

Superimposition of serial 3-dimensional facial photographs to assess changes over time: A systematic review

Jonathan Johannes Wampfler and Nikolaos Gkantidis
Bern, Switzerland

Introduction: Superimpositions of 3-dimensional photographs enable a thorough and risk-free assessment of facial changes over time. However, the available methods and the evidence supporting them have not been assessed systematically. The paper summarizes and assesses the current evidence on superimposition methods of serial 3-dimensional facial photographs available in the literature. **Methods:** The following databases were searched without time restriction (last updated December 2020): MEDLINE via PubMed, EMBASE, Cochrane Library, and Google Scholar. Unpublished literature was searched on Open Grey and Grey Literature Report. Authors were contacted if necessary, and reference lists of relevant papers were screened. All studies with sample size ≥ 6 that tested the accuracy or precision of a superimposition technique, or agreement between different techniques regarding facial surface changes, were considered. The 2 authors performed data extraction independently using predefined forms. The risk of bias was assessed through the Quality Assessment and Diagnostic Accuracy Tool 2 tool. **Results:** Eight studies fulfilled the inclusion criteria. The total risk of bias of 7 studies was high and of 1 low. Seven studies had high total applicability concerns, and 1 was unclear. There was high heterogeneity among studies, which tested constructed planes through manually selected landmarks, a configuration of 9 landmarks, various surface areas, and the entire facial surface as superimposition references. A small rectangular area on the forehead combined with one on the middle part of the nose and the lower wall of the orbital foramen showed promising results. **Conclusions:** The limited available evidence suggests that surface-based registration is superior to landmark-based registration. Further research in the field is mandatory. (Am J Orthod Dentofacial Orthop 2021; ■ : ■ - ■)

In several disciplines of medicine and dentistry, facial photographs of individual patients are required for documentation, diagnosis, planning of the treatment, and treatment progress or outcome evaluation. Two-dimensional (2D) photographic analysis is still the most common imaging technique to evaluate facial soft tissues in clinical practice. The drawbacks of this 2D technique, when used to assess 3-dimensional (3D) structures, namely the face, include restricted validity and reliability errors because of head positioning, image distortion, and reduction of a 3D complex surface to a

2D image. Furthermore, the 2D images do not represent the real size because there is always a magnification factor.^{1,2}

Because of these limitations and the new possibilities available in recent years, 3D imaging techniques are gaining space in many disciplines, including plastic surgery, orthodontics, implantology, and maxillofacial surgery.³ Three-dimensional images include an enormous amount of information compared with conventional 2D images, which is needed for the proper morphologic assessment of complex structures.

Three-dimensional data of the facial surface can be obtained by different methods, including laser scanning, 3D stereophotogrammetry, patterned light techniques, conventional computed tomography, 3D cone-beam computed tomography (CBCT), and 3D magnetic resonance images. Compared with other methods, stereophotogrammetry is noninvasive, risk-free, accurate, user-friendly, and fast.⁴

To objectively evaluate the effects of treatment, growth, aging, or pathology in the morphology of anatomic structures, superimposition of 2 or more serial

From the Department of Orthodontics and Dentofacial Orthopedics, School of Dental Medicine, University of Bern, Bern, Switzerland.

All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and none were reported.

Address correspondence to: Nikolaos Gkantidis, Department of Orthodontics and Dentofacial Orthopedics, School of Dental Medicine, University of Bern, CH-3010, Freiburgstrasse 7, Bern, Switzerland; e-mail, nikosgant@yahoo.gr; nikolaos.gkantidis@zmk.unibe.ch.

Submitted, March 2021; revised and accepted, June 2021.

0889-5406

© 2021 by the American Association of Orthodontists. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1016/j.ajodo.2021.06.017>

images obtained at different time points is necessary.⁵ Different techniques have been suggested to superimpose serial 3D facial surface images, usually on various morphologically stable structures represented by areas, landmarks, or planes. In brief, these can be divided into 2 categories: landmark-based⁴ and surface-based methods.⁶ Landmark-based methods register serial 3D images on ≥ 3 manually selected corresponding anatomic landmarks.³ Surface-based registrations use anatomic areas as superimposition references, comparing the triangular representations of corresponding 3D surface geometries. Then, using appropriate software, automated matching of the serial images on the selected reference landmarks or areas is performed through specific mathematical algorithms that aim to minimize the distance between the corresponding landmarks or areas.^{3,7,8}

The information on 3D superimposition techniques to assess facial surface changes is scarce. So far, there is no widely accepted method for the superimposition of serial facial images. Many factors might affect the superimposition outcome, such as the image quality and the superimposition methods.^{5,9-12} However, an assessment of the available methods and the evidence supporting them in the context of a systematic review has not been undertaken.

OBJECTIVES

The purpose of this study was to summarize and critically assess all the relevant studies on the validity, reliability, and reproducibility of serial 3D facial photograph superimposition techniques through a systematic review of the literature.

MATERIAL AND METHODS

Protocol and registration

The study protocol was registered in PROSPERO (registration no. CRD42019134784) before study implementation.

Eligibility criteria

The following criteria were applied in this review: (1) prospective and retrospective study designs were considered eligible; (2) study sample: studies with sample size ≥ 6 ; (3) index test: surface-based or landmark/reference plane-based superimposition of serial 3D facial photographs; (4) types of participants: patients who underwent any change in their facial morphology because of growth, aging, pathology, or medical intervention. Natural changes of healthy patients assessed at very short term (eg, through a change in facial

expression) were also considered if used to test superimposition outcomes; (5) type of intervention: 3D superimposition to assess any change in facial morphology; (6) primary outcome: accuracy or precision of a superimposition technique, or agreement between different techniques measured in corresponding surface landmarks or areas as soft-tissue change; (7) studies that tested any of the above as secondary outcomes were also considered eligible; (8) comparator/control group: different superimposition techniques, direct measurements, or repeated measurements; (9) unit of analysis: in all cases, the unit of analysis was the measured change in a facial surface area or landmark; (10) follow-up: all observation periods between subsequent photographs were accepted if a morphologic change was expected; and (10) exclusion criteria: nonhuman-derived data.

Information sources, search strategy, and study selection

Search strategy. Relevant studies were collected through specifically developed search strategies in the following databases: MEDLINE via PubMed, EMBASE, Cochrane Library (all databases), and Google Scholar. No time restriction was applied to the search (last update: December 20, 2010). Reference lists of relevant papers, including reviews, were searched through a hand search. Unpublished literature was searched on the following databases: Open Grey, Grey Literature Report, and Google Scholar. The detailed search strategy is shown in [Supplementary Table 1](#). The authors of the included studies were contacted via e-mail for ongoing studies in the field that might be eligible.

Study selection. The 2 authors of this review searched for eligible studies independently. They were not blinded to the identity of the authors of the studies, their institutions, or the results of their research. Included studies were selected by the title and the abstract, and when needed, the full text was read. Noneligible papers were excluded at all stages. If there was a disagreement, the eligibility was discussed between the authors until a consensus was reached. A record of all decisions on study identification was kept.

Data items and collection

Data extraction was performed independently and duplicated by the 2 authors using a custom Excel sheet (Microsoft Excel; Microsoft, Richmond Wash) formed after pilot testing on 3 relevant studies. Disagreements were resolved by reexamining the source of information until a consensus was reached. The following information was extracted from all eligible studies, if available:

(1) methods: author, title, year, objectives, and design of the study; (2) participants: number, age, and gender of patients recruited; (3) materials: type of 3D facial soft-tissue acquisition method and time between serial models; (4) superimposition method: type of superimposition reference areas or points and software used; (5) comparison/control group: type and characteristics; and (6) outcome: type of outcome(s) and method of outcome assessment.

When needed, in the presence of missing information from a study, the authors were contacted by e-mail to request clarifications. In case that the authors did not respond or the data were not receivable, only the available information was regarded.

Risk of bias in individual studies

The quality of the selected studies was evaluated through the Quality Assessment and Diagnostic Accuracy Tool 2 (QUADAS-2) tool.¹³ This tool is recommended by the Cochrane Collaboration to assess the risk of bias and the applicability concerns of diagnostic accuracy studies in systematic reviews. It subdivides each study into 4 key domains: patient selection, index test, reference standard, and flow and timing. Those 4 domains are evaluated for the risk of bias and only the first 3 for applicability concerns. The results of the QUADAS-2 tool are commonly presented in a table using happy (low risk) or sad smiles (high risk). If it is not possible to evaluate a domain, an interrogation mark is set, which indicates an unclear risk. In case a domain is evaluated with a high risk, relevant justification is provided.

The risk of bias was performed independently and in duplicate by the 2 authors. In cases of disagreement, the judgment was discussed until a consensus was reached. In the case of a meta-analysis, studies with a high risk of bias will not be included.

Summary measures and approach to synthesis

Assessment of heterogeneity. The studies were assessed on the basis of the similarity between characteristics, participants, methods, and outcomes as reported in the inclusion criteria.

Data synthesis. A meta-analysis was conducted if there were >2 unclear or low risk of bias studies of similar comparisons, reporting the same outcomes at similar follow-up periods.

Assessment of reporting bias

We attempted to minimize potential reporting biases, including publication bias and multiple (duplicate reports) publication bias, by conducting an accurate but

sufficiently broad search of multiple sources. We also searched for ongoing studies.

Additional analysis. Potential subgroups that will be further assessed whenever possible include patients with vs without growth, short-term (within a week) vs medium-/long-term (>1 week) interval between serial photographs and surface-based vs landmark/reference plane-based techniques.

RESULTS

Study selection and characteristics of the included studies

Out of 1715 studies initially identified by the search, 8 studies fulfilled the eligibility criteria and were included in this review (Fig). From these, 6 studies were published articles,^{6,14-18} 1 was a master thesis,¹⁹ and 1 was completed and published as a preprint.²⁰ In all studies, 3D landmark-based or surface-based superimposition techniques were used to assess actual or simulated facial changes. The included studies evaluated the accuracy or precision of a superimposition technique, or agreement between different techniques, as a primary or a secondary outcome.

All included studies were prospective in terms of superimposition data generation. Three studies tested only growing patients,^{17,18,20} 4 studies nongrowing patients,^{6,14,15,19} and another study did not report this information.¹⁶

Regarding superimposition references, one study used constructed planes on the basis of manually selected landmarks,⁶ one study used the anterior forehead surface,¹⁴ another study used 9 landmarks,¹⁵ 2 other studies used the whole cropped image surface,^{16,17} 2 studies used the nasal bridge and forehead area,^{18,19} and, finally, 1 study tested 5 different surfaces as references.²⁰

All relevant details of the included studies are shown in Tables 1 and II.

Quality assessment

Quality assessment was performed through the QUADAS-2 tool and results are presented in Table III, including the reasons supporting judgments.

Seven included studies had a high total risk of bias,^{6,14-19} whereas 1 study²⁰ had a low total risk of bias. Regarding the individual domains, studies were mainly characterized as high risk on the basis of index tests and reference standards. In these domains, low risk was evident only for 1 study²⁰ and 2 studies,^{6,20} respectively.

Regarding the total applicability concerns, 7 studies had high,^{6,14-19} and 1 study had an unclear risk.²⁰

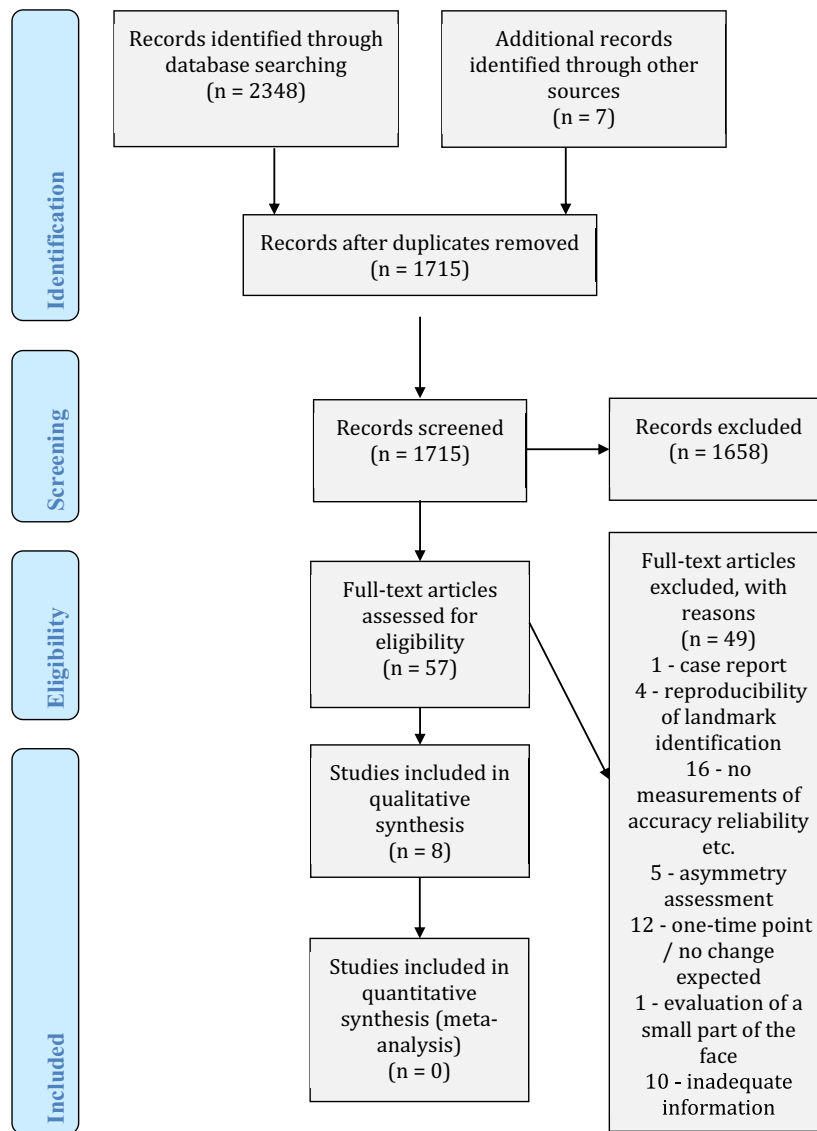


Fig. Flowchart of study selection according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Concerning individual items, studies were judged as having high concerns primarily because of index tests and reference standards. Only 1 study had low risk considering index test²⁰ and 2 for reference standard.^{6,20}

Results of individual studies and qualitative synthesis

The results, conclusions, and limitations of all included studies are provided in [Tables IV](#) and [V](#).

Landmark-based registration. There is only 1 study in this category, which assessed the superimposition of 3D facial photographs on 9 corresponding soft-tissue landmarks.¹⁵ The study focused on intraobserver and

interobserver variability of superimposition outcomes after facial alteration of 10 adult males because of mimicry. For this, 3D photographs of each participant with a happy, sad, angry, and surprised expression were superimposed with a neutral one. The study does not provide information on the reproducibility of the technique in individual patients. The presented approach seems to differentiate between different facial expressions in most patients. The results of this study should be interpreted with caution because of the absence of individual outcome assessment, no adequate landmark identification error, and no testing of repeated images with the same expression.

Table I. Main general characteristics of the included studies

Study	Main objective	Study design	Sample size	Age, y	Growth	Type of participants
Incrapera et al ¹⁴	To compare linear changes (preorthognathic and postorthognathic surgery) measured after superimpositions of lateral cephalometric radiographs and 3D images from the same patient at 5 soft-tissue landmarks	Prospective	n = 34	Mean: 37 ± 8.3	Nongrowing	Orthognathic patients (Class I, n = 13; Class II, n = 16; Class III, n = 5) Preorthognathic and postorthognathic surgery images (lateral cephalograms and 3D photographs) over 6 mo
Maal et al ⁶	To evaluate the agreement between surface-based and landmark-based registrations of serial 3D patient images acquired within 3 wk	Prospective	n = 15	Mean: 23.6 (range, 21-26)	Nongrowing	Three-dimensional photographs at 3 points; baseline, after 1 minute, 3 wk later, and 1 smiling 3D photograph at baseline. No intervention or extensive morphologic change in-between
Gibelli et al ¹⁵	To determine the intraobserver and interobserver variability of a facial soft-tissue landmark-based superimposition after facial alteration because of mimicry	Prospective	n = 10 (males)	Range: 30-40	Nongrowing	Five 3D photographs of each participant with different expressions: happy, sad, angry, surprised, and neutral
Adriaens et al ¹⁶	To determine the intraobserver and interobserver variability of a facial soft-tissue surface-based superimposition to measure volumetric changes at the masseter region after botulinum toxin type A therapy	Prospective	n = 10	NA	NA	Three-dimensional photographs of 3 patients with unilateral and 7 with bilateral masseter hypertrophy, before and 3 mo after injection
Altındış et al ¹⁸	To test the reproducibility of a surface-based superimposition on the volumetric assessment of upper lip changes after RPE	Prospective	n = 42	Range: 11.6-13.3 at baseline	Growing	Three-dimensional photographs of patients with moderate maxillary constriction (approximately 5 mm expansion) at 2 points: baseline and after RPE appliance removal at the end of the 3-mo retention period
Fay ¹⁹	To test the reproducibility of a surface-based superimposition on the volumetric assessment of postsurgical edema after orthognathic surgery	Prospective	n = 30	Mean: 26.2 ± 12.0	Nongrowing or with minimal growth	Three-dimensional photographs of orthognathic surgery patients receiving steroids to control postoperative swelling: before hospital discharge and 4 wk postoperative
Häner et al ²⁰	To test various 3D facial superimposition reference areas in growing patients and compare them to a voxel-based anterior cranial base superimposition, which is considered the gold standard approach	Prospective methodologic (retrospective data)	n = 18 (9 males, 9 females)	Mean: 11.7 ± 0.6 at baseline	Growing	Serial CBCT derived models of orthodontics patients with a period of 1-3 y
Dindaroğlu et al ¹⁷	To determine the intraobserver variability of a facial soft-tissue surface-based superimposition to measure facial changes after RPE	Prospective	n = 10	Range: 8.1-12.6	Growing	Two photographs of patients with unilateral or bilateral posterior crossbite at 2 points: before appliance placement and after the end of expansion

NA, not available; RPE, rapid palatal expansion.

Table II. Main superimposition-related characteristics of the included studies

<i>Study</i>	<i>Machines</i>	<i>Acquisition parameters</i>	<i>Superimposition method</i>	<i>References for superimposition</i>	<i>Superimposition software</i>	<i>Comparison/control</i>
Incrapera et al ¹⁴	Kodak 8000C digital panoramic and cephalometric system machine (Kodak, Rochester, NY) 3dMD stereophotogrammetric camera capture system	3dMD acquisition time: 1.5 ms Both machines: same sitting position, natural head position, images taken in similar matching facial poses	Cephalometry: cranial base (hand tracing, manual best fit) Three-dimensional photographs: automatic best-fit surface registration	Cephalometry: Cranial base (SN line), anterior clinoid process, and sphenothmoid plane Three-dimensional photographs: nasal bridge and forehead area	Three-dimensional photographs: Rapidform 2006 (Rapidform Inc, Sunnyvale, Calif)	Three-dimensional photographs vs cephalometric radiographs regarding the superimposition outcome at 5 soft-tissue landmarks (STN, TN, UL, LL, STP)
Maal et al ⁶	3dMD stereophotogrammetric camera capture system	Bite in centric occlusion, relaxed lips, open eyes.	Surface-based registration (best fit) on original and cropped images and landmark-based registration	Surface-based registration: whole 3D image (original and cropped) Landmark (constructed planes)-based registration: left and right exocanthion, right supraaurale, pupil reconstructed point (in the middle of the line through the pupils); the horizontal plane was defined as a plane created through a 6.6° downward rotation of the canthion-superaural line passing through the pupil reconstructed point; the vertical plane was the coronal plane constructed perpendicular to the horizontal plane and along the horizontal direction of the natural head position; the median plane was perpendicular to the horizontal and vertical planes; the superimposition reference frame was formed by the horizontal, the vertical, and the medial plane	Surface-based registration: Maxilim (version 2.2.1; Medicim NV, Mechelen, Belgium) and 3dMD (version 3.0, 3dMDpatient Software Platform). Landmark-based registration: Maxilim	Surface-based vs landmark-based registration

Table II. Continued

Study	Machines	Acquisition parameters	Superimposition method	References for superimposition	Superimposition software	Comparison/control
Gibelli et al ¹⁵	VECTRA-3D (Canfield Scientific, Inc, Fairfield, NJ)	NA	Landmark-based registration (best fit of 9 landmarks) of all facial expressions with 1 neutral landmark	Nine landmarks: right and left endocanthion, right and left exocanthion, right and left cheilion, sellion, pronasale, subnasale	VAM software (Canfield Scientific Inc)	Intraobserver and interobserver error on the whole sample
Adriaens et al ¹⁶	3dMD stereophotogrammetric camera capture system	Acquisition time: 10 ms	Surface-based registration (best fit)	Surface-based registration: whole 3D image (cropped)	Maxilim	Intraobserver and interobserver error
Altındış et al ¹⁸	3dMD stereophotogrammetric camera capture system	Acquisition time: 1.5 ms	Surface-based registration (best fit)	Nasal bridge and forehead area (RMSD \leq 0.5 mm considered acceptable)	3dMD Vultus software	Intraobserver error
Fay ¹⁹	3dMD stereophotogrammetric camera capture system	Acquisition time: 1.5 ms	Surface-based registration (best fit)	Nasal bridge and forehead area (RMSD \leq 0.5 mm considered acceptable)	3dMD Vultus software	Intraobserver and interobserver error
Häner et al ²⁰	KaVo 3D eXam (Hatfield, Pa)	Voxel size: 0.4 mm; 3.5 mA tube current; 120 kVa tube voltage; field of view of 170-mm height \times 232-mm diameter; scan time 8.9 s, exposure time 3.7 s	CBCT derived facial surface models: surface-based registration (best fit) CBCT radiographs: voxel-based registration (best fit)	Facial surface models: area 1, whole facial surface excluding the eyes, the mouth, and the tip of the nose; area 2, forehead glabella and base of the nose; area 3, upper half of the face, excluding the eyes and the tip of the nose; area 4, a small rectangular area on the forehead and an area including the middle part of the nose and the lower wall of the orbital foramen; area 5, same as area 4 but without the area on the forehead CBCT radiographs: anterior cranial base	Facial surface models: Viewbox 4 Software, Athens, Greece) CBCT radiographs: Dolphin 3D software (Dolphin Imaging and Management Solutions, Chatsworth, Calif)	Facial surface model superimposition outcomes vs anterior cranial base superimposition outcome
Dindaroğlu et al ¹⁷	3dMD Flex System	NA	Surface-based registration (best fit)	Surface-based registration: whole face cropped 3D image (the hair, ears, and the below-neck region were removed)	Rapidform 2006	Intraobserver error

NA, not available; RMSD, root mean squared distance.

Table III. Quality assessment of the included studies through the QUADAS-2 tool

Study	Risk of bias					Applicability concerns			
	Patient selection	Index test	Reference standards	Flow and timing	Total risk of bias	Patient selection	Index test	Reference standards	Total applicability concerns
Incrapera et al ¹⁴	⊖	⊖ (No assessment of individual differences)	⊖ (No assessment of individual differences)	⊖	⊖ (No assessment of individual differences)	⊖	⊖ (No assessment of individual differences)	⊖ (No assessment of individual differences)	⊖ (No assessment of individual differences)
Maal et al ⁶	⊖ (No intervention or expected change in-between the serial images)	⊖ (No assessment of individual differences, no color maps)	⊖	⊖	⊖	⊖ (No intervention or expected change in-between the serial images)	⊖ (No assessment of individual differences)	⊖	⊖
Gibelli et al ¹⁵	⊖	⊖ (Small sample size, no assessment of individual differences)	⊖ (No assessment of superimposition outcomes on repeated images with the same expression)	⊖	⊖	⊖	⊖ (No assessment of individual differences)	⊖ (No assessment of individual differences)	⊖
Adriaens et al ¹⁶	⊖	⊖ (Small sample size, no assessment of individual differences)	⊖ (No assessment of individual differences)	⊖	⊖	⊖	⊖ (No assessment of individual differences)	⊖ (No assessment of individual differences)	⊖
Altındış et al ¹⁸	⊖	⊖ (No assessment of individual differences, no color maps)	⊖ (No assessment of individual differences, no color maps)	⊖	⊖	⊖	⊖ (No assessment of individual differences)	⊖ (No assessment of individual differences)	⊖
Fay ¹⁹	⊖	⊖ (No assessment of individual differences, no color maps)	⊖ (No assessment of individual differences, no color maps)	⊖	⊖	⊖	⊖ (No assessment of individual differences)	⊖ (No assessment of individual differences)	⊖
Häner et al ²⁰	⊖	⊖	⊖	⊖	⊖	? (Moderate sample size, testing of a specific age group)	⊖	⊖	?
Dindaroğlu et al ¹⁷	⊖	⊖ (No assessment of method accuracy, no assessment of individual differences, no detailed data presented, type of intraclass correlation coefficient not specified)	⊖ (No power calculation of reliability testing)	⊖	⊖	⊖	⊖ (No assessment method accuracy, no assessment of individual differences, no detailed data presented, type of intraclass correlation coefficient not specified)	⊖ (No power calculation for reliability testing)	⊖

So far, no solid conclusions can be drawn regarding landmark-based registration of 3D facial photographs because there is only 1 study on the topic, which has severe drawbacks.

Surface-based registration. There are 6 studies that solely tested surface-based registration of serial 3D facial photographs.^{14,16-20} Incrapera et al¹⁴ compared linear facial soft-tissue changes (preorthognathic and postorthognathic surgery) of 34 nongrowing patients, measured through superimpositions of lateral cephalometric radiographs and 3D photographs. Distances were measured between 5 corresponding soft-tissue landmarks. The superimposition of lateral cephalometric radiographs and 3D images was done by manual best fit of hand tracings and automatic best fit of surfaces. The cephalometric superimposition references included the cranial base (SN line), the anterior clinoid process, and the sphenothmoidale plane. The 3D image superimposition references included the anterior forehead, the soft-tissue glabella, the nasion, and the bridge of the nose area. The study showed that the mean linear changes measured through superimpositions of 2D radiographs or 3D photographs were not statistically different. However, the study had several limitations that questioned the validity of the findings. These include the absence of magnification measurement for the lateral cephalograms and of an assessment of individual differences through analysis, such as the Bland Altman test. Furthermore, no method errors were reported.

Adriaens et al¹⁶ tested the intraobserver and interobserver reproducibility of masseter muscle volume differences after botulinum toxin type A therapy in 10 patients with unilateral or bilateral masseter hypertrophy. Three-dimensional photographs were obtained with a stereophotogrammetric camera (3dMD, Atlanta, Ga), before and 3 months after injection. The superimposition reference was the entire face after cropping the surrounding structures. This superimposition might show adequate reproducibility when mean changes in a group of patients are considered. Limitations such as small sample size, absence of Bland Altman or similar method to assess individual differences, and poor reporting weaken the evidence provided by the study.

Häner et al²⁰ analyzed the superimposition outcomes of 5 different facial surface superimposition techniques compared with the voxel-based anterior cranial base superimposition. The study assessed outcomes at 7 soft-tissue areas distributed over the entire facial surface. Eighteen serial CBCT derived facial surface models of growing patients, with a span of 1-3 years, were selected for this study. The superimposition references for the facial surfaces were (1) the whole facial surface excluding the eyes, the mouth, and the tip of the nose;

(2) the forehead and base of the nose; (3) the upper half of the face excluding the eyes and the tip of the nose; (4) a small rectangular area on the forehead and an area including the middle part of the nose and the lower wall of the orbital foramen; and (5) the same area as the latter, but without the area on the forehead. The best agreement of facial surface superimposition outcomes compared with the anterior cranial base technique was evident for technique 4. The study had a low total risk of bias, but unclear applicability concerns because of the moderate sample size, with relatively broad baseline characteristics, such as malocclusion type, and the limited age range.

Dindaroğlu et al¹⁷ evaluated the intraobserver variability of a surface-based superimposition to measure facial changes after rapid maxillary expansion. Three-dimensional photographs of 25 growing patients with posterior crossbite malocclusion were taken before and after expansion. Reproducibility was tested on the repeated superimposition of 10 randomly selected pairs of photographs using intraclass correlation. The whole face was used as a superimposition reference. The method indicated adequate reproducibility, but the study had significant limitations. These include the absence of a control/comparison method, the small sample size, no assessment of accuracy or reproducibility of individual measurements, and poor reporting.

Altındış et al¹⁸ tested the intraobserver reproducibility of a surface-based superimposition on volumetric assessment of upper lip changes after rapid palatal expansion. Forty-two 3D photographs of growing patients with moderate maxillary constriction were taken before and after rapid palatal expansion and were superimposed on the nasal bridge and forehead area. The presented superimposition might show adequate intraobserver reproducibility when mean changes on a group of patients are considered. Limitations of this study are that only mean values were assessed, no color maps were shown, no comparative statistics were applied, and the type of intraclass correlation coefficient was not reported.

Fay¹⁹ examined intraobserver and interobserver reproducibility of facial volume measurements between superimposed facial 3D photographs. Photographs of 30 orthognathic surgery patients receiving steroids to control postoperative swelling were taken before hospital discharge and 4 weeks postoperative. The 2 records were superimposed on the nasal bridge and forehead area to illustrate volume differences. The presented method might show adequate intraobserver and interobserver reproducibility when mean changes on a group of patients are analyzed. The study considers only mean values, which is a significant weakness. Furthermore, no

Table IV. Results of the included studies

Study	Main outcome	Secondary outcomes	Type of analysis	Main results	Secondary results
Incrapera et al ¹⁴	Linear distances of 5 pairs of corresponding soft-tissue landmarks, after superimposition of serial Ceph radiographs or facial 3D photographs	NA	Comparative analysis (paired <i>t</i> test)	STN Ceph: 1.07 ± 0.56 mm; 3D: 1.26 ± 0.64 mm TN Ceph: 2.09 ± 1.44 mm; 3D: 2.65 ± 1.48 mm UL Ceph: 5.39 ± 3.13 mm; 3D: 5.59 ± 3.06 mm LL Ceph: 5.93 ± 2.96 mm; 3D: 5.87 ± 2.83 mm STP Ceph: 8.07 ± 5.16 mm; 3D: 7.34 ± 3.94 mm <i>t</i> tests <i>P</i> > 0.05	NA
Maal et al ⁶	Mean absolute distance between superimposed serial 3D photographs obtained at a short-term interval (baseline, T0; after 1 minute, T1; 3 wk later, T2; smiling at T0, TL), using a surface and a landmark-based method. The whole face comprised the measurement area	Difference between 2 software packages (Maxilim [Medicim NV, Mechelen, Belgium] and 3dMD) used for surface-based registration Difference between original and cropped 3D photographs used for the surface-based registration Difference between rest and smiling photographs (T0 – TL) Intraobserver and interobserver error of landmark-based registration	Descriptive statistics	Mean distances after surface-based superimposition. Whole image; T0 – T1: 0.39 ± 0.50 mm, T0 – T2: 0.52 ± 0.62 mm, T0 – TL: 1.2 ± 0.25 mm Cropped image; T0 – T1: 0.30 ± 0.34 mm; T0 – T2: 0.42 ± 0.42 mm, T0 – TL: 1.2 ± 1.31 mm Mean distances after landmark-based superimposition. T0 – T1: 1.00 mm, T0 – T2: 1.10 mm, T0 – TL: 1.50 mm	Mean distances after surface-based superimposition Whole image 3dMD; T0 – T1: 0.39 ± 0.52 mm, T0 – T2: 0.53 ± 0.61 mm, T0 – TL: 1.24 ± 1.25 mm Whole image Maxilim; T0 – T1: 0.39 ± 0.48 mm, T0 – T2: 0.51 ± 0.61 mm, T0 – TL: 1.18 ± 1.24 mm Cropped image 3dMD; T0 – T1: 0.28 ± 0.33 mm, T0 – T2: 0.42 ± 0.41 mm, T0 – TL: 1.39 ± 1.29 mm Cropped image Maxilim; T0 – T1: 0.32 ± 0.34 mm, T0 – T2: 0.42 ± 0.42 mm, T0 – TL: 1.33 ± 1.32 mm Intraobserver and interobserver error for landmark-based method: 1.0 mm and 1.2 mm, respectively
Gibelli et al ¹⁵	Intraobserver and interobserver error of the RMSD between 9 corresponding landmarks located at superimposed facial 3D photographs	RMSD of the 9 landmarks between neutral vs. happy, sad, angry, and surprised photographs	Mean comparison analysis (Student <i>t</i> test)	Intraobserver error: <i>P</i> = 0.92 Interobserver error: <i>P</i> = 0.73	Neutral vs happy: 4.11 ± 1.13 mm Neutral vs sad: 1.3 ± 0.49 mm Neutral vs angry: 1.21 ± 0.37 mm Neutral vs surprised: 2.74 ± 1.02 mm <i>P</i> < 0.01 for all pairwise comparisons except sad vs angry

Study	Main outcome	Secondary outcomes	Type of analysis	Main results	Secondary results
Adriaens et al ¹⁶	Intraobserver and interobserver reproducibility of measurements of masseter muscle volume difference between superimposed facial 3D photographs	NA	Comparative analysis (Student <i>t</i> test) Pearson correlation coefficient	Intraobserver error: <i>P</i> = NS, <i>r</i> = 0.999. Interobserver error: <i>P</i> = NS, <i>r</i> = 0.990.	NA
Altındış et al ¹⁸	Intraobserver reproducibility of measurements of upper lip volume difference between superimposed facial 3D photographs	NA	ICC	ICC: 0.962-0.999 (95% CI)	NA
Fay ¹⁹	Intraobserver and interobserver reproducibility of measurements of facial volume difference between superimposed facial 3D photographs	RMSD of the reference area	Descriptive statistics and ICC	Intraobserver error: 0.8 ± 6.7 cm ³ ; ICC, 0.95 (95% CI, 0.90-0.98) Interobserver error: 2.1 ± 6.0 cm ³ ; ICC, 0.93 (95% CI, 0.86-0.97)	RMSD ≤ 0.2 mm
Häner et al ²⁰	Differences of the surface-based from the voxel-based ACB superimposition outcomes at the following 7 areas: soft-tissue N-point, soft-tissue A-point, soft-tissue pogonion, soft-tissue zygomatic arch right and left, and soft-tissue gonial angle right and left	Intraoperator agreement	Qualitative analysis of color-coded distance maps Comparative statistics for the 3 surface methods that were more similar to the voxel-based ACB method	Area 3 median: 1.37, range, 0.02-7.38 mm; area 4 median: 0.88, range, 0.03-6.93 mm; area 5 median 0.97, range, 0.03-8.61 mm; Kruskal-Wallis test: <i>P</i> = 0.001	The reproducibility of the tested surface-based superimposition methods, including the segmentation error, was in most patients within 0.5 mm
Dindaroğlu et al ¹⁷	Intraoperator reproducibility on the RMSD of 6 facial areas (total face, upper face, lower face, upper lip, lower lip, nose) between superimposed facial 3D photographs	NA	ICC	ICC, 0.911-0.964	NA

Ceph, cephalometric radiograph; *NA*, not available; *RMSD*, root mean squared distance; *NS*, nonsignificant; *ACB*, anterior cranial base; *ICC*, intraclass correlation coefficient; *CI*, confidence interval.

Table V. Conclusions and limitations of the included studies

Study	Conclusion	Limitations
Incrapera et al ¹⁴	The mean linear changes measured at 5 soft-tissue landmarks through superimpositions of serial cephalograms or 3D photographs were not statistically different	No evaluation of magnification in the lateral cephalograms (same machine); only mean values were assessed; no Bland Altman or similar method to assess individual differences; and no method error
Maal et al ⁶	Distances between registered images were less in cropped photographs and with surface-based registration (identical patients) than with landmark-based registration Differences between the software packages used for surface-based registration were negligible	Only mean values were assessed; no Bland Altman or similar method to assess individual differences. No color maps were shown; no comparative statistics were applied; and no intervention or expected morphologic change in-between the serial images
Gibelli et al ¹⁵	No valid conclusions can be drawn regarding the reproducibility of the used landmark-based superimposition in individual patients; the presented approach seems to differentiate between different facial expressions in most patients	Inadequate method error assessment (<i>t</i> test); small sample size (questionable power for error assessments); no evaluation of individual outcomes; no assessment of landmark identification error; and no assessment of superimposition outcomes on repeated images with the same expression
Adriaens et al ¹⁶	The presented superimposition might show adequate reproducibility when mean changes on a group of patients are considered	Small sample size (questionable power); only mean values were assessed; no Bland Altman or similar method to assess individual differences; and poor reporting
Altındış et al ¹⁸	The presented superimposition might show adequate intraobserver reproducibility when mean changes on a group of patients are considered	Only mean values were assessed; no Bland Altman or similar method to assess individual differences; no color maps were shown; no comparative statistics were applied; and the type of ICC was not reported
Fay ¹⁹	The presented superimposition might show adequate intra and interobserver reproducibility when mean changes on a group of patients are considered	Only mean values were assessed; no Bland Altman or similar method to assess individual differences; no color maps were shown; and the type of ICC was not reported
Häner et al ²⁰	The reference area selection considerably affects the superimposition outcomes of serial facial surface models; area 4, a small rectangular area on the forehead and an area including the middle part of the nose and the lower wall of the orbital foramen, shows comparable outcomes to the standard anterior cranial base superimposition technique, and it also shows adequate reproducibility	Moderate sample size; a specific age group was tested
Dindaroğlu et al ¹⁷	The presented reference for superimposition might show adequate reproducibility	Small sample size; no assessment of method accuracy; individual measurements of reproducibility were not assessed; no detailed data presented; and the ICC type is not specified

ICC, intraclass correlation coefficient.

color-coded distance maps were shown to assess the spatial distribution of changes.

Overall, 6 studies tested a surface-based superimposition approach for 3D facial photographs. However, 4 of them¹⁶⁻¹⁹ tested only reproducibility, and all had important limitations. A fifth study tested the agreement of facial soft-tissue superimposition with cephalometric superimposition, but it also had significant limitations.¹⁴ Finally, 1 low risk of bias study showed promising findings for a surface-based superimposition on a small rectangular area of the forehead plus an area at the middle part of the nose and the lower wall of the orbital foramen.²⁰ Nevertheless, this study also needs testing in different samples to generalize the applicability of the findings.

Landmark-based vs surface-based registration. Finally, 1 study evaluated the agreement between

surface-based and landmark-based registrations⁶ using serial 3D facial photographs acquired within 1 minute and 3-week intervals. Furthermore, 2 software packages and the differences between original and cropped photographs were compared regarding surface-based registration outcomes. Zero distance between corresponding serial 3D photographs was defined as the gold standard value. The superimposition reference for surface-based methods was the whole 3D facial photograph (original and cropped). For landmark (constructed planes)-based registration, certain points were used to construct a horizontal, a vertical, and a median plane. The superimposition reference frame was formed by these planes. The study showed that distances between registered images were less in cropped than uncropped photographs. Surface-based registration also provided smaller distances as compared with landmark-based registration.

The differences between the software packages used for surface-based registration were negligible. Important limitations of the study included that only mean values and mean distances were assessed. Further drawbacks were the absence of color-coded distance maps to localize differences, of comparative statistics, and of considerable morphologic change between the serial images.

Thus, there is only 1 study comparing landmark-based with surface-based registration, and it favored surface-based techniques and cropped facial photographs. Furthermore, the study did not detect any difference between 2 software applying for surface-based registration. However, it showed important limitations.

DISCUSSION

Two-dimensional photographs still comprise the main imaging technique used to illustrate facial morphology. However, they have major limitations deriving from the information reduction of a 3D surface in 2D.^{1,2} The same is true for lateral cephalometric radiographs, and therefore, they are, when indicated, replaced by 3D radiographs.³ Nevertheless, any radiographic technique requires patient exposure to harmful ionizing radiation, which is usually higher for 3D acquisitions. Three-dimensional photographs comprise risk-free tools that offer convenient, fast, accurate, and detailed 3D facial surface imaging. The accuracy of 3D photographs is well documented.²¹⁻²³ In addition, the high quality and considerable amount of information offered by the 3D facial photographs is also quite clinically relevant because it represents the anatomic area that is assessed by patients and is perceived by people during human interactions. This is the primary area people consider in the perception of facial appearance and related treatment outcomes when looking at others or themselves in the mirror. It has been shown that the primary factor affecting patient satisfaction from a given intervention is associated with the perceived esthetic outcomes.^{24,25} Thus, the proper assessment of facial surface changes because of treatment, growth, or pathology is a fundamental part of evaluating treatment effects.

Furthermore, it can determine how objective changes relate to perceived changes, which is a key factor affecting patient satisfaction. However, to objectively assess changes over time, superimposition techniques should be applied to serial images, and improper application or interpretation of such techniques might confound outcomes.^{7,8,10-12} Therefore, we have gathered evidence from the available studies on facial soft-tissue superposition, aiming to provide useful guidelines for applying these techniques.

The systematic literature search identified a limited number of relevant studies, and most represented the low quality of evidence and had high applicability concerns. Furthermore, the heterogeneity among studies was high, and thus, solid conclusions cannot be drawn at the moment. There was only 1 low risk of bias study that compared different surface-based techniques with a standard technique.²⁰ This became feasible through the use of facial surface models extracted from CBCT images. The study concluded that a small rectangular area on the forehead and an area including the middle part of the nose and the lower wall of the orbital foramen provided comparable outcomes to an anterior cranial base superimposition when applied in growing patients. Thus, there is an urgent need for further work in the field so that both researchers and clinicians gain full advantage of this very powerful, risk-free imaging technique.

Though we did not limit our search regarding the time of publication, all eligible studies were published after 2010. This might be related to the fact that the 3D facial photograph acquisition methods comprise recently developed technologies. Furthermore, several studies reached the final stages of the search strategy but were excluded primarily because only 1 point was evaluated—usually to assess facial asymmetry—or there was no method error assessment. All these excluded studies, with the corresponding reasons for this, are listed in [Supplementary Table II](#).

Our review assessed only facial surface models derived from 3D facial photographs or models of equivalent accuracy. Indeed, 7 of the 8 included studies tested 3D facial photographs, whereas 1 study²⁰ used 3D facial models extracted from CBCT images. The latter have equivalent accuracy to that of the 3D photographs.^{21,23,26-28} Thus, studies that tested laser scans, Moiré topography, or other such methods were excluded. These methods are not easily applicable in a clinical environment, especially in children, because of the increased acquisition time, generating increased motion artifacts.²⁹⁻³² Another drawback concerns the difficulty of recording soft-tissue surface texture, which impedes the actual representation of a patient's image and might also affect landmark identification.^{30,31} Therefore, we focused our review on the most clinically relevant method. Our findings should apply to any facial surface model with comparable accuracy to that tested here.

Different studies in the literature introduced various techniques as appropriate to superimpose serial 3D facial images. In 8 included studies, 6 superimposition references were deemed appropriate for superimposition of 3D facial photographs. However, 1 study tested only identical photographs or differences of a smiling from

rest photographs,⁶ whereas 7 studies did not assess individual differences between techniques, but performed only group mean comparisons. Thus, the applicability of the tested techniques in every single patient remains unexplored. Moreover, there was significant heterogeneity among studies in samples, methods, and outcomes making direct comparisons impossible. Therefore, no studies could be combined in a meta-analysis. Furthermore, a recent study that assessed various superimposition reference areas identified significant differences in outcomes, though all areas were primarily located at the upper half of the face.²⁰ A similar finding was previously evident for the superimposition of serial maxillary models on palatal structures⁷ and for the superimposition of tooth crowns to assess tooth wear.^{11,12}

The most widely used techniques to superimpose serial facial images apply for surface-based registrations. According to Incrapera et al,¹⁴ Altındış et al,¹⁸ and Fay,¹⁹ the anterior forehead is a promising superimposition reference area because of the morphologic stability that this area shows over time. However, the low amount and quality of evidence supporting this approach emphasize the need for further studies. A recent low risk of bias study also supported the use of the forehead area but combined with an area consisting of the middle part of the nose and the lower wall of the orbital foramen.²⁰ The primary advantage of this study is that it used a standard reference method to compare the newly introduced facial surface-based methods. The study did not favor the nasal bridge area in growing patients, stating that it often presents a different growth pattern than the forehead,^{33,34} and it might also show increased artifacts in certain patients because of facial hair. The above could confound the superimposition outcomes.⁸ Adriaens et al¹⁶ and Dindaroğlu et al¹⁷ suggested that the whole 3D facial (cropped) photographs can be used as a reference for superimposition. However, both studies had a high risk of bias and applicability concerns.

Furthermore, in most patients, the clinical applicability of the latter approach is questionable. When the entire facial image is used as a superimposition reference, the changes in a specific area cannot be assessed accurately because the superimposition results are skewed by changes in other facial areas. This is inevitable because the application of the best-fit algorithm aimed to minimize the overall distance of the superimposed surfaces. For example, when a face has been primarily changed at the lower third, which is usually the case in preadolescent growing patients, it would make sense to use relatively stable areas at the middle or the upper facial third as references. Then, by superimposing 2 serial images on these relatively stable areas, the actual changes in the lower third can be revealed. If a whole

face approach is adopted, the assessed changes in the lower part of the face will be mediated (appear reduced) by the other relatively stable parts of the face, which will also appear falsely as being changed.

In general, larger reference areas are preferable to smaller areas because they are less prone to operator- or artifact-related errors.^{8,11,12} However, when using larger areas, the chances to include regions that have been changed over time are increasing. These changes might skew outcomes on other regions,¹⁰ as explained above. Thus, relatively large areas are advantageous when they do not compromise the overall stability of the reference structures between serial images. Stable superimposition reference structures are usually a prerequisite in clinically relevant outcomes to assess changes in other areas because of treatment, growth, or pathology.^{5,9,10}

So far, there was only 1 study²⁰ that performed a comparison of different superimposition reference areas with a well-documented superimposition technique on the anterior cranial base.^{3,35,36} The study showed a low risk of bias, but it had unclear applicability concerns because only young adolescents were assessed.²⁰ The findings of this study look promising, but further studies on different samples would be beneficial to confirm and generalize the results.

As an alternative to the most often applied surface-based registration, a landmark-based registration can be performed. The latter is simple to use and may show adequate reproducibility, especially if a considerable number of landmarks is used.^{15,37} However, when few landmarks are used as superimposition reference, small errors in landmark identification considerably affect the outcomes.³ In contrast, with the increase in the number of landmarks, the technique becomes more complex and time-consuming. So far, there is only 1 study that tested 9 landmarks around the eyes, nose, and mouth as reference for facial 3D photographs superimposition¹⁵ and this has important limitations in outcome assessment. Furthermore, the study did not test faces subjected to changes over time, but only faces changed because of facial expressions.

Currently, there is only 1 study in the literature that tested a landmark-based technique compared with a surface-based technique.⁶ The surface-based registration seemed to offer better accuracy than a reference frame created from identified landmarks. These findings are in agreement with previous evidence³ supporting surface-based techniques, which are automated, and thus, less operator-dependent. In addition, the postprocessing capabilities of surface models are also higher because the congruence of a whole area, consisting of thousands of points, is used and assessed instead of a few arbitrarily selected landmarks.

Limitations

According to the inclusion criteria, only studies on 3D photographs or surface models of equivalent accuracy were considered. However, some studies tested other types of facial imaging, such as laser scans or Moiré fringe techniques. We decided to exclude these studies because these techniques are not realistic in daily clinical practice and for a large load of patients because of the time-consuming acquisition and the reduced image quality. Furthermore, the relatively small amount of included studies, the high heterogeneity, and the high risk of bias led to limited synthesis possibilities and no meta-analysis of the results. Thus, no solid conclusions with broad applicability could be drawn at the moment.

CONCLUSIONS

The present study on the superimposition of serial 3D facial photographs revealed the limited amount and the low quality of evidence supporting these newly introduced but quite powerful techniques. Few relevant studies were identified, and most had a high risk of bias and applicability concerns. The heterogeneity in methods and outcomes was also high.

The limited available evidence suggests that surface-based registration performs superiorly to landmark-based registration. A small rectangular area on the forehead and an area including the middle part of the nose and the lower wall of the orbital foramen showed promising results, comparable to the anterior cranial base superimposition. However, there is an urgent need for further studies in the field to confirm and generalize these findings.

AUTHOR CREDIT STATEMENT

Jonathan Johannes Wampfler contributed to data curation, formal analysis, original draft preparation, visualization, investigation, and manuscript review and editing; Nikolaos Gkantidis contributed to conceptualization, methodology, software, data curation, formal analysis, original draft preparation, visualization, investigation, supervision, validation, and manuscript review and editing.

SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ajodo.2021.06.017>.

REFERENCES

1. Farkas LG, Bryson W, Klotz J. Is photogrammetry of the face reliable? *Plast Reconstr Surg* 1980;66:346-55.
2. Biwasaka H, Tokuta T, Sasaki Y, Sato K, Takagi T, Tanijiri T, et al. Application of computerised correction method for optical distortion of two-dimensional facial image in superimposition between three-dimensional and two-dimensional facial images. *Forensic Sci Int* 2010;197:97-104.
3. Gkantidis N, Schauseil M, Pazera P, Zorkun B, Katsaros C, Ludwig B. Evaluation of 3-dimensional superimposition techniques on various skeletal structures of the head using surface models. *PLoS One* 2015;10:e0118810.
4. Brons S, van Beusichem ME, Maal TJ, Plooi JM, Bronkhorst EM, Bergé SJ, et al. Development and reproducibility of a 3D stereophotogrammetric reference frame for facial soft tissue growth of babies and young children with and without orofacial clefts. *Int J Oral Maxillofac Surg* 2013;42:2-8.
5. Mai DD, Stucki S, Gkantidis N. Assessment of methods used for 3-dimensional superimposition of craniofacial skeletal structures: a systematic review. *PeerJ* 2020;8:e9263.
6. Maal TJ, van Loon B, Plooi JM, Rangel F, Ettema AM, Borstlap WA, et al. Registration of 3-dimensional facial photographs for clinical use. *J Oral Maxillofac Surg* 2010;68:2391-401.
7. Vasilakos G, Schilling R, Halazonetis D, Gkantidis N. Assessment of different techniques for 3D superimposition of serial digital maxillary dental casts on palatal structures. *Sci Rep* 2017;7:5838.
8. Henninger E, Vasilakos G, Halazonetis D, Gkantidis N. The effect of regular dental cast artifacts on the 3D superimposition of serial digital maxillary dental models. *Sci Rep* 2019;9:10501.
9. Stucki S, Gkantidis N. Assessment of techniques used for superimposition of maxillary and mandibular 3D surface models to evaluate tooth movement: a systematic review. *Eur J Orthod* 2020;42:559-70.
10. Pazera C, Gkantidis N. Palatal rugae positional changes during orthodontic treatment of growing patients. *Orthod Craniofac Res* 2021;24:351-9.
11. Gkantidis N, Dritsas K, Ren Y, Halazonetis D, Katsaros C. An accurate and efficient method for occlusal tooth wear assessment using 3D digital dental models. *Sci Rep* 2020;10:10103.
12. Gkantidis N, Dritsas K, Katsaros C, Halazonetis D, Ren Y. 3D method for occlusal tooth wear assessment in presence of substantial changes on other tooth surfaces. *J Clin Med* 2020;9:3937.
13. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.
14. Incrapera AK, Kau CH, English JD, McGrory K, Sarver DM. Soft tissue images from cephalograms compared with those from a 3D surface acquisition system. *Angle Orthod* 2010;80:58-64.
15. Gibelli D, De Angelis D, Poppa P, Sforza C, Cattaneo C. An assessment of how facial mimicry can change facial morphology: implications for identification. *J Forensic Sci* 2017;62:405-10.
16. Adriaens J, van Loon B, Ingels KJ, Bergé SJ, Maal TJ. Three-dimensional stereophotogrammetric method of quantifying the effect of botulinum toxin type A injections on masseter hypertrophy. *Plast Reconstr Surg* 2012;130:376e-8e.
17. Dindaroğlu F, Duran GS, Görgülü S. Effects of rapid maxillary expansion on facial soft tissues: deviation analysis on three-dimensional images. *J Orofac Orthop* 2016;77:242-50.
18. Altındış S, Toy E, Başgıftçi FA. Effects of different rapid maxillary expansion appliances on facial soft tissues using three-dimensional imaging. *Angle Orthod* 2016;86:590-8.
19. Fay MB. The prospective evaluation of perioperative steroid dosing on postsurgical edema in orthognathic surgery [master's thesis]. Halifax: Dalhousie University; 2019.

20. Häner ST, Kanavakis G, Matthey F, Gkantidis N. Valid 3D surface superimposition references to assess facial changes during growth. *Sci Rep* 2021;11:16456.
21. Lübbbers HT, Medinger L, Kruse A, Grätz KW, Matthews F. Precision and accuracy of the 3dmd photogrammetric system in craniomaxillofacial application. *J Craniofac Surg* 2010;21:763-7.
22. Eder M, Brockmann G, Zimmermann A, Papadopoulos MA, Schwenzler-Zimmerer K, Zeilhofer HF, et al. Evaluation of precision and accuracy assessment of different 3-D surface imaging systems for biomedical purposes. *J Digit Imaging* 2013;26:163-72.
23. Zhao YJ, Xiong YX, Wang Y. Three-dimensional accuracy of facial scan for facial deformities in clinics: a new evaluation method for facial scanner accuracy. *PLoS One* 2017;12:e0169402.
24. Pachêco-Pereira C, Abreu LG, Dick BD, De Luca Canto G, Paiva SM, Flores-Mir C. Patient satisfaction after orthodontic treatment combined with orthognathic surgery: a systematic review. *Angle Orthod* 2016;86:495-508.
25. Pachêco-Pereira C, Pereira JR, Dick BD, Perez A, Flores-Mir C. Factors associated with patient and parent satisfaction after orthodontic treatment: a systematic review. *Am J Orthod Dentofacial Orthop* 2015;148:652-9.
26. Lin CY, Hsung TC, Khambay B. Reducing cone beam CT scan height as a method of radiation reduction for photorealistic three-dimensional orthognathic planning. *J Craniomaxillofac Surg* 2015;43:907-12.
27. Jayaratne YS, McGrath CP, Zwahlen RA. How accurate are the fusion of cone-beam CT and 3-D stereophotographic images? *PLoS One* 2012;7:e49585.
28. Naudi KB, Benramadan R, Brocklebank L, Ju X, Khambay B, Ayoub A. The virtual human face: superimposing the simultaneously captured 3D photorealistic skin surface of the face on the untextured skin image of the CBCT scan. *Int J Oral Maxillofac Surg* 2013;42:393-400.
29. Krimmel M, Kluba S, Bacher M, Dietz K, Reinert S. Digital surface photogrammetry for anthropometric analysis of the cleft infant face. *Cleft Palate Craniofac J* 2006;43:350-5.
30. Germec-Cakan D, Canter HI, Nur B, Arun T. Comparison of facial soft tissue measurements on three-dimensional images and models obtained with different methods. *J Craniofac Surg* 2010;21:1393-9.
31. Gibelli D, Pucciarelli V, Poppa P, Cummaudo M, Dolci C, Cattaneo C, et al. Three-dimensional facial anatomy evaluation: reliability of laser scanner consecutive scans procedure in comparison with stereophotogrammetry. *J Craniomaxillofac Surg* 2018;46:1807-13.
32. Gibelli D, Dolci C, Cappella A, Sforza C. Reliability of optical devices for three-dimensional facial anatomy description: a systematic review and meta-analysis. *Int J Oral Maxillofac Surg* 2020;49:1092-106.
33. van der Heijden P, Korsten-Meijer AG, van der Laan BF, Wit HP, Goorhuis-Brouwer SM. Nasal growth and maturation age in adolescents: a systematic review. *Arch Otolaryngol Head Neck Surg* 2008;134:1288-93.
34. Enlow DH. A morphogenetic analysis of facial growth. *Am J Orthod* 1966;52:283-99.
35. Häner ST, Kanavakis G, Matthey F, Gkantidis N. Voxel-based superimposition of serial craniofacial CBCTs: reliability, reproducibility and segmentation effect on hard-tissue outcomes. *Orthod Craniofac Res* 2020;23:92-101.
36. Friedli L, Kloukos D, Kanavakis G, Halazonetis D, Gkantidis N. The effect of threshold level on bone segmentation of cranial base structures from CT and CBCT images. *Sci Rep* 2020;10:7361.
37. Chortrakarnkij P, Lonic D, Lin HH, Lo LJ. Establishment of a reliable horizontal reference plane for 3-dimensional facial soft tissue evaluation before and after orthognathic surgery. *Ann Plast Surg* 2017;78(3 Suppl 2):S139-47.

Supplementary Table I. Detailed description of the search performed in various databases

<i>Database</i>	<i>Search strategy</i>	<i>Limits, results, and inclusion</i>
PubMed (https://www.ncbi.nlm.nih.gov/pubmed)	((3D image* OR 3-D image* OR three dimens* image* OR 3D photo* OR 3-D photo* OR three dimens* photo* OR surface*)) AND (regist* OR superimpos*) AND (soft tissue* OR soft-tissue* OR facial OR face* OR craniofac*)	Search date: 20.12.2020 Limits: humans Publication date: from 0001/01/01 to current Search Builder: 'All Fields' Results: 1.251
EMBASE (www.embase.com)	((3D image* OR 3-D image* OR three dimens* image* OR 3D photo* OR 3-D photo* OR three dimens* photo* OR surface*)) AND (regist* OR superimpos*) AND (soft tissue* OR soft-tissue* OR facial OR face* OR craniofac*)	Search date: 20.12.2020 Limits: humans Publication date: From 0001/01/01 to current Search Builder: 'All Fields' Special function used: non Results: 1.145
Google Scholar (www.scholar.google.com)	Advanced search 1 With the exact phrase: 3D superimposition With >1 of the words: facial face craniofacial Advanced search 2 With the exact phrase: 3D registration With >1 of the words: facial face craniofacial	Search date: 20.12.2020 Limits: – Publication date: From to 0001/01/01 to current Search Builder: 'All Fields' Results: 27.900/451.000 (the first 15 pages including the first 150 most relevant results were searched each time)
Cochrane Library (all databases) (www.thecochranelibrary.com)	(3D image* OR three dimension* image* OR 3D photo* OR three dimension* photo*) AND (regist* OR superimpos* OR matching) AND (facial OR face* OR craniofac*)	Search date: 20.12.2020 Limits: – Publication date: from 0001/01/01 to current Search Builder: 'All Fields' Results: 312 (99 reviews, 13 protocols, 200 trials)
Open Grey (http://www.opengrey.eu/)	3D superimposition, 3-D superimposition, 3 Dimensional superimposition, 3-Dimensional superimposition, three-dimensional superimposition, 3D registration, 3-D registration, 3 Dimensional registration, 3-Dimensional registration, three-dimensional registration	Search date: 20.12.2020 Results: 194
Grey Literature Report (www.greylit.org)	3D superimposition, 3-D superimposition, 3 Dimensional superimposition, 3-Dimensional superimposition, three-dimensional superimposition, 3D registration, 3-D registration, 3 Dimensional registration, 3-Dimensional registration, three-dimensional registration	Search date: 20.12.2020 Results: 0

Supplementary Table II. Closely related studies that were not eligible and reasons for this exclusion

<i>Study</i>	<i>Reason for exclusion</i>
Al-Rudainy et al. J Craniomaxillofac Surg. 2018;46:974-78	Facial asymmetry was measured
Brons et al. Int J Oral Maxillofac Surg. 2013;42:2-8.	It superimposed only identical 3D facial photographs
Hoefert et al. J Orofac Orthop. 2010;7:221-34.	It tested only landmark identification error
Chorzakamkji et al. Ann Plast Surg. 2017;78:S139-47.	It had no method error
Verhoeven et al. Int J Oral Maxillofac Surg. 2013;42:19-25.	It tested facial asymmetry
Verhoeven et al. J Craniomaxillofac Surg. 2016;44:1970-76.	It tested facial asymmetry
Rawlani et al. Plast Reconstr Surg. 2019;143:76-85.	No measurements of accuracy or reliability
Maués et al. Int J Oral Maxillofac Surg. 2018;47:1219-25	One time point was tested
Gibelli et al. J Craniomaxillofac Surg. 2018;46:1807-13	One time point was tested
Gibelli et al. J Craniomaxillofac Surg. 2018;46:1493-99	One time point was tested
Gibelli et al. J Oral Maxillofac Surg. 2018;76:1772-84	One time point was tested
Hyer et al. Orbit. 2020; 14:1-9	No changes expected between the superimposed models
De Menezes et al. J Oral Maxillofac Surg. 2010;68:2129-35	It tested only the reproducibility of landmark identification
Verdenik M et al. Br J Oral Maxillofac Surg. 2017; 55:517-23.	No measurement of accuracy or reliability
Demir et al. Am J Orthod Dentofacial Orthop 2020;157:773-82	No measurement of accuracy or reliability
Desmedt et al. Clin Oral Investig. 2015;19:1833-42	It tested facial asymmetry
Kornreich et al. Cleft Palate Craniofac J. 2016;53:126-31	It tested facial asymmetry
Maal et al. Int J Oral Maxillofac Surg. 2012;41:1137-45	It repeated only the measurements and not the superimposition
Krneta et al. Eur J Orthod. 2014;36:133-9	No measurement of accuracy or reliability
Nkenke et al. Oral Maxillofac Surg. 2008;12:5-12	No measurement of accuracy or reliability
Moss et al. Orthod Craniofac Res. 2003;6 Suppl 1:126-31	No measurement of accuracy or reliability
Rasse et al. Int J Oral Maxillofac Surg. 1991;20:163-6	Case report
van Loon et al. Int J Oral Maxillofac Surg. 2010;39:534-40	No image registration
Schwenzer-Zimmerer et al. J Plast Reconstr Aesthet Surg. 2009;62:1181-6	No image registration
Young-Kyun et al. Angle Orthod. 2015;85:833-40	Only reproducibility of landmark identification was tested
Gibelli et al. J Forensic Sci. 2017;62457-461	Inadequate relevant information
Edison et al. Am J Orthod Dentofacial Orthop. 2012;142:410-8	It tested only the reproducibility of landmark identification
Ritschl et al. J Craniomaxillofac Surg. 2018;46:2008-16	It evaluated only a smart part of the face
See et al. J Oral Maxillofac Surg. 2008;66:1410-6	No measurement of accuracy or reliability
Lin et al. Ann plast Surg. 2017;78:S61-9	No measurement of accuracy or reliability
Koparal et al. Exp Ther Med. 2018;15:3820-6.	No measurement of accuracy or reliability
Asutay et al. Niger J Clin Pract. 2018;2:1107-13.	No measurement of accuracy or reliability
Baysal et al. Angle Orthod. 2016;86:934-942.	No image registration
van Loon et al. Rhinology. 2011;49:121-6.	No measurement of accuracy or reliability
Sun et al. J Prost Dent. 2020	No image registration
Van der Meer et al. Br J Oral Maxillofac Surg. 2014;52:922-7.	Very short term and no actual change in facial morphology apart from artificial local change
Elnagar et al. Am J Orthod Dentofacial Orthop. 2017;152:336-47.	Facial impressions were used to create the facial 3D models
Kau et al. Orthod Craniofac Res. 2006;9:31-7.	Surface scans were used instead of 3D photographs; no comparison or method error assessment
Sforza C et al. Br J Oral Maxillofac Surg. 2017; 56:3-7.	Tested very short-term changes because of facial expressions in patients with facial palsy
Linn HH et al. Ann Plast Surg. 2017;78:S61-69.	Inadequate reporting on the assessment of the used superimposition reference area
Asutay F et al. Niger J Clin Pract. 2018; 21:1107-13	No assessment of the used superimposition method
Krneta Đokić B et al. Orthod Craniofac Res. 2020, 23:323-31	No assessment of the used superimposition method
Wong KWF et al. Surgeon. 2019; 17:19-27	n = 1; only the average faces of 2 groups were superimposed; no assessment of a superimposition method
Seo SW et al. J Craniofac Surg. 2017 28:1789-96.	No assessment of the used superimposition method
Pucciarelli V et al. Aesthetic Plast Surg. 2018; 42:456-63.	No assessment of the used superimposition method
Tanikawa C et al. Angle Orthod. 2018; 88:319-28.	Very short-term evaluation (1 wk) with no change on facial expression expected
Brons S et al. PLoS One. 2019; 14:e0217267.	n = 4; average faces were superimposed; no assessment of the used superimposition method
Meyer. Marcotty P et al. J Orofac Orthop. 2012; 73:116-25.	No assessment of the used superimposition method
Resnick CM et al. Plast Reconstr Surg. 2018; 141:540e-6e.	No assessment of the used superimposition method