Orthodontics & Craniofacial Research

REVIEW ARTICLE

P. Pittayapat
N. Limchaichana-Bolstad
G. Willems
R. Jacobs

Three-dimensional cephalometric analysis in orthodontics: a systematic review

Authors' affiliations:
P. Pittayapat, R. Jacobs, Oral Imaging Center, OMFS-IMPATH research group, Department of Imaging & Pathology, Faculty of Medicine, University of Leuven, Leuven, Belgium
P. Pittayapat, Department of Radiology, Faculty of Dentistry, Chulalongkorn University, Bangkok, Thailand
N. Limchaichana-Bolstad, Institute of Clinical Dentistry, University of Tromsø, Tromsø, Norway
G. Willems, Orthodontics, Department of Oral Health Sciences, KU Leuven & Dentistry, University Hospitals Leuven, University of Leuven, Leuven, Belgium

Correspondence to:
P. Pittayapat
Oral Imaging Center, OMFS-IMPATH research group
Dept Imaging & Pathology, Faculty of Medicine, University of Leuven
Kapucijnenvoer 33
3000 Leuven, Belgium
E-mail: p.pittayapat@gmail.com

Date:
Accepted 16 November 2013

DOI: 10.1111/ocr.12034

Pittayapat P., Limchaichana-Bolstad L., Willems G., Jacobs R. Three-dimensional cephalometric analysis in orthodontics: a systematic review
Orthod Craniofac Res 2014; **17**: 69–91. © 2013 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd

Structured Abstract

Context – The scientific evidence of 3D cephalometry in orthodontics has not been well established.

Objective – The aim of this systematic review was to evaluate the evidence for the diagnostic efficacy of 3D cephalometry in orthdontics, focusing on measurement accuracy and reproducibility of landmark identification.

Data Sources – PubMed, EMBASE and the Cochrane library (from beginning to March 13, 2012) were searched. Search terms included: cone-beam computed tomography; tomography, spiral computed; imaging, three-dimensional; orthodontics.

Study Selection – Two reviewers read the retrieved articles and selected relevant publications based on pre-established inclusion criteria. The selected publications had to elucidate the hierarchical model of the efficacy of diagnostic imaging systems by Fryback and Thornbury.

Data Extraction – The data was then extracted according to two protocols, which were based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. Next, levels of evidence were categorized into 3 groups: low, moderate and high evidence.

Data Synthesis – 571 publications were found by database search strategies and 50 additional studies by hand search. A total of 35 publications were included in this review.

Conclusions – Limited evidence for the diagnostic efficacy of 3D cephalometry was found. Only 6 studies met the criteria for a moderate level of evidence. Accordingly, this systematic review reveals that there is still need for methodologically standardized studies on 3D cephalometric analysis.

Key words: cephalometry; cone-beam computed tomography; orthodontics; spiral computed tomography; three-dimensional imaging

Introduction

Cephalometric analysis was first introduced in 1930s by Hofrath (1) in Germany and Broadbent (2) in the United States. The method uses frontal and lateral cephalometric radiographs to evaluate the craniofacial complex, dentofacial proportions, malocclusion and changes related to growth, all of which are crucial for orthodontic treatment planning and evaluation. A conventional cephalometric radiograph is a two-dimensional representation of three-dimensional structures. Although widely accepted as a standard tool for treatment planning, it still has several downsides, such as geometric distortion and superimposition of structures (3–5).

Recently, three-dimensional images have started to play an important role in oral health and diagnosis. Several years ago computed tomography (CT) was introduced into the dental field. However, its high radiation dose has led to controversy. In 1996, dental cone-beam computed tomography (CBCT) was invented, and the technology has been evolving ever since. With relatively lower radiation doses than multi-slice CT (MSCT), CBCT has become very popular in dentistry. Some researchers have also introduced the use of a clinical low-dose CT protocol for 3D cephalometric application (6–8). Both modalities allow orthodontists to visualize craniofacial structures in three dimensions and overcome the drawback of 2D cephalometric analysis.

Several studies have been conducted on cephalometric images derived from CBCTs. These derived lateral cephalometric images were proven to be accurate and comparable with direct measurements on skulls (9–12) and conventional cephalograms of patients (13). This method is the first step towards 3D cephalometry. Nevertheless, it still implies that a patient's anatomy is not evaluated in three dimensions. A combination of measurements on the axial, coronal and sagittal view was also used in several studies (14, 15). This method has been sometimes referred to as 2.5D as it does not allow full access to the patients' structures in real three dimensions (8,15). Three-dimensional cephalometric analysis requires input from 3D images of the patient, either on CBCTs or on MSCTs, and software that offers 3D cephalometric measurement tools (16–19).

An increasing amount of research has been conducted to evaluate the measurement accuracy, reliability and reproducibility of 3D craniofacial landmark identification and to justify whether further elaboration of 3D cephalometry is more beneficial than the standard 2D analysis. To our knowledge, a systematic review specifically focusing on 3D cephalometry for orthodontic diagnosis and treatment planning was not yet available. The aim of this review is therefore to systematically evaluate the current evidence for the diagnostic efficacy of 3D cephalometry, focusing on measurement accuracy and reproducibility of landmark identification for orthodontic diagnosis.

Material and methods

Methods of the analysis and inclusion criteria were specified in advanced and were documented in a protocol.

Eligibility criteria

The selected publications had to elucidate the model of efficacy: diagnostic accuracy efficacy, diagnostic thinking efficacy, therapeutic efficacy, or any combination of the preceding adapted from by Fryback and Thornbury (20).

Diagnostic accuracy efficacy (20) was defined as:

- Observer performance expressed as overall agreement, Kappa Index or correlation coefficients;
- Diagnostic accuracy as percentage of correct landmark identification;
- Diagnostic accuracy as percentage of correct cephalometric linear and/or angular measurement;
- Sensitivity, specificity or predictive values;

Diagnostic thinking efficacy (20) was defined as:

- Percentage of cases in a series in which 3D cephalometry was judged 'helpful' to guide the orthodontic treatment;
- Difference in clinicians' subjectively estimated diagnosis probabilities pre- to post-3D cephalometric information;

Therapeutic efficacy (20) was defined as:

- Percentage of times 3D cephalometry judged helpful in planning management of the patient in a case series;
- Percentage of times therapy-planned pre-3D cephalometry changed after the 3D cephalometric information was obtained;
- Percentage of times clinicians' prospectively stated therapeutic choices changed after 3D cephalometric information.

Information sources

A comprehensive electronic database search was performed in MEDLINE via PubMed (from beginning to 13 March 2012), EMBASE via embase.com (from beginning to 13 March 2012), and the Cochrane library website (from beginning to 13 March 2012). No restrictions were imposed regarding time period or types of study design (i.e. case-controlled, randomized controlled trial). The publications were searched electronically by using controlled index terms and relevant specific free text words. The last search was performed on 13 March 2012. Detailed search strategies for both MEDLINE and EMBASE are shown in Table S1. The Cochrane Central Register of Controlled Trials (CENTRAL), the Cochrane Database of Systematic Reviews (CDSR) and the Cochrane Database of Abstracts of Reviews of Effects (DARE) were searched using the search term 'cephalometry'.

Search strategy

The following Medical Subject Headings (MeSH) were used when searching the literature: Cephalometry; Cone-beam computed tomography; Tomography, Spiral Computed; Imaging, Three-Dimensional; Orthodontics. Table S1 shows the search strategies used in this review.

Study selection

The lists of publications from both databases were imported into EndNote® Web 3.3 (Thomson Reuters, New York, USA). Duplicate articles were deleted, after which two reviewers independently read the resulting collection of titles and abstracts. Book chapters, review studies and animal studies were excluded. Both *in vitro* and *in vivo* studies were included. The full texts of selected publications were then retrieved. When an abstract was considered to be relevant by one of the authors, the publication was then read in full text.

Grey literature were searched but excluded if full texts were not available. When publications elucidated only observer performance, the analysis had to be based on a minimum of two observers.

The second step was to hand search the reference lists of the publications that were found to be relevant in the first step. Titles of the articles should contain a keyword: 'cephalometry' 'cephalometric' and 'cone-beam computed tomography' 'CBCT' or 'computed tomography 'CT' or 'three-dimensional' or '3D'. Similar to the first step, when an abstract was considered to be relevant by one of the reviewers, the full text was then retrieved.

Data extraction

Two authors independently extracted the data into protocol 1 (Table S2) which was formulated based on the Cochrane Handbook for Diagnostic Test Accuracy (DTA) Reviews and the Cochrane Handbook for Systematic Reviews of Interventions (21, 22) and literature describing how to critically appraise studies on diagnostic methods (23, 24).

Secondly, protocol 2 (Table S3) was applied to the included articles to assess the quality of the publications. This protocol is based on the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool and the Cochrane Handbook for DTA Reviews (21, 25). Information was extracted from included studies concerning: type of studies, number of samples, reference method, specific method used in the study, number of observers, statistical method and results according to authors.

Levels of evidence

The quality and internal validity (level of evidence) of each publication was judged to be high, moderate, or low according to the following criteria (23, 24).

A study was assessed to have a *high level of evidence* if it fulfilled all of the following criteria:

- There was an independent blind comparison between test and reference methods.
- The population was described so that the status, prevalence and severity of the condition were clear. The spectrum of patients was similar to the spectrum of patients on whom the test method will be applied in clinical practice.
- The results of the test method being evaluated did not influence the decision to perform the reference method(s).
- Test and reference methods were well-described concerning technique and implementation.
- The judgments (observations, measurements) were well described considering diagnostic criteria applied and information and instructions to the observers.
- The reproducibility of the test method was described for 1 observer (intra-observer performance) as well as for several (minimum 2) observers (interobserver performance).
- The results were presented in terms of relevant data needed for necessary calculations.

A study was assessed to have a *moderate level of evidence* if any of the above criteria were not met. On the other hand, the study was assessed not to have deficits that are described below for studies with a low level of evidence.

A study was assessed to have a *low level of evidence* if it met any of the following criteria:

- The evaluation of the test and reference methods was non-independent.
- The population was not clearly described, and the spectrum of patients was distorted.
- The results of the test method influenced the decision to perform the reference method.
- The test or the reference method or both were not satisfactorily described.
- The judgments were not well described.
- The reproducibility of the test method was not described or was described for only 1 observer.
- The results could have a systematic bias.
- The results were not presented in a way that allowed efficacy calculations to be made.

The scientific evidence on diagnostic efficacy was evaluated according to the scale: strong, moderately strong, limited or insufficient (23, 24) depending on the quality and the level of evidence of the publications.

- Strong research-based evidence: at least two publications with a high level of evidence.
- Moderately strong research-based evidence: one publication with a high level of evidence and two publications with a moderate level of evidence.
- Limited research-based evidence: at least two of the publications with a moderate level of evidence.
- Insufficient research-based evidence: scientific evidence is insufficient or lacking according to the criteria defined in this study.

Synthesis of results

The results were analysed descriptively. No meta-analysis was performed because of the lack of original studies.

Results

Study selection

The results of this systematic review are reported based on the PRISMA statement (26). A total of 524 publications were found from the PubMed database, 175 publications from the EMBASE database and no systematic reviews or clinical trials on 3D cephalometry were found in the Cochrane library. This resulted in 571 publications after removing of duplicates and a total of 77 publications were included in the systematic review after the first assessment (Fig. 1) (26).

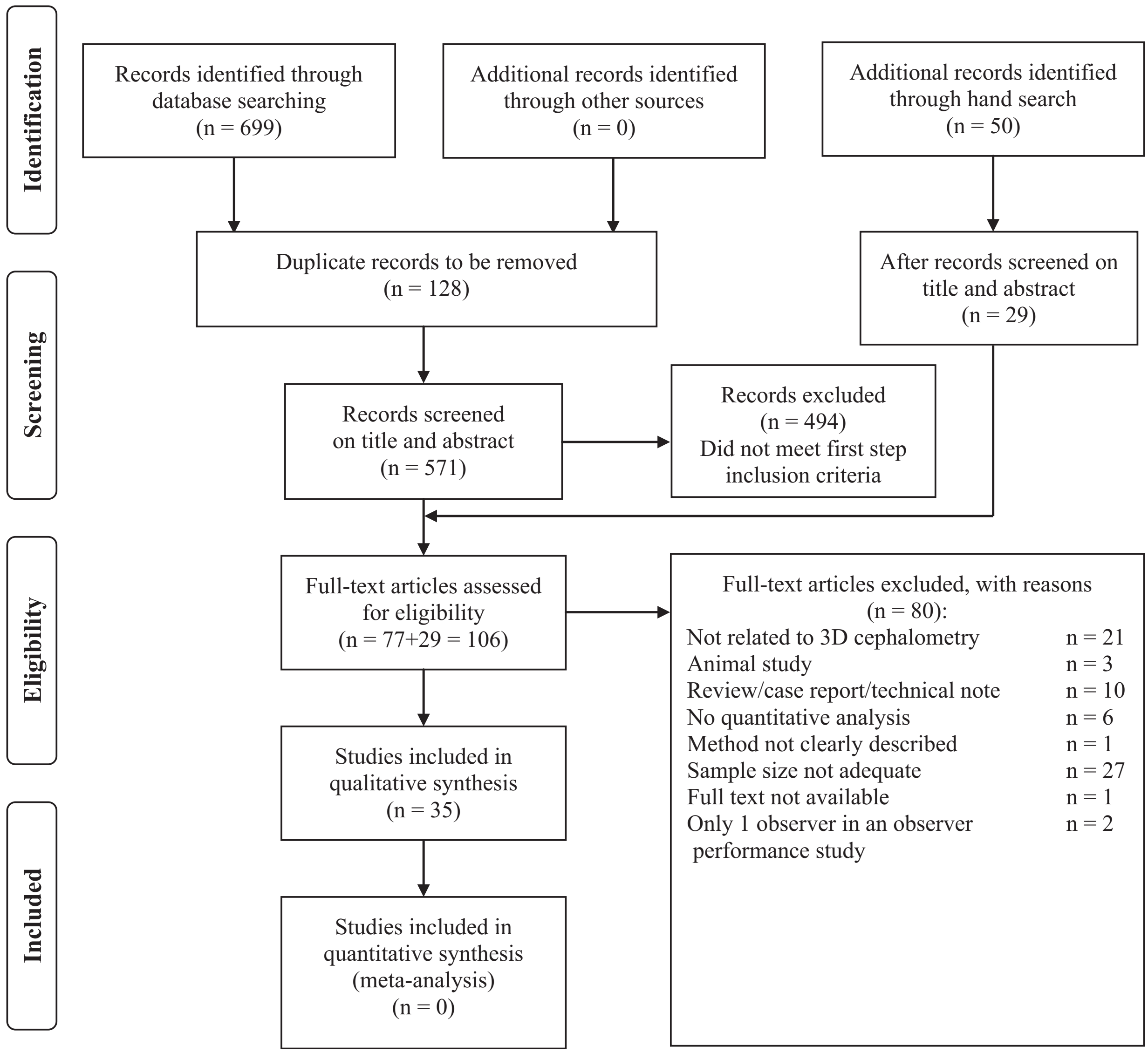

Fig. 1. Flow diagram of the included studies according to the PRISMA.

Data extraction was then performed firstly using protocol 1 (Table S2). Publications that were not relevant to the model of efficacy were excluded in this step, thus a total of 77 articles were read and their quality was assessed. Quality assessment of these studies was evaluated by using protocol 2, based on the QUADAS tool (Table S3).

The second step of the search was carried out by hand searching the reference lists of included publications. Fifty additional articles met the search criteria and were added to the list of protocol 1. As a result of this assessment, 29 original articles were additionally included to the review, setting the total number of articles submitted to the protocol two evaluation at 106 publications (Fig. 1).

In this step, studies that did not have a representative sample spectrum (sample size smaller than 10) were excluded. Studies, in which the observer performance was evaluated but which had only 1 observer, and studies that did not have a valid reference standard were regarded as low quality.

Finally, 35 original articles were included in systematic review (7, 17, 27–58). These included publications were categorized into three different levels: low, moderate and high level of evidence (23, 24).

Study characteristics

To be able to report the diagnostic accuracy of these studies, articles were categorized based on the topics of studies as followed: landmark identification, linear and angular measurement, facial asymmetry and other topics.

The most reported topic was landmark identification. No study met the criteria for high level of evidence. Six publications were qualified as moderate level of evidence (Table 1), and 29 publications were qualified as low level of evidence (Tables 2, 3, S4 and S5). All studies were related to the diagnostic accuracy efficacy. No publication that reported on diagnostic thinking efficacy and therapeutic efficacy was found.

Landmark identification

Fourteen publications regarding landmark identification are shown in Table 2. Two publications from Olszewski et al. (7) and Olszewski et al. (8) reached the moderate level of evidence and both of which are *in vitro* studies on dry human skulls. All studies that were done on patients lacked a gold standard and were therefore regarded as low level of evidence.

In the 2008, study of Olszewsksi et al., the authors compared low- and high-dose CT protocols for landmark identification in 3D cephalometry. They reported that the global intra- and interobserver mean distances for all landmarks were smaller with a high-dose CT protocol ($p = 0.37$) and ($p = 0.03$), respectively (7). Olszewsksi et al. did a similar study in 2013, but this time comparing a low-dose CT protocol and a CBCT for landmark identification in 3D cephalometry. The results revealed that the CBCT showed better reproducibility. The intra- and interobserver mean distance of the CBCT ($p = 0.000075$) were smaller than those from the low-dose CT ($p = 0.00087$) (8).

Regarding the precision of landmark identification, studies reported it in different manners. In 1995, Richtsmeier et al. (48) found that the precision in locating landmarks was less than 0.5 mm for all landmarks. In addition, the accuracy of linear measurements was reported as average difference between 1–2 mm. In 2009, Ludlow et al. reported the precision of landmark identification in another manner. It was found that overall correlation was 0.98 and 13 of 24 landmarks had statistically less variability in at least 1 direction of measurement in the multiplanar reformation (MPR) views (41). Schlicher et al. (50) showed that the average consistency across all 32 landmarks among nine examiners was 1.64 mm, while the average precision (SD) was 0.87 mm. Sella turcica was the most consistent (0.50 mm) and most precise (0.23 mm). Hassan et al. defined the precision as the absolute difference between an observer's repeated measurements and the mean of all measurements per landmark. The 3D surface model together with multi-planar reformation (MPR) images improved the tracing precision in 15 but only statistically significant in 6 of 22 landmarks. The total precision of measurements ranged between 0.29 ± 0.17 mm and 2.82 ± 7.53 mm (36).

The most reported results are about the observer performance, reproducibility and repeatability of the landmarks. Concerning the observer reliability, studies reported the results as mean measurement differences (38, 44, 51) or intraclass correlation coefficient values (ICC) (29, 33, 38, 39). The ICC ranged between more than 0.70 to more than 0.90 for intra-observer and between more than 0.64 to more than 0.90 for interobserver (29, 33, 38, 39). Zamora et al. (58) found ICC higher than 0.99 with the best results for landmark identification in the Z-direction. Chien et al. (29) found lower ICC in 2D: interobserver 0.35 vs. intra-observer 0.57, but Lagravère et al. (39) reported values >0.90. In Chien's study, 2D landmark identifications were generally much less repeatable than in 3D (29).

Linear and angular measurement

Thirteen publications regarding the measurement topic were shown in Table 3. Two publications from Cavalcanti et al. (28) and Lopes et al. (40) reached the moderate level of evidence and both of which are *in vitro* studies. As similar to the studies in the landmark identification group, all studies that were done on patients lacked a

Table 1. **Studies with a moderate level of evidence**

Authors	Study topic	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors
Cavalcanti et al. (2004)	Measurement accuracy & precision	*In vitro*	13 cadaver heads	Electromagnetic 3 Space™ digitizer	Spiral CT Vitrea© version 2.3 software Anthropometric anatomical points	2 observers × 2	Mean absolute error percentage, SD, CI, intra- and interobserver reliability, ANOVA	Intra- vs. interobserver: NS Imaging vs. physical: NS Standard error 0.45–1.44% % Error: physical vs. 3D Bone = 0.83% Soft tissue = 1.78%
Olszewski et al. (2007)	3D ceph craniofacial analysis	*In vitro*	26 dry skulls	3D measuring instrument to assess 3D position (*x*, *y*, *z*)	Lat ceph Spiral CT ACRO 3D© software Delaire's ceph analysis	2 observers × 2	Reproducibility, accuracy	Intra-observer: ICC 2D = 0.06–0.90 ICC 3D CT = 0.97–1.00 Interobserver: ICC 2D = 0.13–0.84 ICC 3D CT = 0.94–1.00 User accuracy of 3D (absolute difference) between 0.75 mm (±0.05) and 0.99 mm (±0.08)
Olszewski et al. (2008)	Landmark accuracy	*In vitro*	15 dry skulls	Calculated reference standard based on the index test	Spiral CT Low-dose CT protocol ACRO 3D© software 12 ceph landmarks	2 observers × 2	Non-parametric test, accuracy, reliability	Intra-observer mean distances: High dose < low dose ($p = 0.37$) Interobserver mean distances: High dose < low dose ($p = 0.03$)
Lopes et al. (2008)	Angular measurements	*In vitro*	28 dry skulls	Beyond Crysta-C9168 series 900 device for 3D coordinates	Multi-slice CT Craniometric anatomical points 6 angular measurements	2 observers × 2	Precision, accuracy	Interobserver: NS Intra-observer: NS Difference between physical and 3D measurement: −3.16% to −0.10%

***Table 1.* (continued)**

Authors	Study topic	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors
Lagravère et al. (2011)	Cranial base foramen	*In vitro*	10 dry skulls	Gutta percha placed at landmarks	CBCT AMIRA© software 4 references foramen landmark	3 observers (Principal × 3 Others × 1)	Accuracy, intra- and interobserver reliability (ICC)	Intra-observer ICC > 0.93 Interobserver ICC >0.92 Accuracy: large average mean difference (max. 3.60 mm for foramen rotundum)
Olszewski et al. (2013)	Landmark reproducibility	*In vitro*	10 dry skulls	Calculated reference standard based on the index test	CBCT Maxilim© software 24 osseous landmarks	2 observers × 2	Reproducibility	Intra-observer mean distance: CBCT < low-dose CT (p = 0.000075) Interobserver mean distance: CBCT < low-dose CT (p = 0.00087)

CT, computed tomography; CBCT, cone-beam computed tomography; ceph, cephalometric; lat ceph, lateral cephalogram; SD, standard deviation; CI, confidence interval; ANOVA, analysis of variance; NS, no statistically significant difference; ICC, Intraclass correlation coefficient.

Table 2. **Landmark Identification: accuracy/reproducibility/precision**

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level of evidence
Richtsmeier et al. (1995)	*In vitro*	10 dry skulls	Polhemus 3Space table-top digitizer	CT Image software version 1.47 35 craniofacial landmarks	1 observer	Precision, repeatability, ANOVA, accuracy	Precision: average difference < 0.5 mm Precision between 2 images: low average error Repeatability: along *X*-axis = 98.5% along *Y*-axis > 99.5% Accuracy of linear measurement: average difference = 1–2 mm	Low
Lagravère et al. (2009)	*In vivo*	24 patients after maxillary expansion	None	CBCT: AMIRA© software 24 craniometric anatomic landmarks	5 observers (Principal × 5 Others × 1)	Intra- and interobserver reliability: ICC, average mean differences, ANOVA	Intra-observer: ICC >0.80 Interobserver: ICC > 0.80 Mean measurement differences: Intra-observer < 1.5 mm, Inter- >intra-observer	Low
Ludlow et al. (2009)	*In vitro*	20 pre-surgical orthodontic patients	None	CBCT Lat ceph Dolphin 3D© software version 10.5 24 ceph landmarks	5 observers × 1	Precision, ODM, DEO, ANOVA, paired *t* test	ANOVA: every effect and the 1° interactions between effects: sig diff ($p < 0.05$) DEO > ODM Overall correlation = 0.98 DEO and ODM of radiographic modality and landmark: sig diff ($p < 0.0001$)	Low
De Oliveira et al. (2009)	*In vivo*	12 orthognathic surgery patients	None	CBCT Dolphin 3D© software 30 landmarks	3 observers × 3	ANOVA, ICC	Intra-observer: ICC ≥ 0.9 (85.55%) Interobserver: ICC ≥ 0.9 (65.55%)	Low

Table 2. **(continued)**

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level of evidence
Chien et al. (2009)	*In vivo*	10 patients	None	CBCT Digital lat ceph Dolphin© Imaging 10.0 software Dolphin 3D 27 landmarks	6 observers × 2	Linear model, the best estimate of true values, absolute value of difference, standard error, ICC	Interobserver difference of 2D > 3D ($p < 0.05$) Interobserver: 3D ICC: 0.64–1.00 2D ICC: 0.35–1.00 Intra-observer: 3D ICC: 0.70–1.00 2D ICC: 0.57–1.00	Low
Lagravère et al. (2010)	*In vivo*	10 adolescent patients	None	CBCT Digital lat ceph AMIRA© software 18 landmarks	3 observers Principal × 3 Others × 1	Intra- and interobserver reliability (ICC)	Interobserver: ICC > 0.90 Intra-observer: ICC > 0.90 Mean landmark difference: Lat ceph ~ 1 mm CBCT > 1 mm	Low
Olszewski et al. (2010)	*In vivo*	13 patients	None actual reference. Stated the possibility of calculated reference standard based on the index test	Spiral CT ACRO 3D© software 44 reference landmarks (22 from 3D-ACRO and 22 from 3D-Swennen)	2 observers × 2	Intra- and interobserver mean distance (reconstructed mean log)	Intra-observer: 3D-ACRO = 1.21 ± 1.04 mm 3D-Swennen = 1.31 ± 1.04 mm (comparison: $p = 0.17$ NS) Interobserver: 3D-ACRO = 1.80 ± 1.04 mm 3D-Swennen = 2.46 ± 1.04 mm (comparison: $p = 0.000000002$) Intra- and interobserver difference: 3D-ACRO = 1.49 ± 1.06 mm 3D-Swennen = 1.88 ± 1.06 mm Reproducibility: 3D-ACRO > 3D-Swennen ($p = 0.0027$)	Low
Titiz et al. (2011)	*In vivo*	20 patients	None	CT VoXim© 6.1 software 28 landmarks	4 observers × 2	Descriptive statistics, ANOVA	Intra-observer: 0.52 mm Interobserver: 0.22 mm	Low

Table 2. **(continued)**

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level of evidence
Shibata et al. (2012)	*In vivo*	10 patients	None	CBCT Digital lat ceph VG Studio MAX 1.1 software Anatomical coordinate system 1, 2, 3, 4	6 observers × 2 Principal × 3 and averaged the coordinates (*x*, *y*, *z*) as the initial coordinates.	95% confidence ellips, reproducibility of the coordinate systems (ellipsoid volume)	The volumes of ellipsoids in systems 2,4 < 1, 3: sig diff Reproducibility: >in system 2,4 Less ellipsoid volume: landmarks at greater distances from each other	Low
Schlicher et al. (2012)	*In vitro*	19 patients	None	CBCT Dolphin imaging© 10.1 software 32 landmarks	9 observers	Average coordinates of all observers for each landmark*, Consistency†, precision‡	Average consistency = 1.64 mm Average precision = 0.87 mm NS between observers The most consistent: Sella turcica (0.50 mm) The most precise: Sella turcica (0.23 mm) The most inconsistent: porion-right (2.72 mm) The most imprecise: orbitale-right (1.81 mm)	Low
Zamora et al. (2012)	*In vivo*	15 patients	None	CBCT Dolphin© imaging 10.1 software 41 landmarks	2 observers × 3	Reproducibility: ANOVA, multiple comparison, ICC, Pearson correlation coefficient	Intra-observer: ICC ≥ 0.99 Interobserver: ICC ≥ 0.99 (best frequency on axis *Z*) The average SD of all landmarks = 1.0 mm, average relative error 1.3%.	Low

Table 2. **(continued)**

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level of evidence
Hassan et al. (2012)	*In vivo*	10 patients	None	CBCT Dolphin 3D© software 22 ceph landmarks (Swennen)	11 observersx 2 3Dx 2 3D+MPR	Mixed model design, Fisher's *F*-test, Cronbach's α for interobserver reliability, precision§	Precision: 3D + MPR > 3D (p = 0.0001): sig diff more precise in 6 of 22 landmarks Intra-observer: 3D + MPR > 3D Interobserver: 3D + MPR > 3D Precision: ranged between 0.29 ± 0.17 and 2.82 ± 7.53	Low
Olszewski et al. (2008)	*In vitro*	15 dry skulls	Calculated reference standard based on the index test	Spiral CT Low-dose CT protocol ACRO 3D© software 12 ceph landmarks	2 observers × 2	Non-parametric test, accuracy, reliability	Intra-observer mean distances: High dose < low dose (p = 0.37) Interobserver mean distances: high-dose < low-dose (p = 0.03)	Moderate
Olszewski et al. (2013)	*In vitro*	10 dry skulls	Calculated reference standard based on the index test	CBCT Maxilim© software 24 osseous landmarks	2 observers × 2	Reproducibility	Intra-observer mean distance: CBCT < low-dose CT (p = 0.000075) Interobserver mean distance: CBCT < low-dose CT (p = 0.00087)	Moderate

CT, computed tomography; CBCT, cone-beam computed tomography; ceph, cephalometric; lat ceph, lateral cephalogram; SD, standard deviation; CI, confidence interval; ANOVA, analysis of variance; sig diff, statistically significant difference, NS, no statistically significant difference; ICC, intraclass correlation coefficient; ODM, average observer difference of the mean; DEO, difference for every observer; SEM, standard error of the mean.
*Average coordinates of all observers for each landmark = the centroid for the particular landmark.
†Consistency = the mean of the measurements of how far the landmarks were from the centroid by all observers.
‡Precision = SD of the distances from the centroid.
§Precision = the absolute difference between an observer's repeated measurements and the mean of all measurements per landmark.

Table 3. **Measurement: accuracy/reliability**

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level ofevidence
Waitzman et al. (1992)	*In vivo*	100 patients	None	CT 15 measurements	2 observer × 2	Pearson product-moment correlation coefficient (*r*)	Intra-observer: $r \geq 0.93$ Interobserver: 85% $r \geq 0.85$ 15% $r \geq 0.74$	Low
Periago et al. (2008)	*In vitro*	23 dry skulls	Electronic digital caliper (3 times by 2 observers)	CBCT Dolphin 3D© software 14 craniometric anatomic landmarks 20 orthodontic linear measurements	2 observer × 3	Accuracy, intra-observer reliability, ICC, paired *t*-test	Intra-observer: ICC skull measurements = 0.99 ± 0.08 ICC CBCT = 0.98 ± 0.02 Skull measurement > CBCT ($p < 0.001$) Mean% measurement error: CBCT = 2.31%±2.11% Skull measurement = 0.63%± 0.51% CBCT > Skull measurement ($p < 0.001$) Mean% difference: skull vs. 3D-based measurements = 1.13% (SD ± 1.47%)	Low
Brown et al. (2009)	*In vitro*	19 dry skulls	Electronic digital caliper (3 times by 2 observers)	CBCT Dolphin 3D© software 24 craniometric anatomic landmarks 16 linear measurements	1 observer × 3	Intra-observer reliability: absolute mean error (± SD), Wilks lambda multivariate test ($P \leq 0.05$), Sidak adjustment for multiple comparisons	Mean absolute error: Between scan settings: NS Average skull absolute error: marked reference points < unmarked reference Mean difference in CBCT: 9 ↓ measurements: 3.1 ± 0.12 mm–0.56 ± 0.07 mm 1 ↑ measurement: 3.3 ± 0.12 mm CBCT sequences: NS	Low

Table 3. **(continued)**

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level ofevidence
Nalcaci et al. (2010)	*In vivo*	10 patients	None	Spiral CT Mimics© 9.0 software Lat ceph 18 ceph landmarks Angular measurements	2 observers × 2	Intra-observer reproducibility: Dahlberg's formula, Wilcoxon matched-pairs signed rank test (α = 0.05)	Interobserver: method error 0.35°–0.65°, NS 2D vs. 3D: sig diff in some measurements (P < 0.05)	Low
Van Vlijmen et al. (2010)	*In vitro*	40 dry skulls	None	CBCT Lat ceph Maxilim© software 17 hard-tissue landmarks 10 angular and 2 linear ceph measurements	1 observer × 5	Pearson correlation coefficient, paired *t*-test	Intra-observer: 0.69–0.98 Standard error: Lat ceph < 3D model (9 of 12measurements) Reproducibility: Lat ceph > 3D models Average difference: –3.12 to 0.83.	Low
Van Vlijmen et al. (2011)	*In vitro*	40 dry skulls	None	2 CBCTs Maxilim© software 14 ceph variables (12 angles and 2 linear ratios)	1 observer × 5	Pearson correlation coefficient, measurement error, non-parametric statistics	Intra-observer: i-CAT = 0.42–0.98 Iluma 0.43–0.99 Reproducibility: Iluma > i-Cat, sig diff (8 measurements)	Low
Damstra et al. (2011)	*In vivo*	25 patients	None	CBCT SimPlant Ortho© software 13 unilateral landmarks 5 bilateral landmarks 3 planes	2 observers × 2	The standard error, ANOVA, SDD for intra-observer and interobserver measurement errors, intra- and interobserver reliability: ICC, Shapiro–Wilks tests, non-parametric tests, Wilcoxon signed rank tests	Intra-observer: ICC 0.86–0.99. Interobserver: ICC > 0.88 except for palatal plane-mandibular plane (ICC, 0.76) SDD between observers: similar Measurement error: Angular = 0.88–6.29 mm Linear = 1.33–3.56 mm	Low

***Table 3.* (continued)**

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level ofevidence
Olmez et al. (2011)	*In vitro*	13 dry skulls	6-inch digital caliper	Multi-slice CT Mimics© 9.0 software Lateral and frontal cephalograms with metal markers 18 ceph landmarks 29 measurements (17 lateral and 12 frontal) 3 groups	1 observer × 1	one-way ANOVA, Tukey's Honestly Significant Difference tests, paired *t*-test, Dahlberg's formula	Computer-assisted 3D vs. physical: NS Conventional 2D vs. physical: sig diff The greatest magnification = Nasion-Menton distance on 2D (14.6%), ($p < 0.01$) The greatest enlargement = distance between the zygomaticomaxillary sutures on 2D (16.2%), ($p < 0.01$)	Low
Damstra et al. (2011)	*In vitro*	10 dry skulls		CBCT Lat ceph SimPlant Ortho© Pro 2.0 software Viewbox© version 3.1.1.13 software Spherical metal markers	1 observer × 2 (on the 2nd radiographs)	Dahlberg's formula for ME, ICC, mean values, Wilcoxon signed rank tests	Intra-observer: ICC 2D > 0.97 ICC 3D > 0.88 2D vs. 3D: NS ($p = 0.41$–1.00). Mean ME < 0.61	Low
Gribel et al. (2011)a	*In vitro*	25 dry skulls	Digital caliper	CBCT Lat ceph SimPlant Ortho© Pro 2.0 software Ten fiducial markers 12 ceph landmarks 10 linear measurements	1 observer × 2	ME: ICC, ANOVA	ME: CBCT ICC = 0.99 Direct measurement ICC = 0.98 lat ceph ICC = 0.98 CBCT vs. direct measurements: NS (ANOVA, $p > 0.05$). Direct vs. ceph: sig diff (Tukey's test, $p < 0.05$)	Low

Table 3. (continued)

Authors	Type of study	Sample size	Reference method	M&M	Observers	Statistical method	Results according to authors	Level ofevidence
Gribel et al. (2011b)	*In vivo*	13 patients	None	CBCT Lat ceph Mimics© version 8.13 software 6 points 2 angular and 4 linear measurements 4 groups: (1) ceph measurement (2) magnification correction (3) algorithm correction (4) CBCT measurement	2 observers × 2	Dahlberg's formula, mean values, R-ANOVA, *post hoc* comparison with Bonferroni corection	Intra-observer: ceph 0.5 mm CBCT 0.2 mm Group 1 vs. group 4: sig diff (for 3 measurements) ($p < 0.05$) Group 2: ↑differences Group 3 vs. group 4: NS	Low
Lopes et al. (2008)	*In vitro*	28 dry skulls	Beyond Crysta-C9168 series 900 device for 3D coordinates	Multi-slice CT Craniometric anatomical points 6 angular measurements	2 observers × 2	Precision, accuracy	Interobserver: NS Intra-observer: NS Difference physical vs. 3D: −3.16% to −0.10%	Moderate
Cavalcanti et al. (2004)	*In vitro*	13 cadaver heads	Electromagnetic 3 Space™ digitizer	Spiral CT Vitrea© version 2.3 software Anthropometric anatomical points	2 observers × 2	Mean absolute error percentage, SD, CI, intra- and interobserver reliability, ANOVA	Intra- vs. interobserver: NS Imaging vs. physical: NS Standard error 0.45–1.44% % Error: physical vs. 3D Bone = 0.83% Soft tissue = 1.78%	Moderate

CT, computed tomography; CBCT, cone-beam computed tomography; ceph, cephalometric; lat ceph, lateral cephalogram; SD, standard deviation; CI, confidence interval; ANOVA, analysis of variance; sig diff, statistically significant difference; NS, no statistically significant difference; ICC, Intraclass correlation coefficient; SDD, smallest detectable difference; ME, method error; ↑, higher; ↓, lower.

gold standard and were consequently considered as low level of evidence.

Studies reporting on observer performance, ICC of 3D measurements ranged from 0.86–0.99 for intra-observer and from 0.76–0.99 for interobserver reliability (31, 32, 34, 47). Some studies used Pearson correlation coefficients and the results showed that the coefficients ranged between 0.42–0.98 (average ICC around 0.89–0.91) (54–56). Two studies reported observer performance as mean difference in measurements (27, 35).

Lopes et al. (40) reported on precision and accuracy of six angular measurements. Difference between physical and 3D measurement ranged between −3.16% and −0.10%. In the study of Cavalcanti et al., (28) the results of the comparison between 2D-CT, 3D-CT volume rendering and physical measurements showed that the error between mean physical measurement and mean 3D-based measurements was 0.83% for bone and 1.78% for soft tissue. Both studies reported no statistically significant differences in the inter- and intra-observer reliability.

Concerning studies with low level of evidence, some studies reported the accuracy of measurements. In 2008, Periago et al. found statistical differences between CBCT means and true dimensions for all of the midsagittal measurements except Nasion-A point and 6 of the 12 bilateral measurements (47). In a study published by Brown et al. in 2009, there were no differences between 3D CBCT and actual skull measurements for six dimensions. CBCT produced generally lower measurements than skull values (27).

Facial asymmetry

Three publications with low level of evidence were found (Table S4). The studies were performed *in vitro* with only 1 observer and without a standard reference. In 2009, Van Vlijmen et al. used 40 dry skulls to test the intra-observer reliability of conventional frontal cephalogram and the 3D cephalometry. The correlation coefficient of the intra-observer reliability ranged from 0.23–0.99 (average = 0.71) for the conventional frontal radiograph and from 0.42–0.93 (average = 0.79) for the 3D models (53).

In 2011, Yanez et al. published a study on asymmetry index, using patients CT data. The results revealed the intra-observer error to be 0.78, 1.05 and 1.07 mm for x, y, and z coordinates. The errors of the linear and angular measurement were 1.36 and 0.91°, respectively (57).

Damstra et al. (30), presented a study on the morphometric method to determine the midsagittal plane on 14 dry human skulls. No statistically significant difference was found between the measurements (p = 0.25–0.97). The agreement was high (r = 0.85–1.00) and the method error was small (mean = 0.39 mm; 95% CI = 0.31–0.47 mm).

Other topics

Five publications that could not be categorized into any topics above are reported in Table S5. Two publications: Olszewski et al. (45) and Lagravère et al. (37) reached the moderate level of evidence. Other studies did not meet the criteria because the reference standard was missing and were therefore considered as low level of evidence.

Olszewski et al. transformed Delaire's 2D cephalometric analysis into a 3D version. The authors validated the system on 26 dry skulls. For the intra-observer reliability, the ICC, found for 2D X-ray was 0.60–0.91 and the ICC for 3D CT was 0.97–1.00. When looking at the interobserver reliability, the ICC varied from 0.13 to 0.84 and from 0.94 to 1.00, respectively. In the 3D CT, the user accuracy (absolute difference) was between 0.75 mm (±0.05) and 0.99 mm (±0.08). There were no statistically significant differences found between the physical measurements and the measurements in ACRO 3D® software (45).

Later on, Lagravère et al. (2011) published an article on the reliability and accuracy in locating several foramina in the cranial base by CBCT images. The ICC values were found to be >0.93 and 0.92 for intra- and interobserver reliability, respectively. From this study, the authors concluded that the foramen spinosum, ovale, and rotundum, as well as hypoglossal canal could be considered as acceptable landmarks to be used in establishing reference coordinate systems for future 3D superimposition analysis (37).

In 2006, Swennen and colleagues presented a new 3D cephalometric reference system and tested the accuracy and reliability of this analysis. The intra-observer measurement error was less than 0.88 mm, 0.76 mm and 0.84 mm for horizontal, vertical and transversal orthogonal measurements, respectively. The interobserver measurement error was less than 0.78 mm, 0.86 mm and 1.26 mm for horizontal, vertical and transversal orthogonal measurements, respectively (17).

Park et al. (46) proposed a new type of cephalometric analysis by using 3D CT in 2006. The authors reported that there was no statistically significant difference when the data were compared with the Korean norm values. All landmarks were reproducible and no significant intra-observer error ($p > 0.01$) was found (46).

Tulunoglu et al. (52) compared the consistency of orthodontic measurements performed on lateral and frontal cephalograms and 3D CT images of cleft lip and palate (CLP) patients. The ICC values were very high for both 2D (0.94–0.99) and 3D measurements (0.88–0.99). Significant differences were found between the measurements from 2D and from 3D methods (52).

Discussion

Summary of evidence

No publication reported diagnostic thinking efficacy or therapeutic efficacy.

Meta-analysis was not possible because there was a lack of primary studies and the heterogeneity of the studies. This review was therefore limited to a qualitative descriptive analysis. Based on the QUADAS tools, six publications have reached the moderate level of evidence, and therefore, it is considered that there is limited research-based evidence on 3D cephalometry.

The most common reason for exclusion of publications was adequate sample size (samples less than 10) (Fig. 1). As the sample size is important for meaningful statistics, we determined minimum sample size equal to 10 as an inclusion criterium.

The reported research findings were most frequently on landmark identification and measurement accuracy. The results of the observer performance were reported in almost all studies. The reliability and reproducibility of the methods and the observers were highly interesting as these are the main factors influencing the cephalometric analysis. It was shown that the 3D landmark identification and measurements were as reliable or more reliable than traditional 2D cephalometric measurements (29, 34, 35, 39), but there was not always full agreement (39, 54). This may depend on the method and analysis, selected in the studies. Even different machine selection may play a role in the difference between the measurements found (55). Different landmarks have shown differences in their reliability, reproducibility and precision in the 3D space (49, 50, 58).

Several researchers have reported studies on facial asymmetry, and a few studies were included in this review (30, 53, 57). Three-dimensional imaging offers a better representation of the real morphology of the skulls unlike in the lateral cephalogram, where left and right structures are superimposed on each other. Frontal cephalograms or postero/anterior views have been used for decades to evaluate symmetry of facial structures, but 3D cephalometry has recently been reported to prove its validity although the evidence is still very limited.

Research-based evidence on 3D cephalometry was found to be limited. More evidence was found for measurement accuracy and reproducibility of landmark identification, but still there is not enough evidence about 3D cephalometry in other aspects.

Limitations

Cephalometry is a widely used method in orthodontics and orthognathic surgery. A variety of research topics was identified such as landmark identification, linear and angular measurements, facial asymmetry, cleft lip and cleft palate, introductions of new analyses and the transformation from the traditional 2D analysis to the new 3D analysis. This made it difficult to perform a systematic review as topics, statistical tests and methods were too diverse. Therefore, a meta-analysis could not be performed.

In general, cephalometric analysis is used to analyse the craniofacial structures of the patient and its results have an impact on treatment planning. Cephalometry is not a direct method to diagnose the conditions of the patients, yet it offers the details of the patients' craniofacial structures and thus reveals diagnostic information helpful in determining orthodontic treatment planning. To perform this systematic review, the authors followed the procedure of the systematic review on diagnostic accuracy as it shows the most similarity.

A cephalometric analysis is basically performed on radiographs, either 2D or 3D. Osseous landmarks both on the skull surface and inside the skull are identified on the images. As a result, it is impossible to check the real landmark positions in patients. The reference standard can only be used in an *in vitro* study, which is not an ideal study type to report the diagnostic efficacy evidence. A cadaver study could overcome this problem, but the sample size will be limited. Studies on patients can only show the observer performance and reliability of the methods as direct physical skeletal measurements are not possible. Although in some publications the reference standards were calculated based on the measurements on the images, it is not preferable according to the QUADAS tool (25).

Another limitation in *in vivo* studies is the ethical issue on radiation safety for orthodontic patients. Although CBCTs offer 3D images with less radiation dose than the multislice CT, it is still too high to perform a controlled trial comparing 2D vs. 3D cephalometry.

Within the limitations, mentioned previously, the authors have tried to adapt the inclusion criteria and protocol of the present systematic review to cover all the evidence provided by current publications. The protocol, used in this study is not as restricted as the standard one as described by Fryback and Thornbury (20). When a study did not use a reference standard, it was not excluded immediately but could still meet the low level of evidence if the study reached other criteria.

Non-English articles were searched during the literature search and study selection, but they were not further considered in the present systematic review. From 571 publications that were retrieved, 48 non-English publications were found (9 Chinese, 2 Dutch, 15 French, 12 German, 1 Hungarian, 3 Italian, 5 Japanese, 1 Polish). Titles and abstracts of these publications were checked for the inclusion criteria. It was found that all articles did not pass these criteria with reasons: not related to 3D cephalometry (33), review literature (5), no quantitative analysis (4), method not clearly described (2), abstract not available (4). Therefore, it is very unlikely to identify other relevant non-English publications than included in this review.

3D cephalometry and the future

There are several concerns using 3D cephalometric analysis, and these concerns can also affect how future studies should be conducted. First concern is related to the selection criteria and thus to outlining when to perform 3D cephalometry. Discussion is going on regarding case selection and the necessity of 3D cephalometry because radiation exposure plays a role in this decision making process. With the current evidence, it cannot be concluded that 3D cephalometry should be performed on all orthodontic patients. Guidelines and recommendations on CBCT for dental and maxillofacial radiology by the European Commission are available and should be followed (59). Recently, new guidelines were proposed by the by the American Academy of Oral and Maxillofacial Radiology (AAOMR) to provide clinicians and more specifically orthodontic specialists some guidance and recommendations in using cone-beam CT (60).

The radiation dose is a very important issue as most orthodontic patients are children and adolescents who are more sensitive to radiation exposure (61, 62). The radiation dose, received from the CBCTs, is strongly related to FOV size and also dependent on the specific CBCT machine (61). This also acts as a difficulty in conducting research on 3D cephalometry. To obtain a large FOV, CT or CBCT scan is difficult because of the ethical concern. Most studies with suitable gold standard were performed *in vitro* which is not fully accepted as evidence as already mentioned previously in the *limitations*. In the *in vitro* studies, more effort should be taken to simulate real human conditions such as the use of water as soft-tissue attenuation, which was carried out only in a few studies, or to develop a material that mimics soft-tissue shape and density to make analysis on the soft tissues possible.

At this point, 3D cephalometric analyses were mostly based on their former 2D versions. Further tests should be performed to evaluate the reliability and accuracy of these 2D transferred to 3D methods and to investigate whether the norm values from 2D cephalometry can still be used in these new 3D analyses.

Three-dimensional cephalometry is fairly new as a research topic. More studies are highly required to provide more evidence on the accuracy and the efficacy of this potentially innovative method. As for the future, researchers should concentrate more on the materials and methods of their 3D cephalometric studies, standardizing protocols, using larger sample sizes and employing more optimal statistical methods for data set evaluation. Studies on diagnostic thinking efficacy (testing whether 3D cephalometry is helpful for diagnosis) and therapeutic efficacy (testing whether 3D cephalometry is helpful for treatment planning and the management of the patients) (20) should also be accomplished to offer more concrete evidence on the benefits of 3D cephalometry for orthodontic treatment planning and patients.

Conclusions

This systematic review reveals that there is still limited research-based evidence on 3D cephalometry. Specific research methodology needs to be developed to be able to perform more standardized diagnostic accuracy studies by using patients' data.

Funding sources

This study was accomplished by a doctoral scholarship in the framework of the Interfaculty Council for Development Co-operation (IRO). The funders played no role in study design, collection, analysis, interpretation of the data, writing of the report, or in the decision to submit the paper for publication. They accept no responsibility for the contents.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements: The authors would like to thank Jens De Groot, librarian at the Biomedical Sciences Campus Library, Katholieke Universiteit Leuven for assisting the development of the search strategies.

Also, we would like to thank Professor Filip Cools and Trudy Bekkerring from the Belgian Branch of the Dutch Cochrane Centre and the Belgian Centre of Evidence-Based Medicine for their advice on the systematic review process.

References

1. Hofrath H. Bedeutung der Röntgenfern und Abstands Aufnahme fur die Diagnostik der Kieferanomalien. *Fortschr Orthod* 1931;1:231–58.
2. Broadbent BH. A new x-ray technique and its application to orthodontia. *Angle Orthodontists* 1931;1:45–66.
3. Baumrind S, Frantz RC. The reliability of head film measurements. 2. Conventional angular and linear measures. *Am J Orthod* 1971;60:505–17.
4. Baumrind S, Frantz RC. The reliability of head film measurements. 1. Landmark identification. *Am J Orthod* 1971;60:111–27.
5. Bjork A, Solow B. Measurement on radiographs. *J Dent Res* 1962;41:672–83.
6. Connor SE, Arscott T, Berry J, Greene L, O'Gorman R. Precision and accuracy of low-dose CT protocols in the evaluation of skull landmarks. *Dentomaxillofac Radiol* 2007;36:270–6.
7. Olszewski R, Reychler H, Cosnard G, Denis JM, Vynckier S, Zech F. Accuracy of three-dimensional (3D) craniofacial cephalometric landmarks on a low-dose 3D computed tomograph. *Dentomaxillofac Radiol* 2008;37:261–7.

8. Olszewski R, Frison L, Vynckier S, Cosnard G et al. Reproducibility of three-dimensional cephalometric landmarks in cone-beam and low-dose computed tomography. *Clin Oral Investig* 2013;17: 285–92.
9. Kumar V, Ludlow JB, Mol A, Cevidanes L. Comparison of conventional and cone-beam CT synthesized cephalograms. *Dentomaxillofac Radiol* 2007;36:263–9.
10. Moshiri M, Scarfe WC, Hilgers ML, Scheetz JP, Silveira AM, Farman AG. Accuracy of linear measurements from imaging plate and lateral cephalometric images derived from cone-beam computed tomography. *Am J Orthod Dentofacial Orthop* 2007;132:550–60.
11. Cattaneo PM, Bloch CB, Calmar D, Hjortshoj M, Melsen B. Comparison between conventional and cone-beam computed tomography-generated cephalograms. *Am J Orthod Dentofacial Orthop* 2008;134:798–802.
12. van Vlijmen OJ, Berge SJ, Swennen GR, Bronkhorst EM, Katsaros C, Kuijpers-Jagtman AM. Comparison of cephalometric radiographs obtained from cone-beam computed tomography scans and conventional radiographs. *J Oral Maxillofac Surg* 2009;67:92–7.
13. Kumar V, Ludlow J, Soares Cevidanes LH, Mol A. In vivo comparison of conventional and cone beam CT synthesized cephalograms. *Angle Orthod* 2008;78:873–9.
14. Grayson B, Cutting C, Bookstein FL, Kim H, McCarthy JG. The three-dimensional cephalogram: theory, technique, and clinical application. *Am J Orthod Dentofacial Orthop* 1988;94:327–37.
15. Adams GL, Gansky SA, Miller AJ, Harrell WE Jr, Hatcher DC. Comparison between traditional 2-dimensional cephalometry and a 3-dimensional approach on human dry skulls. *Am J Orthod Dentofacial Orthop* 2004;126:397–409.
16. Chenin DL. 3D cephalometrics: the new norm. *Alpha Omegan* 2010;103:51–6.
17. Swennen GR, Schutyser F, Barth EL, De Groeve P, De Mey A. A new method of 3-D cephalometry Part I: the anatomic Cartesian 3-D reference system. *J Craniofac Surg* 2006;17:314–25.
18. Farronato G, Farronato D, Toma L, Bellincioni F. A synthetic three-dimensional craniofacial analysis. *J Clin Orthod*. 2010;44:673–8; quiz 88.
19. Kau CH, Richmond S, Palomo JM, Hans MG. Three-dimensional cone beam computerized tomography in orthodontics. *J Orthod* 2005;32:282–93.
20. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88–94.
21. Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 1.0: The Cochrane Collaboration; 2010. Available from: http://srdta.cochrane.org/handbook-dta-reviews.
22. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration; 2011. Available from: http://www.cochrane-handbook.org/.
23. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994;271:389–91.
24. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703–7.
25. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
27. Brown AA, Scarfe WC, Scheetz JP, Silveira AM, Farman AG. Linear accuracy of cone beam CT derived 3D images. *Angle Orthod* 2009;79:150–7.
28. Cavalcanti MG, Rocha SS, Vannier MW. Craniofacial measurements based on 3D-CT volume rendering: implications for clinical applications. *Dentomaxillofac Radiol* 2004;33:170–6.
29. Chien PC, Parks ET, Eraso F, Hartsfield JK, Roberts WE, Ofner S. Comparison of reliability in anatomical landmark identification using two-dimensional digital cephalometrics and three-dimensional cone beam computed tomography in vivo. *Dentomaxillofac Radiol* 2009;38:262–73.
30. Damstra J, Fourie Z, De Wit M, Ren Y. A three-dimensional comparison of a morphometric and conventional cephalometric midsagittal planes for craniofacial asymmetry. *Clin Oral Investig* 2012;16:285–94.
31. Damstra J, Fourie Z, Huddleston Slater JJ, Ren Y. Reliability and the smallest detectable difference of measurements on 3-dimensional cone-beam computed tomography images. *Am J Orthod Dentofacial Orthop* 2011;140:e107–14.
32. Damstra J, Fourie Z, Ren Y. Comparison between two-dimensional and midsagittal three-dimensional cephalometric measurements of dry human skulls. *Br J Oral Maxillofac Surg* 2011;49:392–5.
33. de Oliveira AE, Cevidanes LH, Phillips C, Motta A, Burke B, Tyndall D. Observer reliability of three-dimensional cephalometric landmark identification on cone-beam computerized tomography. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2009;107:256–65.
34. Gribel BF, Gribel MN, Frazao 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.
35. Gribel BF, Gribel MN, Manzi FR, Brooks SL, McNamara JA Jr. From 2D to 3D: an algorithm to derive normal values for 3-dimensional computerized assessment. *Angle Orthod* 2011;81:3–10.
36. Hassan B, Nijkamp P, Verheij H, Tairie J, Vink C, Stelt PV et al. Precision of identifying cephalometric landmarks with cone beam com-

puted tomography in vivo. *Eur J Orthod* 2013;35:38–44.
37. Lagravère MO, Gordon JM, Flores-Mir C, Carey J, Heo G, Major PW. Cranial base foramen location accuracy and reliability in cone-beam computerized tomography. *Am J Orthod Dentofacial Orthop* 2011;139:e203–10.
38. Lagravère MO, Gordon JM, Guedes IH, Flores-Mir C, Carey JP, Heo G et al. Reliability of traditional cephalometric landmarks as seen in three-dimensional analysis in maxillary expansion treatments. *Angle Orthod* 2009;79:1047–56.
39. Lagravère MO, Low C, Flores-Mir C, Chung R, Carey JP, Heo G 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.
40. Lopes PM, Moreira CR, Perrella A, Antunes JL, Cavalcanti MG. 3-D volume rendering maxillofacial analysis of angular measurements by multislice CT. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2008;105:224–30.
41. 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 e1–10; discussion 12-3.
42. Tyndall DA, Kohltfarber H. Application of cone beam volumetric tomography in endodontics. *Austral Dent J* 2012;57(Suppl 1):72–81.
43. Fullmer JM, Scarfe WC, Kushner GM, Alpert B, Farman AG. Cone beam computed tomographic findings in refractory chronic suppurative osteomyelitis of the mandible. *Br J Oral Maxillofac Surg* 2007;45:364–71.
44. Olszewski R, Tanesy O, Cosnard G, Zech F, Reychler H. Reproducibility of osseous landmarks used for computed tomography based three-dimensional cephalometric analyses. *J Craniomaxillofac Surg* 2010;38:214–21.
45. Olszewski R, Zech F, Cosnard G, Nicolas V, Macq B, Reychler H. Three-dimensional computed tomography cephalometric craniofacial analysis: experimental validation in vitro. *Int J Oral Maxillofac Surg* 2007;36:828–33.
46. Park SH, Yu HS, Kim KD, Lee KJ, Baik HS. A proposal for a new analysis of craniofacial morphology by 3-dimensional computed tomography. *Am J Orthod Dentofacial Orthop* 2006;129(600):e23–34.
47. Periago DR, Scarfe WC, Moshiri M, Scheetz JP, Silveira AM, Farman AG. Linear accuracy and reliability of cone beam CT derived 3-dimensional images constructed using an orthodontic volumetric rendering program. *Angle Orthod* 2008;78:387–95.
48. Richtsmeier JT, Paik CH, Elfert PC, Cole TM 3rd, Dahlman HR. Precision, repeatability, and validation of the localization of cranial landmarks using computed tomography scans. *Cleft Palate Craniofac J* 1995;32:217–27.
49. Shibata M, Nawa H, Kise Y, Fuyamada M, Yoshida K, Katsumata A et al. Reproducibility of three-dimensional coordinate systems based on craniofacial landmarks. *Angle Orthod* 2012;82:776–84.
50. Schlicher W, Nielsen I, Huang JC, Maki K, Hatcher DC, Miller AJ. Consistency and precision of landmark identification in three-dimensional cone beam computed tomography scans. *Eur J Orthod* 2012;34:263–75.
51. Titiz I, Laubinger M, Keller T, Hertrich K, Hirschfelder U. Repeatability and reproducibility of landmarks–a three-dimensional computed tomography study. *Eur J Orthod* 2012;34:276–86.
52. Tulunoglu O, Esenlik E, Gulsen A, Tulunoglu I. A comparison of three-dimensional and two-dimensional cephalometric evaluations of children with cleft lip and palate. *Eur J Dent* 2011;5:451–8.
53. van Vlijmen OJ, Maal TJ, Berge SJ, Bronkhorst EM, Katsaros C, Kuijpers-Jagtman AM. A comparison between two-dimensional and three-dimensional cephalometry on frontal radiographs and on cone beam computed tomography scans of human skulls. *Eur J Oral Sci* 2009;117:300–5.
54. van Vlijmen OJ, Maal T, Berge 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.
55. van Vlijmen OJ, Rangel FA, Berge SJ, Bronkhorst EM, Becking AG, Kuijpers-Jagtman AM. Measurements on 3D models of human skulls derived from two different cone beam CT scanners. *Clin Oral Investig* 2011;15:721–7.
56. Waitzman AA, Posnick JC, Armstrong DC, Pron GE. Craniofacial skeletal measurements based on computed tomography: Part II. Normal values and growth trends. *Cleft Palate Craniofac J* 1992;29:118–28.
57. Yanez-Vico RM, Iglesias-Linares A, Torres-Lagares D, Gutierrez-Perez JL, Solano-Reina E. Three-dimensional evaluation of craniofacial asymmetry: an analysis using computed tomography. *Clin Oral Investig* 2011;15:729–36.
58. Zamora N, Llamas JM, Cibrian R, Gandia JL, Paredes V. A study on the reproducibility of cephalometric landmarks when undertaking a three-dimensional (3D) cephalometric analysis. *Med Oral Patol Oral Cir Bucal* 2012;17:e678–88.
59. European Commission. Guidelines on CBCT for Dental and Maxillofacial Radiology: Evidence Based Guidelines: Radiation Protection Publication; 2012. Available from: http://www.sedentexct.eu/content/guidelines-cbct-dental-and-maxillofacial-radiology.
60. American Academy of Oral and Maxillofacial Radiology. Clinical recommendations regarding use of cone beam computed tomography in orthodontic treatment. Position statement by the American Academy of Oral and Maxillofacial Radiology. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2013;116:238–57.
61. Pauwels R, Beinsberger J, Collaert B, Theodorakou C, Rogers J, Walker A et al. Effective dose range for dental cone beam computed tomography scanners. *Eur J Radiol* 2012;81:267–71.
62. Theodorakou C, Walker A, Horner K, Pauwels R, Bogaerts R, Jacobs R et al. Estimation of paediatric organ and effective doses from dental cone beam CT using anthropomorphic phantoms. *Br J Radiol* 2012;85:153–60.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Search strategies for MEDLINE via PubMed and EMBASE via embase.com (database search was last performed on 13 March 2012).

Table S2. Protocol 1 for primary data extraction.

Table S3. Protocol 2 for quality assessment of publications.

Table S4. Facial asymmetry.

Table S5. Other topics.