J Orthod Dentofacial Orthop*. 2006;129:611–618.
3. Weissheimer A, Menezes LM, Koerich L, Pham J, Cevidanes LH. Fast three-dimensional superimposition of cone beam computed tomography for orthopaedics and orthognathic surgery evaluation. *Int J Oral Maxillofac Surg*. 2015;44:1188–1196.
4. Park JH, Tai K, Owtad P. 3-Dimensional cone-beam computed tomography superimposition: a review. *Semin Orthod*. 2015;21:263–273.
5. Hatcher D. Maxillofacial imaging. In: McNeill C, ed. *Science and Practice of Occlusion*. Chicago, IL: Quintessence; 1997:349–364.
6. Lagravère MO, Major PW, Carey J. Sensitivity analysis for plane orientation in three-dimensional cephalometric analysis based on superimposition of serial cone beam computed tomography images. *Dentomaxillofac Radiol*. 2010;39:400–408.
7. DeCesare A, Secanell M, Lagravère MO, Carey J. Multi-objective optimization framework for landmark measurement error correction in three-dimensional cephalometric tomography. *Dentomaxillofac Radiol*. 2013;42:1–10.
8. Adams GL, Gansky SA, Miller AJ, et al. Comparison between traditional 2-dimensional cephalometry and a 3-dimensional approach on human dry skulls. *Am J Orthod Dentofacial Orthop*. 2004;126:397–409.
9. Kapila S, Conley RS, Harrell WE Jr. The current status of cone beam computed tomography imaging in orthodontics. *Dentomaxillofac Radiol*. 2011;40:24–34.
10. Redmond R, Huang J, Bumann A, Mah J. Three-dimensional radiographic analysis in orthodontics. *J Clin Orthod*. 2005;XXXIX:421–428.
11. Angelopoulos C. Cone beam tomography imaging anatomy of the maxillofacial region. *Dent Clin North Am*. 2008;52:731–752.
12. Nada RM, Maal TJJ, Breuning KH, Berge SJ, et al. Accuracy and reproducibility of voxel based superimposition of cone beam computed tomography models on the anterior cranial base and the zygomatic arches. *PLoS ONE*. 2011;6:e16520.
13. Grauer D, Cevidanes LSH, Proffit WR. Working with DICOM craniofacial images. *Am J Orthod Dentofacial Orthop*. 2009;136:460–470.
14. Terajima M, Yanagita N, Ozeki K, et al. Three dimensional analysis system for orthognathic surgery patients with jaw deformities. *Am J Orthod Dentofacial Orthop*. 2008;134:100–111.
15. Naji P, Alsufyani N, Lagravere M. Reliability of anatomic structures as landmarks in three-dimensional cephalometric analysis using CBCT. *Angle Orthod*. 2014;84:762–772.
16. Cevidanes LH, Heymann A, Cornelis M, DeClerck HJ, Tulloch JFC. Superimposition of 3-dimensional cone-beam computed tomography models of growing patients. *Am J Orthod Dentofacial Orthop*. 2009;136:94–99.
17. Cevidanes LH, Motta A, Proffit WR, Ackerman JL, Stynere M. Cranial base superimposition for 3-dimensional evaluation of soft-tissue changes. *Am J Orthod Dentofacial Orthop*. 2010;137:S120–S129.
18. Cevidanes LH, Bailey L'TJ, Tucker SF, et al. Three-dimensional cone-beam computed tomography for assessment of mandibular changes after orthognathic surgery. *Am J Orthod Dentofacial Orthop*. 2007;131:44–50.
19. Viola P, Wells III W. International Journal of Computer Vision. 1997;24:137–154.
20. Lagravère M, Secanell M, Major P, Carey J. Optimization analysis for plane orientation in 3-dimensional cephalometric analysis of serial cone-beam computerized tomography images. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 2011;111:771–777.
21. Gkantidis N, Schauseil M, Pazera P, et al. Evaluation of 3-dimensional superimposition techniques on various skeletal structures of the head using surface models. *PLoS ONE*. 2015;10:e0118810.
22. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Syst Rev*. 2015;4:3–9.
23. Cochrane. Consumers and communication group resources for authors. Available at: <https://cccr.org/cochrane.org/author-resources>. Accessed May 7, 2016.
24. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21:651–657.
25. Cevidanes LH, Bailey LJ, Tucker GR Jr, et al. Superimposition of 3D cone-beam CT models of orthognathic surgery patients. *Dentomaxillofac Radiol*. 2005;34:369–375.
26. Ruellas AC, Tonello C, Alonso N, et al. Author's response. *Am J Orthod Dentofacial Orthop*. 2016;150:398–400.
27. Heymann G, Cevidanes L, Cornelis M, et al. Three-dimensional analysis of maxillary protraction with intermaxillary elastics to miniplates. *Am J Orthod Dentofacial Orthop*. 2010;137:274–284.
28. Ribeiro Carvalho F, Cevidanes LH, Motta AT, Oliveira Almeida MA, Phillips C. Three-dimensional assessment of

- mandibular advancement 1 year after surgery. *Am J Orthod Dentofacial Orthop.* 2010;137:S53.e51–S53.e12.
29. Toyama Hino C, Cevidanes L, Nguyen T, et al. Three-dimensional analysis of maxillary changes associated with facemask and rapid maxillary expansion compared with bone anchored maxillary protraction. *Am J Orthod Dentofacial Orthop.* 2013;144:705–714.
  30. Almukhtar A, Ju X, Khambay B, McDonald J, Ayoub A. Comparison of the accuracy of voxel based registration and surface based registration for 3D assessment of surgical change following orthognathic surgery. *PLoS ONE.* 2014;9:e93402.
  31. Kang AH, Kim MK, Kim HJ, Zhengguo P, Lee SH. Accuracy assessment of image-based surface meshing for volumetric computed tomography images in the craniofacial region. *J Craniofac Surg.* 2014;25:2051–2055.
  32. Almeida MA, Cevidanes L, Phillips C, et al. Observer reliability of three-dimensional cephalometric landmark identification on cone-beam computerized tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2009;107:256–265.
  33. Lagravère MO, Low C, Flores-Mir C, et al. Intraexaminer and interexaminer reliabilities of landmark identification on digitized lateral cephalograms and formatted 3-dimensional cone-beam computerized tomography images. *Am J Orthod Dentofacial Orthop.* 2010;137:598–604.
  34. Lagravère MO, Major PW. Proposed reference point for 3-dimensional cephalometric analysis with cone-beam computerized tomography. *Am J Orthod Dentofacial Orthop.* 2005;128:657–660.
  35. Lagravère MO, Hansen L, Harzer W, Major PW. Plane orientation for standardization in 3-dimensional cephalometric analysis with computerized tomography imaging. *Am J Orthod Dentofacial Orthop.* 2006;129:601–604.
  36. Denis K, Hassam B, Li L, et al. A comparative evaluation of cone beam computed tomography (CBCT) and multi-slice CT (MSCT). Part II: on 3D model accuracy. *Eur J Radiol.* 2010;75:270–274.