Systematic Review Article

Reliability of different three-dimensional cephalometric landmarks in cone-beam computed tomography: *A systematic review*

Alycia Sam[a]; Kris Currie[b]; Heesoo Oh[c]; Carlos Flores-Mir[d]; Manuel Lagravére-Vich[e]

ABSTRACT

Objectives: Conventional two-dimensional (2D) cephalometric radiography is an integral part of orthodontic patient diagnosis and treatment planning. One must be mindful of its limitations as it indeed is a 2D representation of a vaster three-dimensional (3D) object. Issues with projection errors, landmark identification, and measurement inaccuracies impose significant limitations, which may now be overcome with the advent of cone-beam computed tomography (CBCT). A systematic review of the reliability of different 3D cephalometric landmarks in CBCT imaging was conducted.

Materials and Methods: Electronic database searches were administered until October 2017 using PubMed, MEDLINE via OvidSP, EBMR and EMBASE via OvidSP, Scopus, and Web of Science. Google Scholar was used as an adjunctive search tool.

Results: Thirteen articles considering CBCT scans of human subjects from preexisting data sets were selected and reviewed. Most of the studies had methodological limitations and were of moderate quality. Because of their heterogeneity, key data from each could not be combined and were reported qualitatively. Overall, in 3D, midsagittal plane landmarks demonstrated greater reliability compared with bilateral landmarks. A minimum number of dental landmarks were reported, although most were recommended for use.

Conclusions: Further research is required to evaluate the reliability of 3D cephalometric landmarks when evaluating 3D craniofacial complexes. (*Angle Orthod.* 2019;89:317–332.)

KEY WORDS: Cone-beam computed tomography; Cephalometrics; Reliability

INTRODUCTION

Cephalometric radiography is a standardized radiographic technique employed to provide a better understanding of an individual's craniofacial structures in three planes of space: anteroposteriorly (AP), vertically, and transversely. Landmarks routinely used in two-dimensional (2D) lateral cephalometric analyses are chosen based on their ability to be reliably identified.[1] Distances/angles between these landmarks are measured and then compared with one or various sets of standardized norms that provide an indication of relationships shared within the craniofacial complex of an individual at a given time. Radiographic findings are then compared with clinical findings. Issues with image distortion and superimposition of bilateral structures may pose significant limitations to the interpretation of these data.[2,3] Sometimes, a 2D PA cephalogram is a valuable adjunct to routine orthodontic diagnosis and treatment planning as it provides invaluable information, especially in the transverse direction, eliminating superimposition of certain bilateral structures that

[a] Graduate student, Orthodontics, Department of Dentistry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada.

[b] Private practice, Saskatoon, Saskatchewan, Canada.

[c] Professor and Program Director, Arthur A. Dugoni School of Dentistry, University of Pacific, San Francisco, Calif.

[d] Professor and Head, Orthodontics, Department of Dentistry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada.

[e] Assistant Professor, Orthodontics, Department of Dentistry, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Canada.

Corresponding author: Dr Manuel Lagravére-Vich, Department of Dentistry, Faculty of Medicine and Dentistry, University of Alberta, ECHA 5-524, 11405-87 Avenue, Edmonton, Alberta, T6G 1C9, Canada
(e-mail: mlagravere@ualberta.ca)

Accepted: September 2018. Submitted: April 2018.
Published Online: November 13, 2018

eases detection of potential facial asymmetries. Even though chosen landmarks may be easily identified and reproducible, it is imperative to question their true meaningfulness as this transverse dimension is often unaccounted for without additional imaging.

Limitations once imposed by 2D may now be overcome: volumetric data contained within voxels of a single 360° cone-beam computed tomography (CBCT) scan is instrumental in reconstructing and understanding skeletal, dental, and soft tissue drape relationships in three dimensions (3D).[3] The accuracy of landmark identification and placement is now enhanced as each occupies a specific location along a coordinate system of x-y-z axes. As such, it is possible that theoretical discrepancies existing between 2D and 3D cephalometric analyses are attributed to the fact that measurements are made between two lines in the prior whereas, alternatively, CBCT imaging affords the possibility for measurements to be made between two planes.[4]

While most "old" 2D landmarks are reliable for use in 3D cephalometric analyses, specific nerve foramina in the maxilla and mandible provide better landmarks in 3D imaging.[5] These include mental foramina and infraorbital foramina, which are more reliable and reproducible than others. However, the obliquity of infraorbital foramina and oral incisive foramina tends to pose challenges as it can make locating their center point difficult.[5]

A prior systematic review (SR) examining the reliability and reproducibility of 3D cephalometric landmarks using CBCT was published in 2014 by Lisboa et al.[6] Their search ended much earlier, in October 2014; hence, the decision was made to further explore this area based on the increasing popularity of CBCT imaging and corresponding significance demonstrated by the abundance of scientific studies increasingly available every day. As such, the search period for the current SR was vaster, inclusive from 1998 (first introduction of CBCT into dentistry) to October 2017. Databases intended for the search were also more widespread than those previously considered. At least 17 additional articles were reviewed in the second selection phase, published between October 2014 and October 2017. The purpose of this SR was to investigate the available scientific literature to evaluate the reliability of different 3D cephalometric landmarks in CBCT imaging.

MATERIALS AND METHODS

Protocol and Registration

This SR followed as closely as possible the methodology detailed by the PRISMA guidelines[7] for the transparent reporting of SRs and meta-analyses.

Eligibility Criteria

An extensive search of available scientific literature was carried out electronically, with only those studies that examined the reliability of 3D cephalometric anatomic landmarks using CBCT considered for review. No language or study restrictions were placed. Unpublished materials were not excluded.

Information Sources

Databases searched included PubMed, MEDLINE via OvidSP, EBMR and EMBASE via OvidSP, Scopus, and Web of Science. To ensure that a wide range of academic literature was well represented, Google Scholar was used as an adjunctive search tool to discover other scholarly sources that may have existed. The first 100 relevant hits were evaluated from this "gray literature" and considered for inclusion.

Search

Strategic design was developed through consultation with a health sciences research librarian using appropriate keywords and their combinations. The full electronic search strategy for each database is illustrated in Table 1.

Study Selection

Evaluation of selected articles was staged in a two-step process to determine eligibility. First, each individual article title and abstract was screened by two reviewers (Dr Sam, Dr Currie) independently. The aim of this step was to ensure each article pertained to the following topics: 3D imaging, anatomic landmarks, cephalometric analysis, and accuracy and/or reliability of findings. Next, decisions for final eligibility were made based on full-text assessments by the same reviewers. They were not blinded to the authors nor results of the studies. Any disagreements between reviewers were resolved by discussion or by introduction of a third reviewer (Dr Lagravére-Vich) to mediate when deemed necessary.

Data Collection

Data collection was done in duplicate. Key features of eligible articles were documented by each reviewer. Statistical results and conclusions of every study were also retrieved.

Risk of Bias Among Included Studies

Individual articles then underwent a methodological quality scoring, adapted from a process described in a previous related study with modifications based on a

Table 1. Search Strategies for Various Electronic Databases

Database	Keywords	Results
PubMed	(((anatomic landmarks[MeSH Terms]) AND cephalometry[MeSH Terms]) AND (cone beam computed tomography OR imaging, three-dimensional OR anatomy, cross-sectional[MeSH Terms])) AND (dimensional measurement accuracy OR reproducibility of results[MeSH Terms]) AND ((“1998/01/01”[PDat] : “2017/10/01”[PDat]) AND Humans[Mesh])	68
Medline (via OvidSP)	Search group 1: anatomic landmarks.mp. OR exp Anatomic Landmarks/ Search group 2: cephalometry.mp. OR exp Cephalometry/ OR craniometry.mp. Search group 3: cone beam computed tomography.mp. OR exp Cone-Beam Computed Tomography/ or cone-beam CAT scan.mp. OR cone beam computerized tomography.mp. OR volumetric computed tomography.mp. OR CBCT.mp. or digital volume tomography.mp. OR DVT.mp. or imaging, three-dimensional.mp. OR exp Imaging, Three-Dimensional/ or three dimensional image.mp. OR 3D imaging.mp. OR three-dimensional computer assisted.mp. Search group 4: dimensional measurement accuracy.mp. OR exp Dimensional Measurement Accuracy/ or reproducibility of results.mp. OR exp “Reproducibility of Results”/ or Bland-Altman.mp. OR reliability.mp. OR validity.mp. OR precision.mp. OR reproducibility of findings.mp OR intraclass correlation coefficient.mp. Search group 5: 1 AND 2 AND 3 AND 4	67
EBMR (via OvidSP)	Same as for Medline (via OvidSP)	0
EMBASE (via OvidSP)	Same as for Medline (via OvidSP)	54
Scopus	((“imaging, three-dimensional” OR “three dimensional image” OR “3D imaging” OR “three-dimensional computer-assisted” OR three-dimensional) AND PUBYEAR > 1997) AND ((“cone-beam computed tomography” OR “cone-beam computerized tomography” OR cbct OR “volumetric computed tomography” OR “digital volume tomography” OR dvt OR cone-beam OR cone) AND PUBYEAR > 1997) AND ((cephalometr* OR craniometr*) AND PUBYEAR > 1997) AND ((“anatomic* landmark” OR landmark OR structure) AND PUBYEAR > 1997)	976
Web of Science	Set #1: TS=(cephalometr*) OR TS=(craniometr*) *DocType=All document types; Language=All languages;* Set #2: TS=(anatomic* landmark) *DocType=All document types; Language=All languages;* Set #3: TS=(reproducibil*) OR TS=(reliabil*) OR TS=(precision) OR TS=(valid*) OR TS=(accura*) OR TS=(intraclass correlation coefficient) OR TS=(Bland-Altman) *DocType=All document types; Language=All languages;* Set #4: TS=(cone-beam computed tomography) OR TS= (cone beam computed tomography) OR TS =(cone-beam CAT scan) OR TS= (volumetric computed tomography) OR TS=(CBCT) OR TS =(digital volume tomography) OR TS=(DVT) OR TS=(imaging, three-dimension*) OR TS =(imaging) OR TS=(three-dimension*) OR TS=(three dimension* image) OR TS=(3D imaging) OR TS=(three-dimension*) *DocType=All document types; Language=All languages;* Set #5: #4 AND #3 AND #2 AND #1 *DocType=All document types; Language=All languages;*	221

research methodology series for reliability articles.[8,9] Each criterion for judgment was open to discussion among reviewers with the aim of limiting the risk of bias and serving as a baseline for assessments. The way in which points were awarded is detailed in Table 2. Each article received a grading score and was then categorized per its overall quality of evidence and strength of its recommendations. Articles were categorized into groupings based on the methodological quality/magnitude of scoring: excellent or high (76% or more), good or moderate (51%–75%), and poor or low (50% or less). It must be noted that this is a nonvalidated assessment tool.

Synthesis of Results

A meta-analysis was not justifiable for this topic as studies were very diverse, both in study design and report of relevant findings. However, it may be possible to complete one in the future if reliability measures of only a limited number of landmarks are combined for applicable included studies.

RESULTS

Study Selection

A final total of 13 articles satisfied the selection criteria and were included in this review. A detailed

Table 2. Methodological Scoring for Reliability Articles[a]

I. Study design (15✓)
A. Objective: description of measurement or procedure under investigation (✓)
B. Objective: outline of what is known from previous studies (✓)
C. Sample characteristics: subjects described (✓)
D. Sample characteristics: assessors described (✓)
E. Sample size: subjects adequate (✓)
F. Sample size: assessors adequate (✓)
G. Sample representation: subjects representative of population (✓)
H. Sample representation: assessors representative of population (✓)
I. Sample qualifications/experience: all assessors with necessary experience (✓)
J. Sample subject variability: heterogeneous subjects (✓)
K. Minimization of random error: equipment described (✓)
L. Minimization of random error: subjects described (✓)
M. Minimization of random error: assessors described (✓)
N. Clinically stable subjects: yes (✓)
O. Period of time between measurements: adequate (✓)
II. Study measurements (4✓)
P. Measurement method: appropriate to the objective (✓)
Q. Blind measurement: blinding (✓)
R. Reliability: adequate level of intraobserver agreement (✓)
S. Reliability: adequate level of interobserver agreement (✓)
III. Statistical analysis (2✓)
T. Statistical analysis: appropriate for data (✓)
U. Confounders: confounders included in analysis (✓)
IV. Results (2✓)
V. Meaningfulness (eg, ICC, SEM, CI, kappa): provided (✓)
W. Generalized to clinical/research context: yes (✓)
MAXIMUM NUMBER OF ✓s = 23

[a] Adapted from Lagravere et al. (2005),[8] with modifications based on Bialocerkowski et al. (2010).[9]

outline of the selection process, from identification through articles included, is illustrated Figure 1.[10]

In comparison to the previously published SR,[6] this follow-up gained three additional articles[11–13] and excluded four considered in the previous. The search criteria of the previous study ended much earlier (October 2014) and combined reliability studies using both human patient scans and dry human skulls. Thus, any discrepancy between the prior and this latter SR reflect these differences. One of three additional articles[11] retrieved by this review yielded excellent or high methodological quality scoring. Considering it was only one of four included articles to obtain this scoring overall, it had the potential to offer very useful insight into this area of study. The second[12] introduced new landmarks and measurements to shift the traditional 2D cephalometric analysis paradigm toward a novel 3D one. Lastly, the third offered insights into the use of landmark-based superimposition in 3D.[14]

Study Characteristics

Selected articles were published between 2008 and 2017 in several diverse medical/dental journals. All were written in English apart from two: one was in French and another in Korean. These articles were obtained, although English versions were not accessible at the time and the decision was made for them to be excluded. All were retrospective and cross-sectional in nature (data collected before the research project). Summary characteristics of the included articles are described in detail in Table 3.[1–3,5,11–19]

The aim of all studies was to investigate first the reliability (intra- and/or interrater measurements) of anatomic landmarks in 3D cephalometric analysis, reported statistically with one or more of the following: intraclass correlation coefficients (ICCs), Bland-Altman testing, mean error and standard deviations, 3D scatterplots, and Pearson correlation coefficient.

The methodological assessment tool used is outlined in Table 2. The summary of the scores imparted to reliability articles is found in Table 4.[1–3,5,11–19] In general, weaknesses included inadequate description of sample characteristics of subjects (eg, sex, age, inclusion criteria, exclusion criteria, specific database used), no justification or calculation for sample sizes, and lack of explanation regarding dealing with cofounders such as exclusion criteria and employment of randomization.

Risk of Bias Within Studies

A possible source of bias within each article was based on timing of records. As all studies were retrospective and it is unethically sound to expose patients to radiation solely for research purposes, investigators were reliant on the use of preexisting data sets for subject populations. As this review was interested in reliability, it became problematic if a study utilized a set not representative of the spectrum of individuals to generalize findings in a research or clinical context. Although a few studies mentioned the use of randomization, few described how, and none reported using sequence generation within their data set to ensure randomization was somewhat reflected when extracting their subject sample.

Results of Individual Studies

Because of the heterogeneity of studies, specific characteristics of each and key data are reported in Table 3. Notable statistical results, as detailed in Tables 5 through 17, encompass a summary of pertinent statistical reliability measures for various landmarks listed by included studies. Typically, intraexaminer reliability was higher than interexaminer reliability in landmark identification. Skeletal landmarks presented similar reliabilities compared with dental ones; variability was dependent on challenges a specific location posed.

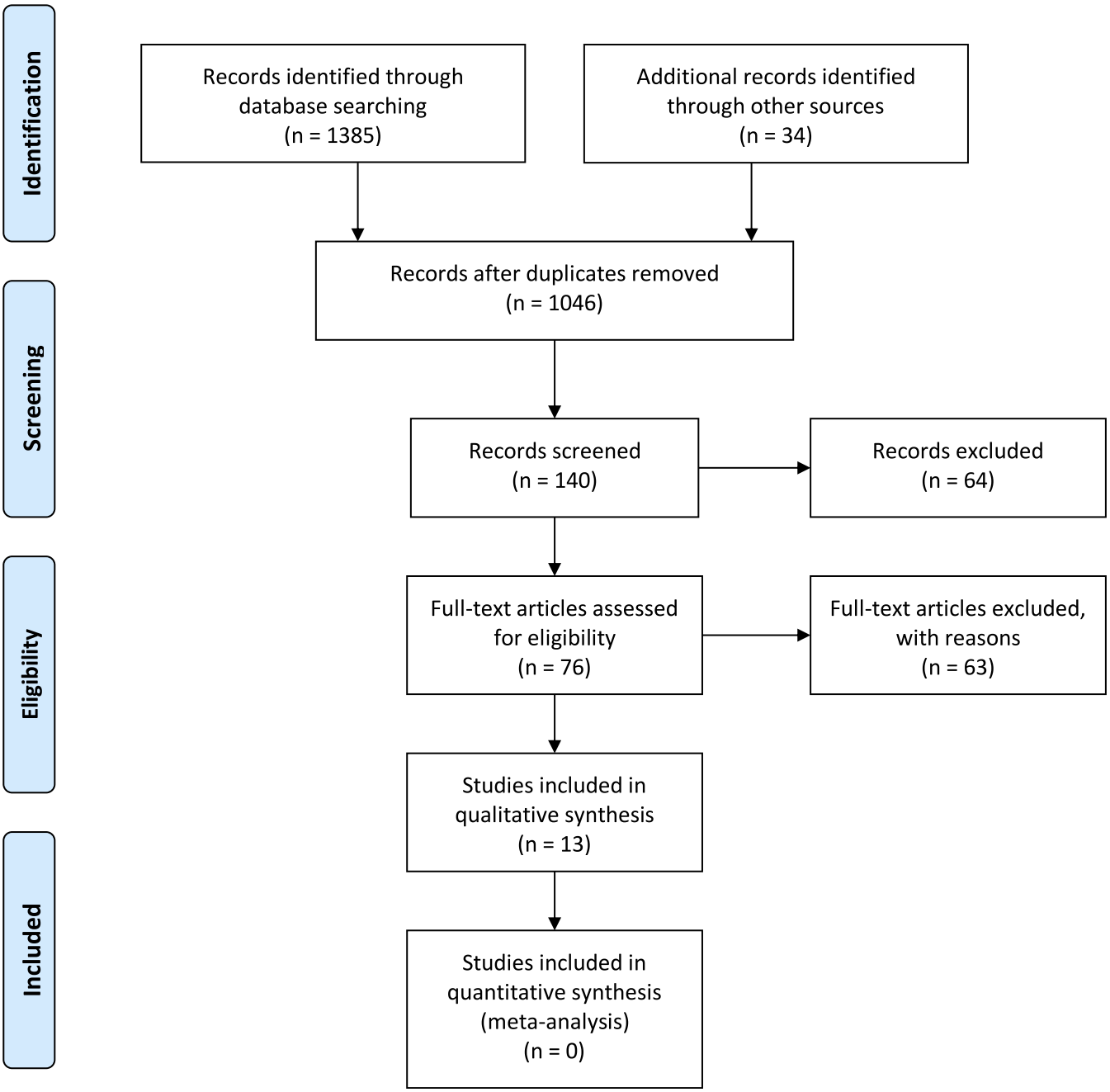

Figure 1. PRISMA flow diagram.

In general, midsagittal plane landmarks tended to demonstrate better consistency in identification compared with bilateral landmarks. The ease of locating landmarks along midlines may come naturally to most clinicians, as manipulation and interpretation of CBCT sagittal views are quite like 2D lateral cephalograms. Midsagittal plane landmarks recommended for use in 3D included Sella, basion, nasion, anterior nasal spine, A-point, B-point, pogonion, gnathion, and menton. Bilateral landmarks demonstrating variable consistency in identification included those on the condyles, orbitale, porion, and lingula. This was further complicated by the fact that some located along broad curvatures or that had indistinct boundaries were more difficult to locate and thus were more erroneous in identification. Dental landmarks demonstrating the greatest consistency were incisor crown tips, tooth root apices, and defined points on teeth. Some nontraditional landmarks recommended for use were infraorbital foramina, mental foramina, and possibly frontozygomatic sutures. Novel 3D landmarks, maxil-

Table 3. Summary Characteristics of Included Articles[a]

	Article	Sample Size, Type of Data, Sex, and Age	Observers (Number, Experience)	Repetitions and Intervals	Type of CBCT (Scanner/Software)	Method of Viewing 3D Imaging
1	Ghoneima et al. (2017)[13]	N = 20 Pre- and post-Herbst treatment CBCTs, T1 and T2 F = 11 M = 9 Ages: 8–15 y Mean age: 11 y	Number not specified Experience not specified	Each pre- and post-treatment CBCT: T_1, T_2 Interval not specified	i-CAT 3D (Xoran Technologies, Ann Arbor, Mich) 0.3-mm voxel size Dolphin Imaging 11.8 Premium (Dolphin, Chatsworth, Calif)	MPRV and 3D-VRV simultaneously
2	Neiva et al. (2015)[11]	N = 12 Human CBCT scans F = 8 M = 4 Ages: 20–43 y	N = 3; 1: Student with undergraduate degree in dentistry 2: Certification in orthodontics 3: Master's degree in orthodontics	Each type of visualization: T_1, T_2, T_3 1-wk intervals	i-CAT 3D (Xoran Technologies, Ann Arbor, Mich) 0.4-mm voxel size; InVivo Dental 5.1 (Anatomage Inc, San Jose, Calif)	MPRV individually 3D-VRV individually
3	Lee et al. (2015)[12]	N = 100 Human CBCT scans Sex not specified Mixed dentition with all permanent molars erupted in occlusion	N = 2; Not specified	Primary: Each type of visualization: T_1, T_2 1-wk interval Other: T_1 Not repeated Randomly selected scans (N = 30)	ILUMA Ultra (Imtec 3M, Ardmore, Okla) 0.3-mm voxel size; InVivo Dental 5.1 (Anatomage Inc, San Jose, Calif)	Not specified whether 3D-VRV seen simultaneously with MPRV
4	Naji et al. (2014)[5]	N = 30 Human CBCT scans Sex not specified Ages: 12–17 y	N = 2; Not specified	Primary: T_1, T_2, T_3 1-wk intervals Others: T_1 Not repeated	Next Generation i-CAT (Imaging Science International, Hatfield, Pa) 0.3-mm voxel size; Avizo 7.0 (Visualization Sciences Group, Burlington, Mass)	Not specified whether 3D-VRV seen simultaneously with MPRV
5	Zamora et al. (2012)[14]	N = 15 Human CBCT scans F = 73.4% M = 26.6% Ages: 8–27 y Presurgical orthodontic and impacted maxillary canine patients	N = 2; 1, 2: Six years' experience or background in orthodontics Previously trained in locating cephalometric landmarks	Each type of visualization: T_1, T_2 1-wk intervals	CBCT i-CAT (Imaging Sciences International, Hatfield, Pa) 0.4-mm voxel size; Beta NemoStudio (Software Nemotec SL, Madrid, Spain)	MPRV and 3D-VRV simultaneously Explanation of standardized protocol for landmark identification
6	Frongia et al. (2012)[15]	N = 10 Human CBCT scans F = 5 M = 5 Ages: 18.9 ± 1.2 y Orthognathic patients	N = 2; 1,2: Experts in orthodontics	Each type of visualization: T_1, T_2, T_3 1-wk intervals	Not specified 0.133-mm voxel size; Simplant OMS Software 13.0 (Materialise Dental NV, Leuven, Belgium)	MPRV and 3D-VRV simultaneously
7	Schilicher et al. (2012)[16]	N = 19 Human CBCT scans F = 13 M = 6 Ages: 18–35 y; 21.2 ± –7.9	N = 9; 1–9: Second- or third-year orthodontic residents Calibrated; definition of landmarks	Each type of visualization: T_1 No intervals 6 mo to complete	Hitachi CB MercuRay (Hitachi Medico Technology, Tokyo, Japan) 0.2–0.376 mm voxel size; Dolphin Imaging 10.1 (Dolphin, Chatsworth, Calif)	MPRV and 3D-VRV simultaneously
8	Hassan et al. (2011)[17]	N = 10 Human CBCT scans F = 6 M = 4 Ages: 18–23 y	N = 11; 1–11: Orthodontic residents	Each type of visualization: T_1, T_2 At least 1 d between intervals Maximum 2 h per individual session	NewTom 3G CBCT (QR-SRL, Verona, Italy) 0.3-mm voxel size; Dolphin-3D v. 11 (Dolphin Imaging and Management Systems, Chatsworth, Calif)	First: 3D-VRV only Second: MPRV and 3D-VRV simultaneously (starting from landmark coordinates determined from 3D-VRV only)

Table 3. Extended

Landmark Identification	Reliability Statistical Analysis	Results	Methodological Quality Score
Skeletal and dental landmarks Landmarks = 7	Reliability of 3D landmark-based superimposition methods ICC	All landmarks had an ICC >0.90, except ACP-x and PNS-y Most reliable landmarks in x- and y-coordinates: Ba, Na, A point, ANS, B point, Pg, Me, U1, L1 Landmark-based superimposition method reliable, although less than surface based and voxel based	Moderate
Skeletal and dental MPRV = (30–2) 3D-VRV = 30 *Zygomatic-maxillary suture right and left could not be correctly displayed in this view	3D landmark identification in CBCT, using two different visualization techniques (MPRV vs 3D-VRV); ICC	More highly reliable values in intra- than interobserver assessment MPRV more highly reliable values in landmark identification than 3D-VRV Landmarks on midsagittal plane demonstrated higher reliability Landmarks on condyle demonstrated lower reliability Clinically reliable in both techniques (ICC >0.75): B, Pg, Me, ANS, rGo, IMCo, and IUM1 Poor clinical reliability in both techniques (ICC <0.45): rCo, IRP, rZS, lZS	High
Skeletal; estimation of maxillary and mandibular basal bone Nontraditional cephalometric landmarks = 2	Reliability of two novel 3D cephalometric landmarks; Mean and standard deviation 3D scatterplots	Overall, favorable intra- and interexaminer reliability for both maxillary and mandibular centroid landmark	Moderate
Skeletal and dental Nontraditional cephalometric landmarks = 42	Reliability of several anatomic 3D cephalometric landmarks; Mean and standard deviation ICC	Intra- and interexaminer reliability for x, y, and z coordinates for all landmarks ICC >0.95 with CI of 0.88–0.99 Mean differences of measurements for landmarks: •Intraexaminer, mostly <0.5 mm •Interexaminer, all <1.4 mm Landmarks reliable to be used in 3D analysis: Mental foramina, infraorbital foramina, inferior hamulus, dens axis, foramina transversium of atlas, medial and lateral condyles of the mandible, superior clinoid processes, and mid-clinoid	Moderate
Skeletal only Landmarks = 41	Intra- and interobserver reliability for landmark identification; Mean and standard deviation ICC	Intra- and interexaminer reliability for x, y, and z coordinates for all landmarks ICC >0.95 Highest values in z-axis (ICC >0.996) Landmarks with no errors in determination: Nasion, Sella, left porion, point A, anterior nasal spine, pogonion, gnathion, menton, frontozygomatic sutures, first lower molars, upper and lower incisors Landmarks with more than 6 errors in determination: supraorbital, right zygion, posterior nasal spine Most reliable landmarks: Na, S, Ba, PoL, A, Ans, FzR, FzL, Pg, Me, Gn, B36, B46, UIR, LIR	Moderate
Skeletal and dental Hard tissue landmarks = 21	Reliability and repeatability of landmark identification in 3D; Mean and standard deviation Pearson's correlation coefficient	Highest SD values in z-axis (0.20–0.24 mm) Pearson's correlation coefficient demonstrated a strong correlation (>0.7) for both intra- and interobserver repeatability of landmark identification	Moderate
Skeletal and dental Landmarks = 32	Interexaminer consistency and precision of landmark identification; Standard deviation Pearson's correlation coefficient	Average consistency across all landmarks was 1.64 mm Landmarks with greatest consistency: Sella, left and right maxillary incisor crown tip, basion, right mandibular incisor crown tip Landmarks with poorest consistency: •Left and right maxillary cant point, left and right orbitale, right porion •Landmarks located along curves do not have clear anatomic boundaries and are more erroneous	Moderate
Skeletal and dental Landmarks = 22	Intra- and interexaminer precision of landmark identification; Mean and standard deviation Cronbach's α	Range for precision of landmark measurements: 0.29 ± 0.17 mm (upper incisor right) and 2.82 ± 7.53 mm (porion right) Adding MPRV to 3D-VRV improved precision of identifying cephalometric landmarks; exceptions are: Nasion, menton, orbitale left, and Sella turcica Intraobserver total reliability, mean and standard deviation, ranged from 0.48 (0.43) to 1.92 (4.69) mm Interobserver reliability, Cronbach's α, of all landmarks below 0.7	Moderate

Table 3. Continued

	Article	Sample Size, Type of Data, Sex, and Age	Observers (Number, Experience)	Repetitions and Intervals	Type of CBCT (Scanner/Software)	Method of Viewing 3D Imaging
9	Lagravere et al. (2010)[3]	N = 10 Human CBCT scans Sex not specified Adolescents	N = 3; Not specified	Primary: T_1, T_2, T_3 Intervals not specified Others: T_1 Not repeated	NewTom 3G (AFP Imaging, Elmsford, NY) Voxel size not specified; AMIRA (Mercury Computer Systems, Berlin, Germany)	No specification if MPRV and 3D-VRV simultaneously viewed Conventional 2D cephalograms
10	Ludlow et al. (2009)[1]	N = 20 Human CBCT scans Sex not specified Ages not specified Presurgical orthodontic patients	N = 5; 1, 2: Experienced oral and maxillofacial radiologists 3: Third-year radiology resident 4: Experienced orthodontist 5: Second-year orthodontic resident	Each type of visualization: T_1, T_2, T_3, T_4 No intervals specified Observation sessions spread over 2-wk period	NewTom 3G (QR-NIM s.r.l., Verona, Italy) 0.4-mm voxel size; Dolphin Imaging 10.1 (Dolphin, Chatsworth, Calif)	MPRV and 3D-VRV simultaneously Conventional 2D cephalograms
11	Chien et al. (2009)[18]	N = 10 Human CBCT scans Sex not specified Ages not specified	N = 6; 1–6: Second-year orthodontic residents	Each type of visualization: T_1, T_2 At least 1-wk intervals	i-CAT (Imaging Sciences International, Sacramento, Calif) Voxel size not specified; Dolphin Imaging 10.0 (Chatsworth, Calif)	MPRV and 3D-VRV simultaneously Conventional 2D cephalograms
12	Lagravere et al. (2009)[2]	N = 24 (12 baseline, 12 in treatment for 6 mo) Human CBCT scans Sex not specified Ages not specified Maxillary expansion treatments	N = 5 Not specified	Primary: T_1, T_2, T_3 At least 1-wk intervals Others: T_1 Not repeated	NewTom 3G (AFP Imaging, Elmsford, NY) Voxel size not specified; AMIRA (Mercury Computer Systems, Berlin, Germany)	MPRV and 3D-VRV simultaneously
13	De Oliveira et al. (2008)[19]	N = 12 Human CBCT scans Sex not specified Ages not specified; inclusion criteria ages 13–50 y Presurgical patients; 6 skeletal Class II and 6 skeletal Class III	N = 3; 1: Orthodontist 2: Dental radiologist 3: Third-year dental student	Each type of visualization: T_1, T_2, T_3 3-d intervals	NewTom 3G (AFP Imaging, Elmsford, NY) 0.4-mm voxel size; Dolphin 3D pre-release version 1 (Dolphin Imaging & Management Systems, Chatsworth, Calif)	MPRV (slices) and 3D-VRV simultaneously

[a] N indicates number; T, time point; F, female; M, male; MPRV, multiplanar reconstruction view; 3D-VRV, 3D-virtual reconstruction view; LCR, lateral cephalometric radiograph; ICC, intraclass correlation coefficient; M measurement, a linear 3D analysis; r, correlation coefficient.

Table 3. Continued, Extended

Landmark Identification	Reliability Statistical Analysis	Results	Methodological Quality Score
Skeletal and dental Landmarks = 18	Intra- and interobserver reliability of landmark identification using 3D imaging and 2D digital cephalometrics; Mean and standard deviation ICC	High reproducibility of all measurements (ICC >0.9) in x, y, and z coordinates Intraobserver variability in CBCT: •Generally <1.0 mm •x-axis: orbitale left, Sella, basion, anterior nasal spine, posterior nasal spine, condylion right (1.0–2.0 mm) •x-axis: porion right and left (2.62 and 3.37 mm, highest) •y-axis: gonion right and left, porion left and posterior nasal spine (1.0–2.0 mm) •z-axis: B-point, mandibular incisor root apex left (1.0–2.0mm) Interobserver variability in CBCT: •Generally >1.0 mm •x-axis: orbitale right and left, porion right and left, condylion right and left (>2.0 mm, highest) •y-axis: gonion right and left, anterior nasal spine (>2.0 mm, highest) •z-axis: gonion right and left, mandibular incisor root apex (>2.0 mm, highest)	Moderate
Skeletal and dental Landmarks = 22	Comparison of landmark identification between MPRV derived from CBCT and conventional 2D cephalograms; Difference of the mean (ODM) Difference of each observer from each other (DEO) Analysis of variance (ANOVA) Paired *t*-tests	Generally, landmark identification more precise with MPRV than conventional 2D cephalograms (13 of 22 subjects) Average DEO variability in MPRV: •Sella (0.7 mm, lowest) •Soft tissue pogonion (2.6 mm, highest) DEO variability by 3 directional axes in MPRV: •Anteroposterior (ANS) •Caudal-cranial (A-point, pogonion, porion, soft tissue A-point, and soft tissue pogonion) •Mediolateral (condylion, mandibular incisor tip, maxillary incisor tip, orbitale, and porion)	Moderate
Skeletal and dental Landmarks = 27	Intra- and interobserver reliability of landmark identification using 3D CBCT imaging and 2D digital cephalogram; Mean and standard deviation ICC	Intraobserver reliability, standard error >1 mm in 3D: •x-direction (L1 lingual gingival border, L1 root, orbitale, porion, sigmoid notch) •y-direction (supramentale, gonion, L1 lingual gingival border, L1 root, midramus, ramus point, U1 root) Intraobserver reliability, more than 15% variation for one observer: •x-direction (orbitale, porion) •y-direction (supramentale, L1 lingual gingival border, L1 root, ramus point) Interobserver reliability, standard error >1 mm in 3D: •x-direction (condylion, orbitale) •y-direction (gonion, midramus point, ramus point) Interobserver reliability, more than 15% variation in 3D: y-direction (gonion, midramus, ramus point)	High
Skeletal and dental Landmarks = 44	Intra- and interexaminer reliability of landmark identification using 3D CBCT imaging; those landmarks previously used in traditional 2D imaging; Mean and standard deviation ICC ANOVA	High reproducibility of all measurements (ICC >0.8) in x-, y-, and z- coordinates Intraexaminer reliability: •ICC >0.97 for all landmarks •Variability generally <1.5 mm; exception is in z-axis: piriform right (1.53 mm, highest) •Variability >1.0 mm in x-axis: auditory external meatus right and left, zygomaxillary left •Variability in >1.0 mm in y-axis: none •Variability in >1.0 mm in z-axis: A point, B point, piriform right and left, ectomolare right and left (>1.0 mm) Interexaminer reliability: •ICC >0.92 for most landmarks; exceptions are in x-axis: auditory external meatus right and left, orbit right and left (0.8–0.9, lowest) •Variability greater than in intraexaminer; orbit left (3.61 mm, highest) •Variability in x-axis: orbit right and left (>2.5 mm), zygomaxillary right and left (>1.5 mm) •Variability in y-axis: auditory external meatus left, piriform left, orbit right and left, MB 36 apex, MB 46 apex, anterior nasal spine (>1.5 mm), none >2.5 mm •Variability in z-axis: piriform right and left (>2.5 mm), ectomolare right and left (>1.5 mm) Landmarks presenting with statistical differences with other landmarks in same region: •Auditory external meatus right and left (x-axis and y-axis) •Most landmarks in skeletal facial region •Upper first molar (26B) and mesiobuccal apex (26A), mesiobuccal apex (36A and 46A)	Moderate
Skeletal and dental Landmarks = 30	Intra- and interobserver reliability of landmark identification using 3D imaging; ICC ANOVA	Reliability for x-, y-, and z- coordinates with ICC ≥0.90, intra- and interobserver, respectively: •x (80%, 66.6%) •y (83.33%, 50%) •z (93.33%, 80%) Least reliable landmarks: •x-coordinate (right and left condylion, ICC = 0.46–0.66) •y-coordinate (right and left ramus point, ICC = 0.29–0.68 and right and left tuberosity, ICC = 0.48–0.77) •z-coordinate (right and left condylion, ICC = 0.28–0.5)	Moderate

Table 4. Methodological Scores for Reliability Articles[a]

Article	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Total	%
Ghoneima et al. (2017)[13]	✓	✓	✓	×	×	×	○	×	×	✓	✓	×	○	✓	×	✓	×	✓	×	✓	○	✓	✓	12.5	54.3
Neiva et al. (2016)[11]	✓	✓	✓	✓	×	✓	○	✓	✓	○	✓	✓	○	✓	○	✓	×	✓	✓	✓	○	✓	✓	18.5	80.4
Lee et al. (2015)[12]	✓	✓	○	○	✓	○	○	×	×	×	✓	○	×	✓	✓	✓	×	✓	✓	✓	○	✓	✓	15.0	65.2
Naji et al. (2014)[5]	✓	✓	○	×	×	○	✓	×	×	○	✓	○	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓	15.0	65.2
Zamora et al. (2012)[14]	✓	✓	✓	✓	×	○	○	✓	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓	18.0	78.2
Frongia et al. (2014)[15]	✓	✓	✓	○	×	○	○	✓	○	○	○	×	○	✓	✓	✓	×	✓	✓	✓	×	○	○	13.5	58.7
Schilcher et al. (2012)[16]	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	○	✓	✓	✓	✓	✓	○	✓	✓	20.5	89.1
Hassan et al. (2011)[17]	✓	✓	✓	○	×	✓	○	✓	✓	✓	○	✓	✓	✓	×	✓	×	○	×	✓	○	○	○	15.5	67.4
Lagravere et al. (2010)[3]	✓	○	○	×	✓	✓	✓	×	×	×	✓	×	✓	✓	×	✓	✓	✓	✓	✓	×	✓	✓	15.0	65.0
Ludlow et al. (2009)[1]	✓	✓	○	✓	×	✓	○	✓	✓	×	○	✓	✓	✓	○	○	×	✓	✓	✓	×	○	✓	16.0	69.6
Chien et al. (2009)[18]	✓	✓	○	✓	×	✓	○	✓	✓	×	✓	×	○	✓	✓	✓	○	✓	✓	✓	○	✓	✓	17.5	76.1
Lagravere et al. (2009)[2]	✓	○	○	×	×	✓	○	×	×	×	✓	×	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓	13.5	58.7
De Oliveira et al. (2008)19	✓	○	✓	✓	×	✓	○	✓	✓	○	○	×	✓	✓	○	✓	×	✓	✓	✓	○	✓	✓	17.0	73.9

[a] ✓ indicates satisfactorily fulfilled the methodological criteria (1.0 check point); ○, partially fulfilled the methodological criteria (0.5 check point); ×, fulfillment of no methodological criteria (0.0 check point); A-W, criteria used in the methodological scoring in Table 2.

lary and mandibular centroid landmarks, also showed favorable reliability.

Synthesis of Results

A meta-analysis was not possible. Methodologies of the selected studies were highly heterogeneous, posing a challenge to the consideration of combining results together. In addition, not all studies evaluated the same landmarks, making the comparison more challenging. Some of these were traditionally used cephalometric landmarks, whereas others were non-traditional in nature.

Risk of Bias Across Studies

The more observers involved in measurements of a single study, the greater potential for measurement error due to individual expertise. Also, as some authors were involved in more than one study of this sort, it was possible to use the same preexisting database for patient CBCT scans across multiple studies. If this were the case, it could pose a significant problem as it would artificially inflate the reliability values.

Additional Analysis

No additional analyses were performed.

DISCUSSION

Summary of Evidence

Bilateral landmarks, including midramus, orbitale, ramus point, and sigmoid notch, demonstrated more consistent identification in 3D than 2D. This is likely explained by the 2D limitation of structural superimposition being overcome. Each left and right side of a landmark could be evaluated independently, in a specific location in all three planes of space, without any other structures impeding its interpretation. Because of the unfamiliarity of routine landmarking along a transverse axis, as in 2D, bilateral landmarks tended to show more variability than those located in the midline. De Oliveira et al.[19] found that two bilateral landmarks demonstrated poor reliability in one of the three axes: the ramus in the y-coordinate and the condylion in the z-coordinate. Many bilateral landmarks are located along broad curvatures and pose a challenge for the eye to detect the most prominent point or depression of the structure at hand. Differences in landmark identification error in the axes may differ, and as such, certain landmarks were useful in detecting changes in one axis but not another.[2] Landmarks that demonstrated considerable variability in the x-coordinate were not suitable for use in width (transverse) measurements of the dentofacial complex. For example, condylion, orbitale, and porion demonstrated statistically greater variability in the mediolateral direction, or x-axis, in multiplanar reconstruction views (MPRV) and may not be suitable for use in taking width measurements. A possible explanation for

Table 5. Summary of Notable Statistical Results for Ghoneima et al. (2017)[13]

	Landmark ICC (95% CI)	
Landmark	X	Y
Anterior clinoid process of Sella (ACP)	0.83 (0.35, 0.99)	0.99 (0.95, 0.99)
Basion (Ba)	0.99 (0.99, 1.00)	1.00 (0.99, 1.00)
Nasion (Na)	0.99 (0.98, 1.00)	0.96 (0.86, 0.99)
A point	0.99 (0.99, 1.00)	0.98 (0.94, 0.99)
Anterior nasal spine (ANS)	0.98 (0.96, 0.99)	0.99 (0.98, 1.00)
Posterior nasal spine (PNS)	0.94 (0.74, 0.99)	0.78 (0.27, 0.99)
B point	0.99 (0.97, 1.00)	0.98 (0.94, 0.99)
Pogonion (Pg)	1.00 (0.98, 1.00)	0.98 (0.94, 0.99)
Mention (Me)	0.99 (0.97, 1.00)	0.99 (0.97, 1.00)
Incisal tip of the upper central incisor (U1)	0.94 (0.81, 1.00)	0.99 (0.98, 1.00)
Incisal tip of the lower central incisor (L1)	0.99 (0.98, 1.00)	1.00 (0.98, 1.00)

Table 6. Summary of Notable Statistical Results for Neiva et al. (2016)[11]

Landmark	3D Volume Rendered Intraobserver (ICC)			3D Volume Rendered Interobserver (ICC)			Multiplanar Slice Intraobserver (ICC)			Multiplanar Slice Interobserver (ICC)			Clinical Reliability
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	
B point (B)	0.96	0.95	0.86	0.96	0.84	0.79	0.98	0.99	0.92	0.98	0.99	0.91	Reliable both
Pogonion (Pog)	0.95	0.94	0.90	0.93	0.84	0.88	0.98	0.99	0.97	0.98	0.99	0.86	Reliable both
Menton (Me)	0.95	0.94	0.96	0.94	0.85	0.91	0.98	1.0	0.98	0.98	0.99	0.99	Reliable both
Anterior nasal spine (ANS)	0.95	0.94	0.97	0.92	0.92	0.93	1.0	0.98	0.99	1.0	0.96	0.99	Reliable both
Left mandibular gonion (lGo)	0.98	0.95	0.92	0.96	0.92	0.93	0.99	0.95	0.93	0.99	0.95	0.90	Reliable both
Left medial mandibular condyle (lMco)	0.93	0.95	0.96	0.92	0.93	0.96	0.99	0.99	1.0	0.98	0.99	0.99	Reliable both
Left upper molar point (lUM1)	0.89	0.95	0.97	0.89	0.90	0.94	0.99	1.0	0.98	0.97	1.0	0.97	Reliable both

bilateral landmarks showing more variability in the x-axis was perhaps related to the inadequacy of landmark definition in this dimension.[1]

A main limitation of 2D imaging is that a 3D object is reduced to two planes of space. More precise landmark identification was obtained with most MPRV in 3D than in 2D cephalograms. Of these landmarks, Sella demonstrated the lowest variability of 0.7 mm, whereas soft tissue pogonion showed the highest variability of 2.6 mm.[1] This was in agreement with another study that found a high reproducibility (ICC >0.9) of all measurements traditionally employed in 2D cephalometric analyses.[3] In contrast, another study found that intraobserver identification for only specific landmarks was greater in 3D than in 2D.[18] This could be explained by the fact that this study included a total of 27 landmarks, which was more than included in the prior two studies. In fact, the higher number of landmarks analyzed afforded a benefit, as meaningful errors were more likely noted. It is important to realize that ease of identifiability of points does not necessarily translate to meaningful implications. This may provide an artificial sense of reliability.

Sometimes a discrepancy between the reliability of identifying left vs right structures is apparent. The manifestation may be attributed purely to the individual examiner's systematic error. Another hypothetical and plausible explanation is that this could be the neuropsychological linkage between left- and right-handedness and its effect on preferences of the human brain. Right-handed artists have been shown to prefer their subjects on the right with light sources from the left. Left-handed artists tend to demonstrate the opposite trend.[20] Extrapolating, this may imply some influence of handedness in an evaluator's spatial orientation of CBCT scans and their identification of 3D landmarks.

There was a recognizable trend that midline landmarks such as Sella and A-point showed the same consistency, if not greater, in landmark identification as in 2D. In contrast to bilateral structures, this may actually be facilitated by the familiarity of observers with radiographic interpretation in the sagittal plane, used in lateral head films. The MPRV display in 3D software provides an avenue for limiting the magnitude of superimposition of multiple structures, as slices can be set to a particular thickness when investigating an area of interest.

Three-dimensional objects occupy a specific location on an x-y-z coordinate system. Although the maximum mean difference was minimal, one study noted that the y-coordinate was more reliable than the x- and z-coordinates among observers. The least reliable landmark identifications in these axes were as follows: condylion in the x-coordinate or mediolateral direction, ramus point closely followed by tuberosity in the y-coordinate or anteroposterior direction, and condylion

Table 7. Summary of Notable Statistical Results for Lee et al. (2015)[12]

Landmark	Intraexaminer 3D Distance, mm, Mean (SD)	Interexaminer 3D Distance, mm, Mean (SD)
Average maxillary centroid	0.76 (0.32)	0.89 (0.49)
Average mandibular centroid	0.57 (0.39)	0.82 (0.43)

Table 8. Summary of Notable Statistical Results for Naji et al. (2014)[5]

Landmark	x-Axis, mm, Mean (SD)	y-axis, mm, Mean (SD)	z-axis, mm, Mean (SD)
Mental foramen right	0.48 (0.32)	0.39 (0.26)	0.31 (0.27)
Mental foramen left	0.40 (0.20)	0.44 (0.28)	0.40 (0.36)
Dens axis	0.46 (0.28)	0.42 (0.20)	0.13 (0.18)
Right transversium atlas	0.34 (0.19)	0.47 (0.27)	0.50 (0.37)
Left transversium atlas	0.41 (0.20)	0.38 (0.26)	0.39 (0.34)
Inferior right hamulus	0.47 (0.35)	0.50 (0.28)	0.37 (0.41)
Inferior left hamulus	0.54 (0.37)	0.45 (0.28)	0.36 (0.35)
Right infraorbital	0.51 (0.47)	0.51 (0.20)	0.57 (0.54)
Left infraorbital	0.51 (0.31)	0.54 (0.37)	0.66 (0.43)
Superior right clinoid process	0.60 (0.51)	0.33 (0.18)	0.17 (0.23)
Superior left clinoid process	0.64 (0.68)	0.41 (0.20)	0.16 (0.23)
Mid-clinoid	0.56 (0.45)	0.43 (0.25)	0.56 (0.92)
Lateral right condyle	0.22 (0.22)	0.53 (0.27)	0.59 (0.51)
Medial right condyle	0.25 (0.24)	0.52 (0.30)	0.51 (0.36)
Medial left condyle	0.32 (0.19)	0.55 (0.29)	0.71 (0.35)
Lateral left condyle	0.21 (0.17)	0.54 (0.32)	0.54 (0.35)

Table 9. Summary of Notable Statistical Results for Zamora et al. (2012)[14]

Region	Landmark	SD_X, mm	SD_Y, mm	SD_X, mm	% of Error
Cranial	Nasion (Na)	0.36	0.32	0.49	0.00
	Sella turcica (S)	0.89	0.41	0.61	0.00
	Basion (Ba)	0.79	0.51	0.49	0.00
	Left porion (PoL)	0.90	0.46	0.37	0.00
Maxillary	Point A (A)	0.86	0.41	0.93	0.00
	Anterior nasal spine (Ans)	0.96	0.66	0.30	0.00
	Incisal edge upper right central incisor (UIR)	0.86	0.31	0.28	0.00
Orbital-zygomatic	Right frontozygomatic suture (FzR)	0.28	0.41	0.37	0.00
	Left frontozygomatic suture (FzL)	0.32	0.57	0.38	0.00
Mandibular	Pogonion (Pg)	0.16	0.23	0.67	0.00
	Menton (Me)	0.50	0.63	0.24	0.00
	Gnathion (Gn)	0.15	0.44	0.45	0.0
	First lower left molar (B36)	0.55	0.52	0.83	0.0
	First lower right molar (B46)	0.54	0.37	0.94	0.0
	Incisal edge lower right central incisor (LIR)	0.75	0.31	0.31	0.0

in the z-coordinate or caudal-cranial direction.[19] In contrast, Chien et al.[18] highlighted that some difficulty arose in determining the best estimates of the y-locations for gonion, L1 tip, Sella, and U1 tip. Difficulty arose locating the y-location of structures such as gonion, midramus, and ramus point, since a precise vertical position must be established along these broadly curved structures in 2D and 3D. Most of these inaccuracies were linked to a line parallel to its curvature. Specifically for 3D, erroneous measures may be attributed to the inappropriate use of surface display shading used by the operator.[18]

Most traditionally used cephalometric landmarks were reproducible both in 2D and 3D imaging modalities. Since 3D has the enhanced ability to fulfill the precision of a third dimension, it makes one wonder if there were also nontraditional landmarks that are reproducible for use in 3D analyses. Using 42 newly defined anatomic landmarks, Naji et al.[5] concluded that the mean differences of all measurements were less than 1.4 mm. Moreover, if a center coordinate point was chosen using the x-, y-, and z-coordinates to locate a specific landmark, the analysis of its reliability among evaluators was maximized, and these differences were more impactful clinically. In fact, bilateral mental foramina, dens axis, bilateral transversium atlas, bilateral inferior hamulus, right infraorbital foramen, medial right condyle, and lateral left condyle showed 0.5 mm or even less of a difference. However, one should be mindful of its application, as not all nontraditional landmarks should be routinely used.[5]

Since 3D cephalometric landmarking is still a new concept, labeling landmarks with a variability of clinical significance is not necessarily concrete per se. The clinical significance of cephalometric landmarks with a variability of less than 0.5 mm is unlikely, whereas variability between 0.5 mm and 1.0 mm may be likely. This differs from 2D, as cephalometric landmarks less than 1.0 mm are unlikely to have clinical significance. Thus, if linear and angular measurements were taken using these landmarks, their clinical implications may be considered to be reduced.

When evaluating the effects of software MPRV vs 3D-virtual reconstruction view (3D-VRV) for anatomic landmark identification, two included studies offered valuable insight. MPRV has been shown to be more highly reliable than 3D-VRV when considering these two types of visualizations independently.[11] However, most software used to import and view DICOM image formats of CBCT scans have the capability for simultaneous viewing of both modalities.

It was an interesting finding that viewing MPRV and 3D-VRV did not improve the precision of identifying the upper right and left central incisors and that MPR alone

Table 10. Summary of Notable Statistical Results for Frongia et al. (2014)[15]

	Intra- and Interobserver SD of 3D Measurements, mm			Intra- and Interobserver Pearson's Correlation Coefficient of 3D Cephalometric Measurements		
Operator	x-Axis	y-Axis	z-Axis	T1–T2	T2–T3	T1–T3
A	0.17	0.18	0.20	0.9997	0.9998	0.9998
B	0.18	0.24	0.24	0.9997	0.9997	0.9997
				T1–T1	T2–T2	T3–T3
A-B	0.18	0.22	0.23	0.7908	0.7913	0.7897

Table 11. Summary of Notable Statistical Results for Schlicher et al. (2012)[16,a]

Landmark	Overall Consistency, mm	x-Axis Consistency, mm	y-Axis Consistency, mm	z-Axis Consistency, mm
Sella	0.50 ± 0.24 (1)	0.14 ± 0.13 (1)	0.31 ± 0.23 (4)	0.23 ± 0.16 (1)
Left maxillary incisor crown tip	0.58 ± 0.28 (2)	0.39 ± 0.32 (3)	0.23 ± 0.16 (2)	0.24 ± 0.15 (2)
Right maxillary incisor crown tip	0.59 ± 0.25 (3)	0.39 ± 0.31 (4)	0.17 ± 0.14 (1)	0.31 ± 0.21 (3)
Basion	0.85 ± 0.32 (4)	0.33 ± 0.25 (2)	0.35 ± 0.26 (5)	0.32 ± 0.28 (4)
Right mandibular incisor crown tip	0.91 ± 0.60 (5)	0.54 ± 0.43 (10)	0.37 ± 0.26 (8)	0.37 ± 0.26 (8)
Nasion	1.02 ± 0.50 (6)	0.48 ± 0.35 (8)	0.62 ± 0.49 (16)	0.33 ± 0.20 (5)
Right maxillary incisor root apex	1.05 ± 0.46 (7)	0.56 ± 0.44 (11)	0.63 ± 0.38 (18)	0.36 ± 0.31 (7)
Left mandibular incisor crown tip	1.13 ± 0.69 (8)	0.50 ± 0.41 (9)	0.46 ± 0.35 (11)	0.53 ± 0.43 (11)
ANS	1.15 ± 0.49 (9)	0.47 ± 0.35 (7)	0.36 ± 0.25 (6)	0.76 ± 0.52 (17)
Point A	1.20 ± 0.59 (10)	0.47 ± 0.36 (5)	1.07 ± 0.60 (27)	0.34 ± 0.24 (6)
Left maxillary incisor root apex	1.20 ± 0.53 (11)	0.47 ± 0.36 (6)	0.71 ± 0.40 (19)	0.38 ± 0.30 (9)
Gnathion	1.35 ± 0.61 (12)	0.67 ± 0.44 (15)	0.72 ± 0.38 (20)	0.78 ± 0.59 (19)
Left mandibular incisor root apex	1.50 ± 0.73 (13)	0.58 ± 0.46 (12)	0.76 ± 0.61 (21)	0.70 ± 0.58 (16)
Point B	1.50 ± 0.72 (14)	0.65 ± 0.43 (14)	0.87 ± 0.63 (22)	0.55 ± 0.39 (13)

[a] The number in parentheses is that landmark's rank for each column.

Table 12. Summary of Notable Statistical Results for Hassan et al. (2011)[17,a]

Landmark	Total Precision: 3D, 3D + MPR, mm	*P*-Value: Statistically Significant Differences Between 3D and 3D + MPR	Interobserver Reliability (Cronbach's α)
Orbitale right	1.00 (1.14)	.11	0.06
Orbitale left	1.00 (1.07)	**.04**	0.37
Nasion	0.64 (0.70)	**.001**	0.66
Anterior nasal spine	0.93 (2.21)	.06	0.55
Posterior nasal spine	0.96 (1.54)	**.04**	0.58
A-point	0.71 (0.80)	.19	0.28
Upper incisor right	0.29 (0.17)	.07	0.58
Upper incisor left	0.30 (0.20)	.48	0.40
Lower incisor right	0.48 (0.50)	.87	0.69
Lower incisor left	0.62 (1.35)	.19	0.67
B-point	0.74 (0.50)	.3	0.27
Pogonion	0.77 (0.81)	.42	0.57
Menton	1.00 (1.58)	**.03**	0.49
Gonion right	0.88 (0.62)	.77	0.21
Upper right molar	0.98 (3.39)	.95	−0.33
Upper left molar	0.62 (0.90)	.58	0.25

[a] Statistically significant *P*-values are bolded.

Table 13. Summary of Notable Statistical Results for Lagravere et al. (2010)[3]

	Intraexaminer Mean Differences of Coordinates, mm, Mean (SD)		
Landmark	x	y	z
Nasion (N)	0.68 (0.48)	0.86 (0.72)	1.78 (1.15)
A-point (A)	0.92 (0.24)	0.80 (0.35)	0.77 (0.60)
B-point (B)	1.51 (1.03)	0.54 (0.32)	1.81 (1.69)
Pogonion (Pg)	1.44 (1.03)	0.71 (0.33)	1.22 (0.74)
Gnathion (Gn)	1.42 (1.05)	0.93 (0.75)	0.73 (0.84)
Menton (Me)	1.51 (0.94)	1.21 (1.10)	0.55 (0.46)
Sella (S)	1.21 (0.80)	0.41 (0.31)	0.57 (0.25)
Basion (Ba)	1.23 (0.78)	0.97 (0.60)	1.03 (0.33)
Posterior nasal spine (PNS)	1.56 (1.11)	1.03 (0.84)	0.47 (0.21)
Upper central incisor tip right (U1T right)	0.61 (0.29)	0.53 (0.30)	0.53 (0.35)
Upper central incisor root apex right (U1R right)	0.52 (0.29)	0.98 (0.87)	1.24 (1.16)
Lower central incisor tip right (L1T right)	1.53 (1.06)	0.72 (0.45)	0.65 (0.58)
Lower central incisor root apex right (L1R right)	1.30 (0.95)	1.30 (0.90)	1.38 (0.64)
Upper central incisor tip left (U1T left)	0.78 (0.60)	0.44 (0.12)	0.58 (0.34)
Upper central incisor root apex left (U1R left)	1.11 (1.07)	0.79 (0.72)	1.21 (0.97)
Lower central incisor tip left (L1T left)	1.11 (0.72)	0.43 (0.25)	0.49 (0.26)

Table 14. Summary of Notable Statistical Results for Ludlow et al. (2009)[1]

	Difference for Every Observer (DEO) Variability by Directions of MPR Views, mm		
Landmark	Anteroposterior (AP)	Caudal-Cranial (CC)	Mediolateral (ML)
Sella	0.65	0.66	1.05
ANS	1.43	0.73	0.66
A-point	0.74	2.01	0.68
Pogonion	0.69	1.91	1.35
Soft tissue A-point	0.78	1.93	0.79
Soft tissue pogonion	1.31	3.98	1.44
Condylion	1.82	1.01	2.55
Mandibular incisor tip	0.63	0.67	2.06
Maxillary incisor tip	0.62	0.76	1.99
Orbitale	2.80	0.80	5.76
Porion	1.46	3.46	7.14

Table 15. Summary of Notable Statistical Results for Chien et al. (2009)[18]

	Intraobserver, Mean 3D Error (SD), mm		Interobserver, Mean 3D Error (SD), mm		Interobserver, ICC	
Landmark	x-Coordinate	y-Coordinate	x-Coordinate	y-Coordinate	x-Coordinate	y-Coordinate
Subspinale	0.71 (0.79)	1.16 (0.78)	0.53 (0.56)	0.98 (0.79)	0.98	0.97
Condylion	1.08 (0.84)	0.83 (0.62)	1.23 (0.73)	0.91 (0.70)	0.96	0.97
Gonion	1.03 (0.87)	1.34 (1.12)	0.97 (0.83)	1.27 (1.20)	0.96	0.81
Midramus	0.38 (0.38)	1.64 (1.28)	0.24 (0.25)	1.54 (1.02)	0.99	0.75
Orbitale	1.13 (1.59)	0.50 (0.46)	1.06 (1.19)	0.40 (0.33)	0.95	0.99
Ramus point	0.49 (0.39)	1.80 (1.84)	0.48 (0.45)	2.71 (2.11)	0.99	0.64
Supramentale	0.42 (0.40)	1.15 (1.06)	0.28 (0.26)	0.97 (0.83)	1.00	0.86
L1 root	0.93 (1.21)	1.23 (1.50)	0.52 (0.49)	0.89 (0.60)	0.98	0.90
Porion	1.10 (2.13)	0.83 (0.83)	0.88 (0.71)	0.66 (0.69)	0.94	0.99

Table 16. Summary of Notable Statistical Results for Lagravere et al. (2009)[2]

	Intraexaminer Absolute Mean Measurement Difference (SD), mm			Interexaminer Absolute Mean Measurement Difference, mm		
Landmark	x-Coordinate	y-Coordinate	z-Coordinate	x-Coordinate	y-Coordinate	z-Coordinate
Ectomolare left (EkmL)	0.55 (0.27)	0.68 (0.31)	1.45 (0.46)	0.99 (0.56)	1.18 (0.53)	2.44 (0.92)
Ectomolare right (EkmR)	0.60 (0.36)	0.70 (0.38)	1.46 (0.59)	0.92 (0.60)	1.36 (0.59)	2.18 (0.72)
Upper first molar, right (16B)	0.29 (0.39)	0.53 (0.23)	0.46 (0.23)	0.36 (0.38)	0.63 (0.21)	0.58 (0.23)
Mesiobuccal apex (16A)	0.46 (0.19)	0.43 (0.14)	0.55 (0.42)	0.73 (0.31)	0.67 (0.27)	0.95 (0.52)
Buccal apex (14B)	0.43 (0.42)	0.44 (0.33)	0.57 (0.24)	0.58 (0.46)	0.41 (0.25)	0.70 (0.34)
Buccal apex (14A)	0.51 (0.19)	0.47 (0.16)	0.80 (0.41)	0.62 (0.31)	0.51 (0.20)	0.94 (0.51)
Upper canine, right (13B)	0.37 (0.19)	0.37 (0.20)	0.57 (0.19)	0.47 (0.21)	0.42 (0.16)	0.98 (0.29)
Canine apex (13A)	0.51 (0.18)	0.45 (0.18)	0.67 (0.24)	0.63 (0.33)	0.59 (0.24)	0.84 (0.30)
Upper canine, left (23B)	0.36 (0.17)	0.30 (0.14)	0.59 (0.26)	0.56 (0.21)	0.44 (0.18)	1.03 (0.24)
Canine apex (23A)	0.43 (0.18)	0.47 (0.19)	0.69 (0.32)	0.74 (0.48)	0.72 (0.49)	0.98 (0.61)
Upper first premolar, left (24B)	0.41 (0.24)	0.43 (0.36)	0.66 (0.36)	0.52 (0.35)	0.51 (0.55)	0.65 (0.28)
Buccal apex (24A)	0.40 (0.18)	0.46 (0.19)	0.76 (0.58)	0.63 (0.33)	0.50 (0.28)	0.86 (0.50)
Upper first molar, left (26B)	0.22 (0.16)	0.47 (0.28)	0.54 (0.29)	0.35 (0.27)	0.57 (0.33)	0.69 (0.33)
Mesiobuccal apex (26A)	0.56 (0.21)	0.53 (0.45)	0.86 (0.51)	0.70 (0.37)	0.76 (0.34)	1.34 (0.76)
Lower first molar, left (36B)	0.42 (0.24)	0.37 (0.11)	0.41 (0.21)	0.55 (0.29)	0.53 (0.22)	0.69 (0.24)
Lower canine, left (33B)	0.35 (0.15)	0.30 (0.20)	0.67 (0.23)	0.64 (0.44)	0.53 (0.39)	0.85 (0.22)
Lower first molar, right (46B)	0.39 (0.18)	0.41 (0.14)	0.51 (0.14)	0.47 (0.26)	0.63 (0.27)	0.59 (0.27)
Lower canine, right (43B)	0.37 (0.23)	0.38 (0.21)	0.62 (0.26)	0.41 (0.21)	0.43 (0.23)	1.04 (0.28)
Foramen spinosum, left (FSL)	0.39 (0.31)	0.48 (0.36)	0.67 (0.37)	0.74 (0.49)	0.42 (0.10)	0.46 (0.21)
Foramen spinosum, right (FSR)	0.38 (0.29)	0.37 (0.15)	0.40 (0.33)	0.70 (0.47)	0.37 (0.13)	0.52 (0.00)
Center coordinate point (ELSA)	0.48 (0.17)	0.55 (0.25)	0.52 (0.27)	1.04 (0.53)	0.39 (0.26)	0.42 (0.32)
Auditory external meatus, left (AEML)	1.46 (0.60)	0.83 (0.47)	0.40 (0.30)	3.40 (1.30)	0.45 (0.59)	0.48 (0.12)
Auditory external meatus, right (AEMR)	1.22 (0.88)	0.76 (0.29)	0.42 (0.33)	3.09 (1.08)	0.59 (0.40)	0.33 (0.22)
Dorsum foramen magnum (DFM)	0.70 (0.39)	0.66 (0.48)	0.88 (1.28)	0.87 (0.49)	0.82 (0.46)	0.38 (0.21)

Table 17. Summary of Notable Statistical Results for De Oliveira et al. (2008)[19]

	Intraobserver Reliability, ICC			Interobserver Reliability, ICC		
Landmark	x	y	z	x	y	z
Left condylion	0.66	1.00	0.50	0.65	0.97	0.49
Right condylion	0.46	0.99	0.29	0.46	0.98	0.28
Left ramus point	0.95	0.68	1.00	0.94	0.44	0.98
Right ramus point	0.98	0.51	0.99	0.95	0.29	0.99
Left tuberosity	0.73	0.75	0.96	0.71	0.57	0.89
Right tuberosity	0.73	0.77	0.94	0.71	0.48	0.90

demonstrated consistency in accurate landmark identification. A reasonable explanation for this is that the root apex of the mandibular incisors is typically difficult to identify in the sagittal view because of the superimposition of the root apices of the anterior teeth.

Limitations

Since CBCT technology is a recent development and its integration into routine practice in dentistry is relatively recent, all selected studies were from no earlier than 2009. Although there have been prior attempts to synthesize a single document conveying all research done in understanding applications of landmark identification in 3D techniques, studies covered a broad range of topics and should not be automatically unified. As such, to narrow in on the area of interest, it was opted to use more rigid exclusion criteria than those prior.

One of the exclusion criteria that was chosen to be used after the synthesis of selected articles was studies using human dry skulls. This was because the soft tissue attenuation of facial structures could not be accounted for in absolute, despite researchers' best attempts with fluid-filled units.

All selected studies underwent a methodological quality assessment carried out by a single examiner. There is no gold standard methodological quality assessments tool used in reliability studies at this time. This posed difficulties when trying to emphasize the relative weight of certain studies on the overall conclusions.

CONCLUSIONS

- The mid-sagittal plane, followed by bilateral structures, demonstrated the highest reliability.
- Landmarks with the lowest reliability included those marked on the condyle and other anatomic structures with prominent curvatures without definitive boundaries.
- A minimum number of dental landmarks was reported, with many demonstrating good to excellent reliability.

REFERENCES

1. 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.e1–312.e10.
2. Lagravere MO, Gordon JM, Guedes IH, et al. Reliability of traditional cephalometric landmarks as seen in three-dimensional analysis in maxillary expansion treatments. *Angle Orthod*. 2009;79:1047–1056.
3. Lagravere 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.
4. van Vlijmen OJC, Maal T, Bergé SJ, et al. A comparison between 2D and 3D cephalometry on CBCT scans of human skulls. *Int J Oral Maxillofac Surg*. 2010;39:156–160.
5. Naji P, Alsufyani, NA, Lagravere MO. Reliability of anatomic structures as landmarks in three-dimensional cephalometric analysis using CBCT. *Angle Orthod*. 2014;84:762–772.
6. Lisboa Cde O, Masterson D, da Motta AF, Motta AT. Reliability and reproducibility of three-dimensional cephalometric landmarks using CBCT: a systematic review. *J Appl Oral Sci*. 2015;23:112119.
7. Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J Craniomaxillofac Surg*. 2011;39:91–92.
8. Lagravere 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.
9. Bialocerkowski A, Klupp N, Bragge P. How to read and critically appraise a reliability article. *Int J Ther Rehabil*. 2010;17:112–114.
10. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med*. 2009; 6(7):e1000097.
11. Neiva MB, Soares ÁC, Lisboa Cde O, Vilella Ode V, Motta AT. Evaluation of cephalometric landmark identification on CBCT multiplanar and 3D reconstructions. *Angle Orthod*. 2015;85(1):11–17.
12. Lee M, Kanavakis G, Miner RM. Newly defined landmarks for a three-dimensionally based cephalometric analysis: a retrospective cone-beam computed tomography scan review. *Angle Orthod*. 2015;85:3–10.
13. Ghoneima A, Cho H, Farouk K, Kula K. Accuracy and reliability of landmark-based, surface-based and voxel-based 3D cone-beam computed tompography superimposition methods. *Orthod Craniofac Res*. 2017;20:227–236.
14. Zamora N, Llamas JM, Cibrián 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–e688.
15. Frongia G, Piancino MG, Bracco AA, Crincoli V, Debernardi CL, Bracco P. Assessment of the reliability and repeatability of landmarks using 3-D cephalometric software. *Cranio*. 2012;30:255–263.
16. Schlicher W, Nielsen I, Huang JC, Maki K, Hatcher DC, Miller AJ. Consistency and precision of landmark identifica-

tion in three-dimensional cone beam computed tomography scans. *Eur J Orthod*. 2012;34:263–275.
17. Hassan B, Nijkamp P, Verheij H, et al. Precision of identifying cephalometric landmarks with cone beam computed tomography in vivo. *Eur J Orthod*. 2013;35:38–44.
18. 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;38:262–273.
19. de Oliveira AEF, 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 Endodontol*. 2009;107:256–265.
20. Strachan A. In the brain of the beholder: the neuropsychological basis of aesthetic preferences. *The Harvard Brain*, 2000;7:1–10.