A Systematic Review to Uncover a Universal Protocol for Accuracy Assessment of 3-Dimensional Virtually Planned Orthognathic Surgery

Ramy M. Gaber, BDSc, MDSc, *Eman Shabeen, BSc, MSc, PhD, †Bart Falter, DMD, MD, ‡Sebastian Araya, DDS, §Constantinus Politis, MD, DDS, MHA, MM, EBOMFS, PhD, ||Gwen R. J. Swennen, MD, DMD, PhD, FEBOMFS, MSc, ¶ and Reinbilde Jacobs, DDS, MSc, PhD, Dhc#

Purpose: The aim of this study was to systematically review methods used for assessing the accuracy of 3-dimensional virtually planned orthognathic surgery in an attempt to reach an objective assessment protocol that could be universally used.

Materials and Methods: A systematic review of the currently available literature, published until September 12, 2016, was conducted using PubMed as the primary search engine. We performed secondary searches using the Cochrane Database, clinical trial registries, Google Scholar, and Embase, as well as a bibliography search. Included articles were required to have stated clearly that 3-dimensional virtual planning was used and accuracy assessment performed, along with validation of the planning and/or assessment method. Descriptive statistics and quality assessment of included articles were performed.

Results: The initial search yielded 1,461 studies. Only 7 studies were included in our review. An important variability was found regarding methods used for 1) accuracy assessment of virtually planned orthognathic surgery or 2) validation of the tools used. Included studies were of moderate quality; reviewers' agreement regarding quality was calculated to be 0.5 using the Cohen κ test.

Conclusions: On the basis of the findings of this review, it is evident that the literature lacks consensus regarding accuracy assessment. Hence, a protocol is suggested for accuracy assessment of virtually planned orthognathic surgery with the lowest margin of error.

© 2017 American Association of Oral and Maxillofacial Surgeons

J Oral Maxillofac Surg ■:1-11, 2017

*Assistant Lecturer and Doctoral Student, Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Ain Shams University, Cairo, Egypt, and Research Fellow, OMFS-IMPATh Research Group, Department of Imaging and Pathology, Catholic University of Leuven, Leuven, Belgium.

†Clinical engineer, OMFS-IMPATh research group, Department of Imaging and Pathology, Catholic University of Leuven & Department of Oral and Maxillofacial Surgery, University Hospital of Leuven, Leuven, Belgium.

‡Consultant, Department of Oral and Maxillofacial Surgery, University Hospital of Leuven, Leuven, Belgium.

§Private Practice, Santiago, Chile.

||Professor and Head, Department of Oral and Maxillofacial Surgery, Catholic University of Leuven, Leuven, Belgium.

¶Professor, Maxillofacial and Facial Plastic Surgery, Department of Surgery, Cleft and Craniofacial Centre, Bruges, Belgium, and

Three-Dimensional Facial Imaging Research Group, General Hospital St Jan, Bruges, Belgium.

#Professor, OMFS-IMPATh Research Group, Department of Imaging and Pathology, Catholic University of Leuven, Leuven, Belgium.

Conflict of Interest Disclosures: None of the authors have any relevant financial relationship(s) with a commercial interest.

Address correspondence and reprint requests to Dr Gaber: OMFS-IMPATh Research Group, Department of Imaging and Pathology, Faculty of Medicine, University of Leuven, Kapucijnenvoer 33, Leuven 3000, Belgium; e-mail: ramy.m.gaber@asfd.asu.edu.eg

Received January 18 2017

Accepted May 17 2017

© 2017 American Association of Oral and Maxillofacial Surgeons

0278-2391/17/30602X

<http://dx.doi.org/10.1016/j.joms.2017.05.025>

Planning in orthognathic surgery has been largely investigated ever since Downs¹ started performing radiographic analysis of the dentofacial profile.² Since then, more accurate and more realistic treatment objectives have been developed from the cephalometric radiograph tracing in combination with all the data obtained from the clinical patient evaluation. The tracing also can be used after surgery to check whether the predicted surgical results were achieved.³ Model surgery and surgical splint fabrication can be used as additional tools to check the functional and spatial simulation of the intended surgical procedure.

Recent advances in 3-dimensional (3D) technology have resulted in a series of new computerized tools for use in preoperative planning.⁴ In 2003 Gateno et al⁵ introduced the concept of the composite skull model, which combines physical impressions with computed tomography scans to facilitate complete computer-aided orthognathic planning. Since Swennen et al^{6,7} initiated 3D cephalometric analysis and treatment planning, 3D virtual planning has been replacing 2-dimensional cephalometric analysis.⁸ Computer-aided design as well as manufacturing of the surgical splint, on the basis of the virtual planning, is used more and more instead of the classic model surgical procedure.

The introduction of cone-beam computed tomography (CBCT) in combination with planning software tools such as 3D virtual osteotomy, 3D soft tissue simulation, 3D-based surgical splint manufacturing, and finally, 3D superimposition has led to better outcomes and surgical risk reduction.⁹ Yet, not all doubts have been lifted about the presumption that the accuracy of 3D virtual planning is higher than that of model surgery.^{10,11}

Gateno et al¹² were among the first authors who compared virtually planned orthognathic surgery versus conventional model surgery. Since then, many studies have investigated the accuracy of computer-aided planning in orthognathic surgery, but no one has tested and compared the methods or tools used to measure this accuracy.^{13,14}

The purpose of this study was to systematically review the available literature for different assessment methods of 3D virtually planned orthognathic surgery. A secondary aim was to attempt to reach a consensus regarding an assessment protocol that could be universally used. We hypothesized that the assessment methods used would be highly variable, making comparison of the results between different studies difficult. The specific aims of this study were 1) to review methods used for accuracy assessment of virtually planned orthognathic surgery, 2) to critically assess the validation methods used within included studies, and 3) to uncover or suggest a universal

protocol that could be used for assessment of 3D virtually planned orthognathic surgery that offers the lowest margin of error possible.

Materials and Methods

STUDY DESIGN AND SAMPLE

To address the research purpose, we designed and implemented a systematic review. The study population was composed of publications concerned with the accuracy of virtually planned orthognathic surgery. Articles published until September 12, 2016, were considered.

To be included in the study sample, publications had to fulfill the following criteria: The study was published in English, was a prospective clinical trial, and contained a method for 3D virtual planning; the data were original; and the planning or assessment methods were validated. Articles were excluded from our review if they were case reports, case series, or protocol descriptions and if no validation of the planning or assessment methods used was performed. Quality assessment of the included studies was carried out, and investigators' agreement regarding quality was tested.

VARIABLES

The primary predictor variable was the accuracy assessment of orthognathic surgery. The main outcome variable was the assessment of planning accuracy through the measurement of change in the spatial orientation and/or location between the planned and actual location of jaw position, which could be achieved with different tools and techniques mainly using linear and angular measurements. Additional variables included the validation used for planning and/or assessment method.

DATA COLLECTION METHODS

A primary PubMed search was performed by use of the following strategy so that the resulting study should contain 1 Medical Subject Headings (MeSH) term, 1 primary keyword, and 1 secondary keyword.

1. MeSH terms: Image Processing, Computer-Assisted OR Orthognathic Surgical Procedures OR Surgery, Computer-Assisted
2. Primary keywords: orthognathic OR craniofacial OR maxillofacial
3. Secondary keywords: asses* OR accur* OR precis* OR valid* OR reliabil*

Boolean operators were used to merge the search terms in the following manner: ((Image Processing, Computer-Assisted[MeSH Terms]) OR (Orthognathic

Surgical Procedures[MeSH Terms]) OR (Surgery, Computer-Assisted[MeSH Terms])) AND ((orthognathic[tw] OR maxillofacial[tw] OR craniofacial[tw]) AND (asses*[tw] OR accur*[tw] OR precis*[tw] OR valid*[tw] OR reliabil*[tw])).

The second step was to apply filters in PubMed, namely the language and study design. Only studies in English and clinical trials were to be included. Title screening of resulting studies was performed by one the observers (R.M.G.) to fetch only relevant studies. Articles found were retrieved in full text, and their reference lists were examined for related studies.

A secondary search was performed in Google Scholar and Embase using the search strategy: (((virtual plan*) OR (3D plan*) OR simulation OR navigation OR (digital plan*) OR prediction OR (virtual model surgery) OR (CAD/CAM guided surgery)) AND (orthognathic OR maxillofacial OR craniofacial OR lefort OR bssso OR genioplasty) AND (accur* OR valid* OR precis* OR reliab* OR asses*)). An additional search in the Cochrane Database using the following term also was conducted: (orthognathic surgery). The American clinical trial registry (www.clinicaltrials.gov) and the registry of the World Health Organization (www.who.int/ictrp) were searched for any trials that did not appear in the previous searches.

Resulting relevant studies were subjected to abstract screening to determine whether they fulfilled the inclusion criteria; this screening was performed by 2 observers (R.M.G. and S.A.). If the abstract was not available or did not contain sufficient information, the full text of the article was retrieved and read for evaluation (Fig 1).

DATA ANALYSES

Four of the authors (R.M.G., E.S., B.F., and S.A.) shared in the data extraction process for the included studies. Studies were divided among investigators for initial data extraction; then a secondary check for the accuracy of extracted data was performed and a consensus reached by 2 investigators (R.M.G. and E.S.). Because of the heterogeneity of terminology used in individual studies, aside from data extraction, we attempted to unify terms that indicated the same method or assessed variable.

Assessment of the included studies' quality was performed with the Methodological Index for Non-Randomized Studies (MINORS) tool.¹⁵ Quality assessment was performed by 2 reviewers (R.M.G. and B.F.); in the case of disagreement between reviewers, a discussion was undertaken until mutual agreement was reached. Reviewers' agreement was tested with the Cohen κ test by use of MedCalc software (MedCalc Software, Ostend, Belgium). Custom-made forms were

used for data extraction, as well as for the quality assessment of the included studies.

Results

SEARCH RESULTS

The primary PubMed search yielded 1,461 studies; after application of the language and study design filters, the number was reduced to 51 studies. Title screening of these 51 studies uncovered only 7 relevant articles (Fig 1).

The full texts of the 7 studies were retrieved, and the reference lists were screened for additional articles. Eighteen related studies were found from the bibliography search. The secondary searches only found 3 additional studies (Fig 1).

Twenty-eight studies were submitted for abstract screening against the preset inclusion criteria, which was conducted by 2 observers (R.M.G. and S.A.). If the abstract was not available or did not contain sufficient data, the full text of the article was read. In the case of disagreement between the 2 observers, the article in question was discussed thoroughly and an agreement was reached regarding its inclusion or exclusion.

Of the studies, 7 did not report any validation of assessment or planning¹⁶⁻²² and 8 studies were not clinical trials but rather were case reports, case series, or protocol descriptions.^{7,23-29} Another 3 studies did not involve 3D planning,³⁰⁻³² whereas 2 were not orthognathic.^{33,34} A final study was published only as a pilot study,³⁵ with further results used in a later publication.³⁶

Only 7 articles fulfilled the inclusion criteria and were included in this systematic review.^{4,36-41} Details of the included studies are shown in Tables 1-4.

A total of 131 patients were included in the 7 studies. The mean age was 24 years and the gender distribution was 57 male and 64 female patients in 6 of the 7 studies; age and gender distribution were not mentioned by Aboul-Hosn Centenero and Hernandez-Alfaro.⁴ The surgical procedures were distributed as follows: 124 bimaxillary operations, 6 Le Fort I procedures, and a single bilateral sagittal split osteotomy.

PLANNING ACCURACY

Differences in jaw position between the virtual plan and the actual postoperative position were assessed by use of the mean difference between linear and angular measurements in 4 studies,^{36,38,40,41} whereas in 2 studies, the mean distance was used to compare the plan versus the actual outcome.^{38,39} The study by Aboul-Hosn Centenero and Hernandez-Alfaro⁴ used

FIGURE 1. Search strategy and workflow. MeSH, Medical Subject Headings; 3D, 3-dimensional.

Gaber et al. *Virtually Planned Orthognathic Surgery. J Oral Maxillofac Surg* 2017.

the intraclass correlation coefficient to compare between the planned and actual measurements.

VALIDATION

Methods of validating the techniques used were highly variable among the included studies, yet with the different methods used, most of the studies reported high accuracy and reproducibility of their methods. Four studies used the intraclass correlation coefficient to validate their results.^{4,38,40,41} Hernandez-Alfaro and Guijarro-Martinez³⁹ performed an in vivo part in their study and used this for validation, Hsu et al³⁶ reported using the Bland-Altman statistical method to test for agreement as a way to validate their methods, and Tucker et al³⁷ validated their work

through a Student *t* test that compared measurements performed at 2 time intervals.

QUALITY ASSESSMENT AND REVIEWERS' AGREEMENT

Detailed quality assessment is shown in Table 5. With a possible maximum score of 16, the studies' scores ranged between 10 and 12, indicating moderate quality. By use of the Cohen κ test in the MedCalc program, agreement between the 2 observers regarding the quality assessment was moderate ($\kappa = 0.5$).

Discussion

The purpose of this study was to systematically review the available literature on the different methods

Table 1. DEMOGRAPHIC DATA OF INCLUDED STUDIES

Authors, Publication Year, Country of Origin	Sample Size	Mean Age (Range), yr	Gender Distribution	BSSO	Le Fort I	Bimaxillary Surgery
Baan et al, ⁴¹ 2016, Netherlands	10	26.5 (17-45)	4 M and 6 F	0	0	10
Hernandez-Alfaro and Guijarro-Martinez, ³⁹ 2013, Spain	6	23.7 (19-37)	3 M and 3 F	0	0	6
Hsu et al, ³⁶ 2013, United States	65	26.7 (15-51)	31 M and 34 F	0	0	65
Shehab et al, ⁴⁰ 2013, Egypt	6	23.5 (18-30)	6 F	0	0	6
Aboul-Hosn Centenero and Hernandez-Alfaro, ⁴ 2012, Spain	16	Not specified	Not specified	1	0	15
Zinser et al, ³⁸ 2012, Germany	8	22.4 (19-35)	4 M and 4 F	0	0	8
Tucker et al, ³⁷ 2010, United States	20	21 (14-35)	9 M and 11 F	0	6	14

Abbreviations: BSSO, bilateral sagittal split osteotomy; F, female; M, male.

Gaber et al. *Virtually Planned Orthognathic Surgery. J Oral Maxillofac Surg* 2017.

of assessment of virtually planned orthognathic surgery and try to reach a consensus regarding an assessment protocol that could be universally used. We hypothesized that the assessment methods used would be highly variable, making comparison of

the results between different studies difficult. The specific aims of this study were 1) to review assessment methods used for virtually planned orthognathic surgery, 2) to critically assess the validation methods used within included studies, and 3) to

Table 2. RADIOGRAPHIC, PLANNING, ASSESSMENT, AND VALIDATION METHODS APPLIED IN STUDIES

Study	Radiographic Method	Software Used for Planning	Assessment Tool and Software	Validation Method
Baan et al, ⁴¹ 2016	CBCT	Maxilim (Medicim NV, Mechelen, Belgium)	OrthoGnathicAnalyser	ICC
Hernandez-Alfaro and Guijarro-Martinez, ³⁹ 2013	CBCT	Implant PRO OMS (Materialise Dental, Leuven, Belgium)	Mimics (Materialise Dental) The MathWorks	In vitro part of study
Hsu et al, ³⁶ 2013	CT	Implant OMS (Materialise Dental)	3DS Max (Autodesk, San Rafael, CA)	Bland-Altman statistical method
Shehab et al, ⁴⁰ 2013	CT	Onyx Ceph 2/6/4 (Image Instruments, Gottingen, Germany) Voxim (IVS Solutions, Chemnitz, Germany)	Lateral cephalometric radiographs	ICC
Aboul-Hosn Centenero and Hernandez-Alfaro, ⁴ 2012	CT and CBCT	Implant Pro OMS (Materialise Dental)	ICC	ICC
Zinser et al, ³⁸ 2012	MSCT and CBCT	Implant Pro Crystal (Materialise Dental)	Angular and linear measurements Implant Pro Crystal	ICC
Tucker et al, ³⁷ 2010	CBCT	CranioMaxilloFacial (CMF) application software (M.E. Müller Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland)	CMF application software	Repeating surgical simulation for 5 random patients

Abbreviations: CBCT, cone-beam computed tomography; CT, computed tomography; ICC, intraclass correlation coefficient; MSCT, multislice computed tomography.

Gaber et al. *Virtually Planned Orthognathic Surgery. J Oral Maxillofac Surg* 2017.

Table 3. PLANNING ACCURACY

Study	Method	Accuracy
Baan et al, ⁴¹ 2016	Absolute mean difference	AP, 1.41 mm; left-right (ML), 0.49 mm; up-down (SI), 1.85 mm; P, 2.72°; R, 1.04°; Y, 0.97°
Hernandez-Alfaro and Guijarro-Martinez, ³⁹ 2013	Mean distance	AP, 0.25 mm; left-right (ML), 0.15 mm; up-down (SI), 0.5 mm
Hsu et al, ³⁶ 2013	Absolute mean difference	Maxilla: left-right (ML), 0.8 mm; AP, 1 mm; up-down (SI), 0.6 mm; P, 1.5°; R, 0.9°; Y, 1.3° Mandible: left-right (ML), 0.8 mm; AP, 0.8 mm; up-down (SI), 0.6 mm; P, 1.8°; R, 1°; Y, 1.7°
Shehab et al, ⁴⁰ 2013	Mean difference, linear distances	AP, 0.4-1.2 mm; up-down (SI), 0.8 mm
Aboul-Hosn Centenero and Hernandez-Alfaro, ⁴ 2012	ICC for 17 linear and angular measurements	High, n = 10 Medium, n = 4 Low, n = 3
Zinser et al, ³⁸ 2012	Mean angular difference Mean linear distance	Frankfurt horizontal plane, 0.31°; midfacial plane, 0.22°; frontozygomatic line, 0.27° Frankfurt horizontal plane, 0.2 mm; coronal plane, 0.21 mm; midfacial plane, 0.06 mm
Tucker et al, ³⁷ 2010	Mean difference for 11 regions in maxilla and mandible	<0.5 mm except for left and right maxillae, for which it was 0.53 mm

Abbreviations: AP, anteroposterior, z-axis; ICC, intraclass correlation coefficient; ML, mediolateral, x-axis; P, pitch; R, roll; SI, superoinferior, y-axis; Y, yaw.

Gaber et al. *Virtually Planned Orthognathic Surgery. J Oral Maxillofac Surg* 2017.

uncover or suggest a universal protocol that could be used for assessment of virtually planned orthognathic surgery that offers the lowest margin of error possible.

In this review it was evident that the literature lacks sufficiently validated methods; hence only 7 articles were included. There was a striking absence of consensus among different studies regarding the assessment or validation methods. For the assessment

of accuracy, the change in linear and angular measurements was the most commonly used method, yet how this change was calculated varied between different studies. Whereas 4 studies used the intraclass correlation coefficient for validation,^{4,38,40,41} the 3 remaining studies validated their methods through unique methods.^{36,37,39} Nonetheless, despite this variability, most authors indicated high accuracy and reproducibility of methods used.

Table 4. VALIDATION RESULTS

Study	Method	Result
Baan et al, ⁴¹ 2016	ICC (intraobserver and interobserver)	High
Hernandez-Alfaro and Guijarro-Martinez, ³⁹ 2013	Mean difference of in vitro study	Left-right (ML), 0.09 mm; up-down (SI), 1.26 mm; AP, 0.18 mm
Hsu et al, ³⁶ 2013	Bland-Altman statistical method	High agreement
Shehab et al, ⁴⁰ 2013	ICC (interobserver)	High
Aboul-Hosn Centenero and Hernandez-Alfaro, ⁴ 2012	ICC (between CAD/CAM splints and conventional splints)	High, 9 patients Moderate, 6 patients Low, 1 patient
Zinser et al, ³⁸ 2012	ICC (intraobserver and interobserver)	High
Tucker et al, ³⁷ 2010	Student <i>t</i> test	NS

Abbreviations: AP, anteroposterior, z-axis; CAD/CAM, computer-aided design/computer-aided manufacturing; ICC, intraclass correlation coefficient; ML, mediolateral, x-axis; NS, nonsignificant; P, pitch; R, roll; SI, superoinferior, y-axis; Y, yaw.

Gaber et al. *Virtually Planned Orthognathic Surgery. J Oral Maxillofac Surg* 2017.

Table 5. QUALITY SCORE *

Question	Baan et al, ⁴¹ 2016	Hernandez-Alfaro and Guijarro-Martinez, ³⁹ 2013	Hsu et al, ³⁶ 2013	Shehab et al, ⁴⁰ 2013	Aboul-Hosn Centenero and Hernandez-Alfaro, ⁴ 2012	Zinser et al, ³⁸ 2012	Tucker et al, ³⁷ 2010
Clearly stated aim	2	2	2	2	2	2	2
Inclusion of consecutive patients	2	1	1	2	2	1	2
Prospective collection of data	1	2	1	2	2	2	2
Endpoints appropriate to aim of study	2	2	2	1	1	2	2
Unbiased assessment of study endpoint	0	0	0	0	0	0	0
Follow-up period appropriate to aim of study	2	2	2	2	2	2	2
Loss to follow-up <5%	2	2	2	2	2	2	0
Prospective calculation of study size	0	0	0	0	0	0	0
Total score	11	11	11	11	12	11	10

* The items are scored as follows: 0, not reported; 1, reported but inadequate; or 2, reported and adequate. The global ideal score is 16.

Gaber et al. *Virtually Planned Orthognathic Surgery. J Oral Maxillofac Surg* 2017.

Three-dimensional virtual planning has been used in orthognathic surgery for about 25 years.⁴² Recently, attempts have been made to assess the accuracy of the planning in comparison with the actual outcome. Two systematic reviews commenting on the accuracy of virtually planned surgery were published in recent years.^{13,14} The studies included in these reviews matched the studies included in our review. Apart from the study by Baan et al,⁴¹ all other studies were included in at least 1 of the previous reviews.

Even though the accuracy of virtually planned orthognathic surgery has been highly investigated recently, assessment methods are not recognized at the same level. The 3D evaluation of virtually planned orthognathic surgery fully uses the benefits of virtual planning, enriching our knowledge about the effects of orthognathic surgery on both the hard and soft tissues.⁴³ Combining virtual planning and better knowledge improves the overall experience for the surgeon; this in turn enhances the prediction of outcomes and thus yields a better diagnosis and improved preoperative planning.

The workflow for the assessment of accuracy could be broadly divided into 3 main steps:

- Registration (superimposition) of postoperative 3D findings to the planned model
- Assessment
- Validation

REGISTRATION

Currently, 3 methods of rigid registration are available to superimpose the postoperative model onto the preoperatively planned object: point-, surface-, or voxel-based registration.⁴⁴ Superimposition has to be performed in a non-surgically exposed region such as the cranial part.

Point-based registration relies mainly on predefined cephalometric landmarks in unoperated regions.³⁸ This type of registration is considered the least accurate, with an average error of 1 mm.⁴⁵

Surface-to-surface or surface best-fit registration relies on automated superimposition, which is considered more accurate and objective than the subjective manual point selection needed for point registration. It is believed that voxel-based registration, first introduced by Cevdanes et al⁴⁶ in 2005, is the most accurate method, given that it relies on volumetric data rather than the surfaces of 3D objects, which might be affected by the quality of the segmentation performed. Yet, in a study by Al-mukhtar et al,⁴⁷ no statistical significance was found between surface- and voxel-based registration methods.

Before either point- or surface-based registration, a segmentation step has to be performed. Currently, most software tools retrieve the 3D object from the Digital Imaging and Communications in Medicine (DICOM) images through global thresholding, selecting a range of Hounsfield units (HU) or grey values presuming that this range will show the best image quality of bone and exclude undesired artifacts and soft tissue. Multislice computed tomography (MSCT) relies on Hounsfield units, which is the gold standard, whereas CBCT, most commonly used in maxillofacial surgery, relies on grey values, with a considerable difference of up to 1.2 mm of error between CBCT and MSCT.⁴⁸ More errors are introduced into the segmentation procedure, which is sensitive and subject to the cutoff threshold level and streak metal artifacts owing to the presence of orthodontic brackets, thus negatively affecting the accuracy of the surface-based registration. Therefore, voxel-based registration is recommended because it lacks the necessity for segmentation.

Tucker et al³⁷ and Baan et al⁴¹ used voxel-based registration on the cranial part with a positive effect on the results. Hsu et al³⁶ and Hernandez-Alfaro and Guijarro-Martinez³⁹ presented a valid method for superimposition via surface best fit, whereas Zinser et al³⁸ relied on point-based registration, which introduced human error into the results. Shehab et al⁴⁰ and Aboul-Hosn Centenero and Hernandez-Alfaro⁴ adopted a registration-free method in which they compared results using cephalometric tracing on both the preoperative and postoperative images.

ASSESSMENT

The methods of accuracy assessment of the actual output compared with the virtual plan varied. The studies used linear and/or angular measurements, surface-to-surface differences, or virtual triangles created from 3 landmarks to assess the change in 3D space in relation to the axes X, Y, and Z.

Shehab et al⁴⁰ assessed the accuracy of their 3D virtual planning method using linear and angular measurements on 2-dimensional lateral cephalograms preoperatively and postoperatively. The accuracy assessment was therefore evaluated in only 1 direction (pitch) whereas the roll and yaw directions were not examined; thus this evaluation method can be considered incomplete.

The most commonly used approach was linear and angular measurements,^{4,38} which were implemented in different ways. These measurements are based on the identification of cephalometric landmarks, which is susceptible to human error ranging from 0.3 to 2.8 mm,⁴⁹ especially when these landmarks had to

be identified once on the preoperatively planned object and another time on the postoperative object.

Aboul-Hosn Centenero and Hernandez-Alfaro⁴ and Zinser et al³⁸ relied on identifying a number of cephalometric landmarks to measure distances and angles in relation to the axial, sagittal, and coronal axes. Furthermore, Zinser et al proposed another set of measurements described as volumetric changes, but in fact, these only represented surfaces changes measured on specific points, with the conclusion that these also can be categorized as linear point-based measurements.

Hernandez-Alfaro and Guijarro-Martinez³⁹ scanned the intraoperative position of the dentition in the intermediate splint using an intraoral scanner and compared these surfaces with the planned objects using software from The MathWorks (Natick, MA) that provided a color scheme diagram. They reported the mean and standard deviation of the difference in distance between the surfaces. Because the technique only assesses the intermediate position of the 2 jaws while connected with the intermediate wafer, in its current form, it will be far from usable for the assessment of comparing the plan versus the surgical outcome.

A better implementation of the surface comparison concept is evident in the study by Tucker et al.³⁷ They evaluated the accuracy of the virtual plan based on the surface distance differences between the plan and the actual outcome for 11 different regions of the maxilla and mandible. Although this method accurately evaluated the effect of the surgical procedure on the operated regions, it is limited in the clinical application answering the surgeon's question of whether his or her plan was fulfilled.

Hsu et al³⁶ measured the differences between the centroids of 3 landmarks per axis (x, y, z) linearly and angularly to quantify the movement of the jaws in each dimension in 3D space. Baan et al⁴¹ developed a tool called the "OrthoGnathicAnalyser" that relies on the concept of virtual triangles in which 3 landmarks are identified once per object preoperatively; then, the software automatically measures the linear and angular changes for the pitch, roll, and yaw directions in a clinically relevant manner. They eliminated human error for the identification of cephalometric landmarks through the voxel-based registration of the jaw segments.

VALIDATION

Different ideas and technologies to validate planning and assessment methods, as well as checking their reproducibility and reliability, have been proposed. The most widely used method by studies included in our review was interobserver and

intraobserver reliability correlation (r), which was applied by Baan et al⁴¹ and Zinser et al,³⁸ as well as Shehab et al,⁴⁰ who only assessed interobserver agreement. This form of correlation, despite being easy to perform and being widely accepted by surgeons, does not give an indication of the quality or accuracy of the modality used.

An alternative method for the validation of newly introduced measurement or assessment modalities was proposed by Bland and Altman.⁵⁰ Their statistical method, which was introduced in 1986, measures the agreement rather than the correlation between 2 measurement modalities being compared. According to Bland and Altman, the presence of a positive correlation between measurement modalities does not indicate that one could replace the other. This assessment of agreement was used by Xia et al³⁵ to validate the technique used, but instead of comparing 2 measurement modalities, it was used to compare the agreement and accuracy of the coordinates of 3 given points between planned and postoperative skull models with the use of a surface-based fit registration method. It was quite unorthodox to use the Bland-Altman method for the validation of the registration method, and to the best of our knowledge, it was not used for the same purpose in similar other studies.

Tucker et al³⁷ tried a unique idea to validate the planning and check the accuracy of the virtual plan. They repeated the surgical plan for a selected number of patients and compared it with the actual outcomes. This method really tests the workflow of the planning, as well as its reproducibility.

Aboul-Hosn Centenero and Hernandez-Alfaro⁴ compared the accuracy of 3D printed wafers with those produced by a model surgery technique. In our opinion this would be the ideal method to measure and assess the accuracy of the virtual planning protocol if one is still initiating the shift from model surgery to virtual planning. Nonetheless, comparison of the splint accuracy was performed through the assessment of the intraclass correlation between different distances measured on the patient's bone intraoperatively. It would have been more reliable and accurate if the actual measurements and the mean differences were calculated, rather than checking the correlation between the measurements. Yet, Stokbro et al¹³ considered that this change of quantitative data into qualitative categorical forms allows comparison between different studies.

Hernandez-Alfaro and Guijarro-Martinez³⁹ performed an ex vivo validation (dry human skulls), which turned out to be highly predictable. The results of this study allowed them to perform a further analysis in a clinical trial. Such in vitro or ex vivo validation is essential and is a mandatory step toward applying new software and analyses on clinical data.

QUALITY ASSESSMENT

The included articles were all of moderate quality; none used blinded assessment or sample size calculation. We assume that the absence of blinded assessors might be attributed partly to the fact that the use of virtual planning and evaluation is still considered a recent trend by many centers, which probably limits the availability of a sufficient number of practitioners who have mastered the required techniques. This deficiency in the number of trained assessors is what makes it mandatory for the surgeon who performed the planning to also evaluate the outcome. Lack of power analysis for pilot studies might be accepted or a common practice, but the lack thereof for clinical trials renders clinical trials more liable to criticism despite the large sample sizes.^{36,37}

Studies with large sample sizes did not follow the ideal methods for sample reporting or study conduction, either by not commenting on the final number of included patients and mentioning dropouts or by the inclusion of different surgical centers in the trial after the initiation of the study.^{36,37}

In our review, in an attempt to maintain a minimum level of quality among the included studies, we only included prospective clinical trials. Such a restriction might have resulted in missing out on other studies that used alternative assessment or validation methods.

RECOMMENDATIONS

On the basis of the findings of this study, it became clear that the literature is deficient in high-quality clinical trials using validated methods. A lack of consensus between different centers regarding assessment and validation methods also was noticed. The assessment tool designed and used in the study of Baan et al⁴¹ seems to be the best method currently available for assessing planning accuracy because it uses voxel-based registration and the assessment is semiautomated, requiring a minimal number of manual point identification; both steps aim at decreasing the margin of error. Yet, there is a drawback to this as well: It is software dependent, only compatible with the authors' planning software.

Hence, we suggest a protocol or criteria (Fig 2) for ideal (with the lowest margin of error) assessment of the accuracy of virtually planned orthognathic surgery. This protocol could be achieved through a single software tool or multiple software tools, depending on the capabilities and limitations of each available software program:

1. Voxel-based registration, on the cranial base, of planned and postoperative images should be performed to decrease the possibility of human error.

FIGURE 2. Proposed criteria for ideal evaluation of outcome.

Gaber et al. *Virtually Planned Orthognathic Surgery*. *J Oral Maxillofac Surg* 2017.

2. Automated or semiautomated evaluation of the outcome should be performed. The outcome should be indicative of changes in 3D—whether it is translational or rotational based on the different axes (x, y, z).
3. Interobserver and intraobserver reliability should be used to validate the results.

Future studies implementing the proposed assessment protocol are needed to evaluate its applicability and superiority to currently used methods.

Acknowledgments

The authors thank the Egyptian Ministry of Higher Education for funding the short-term research scholarship of Ramy M. Gaber at the OMFS-IMPACT Research Group, KU Leuven.

References

1. Downs W: Analysis of the dentofacial profile. *Angle Orthod* 26: 191, 1956
2. Arnett G, Bergman R: Facial keys to orthodontic diagnosis and treatment planning. Part I. *Am J Orthod Dentofacial Orthop* 103:299, 1993
3. Reyneke J: Diagnosis and treatment planning, in *Essentials of Orthognathic Surgery*. Chicago, IL, Quintessence Books, 2003, pp 69–150
4. Aboul-Hosn Centenero S, Hernandez-Alfaro F: 3D planning in orthognathic surgery: CAD/CAM surgical splints and prediction of the soft and hard tissues results—Our experience in 16 cases. *J Craniomaxillofac Surg* 40:162, 2012
5. Gateno J, Xia J, Teichgraber J, Rosen A: A new technique for the creation of a computerized composite skull model. *J Oral Maxillofac Surg* 61:222, 2003

6. Swennen G, Schutyser F, Hausamen J: *Three Dimensional Cephalometry. A Color Atlas and Manual*. Heidelberg, Springer, 2005
7. Swennen G, Mollemans W, Schutyser F: Three-dimensional treatment planning of orthognathic surgery in the era of virtual imaging. *J Oral Maxillofac Surg* 67:2080, 2009
8. Patel P, Zhao L, Ferraro J: Surgical planning: 2D to 3D, in Taub P, Patel P, Buchman S, Cohen M (eds): *Ferraro's Fundamentals of Maxillofacial Surgery* (ed 2). New York, NY, Springer, 2015, pp 331–368
9. Swennen G: Timing of three-dimensional virtual treatment planning of orthognathic surgery: A prospective single-surgeon evaluation on 350 consecutive cases. *Oral Maxillofac Surg Clin North Am* 26:475, 2014
10. Metzger M, Hohlweg-Majert B, Schwartz U, et al: Manufacturing splints for orthognathic surgery using a three-dimensional printer. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 105:e1, 2008
11. Song K, Baek S: Comparison of the accuracy of the three-dimensional virtual method and the conventional manual method for model surgery and intermediate wafer fabrication. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 107:13, 2009
12. Gateno J, Xia J, Teichgraber J, et al: The precision of computer-generated surgical splints. *J Oral Maxillofac Surg* 61:814, 2003
13. Stokbro K, Aagaard E, Torkov P, et al: Virtual planning in orthognathic surgery. *Int J Oral Maxillofac Surg* 43:957, 2014
14. Haas JO, Becker O, de Oliveira R: Computer-aided planning in orthognathic surgery—Systematic review. *Int J Oral Maxillofac Surg* 44:329, 2015
15. Slim K, Nini E, Forestier D, et al: Methodological Index for Non-Randomized Studies (MINORS): Development and validation of a new instrument. *ANZ J Surg* 73:712, 2003
16. Marchetti C, Bianchi A, Bassi M, et al: Mathematical modeling and numerical simulation in maxillofacial surgery. *J Craniofac Surg* 18:826, 2007
17. Mazzoni S, Badiali G, Lancellotti L, et al: Simulation-guided navigation: A new approach to improve intraoperative three-dimensional reproducibility during orthognathic surgery. *J Craniofac Surg* 21:1698, 2010
18. Li B, Zhang L, Sun H, et al: A novel method of computer aided orthognathic surgery using individual CAD/CAM templates: A combination of osteotomy and repositioning guides. *Br J Oral Maxillofac Surg* 51:e239, 2013
19. Zinser M, Mischowski R, Dreiseidler T, et al: Computer-assisted orthognathic surgery: Waferless maxillary positioning, versatility, and accuracy of image-guided visualisation display. *Br J Oral Maxillofac Surg* 51:827, 2013
20. De Riu G, Meloni S, Baj A, et al: Computer-assisted orthognathic surgery for correction of facial asymmetry: Results of a randomised controlled clinical trial. *Br J Oral Maxillofac Surg* 52: 251, 2014
21. Ying B, Ye N, Jigang Y, et al: Correction of facial asymmetry associated with vertical maxillary excess and mandibular prognathism by combined orthognathic surgery and guiding templates and splints fabricated by rapid prototyping technique. *Int J Oral Maxillofac Surg* 44:1330, 2015
22. Mazzoni S, Bianchi A, Schiariti G, et al: Computer-aided design and computer-aided manufacturing cutting guides and customized titanium plates are useful in upper maxilla waferless repositioning. *J Oral Maxillofac Surg* 73:701, 2015
23. Xia J, Ip H, Samman N, et al: Three-dimensional virtual-reality surgical planning and soft-tissue prediction for orthognathic surgery. *IEEE Trans Inf Technol Biomed* 5:97, 2001
24. Schneider M, Tzscharnke O, Pilling E, et al: Comparison of the predicted surgical results following virtual planning with those actually achieved following bimaxillary operation of dysgnathia. *J Craniomaxillofac Surg* 33:8, 2005
25. Bell R: Computer planning and intraoperative navigation in orthognathic surgery. *J Oral Maxillofac Surg* 69:592, 2009
26. Nadjmi N, Mollemans W, Daelemans A, et al: Virtual occlusion in planning orthognathic surgical procedures. *Int J Oral Maxillofac Surg* 39:457, 2010
27. Olszewski R, Tranduy K, Reyhler H: Innovative procedure for computer-assisted genioplasty: Three-dimensional

- cephalometry, rapid-prototyping model and surgical splint. *Int J Oral Maxillofac Surg* 39:721, 2010
28. McCormick S, Drew S: Virtual model surgery for efficient planning and surgical performance. *J Oral Maxillofac Surg* 69:638, 2011
 29. Sun Y, Luebbers H, Agbaje J, et al: Accuracy of upper jaw repositioning with intermediate splint fabrication after virtual planning in bimaxillary orthognathic surgery. *J Craniofac Surg* 24: 1871, 2013
 30. Donatsky O, Hillerup S, Bjorn-Jorgensen J, Jacobsen P: Computerized cephalometric orthognathic surgical simulation, prediction and postoperative evaluation of precision. *Int J Oral Maxillofac Surg* 21:199, 1992
 31. Donatsky O, Bjorn-Jorgensen J, Hermund N, et al: Immediate postoperative outcome of orthognathic surgical planning, and prediction of positional changes in hard and soft tissue, independently of the extent and direction of the surgical corrections required. *Br J Oral Maxillofac Surg* 49:386, 2011
 32. de Paula L, Ruellas A, Paniagua B, et al: One-year assessment of surgical outcomes in class III patients using cone beam computed tomography. *Int J Oral Maxillofac Surg* 42:780, 2013
 33. Marmulla R, Luth T, Muhling J, Hassfeld S: Markerless laser registration in image-guided oral and maxillofacial surgery. *J Oral Maxillofac Surg* 62:845, 2004
 34. Zhao L, Patel P, Cohen M: Application of virtual surgical planning with computer assisted design and manufacturing technology to cranio-maxillofacial surgery. *Arch Plast Surg* 39:309, 2012
 35. Xia J, Gateno J, Teichgraber J, et al: Accuracy of the computer-aided surgical simulation (CASS) system in the treatment of patients with complex craniomaxillofacial deformity: A pilot study. *J Oral Maxillofac Surg* 65:248, 2007
 36. Hsu S, Gateno J, Bell R, et al: Accuracy of a computer-aided surgical simulation protocol for orthognathic surgery: A prospective multicenter study. *J Oral Maxillofac Surg* 71:128, 2013
 37. Tucker S, Cevdanes L, Styner M, et al: Comparison of actual surgical outcomes and 3-dimensional surgical simulations. *J Oral Maxillofac Surg* 68:2412, 2010
 38. Zinser M, Mischkowski R, Sailer H, Zoller J: Computer-assisted orthognathic surgery: Feasibility study using multiple CAD/CAM surgical splints. *Oral Surg Oral Med Oral Pathol Oral Radiol* 113:673, 2012
 39. Hernandez-Alfaro F, Guijarro-Martinez R: New protocol for three-dimensional surgical planning and CAD/CAM splint generation in orthognathic surgery: An in vitro and in vivo study. *Int J Oral Maxillofac Surg* 42:1547, 2013
 40. Shehab M, Barakat A, Abdelghany K, et al: A novel design of a computer-generated splint for vertical repositioning of the maxilla after Le Fort I osteotomy. *Oral Surg Oral Med Oral Pathol Oral Radiol* 115:e16, 2013
 41. Baan F, Liebrechts J, Xi T, et al: A new 3D tool for assessing the accuracy of bimaxillary surgery: The OrthoGnathicAnalyser. *PLoS One* 11:e0149625, 2016
 42. Ayoub A, Khambay B, Benington P, et al: Prediction planning, *in* Handbook of Orthognathic Surgery: A Team Approach. West Sussex, Wiley Blackwell, 2014, pp 94-110
 43. Swennen G (ed): 3D Virtual Treatment Planning of Orthognathic Surgery. Berlin, Springer, 2017
 44. Swennen G, Schutyser F: Three-dimensional virtual approach to diagnosis and treatment planning of maxillofacial deformity, *in* : Bell W, Guerrero C (eds): Distraction Osteogenesis of the Facial Skeleton, Volume 6. Hamilton, Decker, 2007, pp 55-79
 45. Luebbers H, Messmer P, Obwegeser J, et al: Comparison of different registration methods for surgical navigation in cranio-maxillofacial surgery. *J Craniomaxillofac Surg* 36:109, 2008
 46. Cevdanes L, Bailey L, Tucker G Jr, et al: Superimposition of 3D cone-beam CT models of orthognathic surgery patients. *Dento-maxillofac Radiol* 34:369, 2005
 47. Almukhtar A, Ju X, Khambay B, et al: Comparison of the accuracy of voxel based registration and surface based registration for 3D assessment of surgical change following orthognathic surgery. *PLoS One* 9:e93402, 2014
 48. Loubele M, Maes F, Schutyser F, et al: Assessment of bone segmentation quality of cone-beam CT versus multislice spiral CT: A pilot study. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 102:225, 2006
 49. Makram M, Kamel H: Reeb graph for automatic 3D cephalometry. *Int J Image Process* 8:17, 2014
 50. Bland J, Altman D: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307, 1986