Reliability of cephalometric superimposition for the assessment of craniofacial changes: a systematic review

Carmen Camila Graf[a], Konstantinos Dritsas[a], Mohammed Ghamri[a,b], Nikolaos Gkantidis[a,*]

[a]Department of Orthodontics and Dentofacial Orthopedics, School of Dental Medicine, University of Bern, CH-3010 Bern, Switzerland

[b]Directorate of Health Affairs-Jeddah, Ministry of Health, 11176 Riyadh , Saudi Arabia

Keywords: cephalometry; humans; dental radiography; craniofacial morphology; image registration; superimposition; reproducibility; accuracy; systematic review

***Corresponding author:**

Department of Orthodontics and Dentofacial Orthopedics, University of Bern, CH-3010, Freiburgstrasse 7, Bern, Switzerland

Tel.: +41 (0) 31 632 25 91, Fax.: +41 (0) 31 632 98 69

e-mail: nikosgant@yahoo.gr; nikolaos.gkantidis@zmk.unibe.ch

SUMMARY

Background: Superimposition of serial cephalometric radiographs enables the assessment of craniofacial changes over time, and therefore, several methods have been suggested in the literature.

Objective: The aim of the present study is to summarize and critically evaluate the available evidence on the reliability of methods used to superimpose serial cephalometric radiographs.

Search methods: Electronic searches were performed in MEDLINE, EMBASE, Google Scholar, and Cochrane Databases, without time limit (last update: November 1st, 2020). Unpublished literature was searched on the Open Grey and Grey Literature Report databases.

Selection criteria: Studies that tested the accuracy, precision, or agreement between different cephalometric superimposition techniques, used to evaluate the craniofacial changes due treatment or growth.

Data collection and analysis: Reference lists of relevant papers were screened and authors were contacted, if needed. All study selection steps, data extraction, and risk of bias (QUADAS-2 tool) assessments were performed independently by two authors on predefined forms.

Results: There were twenty-seven eligible studies. From these, seventeen tested superimpositions methods on the anterior cranial base, ten on the maxilla and twelve on the mandible. There were three studies that compared superimpositions on the cranial base with those on the maxilla and one that compared the cranial base with the mandibular superimposition. There was high heterogeneity among studies in terms of sample size, growth, radiographic machines, selection criteria, superimposition methods, references and outcomes measured. Furthermore, almost all studies presented important methodological limitations, with only two studies having unclear risk of bias and the rest twenty-five presenting high risk.

Conclusions: Currently there is no cephalometric superimposition method that has been proved to deliver accurate results. There is an urgent need for further research in this topic, since this is a primary assessment method to assess craniofacial changes over time for several relevant disciplines.

Registration: PROSPERO (CRD42020200349).

Conflict of interest: None.

INTRODUCTION

There was always increased interest from various fields that work in the craniofacial area to quantify changes in craniofacial morphology over time (1). Superimpositions of serial lateral cephalometric radiographs have been used since many years as an essential diagnostic tool (2–4).

Superimposition of lateral cephalometric radiographs in stable structures, facilitates the objective evaluation of changes, which occur during growth in individuals and/or treatment in patients, in other structures of interest. Therefore, structures that cease their growth at an early stage or that are expected to be only slightly altered during the observation period should preferably be used (5, 6). However, the growth and development of the craniofacial configuration is complex, as each structure grows differently in direction and intensity, and thus, this is not always easy (7). Another important factor affecting the outcomes is related to the 2D nature and the inaccuracies of the cephalometric image itself. Therefore, the selected structures are often inaccurately depicted, and also, they cannot be easily identified by the operator (8).

Over the years, various superimposition methods have been suggested in the literature. Most of them are using different structures that are considered stable, as references to register the serial images. Björk and Skieller placed implants as markers in a group of growing patients to identify morphologically stable structures in the craniofacial complex (9–12). Nowadays, Bjork's method is widely accepted as a standard to objectively assess craniofacial changes. There are also several other methods, such as those of Broadbent (1931), Ricketts (1975), Pancherz (1982), and You and Hägg (1999*)* that have been introduced as appropriate for quantifying craniofacial changes over time (2, 13–15). Each method uses a different superimposition reference, but most of them measure changes of anatomical landmarks to assess morphological alterations over time.

In the past, several authors have compared Björk's structural method with other cephalometric superimposition methods (16–18). Other studies tested the agreement (19–21) or the precision of different methods (15, 22, 23). There were also investigators that tested the reproducibility of certain cephalometric superimposition methods (24–26). Various researchers have reported favorable outcomes for Björk's structural method (16, 21, 27) whereas, other researchers have questioned its performance (22, 26, 28). Cook et al. (1994) proposed Ricketts' method, because it is easier (19). Björk's method also showed quite reduced reproducibility in comparison to superimposition using either the mandibular plane or the mandibular outline (26). Furthermore, Houston and Lee. found that Bjork's method showed relatively large errors (22). Thus, there is high controversy among researchers for the important issue of serial cephalometric radiograph superimposition. There are many different methods suggested in the literature as appropriate, that they, however, result in varying outcomes.

The purpose of this systematic review is to critically evaluate and summarize the available evidence on the topic. The study will assess the different cephalometric superimposition techniques on the cranial base, maxilla, and mandible regarding their reproducibility, validity, or agreement between each other, aiming to provide useful guidelines based on the evidence that supports them.

MATERIALS AND METHODS

Protocol and registration

The study protocol was registered in PROSPERO (CRD42020200349). The present methodology is based on previously published protocols by Stucki et al. (2020) and Mai et al. (2020) on topics analogous to, but different from the present topic (1, 29).

Eligibility criteria

- Study design: Any study design was considered eligible, including prospective, and retrospective studies of any type.
- Study sample: Studies with sample size ≥ 8.
- Index test: Landmark-based, straight line-based or outline-based superimposition techniques for serial 2D cephalometric radiographs, applied to assess craniofacial changes over time.
- Types of participants: Serial cephalometric radiographs of individuals who received actual or simulated treatment. Data from individuals where changes in craniofacial morphology are expected due to growth or pathology were also considered.
- Type of intervention: superimposition of serial images to evaluate changes in craniofacial morphology.
- Primary outcome: Accuracy or precision of cephalometric superimposition techniques, or the agreement between different techniques measured in terms of angles or distances

between specific landmarks. Studies where the above parameters were tested as a secondary outcome were also eligible.

- Comparator/control group: Different superimposition techniques, direct measurements, or repeated measurements.
- Unit of analysis: The unit of analysis was the measured distance or angle.
- Follow-up: All observation periods between subsequent radiographs were accepted.
- Exclusion criteria: Non-human-derived data.

Information sources, search strategy, and study selection

Search strategy

Database specific electronic searches were performed on the following databases to detect eligible studies: MEDLINE via Ovid and Pubmed, EMBASE via Ovid, Cochrane Register of Diagnostic Test Accuracy Studies, Google Scholar (last update: November 1st, 2020). No time restriction was applied. The reference lists of relevant reviews and of all studies that were assessed through full-text reading were hand-searched for potentially eligible studies. The Open Grey and Grey Literature Report databases were also searched for unpublished literature. All the detailed search strategies and the individual results are provided as supplementary information (Table S1).

Study selection

Database search and study selection were performed by two authors of this review (C. G. and M.G.) that were not blinded regarding the authors' names, their institutions or the conclusion. The articles were initially selected according to their title and afterwards through summary reading. If necessary, entire articles were also read independently by the two reviewers for eligibility assessment.

In the event that there was a disagreement between the first two authors regarding the eligibility of an article, it was discussed with the last author until an agreement was reached. A record was kept of all decisions about study identification.

Data items and collection

The first two authors performed data extraction from all studies independently. Disagreements were resolved through discussion with the last author till a consensus was reached. Information obtained from all eligible studies, if available, is described below:

- Methods: Author, title, year, objectives, and design of study.
- Participants: Age, gender and number of patients recruited.

• Materials: Superimposition method of serial cephalometric radiographs and time between cephalometric radiographs if available.

• Superimposition method: Type of superimposition reference lines, landmarks or outlines and tracing and superimposition tools used.

• Comparison/control group: Type and characteristics.

• Outcome: Type of outcome(s) and method of outcome assessment.

• If missing data were detected, the authors were contacted by email to request the missing information. If the authors did not respond or the data was not receivable, then only the information available in the article was considered.

Risk of bias in individual studies

We used QUADAS-2 checklist tool to evaluate the quality of the included studies (30). This tool evaluates the potential risk of bias and the applicability of primary diagnostic accuracy studies regarding four key domains: patient selection, index test, reference standard, and flow and timing. The QUADAS-2 results are usually presented in a table, in which a happy face represents low risk, a sad face high risk and a question mark unclear risk.

The first and the last author applied independently and in duplicate the QUADAS-2 tool on the included studies. The signalling questions used for the implementation of the QUADAS-2 tool are provided in Table S2. In case of disagreement, this was discussed among all authors till a consensus was reached. Studies presenting a high risk of bias were not considered eligible for meta-analysis.

Summary measures and approach to synthesis

Assessment of heterogeneity

The heterogeneity of the studies, as specified in the inclusion criteria, was assessed by evaluating the study characteristics, the similarity between the types of participants, the results as well as the methods that were compared.

Data synthesis

A meta-analysis was planned in the case that there were at least two studies with similar methods, that tested comparable outcomes and that were classified as having unclear or low risk of bias.

Assessment of reporting bias

To reduce potential reporting biases, including publication bias and multiple publication bias, we carried out a thorough search of multiple sources, including ongoing studies.

Additional analysis

Where possible, results were evaluated in the following subgroups:

- Patients with versus without growth.
- Patients with versus without orthodontic treatment.
- Superimposition on the anterior cranial base vs. superimposition on maxillary structures vs. superimposition on mandibular structures.

RESULTS

Study selection and characteristics of the included studies

The Flow Chart describing the study selection process is shown in Figure 1. After removing duplicates, 5889 studies were found through several databases. 782 studies remained after title and 86 after abstract reading. The full text of these studies was assessed for eligibility. Following full-text reading, 27 relevant articles were included in this review (3, 5, 6, 15–17, 31–40). Studies that seemed to be relevant, but were not eligible, and the reasons for this are provided in Table S3. All included studies refer to the precision/reproducibility of cephalometric superimposition techniques (15, 22, 23, 25, 26, 31, 32), or to the agreement between different techniques (4, 15–20, 24, 27, 33) measured in terms of angles or distances between specific skeletal or dental landmarks.

All included studies were retrospective in terms of radiograph acquisition and prospective regarding the superimposition data generation and the associated comparisons.

In twenty-four studies the sample included only growing patients (4–6, 15–21, 24–27, 31–40), whereas two studies did not report on this (22, 23) and one study included growing and non-growing patients (3). None of the selected studies included patients with severe craniofacial malformations due to syndromes or other anomalies or diseases.

Regarding the location of superimposition reference areas, seventeen studies superimposed the cephalograms on the cranial base (3, 5, 17, 18, 20, 22–25, 27, 31–36), ten studies on maxillary structures (3, 6, 15, 16, 19, 20, 31, 34, 37, 38) and twelve studies on the mandible (3, 4, 15, 19–21, 25, 26, 31, 34, 38, 39). Three studies compared the superimposition on the cranial base with that on the maxilla (15, 37, 40), and finally, one study compared the superimpositions on the cranial base with those on the mandible (15).

The general information and superimposition related characteristics of the included studies are shown in Tables 1 and S4.

Quality assessment

The results of the quality assessment through the QUADAS-2 tool are shown in Table 2. Downgrades were based on the individual study limitations reported on Table 3.

From the 27 studies that were included, 25 showed high (3, 5, 6, 15–22, 24-27, 32–40), and 2 showed unclear risk of bias (23, 31).

On the individual categories, 5 studies have high (15, 16, 24,33, 39), 18 low (3, 4, 6, 17–22, 25–27, 32, 34–37, 40), and 4 unclear risk of bias regarding patient selection (5, 23, 31, 38). Concerning index test, 25 studies have high (3–6, 15–27, 32, 33, 35–40) and 2 unclear risk of bias (31, 34). The reference standard of 2 studies shows a high risk of bias (32, 33), and the rest studies presented a lowrisk of bias (3–6 ,15–27, 31, 34–40). The flow and timing of all studies has low risk of bias (3–6 ,15–27, 31–40).

25 studies showed high total applicability concerns (3–6, 15–27, 32, 33, 35–40), and 2 unclear (31, 34), whereas no study showed low applicability concerns. Regarding the individual categories, 7 studies had high (15, 16, 21, 22, 24, 25, 39), 5 unclear (5, 23, 31, 33, 38) and 15 low (3, 4, 6, 17–20, 26, 27, 32, 34, 36, 37, 40) applicability concerns in the patient selection. Considering the index test, 25 studies have high (3–6, 15–27, 32, 33, 35–40), and 2 unclear applicability concerns (31, 34). Regarding the reference standard, 2 studies showed high (32, 33) and the rest studies (3–6, 15–27, 31, 34–40) showed low risk applicability concerns.

Results of individual studies and qualitative synthesis

The tested outcomes, the conclusions, and the limitations of the individual studies are presented in Table 3. The outcomes of the included studies are presented in more detail in Table S5 and described in Supplementary Text 1.

There was high heterogeneity among the studies in terms of sample size, growth, radiographic machines, selection criteria, superimposition methods, references and outcomes measured. Consequently, it was decided not to perform a quantitative synthesis.

To carry out the qualitative synthesis, the studies were divided into five categories based on the tested anatomical area where the superimposition reference was located; namely, the cranial base, the maxilla, the mandible, the cranial base compared to the maxilla and the cranial base compared to the mandible.

Superimposition on the cranial base

Seventeen studies superimposed the cephalograms on the cranial base (3, 5, 15, 17, 18, 20, 22–25, 27, 31–36). The conclusion of the vast majority of studies that tested cephalometric superimpositions on the cranial base was that most superimposition methods work properly

(3, 5, 15, 18, 20, 23–25, 32, 34, 36). On the other hand, certain studies questioned the use even of commonly applied methods, such as the best fit on ACB, Viazis' method, Johnston's method, or subtraction method (17, 22, 27, 31, 33, 35). However, most studies presented severe limitations, such as the absence of testing of individual differences between different methods or repeated measurements (3, 5, 15, 17, 18, 20, 22, 24, 25, 27, 32–36), or the use of only good quality radiographs (5, 31 ,34). For example, Bjork's structural superimposition on the ACB is a commonly tested method that was often used as a gold standard for comparisons with other methods (18), but no high-quality study was found to provide an adequate error evaluation of this standard method. Furthermore, certain studies suggested novel approaches as appropriate (33, 35), but the evidence to support them was very weak. Overall, from the 17 studies of this category, only two had unclear risk of bias (23, 31), whereas all other studies were high risk (3, 5, 15, 17, 18, 20, 22, 24, 25, 27, 32–36). Additionally, there were contradictory findings among studies. Thus, despite the considerable number of studies in this category, no solid conclusions can be drawn at present for the trueness or the precision of any tested method.

Superimposition on the maxilla

The ten studies that investigated cephalometric superimposition on the maxilla provide contrasting evidence. Five studies suggested certain superimposition methods as effective (3, 15, 16, 20, 34). However, other five studies questioned the use of maxillary superimposition methods, including even widely used methods, such as the best fit on maxillary structures, Johnston's approach, or Ricketts position 3 (6, 19, 31, 37, 38). It should be noted that the majority of studies had significant flaws, such as the lack of examination of individual differences (3, 6, 15, 16, 19, 20, 34, 37, 38) or the use of only high-quality radiographs (16, 34). For example, the implant superimposition method was frequently used as the gold standard approach (16, 37), but there was no high-quality study that performed an adequate error evaluation of this method. Moreover, there is a study suggesting a new computer based automated method as appropriate, but the data to support this was lacking (3). Only two of the 10 studies in this category had an unclear risk of bias (31. 34), while the rest were all high risk (3, 6, 15, 16, 19, 20, 37, 38). Thus, though there is a relatively large number of studies in this field, they show variable results and present important limitations. Therefore, the available evidence on the topic is quite limited and no firm conclusions for the accuracy of maxillary superimposition methods can be formed.

Superimposition on the mandible

Twelve studies assess mandibular superimpositions and the majority of them suggested various techniques as appropriate (3, 4, 15, 19, 20, 25, 34, 38, 39). There are studies, however, that questioned the use of widely used methods, such as the best fit on mandibular structures, Björk's method and Ricketts position 4 (21, 26 ,31). Almost all studies had serious limitations, such as no testing of individual differences between methods or using only high-quality radiographs (3, 4, 15, 19–21, 25, 26, 38, 39). For example, mandibular Bjork's structural superimposition has been widely studied and was frequently cited as a reliable method (4, 39), but no high-quality study performed an adequate error evaluation of this method. Furthermore, a study proposed a new approach as appropriate (3), but the evidence to support it was lacking. Only two studies out of the 12 in this category had an unclear risk of bias (31, 34), while the rest were all high risk (3, 4, 15, 19–21, 25, 26, 38, 39). As a result, despite the considerable number of studies in this field, the conclusions of the 12 studies listed above are inconsistent and are based on studies with significant limitations. Thus, the available evidence for this category is limited, and no strong conclusions can be drawn.

Superimposition on the cranial base compared to the maxilla

From the three studies (15, 37, 40) that compared the cephalometric superimposition on the cranial base to that on the maxilla, two studies found that these superimposition strategies provided different outcomes, especially in growing patients (37, 40). On the other hand, one study found that Pancherz's ACB superimposition method provided similar outcomes to Bjork's maxillary and Rickets Position 3 method and that all methods were reproducible (15). However, all studies had significant flaws, such as no testing of individual differences between compared methods and high risk of bias. The conclusions of the aforementioned three studies are inconsistent and the evidence provided is weak. Thus, no firm conclusions can be reached.

Superimposition on the cranial base compared to the mandible

There is only one study testing superimpositions on the cranial base compared to the mandible (15), but due to the important limitations, no conclusion can be drawn at present.

DISCUSSION

Craniofacial changes are always expected in an individual due to treatment, growth, ageing or pathology (1). Superimpositions of serial images are used for the assessment of skeletal, dental, and soft-tissue changes that occur over time (1, 29). This has been traditionally performed through superimposition of serial cephalometric radiographs on certain anatomical structures that are considered stable (10–12). The present systematic review summarized and assessed the available evidence on the topic, based on twenty-seven papers that investigated the different cephalometric superimposition methods on the cranial base, the maxilla, and the mandible regarding their reproducibility, trueness, or agreement between each other. Overall, there was significant heterogeneity among studies and the vast majority of them had high risk of bias and provided low quality evidence. Out of twenty-seven studies, there were only two that presented unclear risk of bias. As a result, it was not possible to draw any solid conclusions and combine the outcomes of the studies in a meta-analysis. The present review highlights the urgent need for more research in this topic.

Superimposition on anterior cranial base structures, provides important information regarding the morphological changes of other craniofacial structures over time. This superimposition is fundamental to assess changes in the head, since the cranial base reference structures have a central location at the center of the object of interest, they have been shown to remain stable already at early ages (> 7 years) (1, 41, 42) and they also retain a constant relation to the natural head position (43, 44). There were seventeen studies that tested the superimposition on the cranial base, but most of them presented significant methodological limitations and had high risk of bias. Therefore, there is an urgent need for further research in this field, since this is the main superimposition method used to assess changes in the craniofacial area over time.

Superimposition on the maxilla, also yields important information about changes of the dentoalveolar complex, according to its basal bone, during growth and development or due to orthodontic treatment. This superimposition is also crucial for the orthodontic specialty, since this is the area primarily affected by most orthodontic interventions. Ten studies evaluated the performance of corresponding superimposition methods, but apart from the high heterogeneity among them in methods and outcomes, only two had unclear risk of bias, whereas the rest studies had high risk. Thus, the scarcity of high-quality evidence to support these superimposition methods highlights the need for more research in this topic.

Similarly, superimposition on the mandible is equally important, but strong evidence is also lacking to support any suggested method. Twelve studies investigated this topic, with high heterogeneity and high risk of bias for all but two studies that were of unclear risk.

Furthermore, there were three studies that compared the superimpositions on the cranial base to that on the maxilla. Two of them identified differences in the outcomes, which is an expected outcome, since they tested growing patients (37, 40). However, one study identified similar outcomes between the two methods (15). These findings should be treated with caution, especially when considering growing patients, since it is not expected that a CB method could show comparable results to a maxillary method, considering the normal growth pattern of the human head. The structures between the two reference areas are expected to change over time and this cannot leave the outcome measurement area unaffected. There is only one study testing superimpositions on the CB compared to the mandible and concluded that the methods show good agreement to each other. However, for the reason explained above, these results cannot be considered reliable.

Various studies included in this systematic review suggested different superimposition methods as appropriate. Unfortunately, there is no single cephalometric superimposition method that has been undoubtedly proved to be accurate. The eligible studies were mostly not recent, although our search did not have time restrictions. This may be related to the fact that nowadays certain methods are being used for years, and therefore, are considered standard, and they might no longer be questioned. For example, Björk and Skieller, used implants to identify anatomically stable structures in the craniofacial complex of growing patients (12). These structures, known as "Björk's structures'', can be used in the absence of implants and are considered to provide comparable superimposition outcomes. Since then, these methods have been widely used and are known as ''structural methods''. Many researchers used Björk's structural method as a gold standard for their studies (17, 45). A modification of Bjorks' ''structural method'' is also used by the American Board of Orthodontics (ABO) for evaluating the dental and skeletal changes that occurred over time in the cases presented by the candidates (45). However, the error of this method has never been adequately tested and reported so far, especially considering individual case measurements.

Nowadays, implant studies or studies that expose patients to radiation for research purposes are not allowed due to ethical constraints. Thus, it is impossible to obtain such study material and test the accuracy of cephalometric methods in a similar manner. Therefore, the existing material is quite valuable, but the studies testing it had severe methodological drawbacks. Furthermore, a direct comparison of these historical data to that currently obtained from our patients is not optimal, due to the differences in the quality of the radiographs between that time and today. The images obtained with current machines and methods are superior to those obtained 40 or 50 years ago. Secular trends in craniofacial growth should also be considered when trying to extrapolate findings based on historical data to contemporary practice (46, 47). Alternatively, recently developed 3D surface- or voxel-based superimposition methods can be used to assess the traditional cephalometric superimposition methods (1, 29, 48, 49).

Future studies should include consecutive patient data obtained in the last few years or control otherwise selection bias, assess individual outcomes in addition to mean comparisons, and compare cephalometric superimposition outcomes to a reference standard, which could derive from 3D superimpositions (1, 29, 48, 49). The thorough assessment of outcomes in each individual case is fundamental when testing diagnostic or outcome assessment methods (50). Raw data of each single measurement, e.g. on differences between methods or between repeated measurements, should be provided through scatter plots, Bland Altman plots, or by other means (50). Only then the performance of a method in every single patient can be verified. The need for such studies is high, since cephalometric superimposition is currently the standard method for craniofacial change assessment in orthodontics. It should be note here though that the use of cephalometric superimpositions is likely to decrease in the future to avoid unnecessary exposure to radiation. The benefits of serial lateral cephalograms, including the post-treatment cephalometric evaluation, are usually difficult to justify in regular orthodontic treatments. On the other hand, in case of severe craniofacial anomalies or in need for detailed skeletal growth assessment, a CBCT image may be a better diagnostic option (1, 8). In the near future, for regular orthodontic patients, risk-free 3D images, such as 3D photos or intraoral scans, might be preferrable to any radiographic method for the assessment of changes in craniofacial morphology (29, 51, 52).

Limitations

Despite the large number of included studies, the large heterogeneity among studies and the important methodological drawbacks detected for almost all studies did not allow solid conclusions to be drawn. The most common drawback was the consideration of group mean values only, when assessing the methods in question. Individual measurements should have been considered, using methods such as Bland-Altman plots, since the outcomes of a superimposition technique should be valid on each single patient. When groups of measurements are averaged, opposite values of the same magnitude indicate zero difference between the compared methods, which is misleading when regarding single measurements. Summary measures, such as standard deviations, though more informative, also do not ensure the performance of a diagnostic or outcome assessment method in every single case. Based on this rationale, we rated here the corresponding studies as high risk of bias, which might be considered a strict judgement, but it is a valid one if we are interested on the reliability of a method in each individual measurement.

CONCLUSION

A considerable number of studies (twenty-seven) were assessed in this review. Most studies tested the ACB superimposition, but also several studies tested maxillary and mandibular superimpositions. However, great variability was evident among them regarding hypotheses, samples, designs, and outcomes. Furthermore, almost all studies presented important methodological limitations. Thus, valid comparisons between studies were difficult to perform and solid conclusions impossible to be drawn.

From this review, it was evident that currently there is no cephalometric superimposition method that has been proved to deliver accurate results. There is an urgent need for further research in this topic, since this is currently the primary method to assess craniofacial changes within individuals, for several disciplines working in this area.

ACKNOWLEDGMENTS

No funding was received for this work. Authors declare there is no conflict of interest.

Data availability statement

The data underlying this article are available in the article and in its online supplementary material.

REFERENCES

1. Mai, D. D., Stucki, S., and Gkantidis, N. (2020) Assessment of methods used for 3-dimensional superimposition of craniofacial skeletal structures: a systematic review. *PeerJ*, 8, e9263.
2. Broadbent, BH. (1931) A new X-ray technique and its application to orthodontia. *The Angle Orthodontist*, 1, 45-66.
3. Jiang, Y., Song, G., Yu, X., Dou, Y., Li, Q., Liu, S., Han, B., and Xu, T. (2020) The application and accuracy of feature matching on automated cephalometric superimposition. *BMC Medical Imaging*, 20, 31.
4. Jabbal, A., Cobourne, M., Donaldson, N., and Bister, D. (2016) Assessing lower incisor inclination change: a comparison of four cephalometric methods. *European Journal of Orthodontics*, 38, 184–189.
5. Lenza, M. A., Carvalho, A. A., Lenza, E. B., Lenza, M. G., Torres, H. M., and Souza, J. B. (2015) Radiographic evaluation of orthodontic treatment by means of four different cephalometric superimposition methods. *Dental Press Journal of Orthodontics*, 20, 29–36.
6. Baumrind, S., Korn, E. L., Ben-Bassat, Y., and West, E. E. (1987) Quantitation of maxillary remodeling. 2. Masking of remodeling effects when an "anatomical" method of superimposition is used in the absence of metallic implants. *American Journal of Orthodontics and Dentofacial Orthopedics*, 91, 463–474.
7. Gkantidis, N., and Halazonetis, D. J. (2011) Morphological integration between the cranial base and the face in children and adults. *Journal of Anatomy*, 218, 426–438.
8. Halazonetis D. J. (2005) From 2-dimensional cephalograms to 3-dimensional computed tomography scans. *American Journal of Orthodontics and Dentofacial Orthopedics*, 127, 627–637.
9. Björk, A., and Skieller, V. (1972) Facial development and tooth eruption. An implant study at the age of puberty. *American Journal of Orthodontics*, 62, 339–383.
10. Björk A. (1969) Prediction of mandibular growth rotation. *American Journal of Orthodontics*, 55, 585–599.
11. Björk, A., and Skieller, V. (1977) Growth of the maxilla in three dimensions as revealed radiographically by the implant method. *British Journal of Orthodontics*, 4, 53–64.

12. Björk, A., and Skieller, V. (1983) Normal and abnormal growth of the mandible. A synthesis of longitudinal cephalometric implant studies over a period of 25 years. *European Journal of Orthodontics*, 5, 1–46.
13. Ricketts R. M. (1975) A four-step method to distinguish orthodontic changes from natural growth. *Journal of Clinical Orthodontics*, 9, 208–228.
14. Pancherz H. (1982) The mechanism of Class II correction in Herbst appliance treatment. A cephalometric investigation. *American Journal of Orthodontics*, 82, 104–113.
15. You, Q. L., and Hägg, U. (1999) A comparison of three superimposition methods. *European Journal of Orthodontics*, 21, 717–725.
16. Nielsen I. L. (1989) Maxillary superimposition: a comparison of three methods for cephalometric evaluation of growth and treatment change. *American Journal of Orthodontics and Dentofacial Orthopedics*, 95, 422–431.
17. Arat, Z. M., Rübendüz, M., and Akgül, A. A. (2003) The displacement of craniofacial reference landmarks during puberty: a comparison of three superimposition methods. *The Angle Orthodontist*, 73, 374–380.
18. Arat, Z. M., Türkkahraman, H., English, J. D., Gallerano, R. L., and Boley, J. C. (2010) Longitudinal growth changes of the cranial base from puberty to adulthood. A comparison of different superimposition methods. *The Angle Orthodontist*, 80, 537–544.
19. Cook, A. H., Sellke, T. A., and and BeGole, E. A. (1994) The variability and reliability of two maxillary and mandibular superimposition techniques. Part II. *American Journal of Orthodontics and Dentofacial Orthopedics*, 106, 463–471.
20. Roden-Johnson, D., English, J., and Gallerano, R. (2008) Comparison of hand-traced and computerized cephalograms: landmark identification, measurement, and superimposition accuracy. *American Journal of Orthodontics and Dentofacial Orthopedics*, 133, 556–564.
21. Springate, S. D., and Jones, A. G. (1998) The validity of two methods of mandibular superimposition: a comparison with tantalum implants. *American Journal of Orthodontics and Dentofacial Orthopedics*, 113, 263–270.
22. Houston, W. J., and Lee, R. T. (1985) Accuracy of different methods of radiographic superimposition on cranial base structures. *European Journal of Orthodontics*, 7, 127–135.

23. Gliddon, M. J., Xia, J. J., Gateno, J., Wong, H. T., Lasky, R. E., Teichgraeber, J. F., Jia, X., Liebschner, M. A., and Lemoine, J. J. (2006) The accuracy of cephalometric tracing superimposition. *Journal of Oral and Maxillofacial Surgery*, 64, 194–202.
24. Goel, S., Bansal, M., and Kalra, A. (2004) A preliminary assessment of cephalometric orthodontic superimposition. *European Journal of Orthodontics*, 26, 217–222.
25. Buschang, P. H., LaPalme, L., Tanguay, R., and Demirjian, A. (1986) The technical reliability of superimposition on cranial base and mandibular structures. *European Journal of Orthodontics*, 8, 152–156.
26. Cook, P. A., and Southall, P. J. (1989) The reliability of mandibular radiographic superimposition. *British Journal of Orthodontics*, 16, 25–30.
27. Ghafari, J., Engel, F. E., and Laster, L. L. (1987) Cephalometric superimposition on the cranial base: a review and a comparison of four methods. *American Journal of Orthodontics and Dentofacial Orthopedics,* 91, 403–413.
28. Mathews, J. R., and Payne, G. S. (1980) Quantitative computerized analysis of lower incisor changes: a longitudinal implant study in man. *The Angle Orthodontist*, 50, 218–229.
29. Stucki, S., and Gkantidis, N. (2020) Assessment of techniques used for superimposition of maxillary and mandibular 3D surface models to evaluate tooth movement: a systematic review. *European Journal of Orthodontics*, 42, 559–570.
30. Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A., Bossuyt, P. M., and QUADAS-2 Group (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine,* 155, 529–536.
31. Baumrind, S., Miller, D., and Molthen, R. (1976) The reliability of head film measurements. 3. Tracing superimposition. *American Journal of Orthodontics*, 70, 617–644.
32. Iseri, H., and Solow, B. (1990) Growth displacement of the maxilla in girls studied by the implant method. *European Journal of Orthodontics*, 12, 389–398.
33. Standerwick, R., Roberts, E., Hartsfield, J., Jr, Babler, W., and Kanomi, R. (2008) Cephalometric superimposition on the occipital condyles as a longitudinal growth assessment reference: I-point and I-curve. *Anatomical Record*, 291, 1603–1610.

34. Huja, S. S., Grubaugh, E. L., Rummel, A. M., Fields, H. W., and Beck, F. M. (2009) Comparison of hand-traced and computer-based cephalometric superimpositions. *The Angle Orthodontist*, 79, 428–435.
35. Standerwick, R. G., Roberts, E. W., Hartsfield, J. K., Jr, Babler, W. J., and Katona, T. R. (2009) Comparison of the Bolton Standards to longitudinal cephalograms superimposed on the occipital condyle (I-point). *Journal of Orthodontics*, 36, 23–35.
36. Türköz, Ç., and İşcan, H. N. (2011) Evaluation of extraction and non-extraction treatment effects by two different superimposition methods. *European Journal of Orthodontics*, 33, 691–699.
37. Baumrind, S., Ben-Bassat, Y., Bravo, L. A., Curry, S., and Korn, E. L. (1996) Partitioning the components of maxillary tooth displacement by the comparison of data from three cephalometric superimpositions. *The Angle Orthodontist*, 66, 111–124.
38. Gu, Y., and McNamara, J. A., Jr (2008) Cephalometric superimpositions: A comparison of anatomical and metallic implant methods. *The Angle Orthodontist*, 78, 967–976.
39. Cook, P. A., and Gravely, J. F. (1988) Tracing error with Björk's mandibular structures. *The Angle Orthodontist*, 58, 169–178.
40. Efstratiadis, S. S., Cohen, G., and Ghafari, J. (1999) Evaluation of differential growth and orthodontic treatment outcome by regional cephalometric superpositions. *The Angle Orthodontist*, 69, 225–230.
41. Melsen, B., (1974) The cranial base: The postnatal development of the cranial base studied histologically on human autopsy material. *Acta Odontologica Scandinavica Supplement*, 32, 1-126.
42. Bastir, M., Rosas, A., and O'higgins, P. (2006) Craniofacial levels and the morphological maturation of the human skull. *Journal of Anatomy*, 209, 637–654.
43. Peng, L., and Cooke, M. S. (1999) Fifteen-year reproducibility of natural head posture: A longitudinal study. *American Journal of Orthodontics and Dentofacial Orthopedics*, 116, 82–85.
44. Solow, B., and Siersbaek-Nielsen, S. (1986) Growth changes in head posture related to craniofacial development. American Journal of Orthodontics, 89, 132–140.
45. American Board of Orthodontics. (n.d.). *Superimpositions*. Retrieved May 31, 2021, from https://www.americanboardortho.com/orthodontic-professionals/about-board-certification/clinical-examination/certification-renewal-examinations/mail-in-cre-

submission-procedure/case-report-examination/case-record-preparation/superimpositions/

46. Papageorgiou, S. N., Koretsi, V., and Jäger, A. (2017) Bias from historical control groups used in orthodontic research: a meta-epidemiological study. *European Journal of Orthodontics*, 39, 98–105.
47. Antoun, J. S., Cameron, C., Sew Hoy, W., Herbison, P., and Farella, M. (2015) Evidence of secular trends in a collection of historical craniofacial growth studies. *European Journal of Orthodontics*, 37, 60–66.
48. Kanavakis, G., Häner, S. T., Matthey, F., and Gkantidis, N. (2021) Voxel-based superimposition of serial craniofacial cone-beam computed tomographies for facial soft tissue assessment: Reproducibility and segmentation effects. *American Journal of Orthodontics and Dentofacial Orthopedics*, 159, 343–351.e1.
49. Häner, S. T., Kanavakis, G., Matthey, F., and Gkantidis, N. (2020) Voxel-based superimposition of serial craniofacial CBCTs: Reliability, reproducibility and segmentation effect on hard-tissue outcomes. *Orthodontics & Craniofacial Research*, 23, 92–101.
50. Bland, J. M., and Altman, D. G. (1995) Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*, 346, 1085–1087.
51. Wampfler, J. J., and Gkantidis, N. (2021) Superimposition of serial 3-dimensional facial photographs to assess changes over time: A systematic review. *American Journal of Orthodontics and Dentofacial Orthopedics*, S0889-5406, 00629-6.
52. Häner, S. T., Kanavakis, G., Matthey, F., and Gkantidis, N. (2021) Valid 3D surface superimposition references to assess facial changes during growth. *Scientific Reports*, 11, 16456.

FIGURE CAPTIONS

Figure 1. Flowchart of study selection according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

TABLE CAPTIONS

Table 1. Main general and superimposition-related characteristics of the included studies.

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

Table 3. Tested outcomes, conclusions and limitations of the included studies.

SUPPLEMENTARY TABLE CAPTIONS

Table S1. Detailed description of the search performed in various databases.

Table S2. Signalling questions used for the QUADAS-2 tool.

Table S3. Relevant studies that were not eligible and reasons for this.

Table S4. Further general and superimposition-related characteristics of the included studies.

Table S5. Detailed main and secondary results of the included studies.

Figure 1. Flowchart of study selection according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

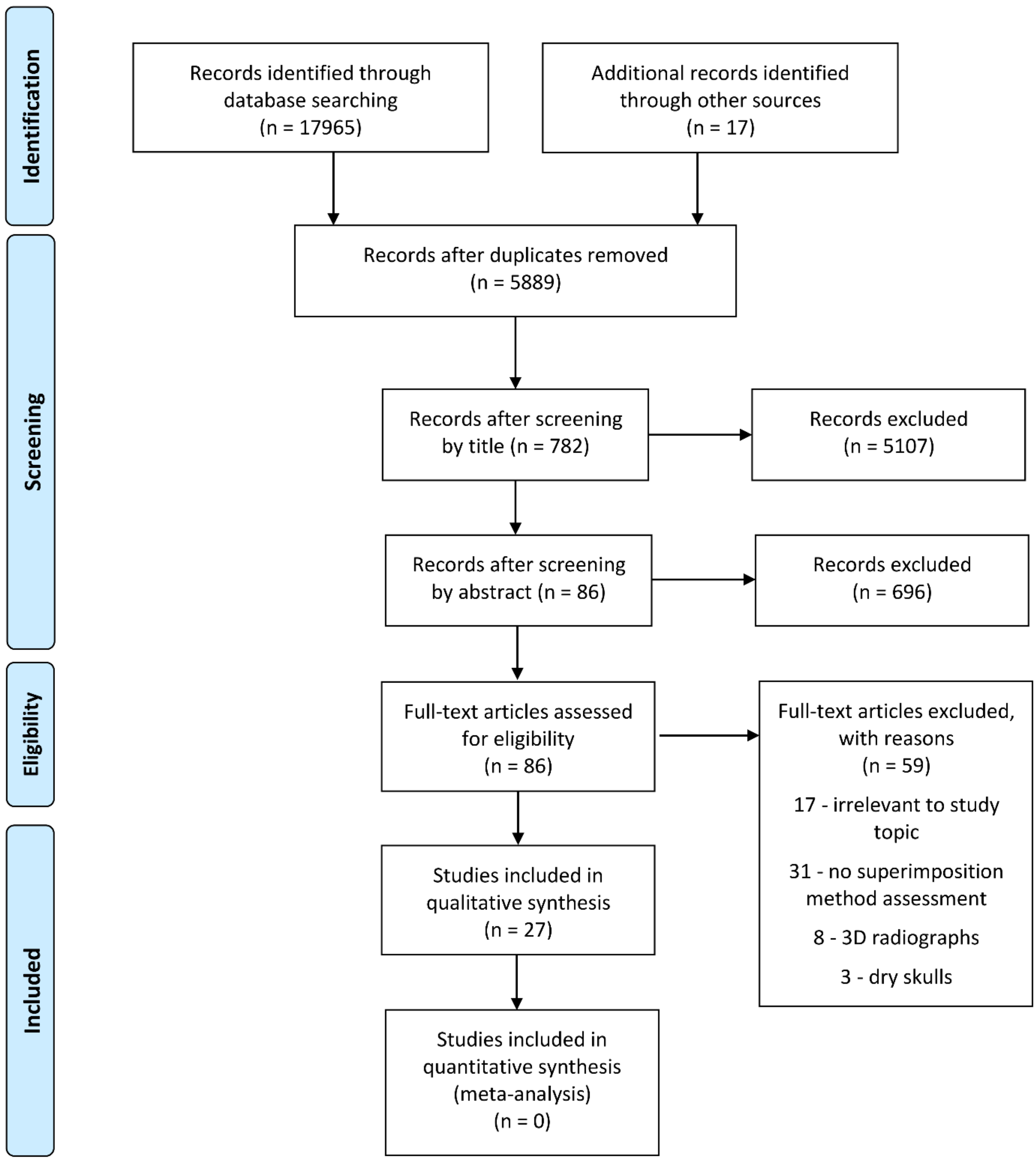

Table 1. Main general and superimposition-related characteristics of the included studies.

Study name	Objectives	Sample size and sex	Mean age at baseline and growth status	Average time between serial Rx	Superimposition method
Baumrind et al. (1976)	To test the precision of tracing superimposition using measurements.	25 patients (25 Rx pairs), 11 males and 14 females	9.2 yrs Growing	2 yrs	Best fit method (manual) 1. S-N Method (S fixed) 2. ACB 3. Palatal plane 4. Mandibular border
Houston and Lee (1985)	To test the precision of different methods of Rx superimposition on CB structures.	24 patients (24 Rx pairs)	NA Growing	1 - 7 yrs	Best fit ACB method (manual) 1. Directly 2. Blink method 3. Subtraction method 4. Tracing of CB 5. Tracing on S-Na line at Sella
Buschang et al. (1986)	To test the reproducibility of superimposition on the ACB and Björk's mandibular structures.	20 male patients (3 Rx pairs)	6 yrs Growing	2, 5 and 8 yrs	Best fit method (manual) 1. ACB 2. Mandible
Ghafari et al. (1987)	To test 4 different methods of CB Rx superimposition.	26 patients (26 Rx pairs), 13 males and 13 females	Males: 12.5 yrs Females: 12.2 yrs Growing	Males: 2.9 yrs Females: 2.7 yrs	Best fit method (manual) 1. ACB 2. SN (S fixed) 3. R (Bolton-Nasion plane parallel) 4. a: Ba-N at CC, b: Ba-N at N
Baumrind et al. (1987)	To test the difference between superimposition on anatomical structures versus metallic implants.	31 patients (31 sets of four), 11 males and 20 females	8.5 yrs Growing	2, 4 and 7 yrs	Best fit method 1. Superimposition on maxillary implants 2. Maxillary anatomic structures
Cook et al. (1988)	To assess the horizontal and vertical errors associated with superimposition on Björk's mandibular structures.	50 patients (each Rx traced multiple times)	NA Growing	-	Best fit method (manual) 1. All Björk structures (A) 2. Anterior structures (B) 3. Posterior structures (C) 4. All structures without 3rd molar (D) 5. All structures without inferior dental canal (E)
Cook et al. (1989)	To test the reproducibility of serial cephalometric tracing superimpositions on the mandibular plane, the mandibular outline and Björk's mandibular structures.	30 patients (30 Rx pairs)	12.3 yrs Growing	1.6 (range: 1 – 2.8) yrs	Best fit method (manual) 1. Mandibular plane 2. Mandibular outline 3. Björk's mandibular structures

Nielsen et al. (1989)	To test 3 different methods of maxillary radiographic superimposition.	18 patients (18 Rx pairs), 6 males and 12 females	10 ± 0.4 yrs Growing	4 ± 0.6 yrs	Best fit method (manual) 1. Palatal plane, along ANS 2. Implant method 3. Björk structural method
Iseri et al. (1990)	To test the reproducibility of Bjork's ACB superimposition method, supplemented by a computerized procedure, on measuring maxilla displacement during growth.	14 female patients (123 Rx)	Range: 8 – 21 yrs Growing	One Rx every 12 months starting from the age of 8	Best fit method (manual) 1. ACB: Björk structural method
Cook et al. (1994)	To test the agreement of different methods of Rx superimposition on maxillary and mandibular structures.	90 Rx divided into 3 groups of 30 patients each	NA Growing	1.5 yrs	Best fit method Maxilla: 1. Palate, manual 2. Ricketts' position 3, digital Mandible: 1. Ricketts' position 4, digital 2. Björk's method, manual
Baumrind et al. (1996)	To test the outcomes of 3 superimposition methods regarding tooth displacement.	30 patients, 11 males and 19 females	8.5 yrs Growing	NA	Best fit method (manual) 1. ACB 2. Maxillary implants (IMP_MAX) 3. Maxillary anatomic structures (MAX)
Springate and Jones (1998)	To test the agreement of Björk and Ricketts superimposition method with an implant superimposition method in the mandible.	23 patients (23 Rx pairs), 11 males and 12 females	11.6 ± 0.8 yrs Growing	3.2 ± 0.3 yrs	Best fit method (manual) 1. Björk method 2. Rickett's method 3. Implant methods
Efstratiadis et al. (1999)	To test differences of 2 superimposition methods on measuring mandibular displacement.	22 patients (22 Rx pairs), 10 males and 12 females	11.8 yrs Range: 7.4 – 15.9 yrs Growing	NA	Best fit method (manual) 1. ACB 2. Maxillary structures
You et al. (1999)	To test the precision of three different superimposition methods for the maxilla and three for the mandible.	14 patients (14 Rx pairs)	NA Growing	7.5 ± 2.3 months	Best fit method (manual) 1. Björk method (maxillary and mandibular) 2. Ricketts' positions 3 and 4 3. Pancherz' s method
Arat et al. (2003)	To test the displacement of CB and face landmarks during growth by 3 different superimposition methods.	40 patients (40 Rx pairs), 12 males and 28 females	12.0 ± 0.2 yrs Growing	2 yrs	Best fit method (manual) 1. Björk structural (Method A) 2. Steiner (Method B) 3. Ricketts (Method C)
Goel et al. (2004)	To test the agreement and reproducibility of two cephalometric superimposition methods.	12 patients (12 Rx pairs)	11.1 yrs Growing	1 year	Best fit method (manual) 1. T-method 2. Viazis' method
Gliddon et al. (2006)	To test the precision of 4 methods for CB superimposition.	8 patients	NA NA	NA	Best fit method (manual) 1. FH method

					2. S-N method 3. 5 landmarks (Sella, Nasion, Porion, Orbitale, and Basion) LS-5 method 4. Geometric method
Roden-Johnson et al. (2008)	To test the agreement between hand and digital (Quick ceph) superimposition of the ACB, the maxilla and the mandible.	30 patients (30 Rx pairs)	NA Growing	At least 2 yrs	Björk structural method (manual and digital) 1. ACB 2. MAX 3. Mandible
Gu et al. (2008)	To test if there is a difference between cephalometric superimposition on anatomical structures and on metallic implants according to Björk.	10 patients (60 Rx pairs), 4 males and 6 females	8.9 ± 1.1 Growing	1.5, 3, 4.2, 5.6 and 7.8 yrs (1 at every CVM stage)	Best fit method (manual) 1. ABO method (maxillary and mandibular) 2. Lower mandibular border 3. Best fit on implants
Standerwick et al. (2008)	To test the agreement of two cephalometric superimposition methods with growth patterns defined by Melsen necropsy specimens and the Bjork implant studies.	28 patients (28 Rx pairs), 9 males and 19 females	8 yrs Growing	NA	Best fit method (manual) 1. ACB registered on the anterior curvature of sella turcica 2. Registration on I-point with ACB parallel
Huja et al. (2009)	To test the difference between hand and digital superimposition, as well as between the best-fit cranial base and S-N superimpositions using the digital method.	64 patients (64 Rx pairs), 33 males and 31 females	12.9 yrs, Range: 9.3 – 18.5 yrs Growing	2.7 yrs	Best-fit method (digital): 1. Cranial base superimposition method 2. S-N superimposition method
Standerwick et al. (2009)	To compare superimposition on sella turcica and the anterior cranial base (S-ACB) to superimposition referenced at the occipital condyle (I-point) for demonstrating craniofacial growth and development.	32 patients (160 Rx), 16 males and 16 females	8 yrs Growing	10 yrs	Best fit method (manual) 1. ACB registered on the anterior curvature of sella turcica and anterior cranial base. 2. I-point with on the antero-inferior contour of the occipital condyles in norma lateralis
Arat et al. (2010)	To compare the results of 4 superimposition methods on the displacement of cranial landmarks during growth and adulthood.	30 patients (90 Rx), 12 males and 18 females	12 yrs Growing and non-growing	3.3 and 20 yrs	Best fit method (manual) 1. Björk structural method 2. Ricketts' method 3. Steiner method 4. T-W method
Türköz et al. (2011)	To test if there is a difference in cephalometric superimposition outcomes between Björk's anterior cranial base structural method and Steiner's method of sella–nasion line registered at sella.	70 patients (70 Rx pairs) Ex: N=35 (15 males, 20 females) Non-Ex: N=35 (13 males, 22 females)	Ex: 14.7 ± 1.8 yrs Non-Ex: 15 ± 2.3 yrs Growing	Ex: 32.8 ± 10.5 months Non-Ex: 20 ± 7.6 months	1. Björk's structural method 2. Steiner's method of sella–nasion line registered at sella.
Lenza et al. (2015)	To test 4 different methods of cranial base Rx superimposition.	31 patients (31 Rx pairs),	13.3 yrs Growing	4.2 yrs	Best fit method 1. Björk's structural method

		11 males and 20 females			2. Steiner/Tweed method 3. Ricketts' method (N) 4. Ricketts' method (CC)
Jabbal et al. (2016)	To test if there is a difference between the change in incisor inclination between Björk's mandibular cephalometric superimposition and Me-Go, Go-Gn and the tangent to the lower border of the mandible.	39 patients (39 Rx pairs) 13 males and 26 females	13.7 yrs time span Growing	2.3 yrs	Best fit method (manual) 1. Björk's mandibular structures 2. Menton–Gonion method 3. Gnathion–Gonion method 4. Lower border of the mandible method
Jiang et al. (2020)	To test the interoperator error of a computer-aided automated method for superimposition compared with Johnston's free-hand tracing superimposition method.	28 patients (28 Rx pairs) 7 males and 21 females	15.3 yrs Range: 12–27 yrs Growing and non-growing	2.7 yrs	1. Computer based automated method for anterior cranial base, maxilla, and mandible. 2. Structural superimposition method by Johnston for the anterior cranial base, maxilla, and mandible

ACB: anterior cranial base, Rx: radiograph, ANS: anterior nasal spine, PNS: posterior nasal spine, NA: Not applicable. IMP: Implant, MAX: Maxilla, MP: Mandibular plane, CVM: cervical vertebral maturation, yrs: years

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

Study	Risk of Bias					Applicability Concerns			
	Patient Selection	Index Test	Reference Standard	Flow & Timing	Total Risk of Bias	Patient Selection	Index Test	Reference Standard	Total Applicability Concerns
Baumrind et al. (1976)	?	?	☺	☺	?	?	?	☺	?
Houston and Lee. (1985)	☺	☹	☺	☺	☹	☹	☹	☺	☹
Buschang et al. (1986)	☺	☹	☺	☺	☹	☹	☹	☺	☹
Ghafari et al. (1987)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Baumrind et al. (1987)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Cook et al. (1988)	☹	☹	☺	☺	☹	☹	☹	☺	☹
Cook et al. (1989)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Nielsen et al. (1989)	☹	☹	☺	☺	☹	☹	☹	☺	☹
Iseri et al. (1990)	☺	☹	☹	☺	☹	☺	☹	☹	☹
Cook et al. (1994)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Baumrind et al. (1996)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Springate and Jones (1998)	☺	☹	☺	☺	☹	☹	☹	☺	☹
Efstratiadis et al. (1999)	☺	☹	☺	☺	☹	☺	☹	☺	☹
You et al. (1999)	☹	☹	☺	☺	☹	☹	☹	☺	☹
Arat et al. (2003)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Goel et al. (2004)	☹	☹	☺	☺	☹	☹	☹	☺	☹
Gliddon et al. (2006)	?	☹	☺	☺	☹	?	☹	☺	☹
Roden-Johnson et al. (2008)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Gu et al. (2008)	?	☹	☺	☺	☹	?	☹	☺	☹
Standerwick et al. (2008)	☹	☹	☹	☺	☹	?	☹	☹	☹
Huja et al. (2009)	☺	?	☺	☺	?	☺	?	☺	?
Standerwick et al. (2009)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Arat et al. (2010)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Türköz et al. (2011)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Lenza et al. (2015)	?	☹	☺	☺	☹	?	☹	☺	☹
Jabbal et al. (2016)	☺	☹	☺	☺	☹	☺	☹	☺	☹
Jiang et al. (2020)	☺	☹	☺	☺	☹	☺	☹	☺	☹

Happy face: low risk of bias/low applicability concerns.
Sad face: high risk of bias/high applicability concerns.
?: unclear risk of bias/unclear applicability concerns.

Table 3. Tested outcomes, conclusions and limitations of the included studies.

Study name	Main outcomes	Secondary outcomes	Conclusions	Limitations
Baumrind et al. (1976)	Standard deviations in mm of horizontal and vertical typical landmark displacements due to errors of tracing superimposition.	Standard deviation of angles between levels used to perform superimpositions by four different operators.	The errors attributed to the superimposition procedure (no landmark identification error) are relatively large, when considering individual cases. Different superimpositions have different impacts on the reliability of the estimated changes through time.	1. Good quality Rx. 2. No detailed error. Assessment/reporting regarding individual cases.
Houston and Lee (1985)	Reproducibility of the ACB superimposition with different methods measured through the standard deviations of the displacement (mm) of the midpoints of the registration lines.	Relation of the errors made by the two operators.	All superimposition methods have large errors.	1. No assessment of an actual superimposition outcome. 2. No testing of individual differences. Mean values were only considered. 3. No blinding (the registration lines of the first superimposition were present when the 2nd was performed).
Buschang et al. (1986)	Reproducibility of the vertical and anteroposterior position of the ACB and mandibular reference line points following repeated superimpositions.	Change in reproducibility by time (experience).	1. ACB superimposition was more reproducible than mandibular superimposition. 2. Both ACB and mandible superimposition reference lines showed clinically acceptable reproducibility.	1. No assessment of an actual superimposition outcome. 2. No testing of individual differences. Mean values were only considered 3. No comparative statistics reported. 4. Expertise in cephalometry/ then extensive training on superimpositions 5. Only one operator. 6. Poor reporting.
Ghafari et al. (1987)	Agreement of the four methods measured as horizontal and vertical displacement of the following 6 reference landmarks: ANS, PNS, Point A, Point B, Pogonion, Gonion.	Clinically significant differences (> 1 mm) between the 4 methods.	The different methods show varying results on the measured outcomes that can be considered clinically significant. Method 4a tended to show the biggest differences from the other methods.	1. No testing of individual differences. Mean values were only considered. 2. Only one operator. 3. No reproducibility assessment.
Baumrind et al. (1987)	Differences between the two superimposition methods in mean vertical and horizontal displacement of 3 reference landmarks: ANS, PNS and A point.	NA	The two methods show differences on the measured outcomes regarding the vertical and horizontal displacement of ANS, PNS and A point in growing patients.	1. No testing of individual differences. Mean values were only considered. 2. No method error.

Cook et al. (1988)	Reproducibility of the Björk's mandibular structure superimposition (intrarater and interrater) measured as differences in SNB and MMPA at the same Rx traced different times.	Operator with vs. operator without experience.	Horizontal error was much less than the vertical. Midline structures were more reliable than bilateral and the lower molar tooth germ more reliable than the inferior nerve canal. The Björk structures identified in a single Rx show clinically adequate precision.	1. No assessment of an actual superimposition outcome. Only one time point used. 2. No testing of individual differences. Mean values were only considered.
Cook et al. (1989)	Reproducibility of the three techniques measured as the displacement in mm of a constructed registration point (C), after registering the same tracings at two points, constructed following the 1st and the 2nd superimposition.	Reproducibility level achieved by two operators.	1. All 3 techniques showed sizeable errors. 2. Björk's structure is the least reproducible method of the 3.	1. No assessment of an actual superimposition outcome. 2. No testing of individual differences. Mean values were only considered. 3. No blinding (the registration points of the first superimposition were present when the 2nd was performed).
Nielsen et al. (1989)	Agreement between different methods of maxillary superimposition measured through the displacement of the maxillary skeletal and dental landmarks: ANS, ANS (Harvold), PNS (Posterior nasal spine), A, U6Cusp, U6Apex, U1Edge, U1Apex (measured on Downs occlusal plane X und perpendicular to this one Y).	Interrater agreement for the structural method.	The implant method and the structural method show considerable differences with the PP best fit method. The implant method showed limited mean differences with the structural method. The structural method might show adequate reproducibility.	1. No testing of individual differences. Mean values were only considered. 2. No method error evaluation apart from the structural method. 3. Good quality Rx with clearly visible zygomatic process. 4. In some cases only unilateral implants were used.
Iseri et al. (1990)	Mean difference of the maxillary displacement measured as linear and angular displacement of the following landmarks: s-n-ia,st-nt-ia, s-ia hor, s-ia ver, s-ip hor, s-ip ver, RWFcrb/IPLs, ia-ip.	NA	The mean changes measured repeatedly by the superimposition technique were not statistically different.	1. No testing of individual differences. Mean values were only considered. 2. The anatomical stability of the points used as superimposition reference is not verified.
Cook et al. (1994)	1. Mean difference of maxillary superimposition using best fit of the palate and Position 3. 2. Reproducibility of MAX superimpositions. 3. Mean difference of mandibular superimposition using best fit of the mandible and Position 4. 4. Reproducibility of mandible superimpositions.	NA	The maxillary superimposition methods differ significantly in terms of vertical and horizontal displacement of U1, while they showed similar reproducibility. All other measurements in the maxilla and those in the mandible provided similar mean values. All methods also showed similar reproducibility.	1. No testing of individual differences. Mean values were only considered.

Baumrind et al. (1996)	Mean vertical and horizontal displacement of 4 reference landmarks (U6C, U6A, U1C, U1A) detected by ACB, IMP_MAX and MAX superimpositions.	NA	The compared methods show differences on the measured outcomes for the vertical and horizontal displacement of maxillary first molar and the maxillary central incisor in growing patients. The major differences were detected in the vertical dimension by the maxillary first molar.	1. No testing of individual differences. Mean values were only considered. 2. No method error.
Springate and Jones (1998)	The median difference of Björk's and Rickett's mandibular superimpositions from the implant method, measured through the displacement of six mandibular skeletal and dental landmarks (Me, I, Co, Pog, Mbc, Go).	Inter-operator error.	The structural method shows high trueness, whereas the Rickett's method low, if we assume the implant method as gold standard. No conclusion can be drawn about accuracy and precision of the methods.	1. No testing of individual differences. Mean values were only considered. 2. Results for precision (reproducibility) are not reported for each method separately. 3. Good quality Rx with similarity in mandibular projection.
Efstratiadis et al. (1999)	Mean difference of the ACB & MAX superimpositions measured as linear displacement of pogonion, gnathion and menton.	Horizontal and vertical components of displacement vectors of pogonion, gnathion and menton measured through ACB and MAX.	Cranial and maxillary superimpositions can lead to different results, especially in the vertical dimension and in growing patients.	1. No method error. 2. No testing of individual differences. Mean values were only considered. 3. No blinding.
You et al. (1999)	Reproducibility of the 3 superimpositions methods measured through the displacement of the landmarks A, B, Pg, U1, U6 L1, L6 along or vertical to the occlusal plane.	Agreement between the 3 superimposition methods (mm) of the maxilla and the mandible.	All 3 superimpositions methods may show good reproducibility and agreement to each other.	1. No testing of individual differences. Mean values were only considered. 2. Only one operator. 3. Small sample (inadequate statistical power).
Arat et al. (2003)	Agreement of the three methods in the displacement in mm of specific landmarks (N, PT, S, Ba) from T1 to T2.	1. Intrarater reliability on the main outcome measurements (ICC). 2. Measured displacement at Pg by each method.	Landmarks used for Steiner and Rickett's methods show significant vertical and horizontal displacements according to Björk's structural methods. The three methods measure differently changes at Pg.	1. No testing of individual differences. Mean values were only considered. 2. No adequate reproducibility testing. 3. No gold standard.
Goel et al. (2004)	Reproducibility of 2 superimpositions methods measured through the displacement of the landmarks A, U1, L1, B, Pg along the occlusal plane.	Agreement between the two superimposition methods.	The two superimposition methods may show good reproducibility and agreement to each other.	1. No testing of individual differences. Mean values were only considered. 2. Only one operator. 3. Small sample (inadequate statistical power).

Gliddon et al. (2006)	Reproducibility of the four methods (intrarater and interrater) measured as horizontal and vertical displacement of the following reference landmarks: ANS, Point A, Point B, and Pogonion.	1. Operator with vs. operator without experience. 2. Differences in reproducibility between the four reference landmarks.	Manual geometric method shows the high precision followed by the LS-5, the S-N, and the F-H method in decreasing order.	1. Identical images were superimposed (no actual clinical data). 2. Not adequate sample description.
Roden-Johnson et al. (2008)	Mean difference between computer and hand superimpositions in the assessment of structural changes between T1 and T2.	NA	There might be good agreement between the hand and the digital (Quick ceph) superimposition techniques of the ACB, MAX and the mandible.	1. No testing of individual differences. Mean values were only considered.
Gu et al. (2008)	Mean difference of mandibular and maxillary superimposition methods measured through landmark displacement according to the functional occlusal plane (horizontal) and the pterygomaxillary fissure (vertical).	NA	The ABO maxillary superimposition method seems to overestimate the forward displacement of point A and underestimate the vertical displacement of point A, ANS and PNS. Mandibular superimposition method (ABO) might have a good agreement with the implant superimposition method.	1. No testing of individual differences. Mean values were only considered. 2. Small sample (inadequate statistical power). 3. Inadequate error evaluation.
Standerwick et al. (2008)	Agreement of the two methods measured through the displacement of the landmarks between tracings in degrees.	NA	1. There was no good agreement between measurements using I-point and ACB superimposition. 2. I-point might show better agreement to growth patterns defined by Melsen necropsy specimens and the Bjork implant studies than traditional ACB superimposition.	1. No testing of individual differences. Mean values were only considered. 2. The anatomical stability of I point used as superimposition reference is not verified. 3. Inadequate error evaluation.
Huja et al. (2009)	Mean difference between the best-fit cranial base superimposition and S-N superimpositions using the digital method measured through corresponding T2 landmark distances (mm).	Difference between repeated and between hand and digital superimpositions on the cranial base and S-N measured through corresponding T2 landmark distances (T2 LD).	There might be only small differences between best-fit cranial base and S-N superimpositions. Additionally, it seems that there are no differences between cranial base and regional superimpositions produced by Dolphin Imaging version 10 and those completed by hand.	1. Good quality Rx. 2. No assessment of individual differences. Mean values were only considered.
Standerwick et al. (2009)	Mean distance between corresponding landmarks (x-axis and y-axis) on serial tracings superimposed at I-point and S-ACB.	Repeatability of each superimposition method (ICC).	There is no agreement between the S-ACB to occipital condyle (I-point) superimposition.	1. No testing of individual differences. Mean values were only considered. 2. The anatomical stability of I point used as superimposition reference is not verified. 3. Inadequate reproducibility evaluation.

Arat et al. (2010)	Agreement of the positional changes of ACB landmarks (Nasion, wing, tuberculum sella, sella, basion & pterygomaxillare), detected by 4 ACB superimposition methods. Björk structural superimposition method was used as gold standard.	1. Stability of cranial landmarks according to Björk's method during puberal, postpubertal and overall periods. 2. Repeatability of all procedures in 10 patients.	The T-W method shows the highest agreement with the Björk structural method.	1. No testing of individual differences. Mean values were only considered. 2. Only one operator. 3. The landmarks/method that identified as the best/stable, according to Bjork's structural method were included in the area that was used in Bjork's superimposition.
Türköz et al. (2011)	Mean difference of Björk's and Steiner's superimposition methods in the following landmarks: N, Or, ANS, PNS, A, B, Po, Gn, Me, Go, Co, maxillary rotation and mandibular rotation in extraction and non-extraction cases.	Reproducibility level achieved by one operator.	There might be no significant differences between the outcomes of Björk's and Steiner's anterior cranial base superimposition.	1. No testing of individual differences. Mean values were only considered. 2. No adequate reproducibility testing.
Lenza et al. (2015)	Agreement of the four methods measured as horizontal and vertical displacement (mm) of the following 7 landmarks: ANS, PNS, Gnathion, Gonion, Pogonion, Point A, Point B.	NA	There might be good agreement between the four methods.	1. No testing of individual differences. Mean values were only considered. 2. High quality Rx. 3. Inadequate error evaluation.
Jabbal et al. (2016)	Agreement of three commonly used mandibular superimposition methods with Björk's method in the degrees of incisor inclination change induced by orthodontic treatment (ICC).	Agreement between conventional methods.	1. All methods compared showed a mean change in incisor inclination below 1 degree. 2. All 4 methods may show good agreement to each other.	1. No testing of individual differences. Mean values were only considered.
Jiang et al. (2020)	Mean of T2 landmark distances of paired automated and hand superimposed T1-T2 cephalometric pairs by three different operators.	Differences in hand superimposition among the operators.	Computer-aided cephalometric superimposition provides comparable interoperator error results to those of traditional hand tracing when structural superimposition is concerned.	1. No testing of individual differences. Mean values were only considered. 2. No method error. 3. No gold standard.

ACB: Anterior cranial base, ICC: Intraclass correlation coefficient, ANS: Anterior nasal spine, PNS: Posterior nasal spine, PP: Palatal plane, T(x): Timepoint, S: Sella, N: Nasion, NA: Not applicable, MAX: Maxilla, IMP: Implant, Pg: Pogonion, Rx: Radiograph

Table S1. Detailed description of the search performed in various databases.

Database	Search strategy	Limits and Results
Pubmed (https://www.ncbi.nlm.nih.gov/pubmed)	((cephalometr*[Title/Abstract] OR cephalogr*[Title/Abstract] OR (lateral head film*)[Title/Abstract] OR (lateral head radiograph*))[Title/Abstract] AND (superimpos*[Title/Abstract] OR registrat*)[Title/Abstract] AND ((head[Title/Abstract] OR heads[Title/Abstract] OR cranial[Title/Abstract] OR cranium[Title/Abstract] OR mandib*[Title/Abstract] OR maxilla*[Title/Abstract] OR craniofacial[Title/Abstract] OR zygoma*[Title/Abstract] OR (upper jaw)[Title/Abstract] OR (lower jaw))[Title/Abstract]	Search date: Filters: Humans Publication date: From 0001/01/01 to November 1st, 2020 Search Builder: 'All Fields' Results: 578
EMBASE (www.embase.com)	(cephalometr*:ab,ti OR 'cephalogr* cephalometr*':ab,ti OR cephalogr*:ab,ti OR 'lateral head film*':ab,ti OR 'lateral head radiograph*':ab,ti) AND (superimpos*:ab,ti OR registrat*:ab,ti) AND (head:ab,ti OR heads:ab,ti OR cranial:ab,ti OR cranium:ab,ti OR mandib*:ab,ti OR maxilla*:ab,ti OR craniofacial:ab,ti OR zygoma*:ab,ti OR upper OR lower OR jaw:ab,ti)	Search date: Limits: Humans Publication date: From 0001/01/01 to November 1st, 2020 Search Builder: 'All Fields' Special function used: non Results: 490
Google Scholar (www.scholar.google.com)	**Advanced search 1** With the exact phrase: cephalometric superimposition With at least one of the words: head OR heads OR craniofacial OR cranial OR vs OR cranium OR and OR mandible OR maxilla OR cranial base OR zygoma OR upper jaw OR lower jaw **Advanced search 2** With the exact phrase: cephalometric registration With at least one of the words: head OR heads OR craniofacial OR cranial OR vs OR cranium OR and OR mandible OR maxilla OR cranial base OR zygoma OR upper jaw OR lower jaw	Search date: Limits: - Publication date: From 0001/01/01 to November 1st, 2020 Search Builder: 'All Fields' Results: 16500 (The first 15 pages including the first 150 most relevant results were searched each time)
Cochrane Library (All databases) (www.thecochranelibrary.com)	(cephalometr* OR cephalogr* OR lateral head film* OR lateral head radiograph*) AND (superimpos* OR registrat*) AND (head OR heads OR cranial OR cranium OR mandib* OR maxilla* OR craniofacial OR zygoma* OR upper jaw OR lower jaw)	Search date: Limits: - Publication date: from 0001/01/01 to November 1st, 2020 Search Builder: 'All Fields' Special function used: ''Word variations have been searched'' Results: 72
Open Grey (http://www.opengrey.eu/)	superimposition, superimpositions, cephalometric registration, cephalometric registrations, image registration, image registrations	Search date: November 1st, 2020 Results: 325
Grey Literature Report (www.greylit.org)	superimposition, superimpositions, cephalometric registration, cephalometric registrations, image registration, image registrations	Search date: November 1st, 2020 Results: 0

Table S2. Signalling questions used for the QUADAS-2 tool.

	Domain	Signaling question(s)
Risk of bias	Patient Selection	1. Was the sample adequately described (e.g. age, sex, main characteristics)?
		2. Did the study avoid inappropriate exclusions?
		3. Was the sample size adequate?
	Index Test	1. Were the index test results interpreted without knowledge of the results of the reference standard?
		2. Was the reporting of the index test results adequate (e.g. testing of individual differences)?
		3. Was an appropriate method error evaluation used?
	Reference Standard	1. Is the reference standard likely to correctly measure the outcome of interest?
		2. Were the reference standard results interpreted without knowledge of the results of the index test?
		3. Is the reference standard acceptable?
	Flow and Timing	1. Were all patients included in the analysis?
Applicability	Patient Selection	1. Are there concerns that the included patients and setting do not match the review question?
	Index Test	1. Are there concerns that the index test, its conduct, or its interpretation differ from the review question?
	Reference Standard	1. Are there concerns that the outcome of interest as defined by the reference standard does not match the question?

Table S3. Relevant studies that were not eligible and reasons for this.

Study	Reason for exclusion
Doppel et al. Am J Orthod Dentofacial Orthop. 1994;105;161-168	No assessment of a superimposition method
Pancherz et al. American Journal of Orthodontics. 1984;86:427-434	No assessment of a superimposition method. Superimposition was done to transfer the SN line from T0 to T1 and do cephalometric measurements
Fisher et al. Angle Orthod 1980;50:54-62	No quantitative analysis of the differences of the four superimposition methods
Wellens et al. Eur. J. Orthod. 2016, 38, 569–576	No assessment of a superimposition method
Zhang et al. Am J Orthod Dentofacial Orthop. 2018;153:673-684	No assessment of a superimposition method
Larson et al. Angle Orthod. 2010;80:474-9	In vitro study on dry skulls
Sayinsu et al. Eur J Orthod. 2007;29:105-8	No assessment of a superimposition method
Battagel. Eur J Orthod. 1993;15:305-14	No assessment of a superimposition method
Midtgård et al. Angle Orthod. 1974;44(1):56-61	Only reproducibility of landmark identification was tested
Cooke et al. Aust Dent J. 1991;36:38-43	No assessment of a superimposition method
Chen et al. Angle Orthod. 2004;74:155-61	No assessment of a superimposition method
Cooke et al. Am J Orthod Dentofacial Orthop. 1990;97:489-94	No assessment of a superimposition method. Superimposition used to evaluate changes in NHP
Damstra et al. Am J Orthod Dentofacial Orthop. 2010;138:546.e1-8	No assessment of a superimposition method
Trpkova et al. Am J Orthod Dentofacial Orthop. 1997;112:165-70	Only reproducibility of landmark identification was tested
Kerr.Br J Orthod. 1978;5:51-3	No assessment of a superimposition method
Houston et al. Eur J Orthod. 1986;8:149-51	No assessment of a superimposition method
Richardson. Am J Orthod. 1966;52:637-51	Only reproducibility of landmark identification was tested
Springate. Eur J Orthod. 2010;32:354-362	No assessment of a superimposition method
Hwang et al. PLoS ONE. 2014;9:1-10	3D
Björk. Am J Phys Anthrop. 1968;29:243-254	No assessment of a superimposition method
Davidovitch et al. Eur J Orthod. 2016;38:555-562	No assessment of a superimposition method
Hwang et al. Angle Orthod. 2020;90:69-76	Only reproducibility of landmark identification was tested

Lemieux et al. Am J Orthod Dentofacial Orthop. 2014;146:758-64	3D
Sakima. Am J Orthod Dentofacial Orthop. 2004;126:344-53	Superimposition of oblique cephalometric radiographs
Skieller et al. Am J Orthod. 1984;86:359-370	No assessment of a superimposition method
Steuer et al. Am J Orthod. 1972;61:493-500	No assessment of a superimposition method
He et al. J Int Med Res. 2019;47:2951-2960	No assessment of a superimposition method
Park et al. Angle Orthod. 2019;89:903-909	No assessment of a superimposition method
Leonardi et al. Eur J Orthod. 2010;32:242-7	Only reproducibility of landmark identification was tested
Kane et al. Cleft Palate Craniofac J. 2002;39:219-25	No assessment of a superimposition method
Lew. Br J Orthod. 1989;16:281-3	No assessment of a superimposition method
Choi et al. Korean J Orthod. 2012;42:235-41	3D
Oh et al. Korean J Orthod. 2020;50:170-180	No assessment of a superimposition method
Mathews et al. Angle Orthod. 1980;50:218-29	No assessment of a superimposition method
van der Linden et al. Angle Orthod. 1971;41:119-24	No assessment of a superimposition method. Study on skulls
Halazonetis. Am J Orthod Dentofacial Orthop. 2004;125:571-81	No assessment of a superimposition method
Cevidanes et al. Am J Orthod Dentofacial Orthop. 2009;136:94-9	3D
Naranjilla. Angle Orthod.2005;75: 63–68	No assessment of a superimposition method
Ong et al. Angle Orthod. 2001;71:90-102	No assessment of a superimposition method
Beit et al. Prog Orthod. 2017;18;18(1):44	No assessment of a superimposition method
Weissheimer et al. Int J Oral Maxillofac Surg. 2015;44:1188-96	3D
Adams et al. Am J Orthod Dentofacial Orthop. 2004;126:397-409	3D
Agronin et al. Am J Orthod Dentofacial Orthop. 1987;91:42-8	No assessment of a superimposition method
Ghoneima et al. Orthod Craniofac Res. 2017;20:227-236	3D and reproducibility of landmark identification
Bergersen. Angle Orthod. 1961;31:216–229	No assessment of a superimposition method
Bruntz et al. Am J Orthod Dentofacial Orthop. 2006;130:340-8	No assessment of a superimposition method
Koerich et al. Angle Orthod. 2017;87:473-479	3D

Tollaro et al. Am J Orthod Dentofacial Orthop.1995;108:525-32	No assessment of a superimposition method
Lee. Br J Orthod. 1980;7:121-4	No assessment of a superimposition method. Study on skulls
Moon et al. Angle Orthod. 2020;90:390–396	No serial cephalometric radiographs

Table S4. Further general and superimposition-related characteristics of the included studies.

Study name	Study design	Radiographic machines	Type of participants (Serial cephalograms)	Selection criteria	Tracing and superimposition tools	Number of operators	Area of interest	References for superimposition
Baumrind et al. (1976)	Prospective study on pre-existing radiographs	NA	Untreated patients with Class II, division 1 malocclusions	Good quality Rx. Growing untreated patient with Class II, division 1 malocclusions	Hand (tracing, superimposition) + Computer program (digitized tracing)	4	ACB MAX Mandible	1. SN: registered on sella 2. ACB 3. PP 4. MB
Houston and Lee (1985)	Prospective study on pre-existing radiographs	NA	NA	Pairs of Rx with adequate quality	Hand (tracing, superimposition, pin hole Jig) + Computer program (digitiser for the points of interest)	2	ACB	1. Cribiform plate and anterior wall of sella turcica 2. Tracing on S-N line, registered at Sella
Buschang et al. (1986)	Prospective study on pre-existing radiographs	NA	Untreated patients of similar origin	Children with at least 3 French-Canadian grandparents representing different socioeconomic sections of the population	Hand: A reference line (50mm) marked on the first cephalogram (6years) and transferred to the others after superimposition.	1	ACB Mandible	1. ACB: Björk's stable structures 2. Mandible: Björk's stable structures
Ghafari et al. (1987)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment Rx of Class II, division 1 malocclusion incl. extraction of 4 first premolars	Class II div 1, including ex 4 first premolars and treated by the same orthodontist	Hand (tracing ?) + Tektronik Plot 50 interactive digitizing system	1	CB	Method (M) M1: Best fit of ACB M2: SN line at S M3: Bolton -N plane parallel M4: Ba-N line at CC M4a: Ba-N line at N
Baumrind et al. (1987)	Prospective study on pre-existing radiographs	NA	Treated and untreated patients with maxillary and mandibular implants of Björk type	Patients with maxillary and mandibular implants of Björk type	Hand?	2	MAX	1. Maxillary implants 2. Outline of the palate
Cook et al. (1988)	Prospective study on pre-existing radiographs	NA	Routine cephalograms showing at least one lower third molar with some degree of crown	Each Rx showed at least one lower third molar with some degree of crown calcification but without root formation	Hand (tracing) + Computer program (superimposition, online digitiser for the points of interest)	2 (1 experienced and 1 unexperienc ed)	Mandible	Outlines of Björk¨s mandibular structures: i. anterior contour of chin ii. inner contour at the lower border of the symphysis iii. mandibular canal iv. lower contour of the lower third molar germ

			calcification but without root formation					
Cook et al. (1989)	Prospective study on pre-existing radiographs	Philips. Cephoralix CX 90/20, Focus-film distance 1.5m /Jig	Pre- & post orthodontic treatment Rx showing at least one lower third molar with some degree of crown calcification but without root formation	Presence of at least one lower third molar with some degree of crown calcification, but before root formation	Hand (tracing, superimposition, pin hole Jig) + Computer program (digitiser for the points of interest)	2	Mandible	1. Me-Go, registered at Me 2. Mandibular outline (registered on the symphysis) 3. Björk's mandibular structures (1969)
Nielsen et al. (1989)	Prospective study on pre-existing radiographs	NA	Treated and untreated patients with Class I or II malocclusions and at least 3 metallic implants	Good quality film. Clear visibility of the zygomatic process, ANS and orbital floor. No double contours. At least 3 to 4 stable implants in the zygomatic process (in both Rx)	Hand (tracing) + Computer Aided head film analysis (digitiser for the points of interest)	2	MAX	1. ANS: Alignment of hard palate and the nasal floor (PP) 2. Implant S: Bisected distance between implants 3. Anterior surface of the zygomatic process
Iseri et al. (1990)	Prospective study on pre-existing radiographs	Lumex type B cephalometer /Acquisition parameters: film-focus distance 180 cm, object to film: 10 cm	Treated and untreated patients with metallic implants, but without any craniofacial anomalies	Presence of bilateral posterior maxillary implants and one or two anterior maxillary implants	Hand (tracing) + Computer program (superimposition, digitiser for the points of interest)	?	ACB	Björk's ACB structures
Cook et al. (1994)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment with Class II, division 1 malocclusions Group 1: cervical HG + lower utility arch Group 2: cervical HG Group 3: untreated	Growing patient with Class II, division 1 malocclusions	Hand (tracing, superimposition) + Computer program (specific superimpositions)	1	MAX Mandible	1. MAX: PP registered at ANS) & best fit of the palate (lingual cortical plate of the inferior border of the palate, registered on the internal palatal structures) 2. Mandible: Corpus axis registered at PM & Björk's structural method
Baumrind et al.	Prospective study on	NA	Treated and untreated patients with Class I or	Class I or Class II malocclusions with	Hand?	2 (3 in case of	ACB MAX	1. ACB: anterior cranial fossa and the greater wings of the sphenoid

(1996)	pre-existing radiographs		Class II malocclusions and maxillary and mandibular implants of Björk type	maxillary and mandibular implants of Björk type		disagreement of the 2)		2. IMP_MAX: on 3 maxillary implants 3. MAX: best fit of the outline of the palate with emphasis on ANS and A point areas
Springate and Jones (1998)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment Rx of children with tantalum implants in the mandible	Image quality and similarity in the projection of the mandible in both Rx. Patient with tantalum markers implanted in the mandible	Hand (tracing, superimposition) + Computer program (digitisation of the points of interest and creation of coordinate system)	2	Mandible	1. Anterior contour of chin, inner contour at the lower border of the symphysis, mandibular canal, and lower contour of the lower 3rd molar germ 2. Xi, Pm axis 3. Tantalum implants in left side of the mandible
Efstratiadis et al. (1999)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment Rx of Class II, division 1 treated patients	Class II div 1 treated patients	NA/Hand (?)	NA	ACB MAX	1. ACB 2. Maxilla
You et al. (1999)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment Rx of patients with Herbst appliance	Patients treated with herbst appliance	Hand (tracing + superimposition)	1	ACB MAX Mandible	1a. Mand: Inner contour at the lower border of the symphysis, mandibular canal, and lower contour of the lower 3rd molar germ 1b. Björk's maxillary structural method 2a. MAX: PP at ANS 2b. Mand.: Corpus axis at PM 3. Occlusal plane (x axis) and perpendicular plane through Sella point (y axis)
Arat et al. (2003)	Prospective study on pre-existing radiographs	NA / Acquisition parameters: film-focus distance 155 cm, object to film: 12.5 cm	Cephalograms from a longitudinal growth study. Untreated patients.	No orthodontic treatment. Active growth period. Normal facial profile and vertical growth. Acceptable occlusion (Class I or end-to-end molar relationships, normal OB und OJ, minimal or no crowding)	Hand (tracing) + PorDios + Houston Hipad Digitizer of 1.125 mm resolution	2 (1 traced, 1 checked)	CB	1. ACB 2. SN 3. NBa
Goel et al. (2004)	Prospective study on	Trophy odontorama	Serial cephalograms from students	Students from Bagalkot, India	Hand (tracing + superimposition)	1	CB (anterior	1. Point T (anterior wall of sella turcica) and two planes: true vertical (TVP) and true horizontal

	pre-existing radiographs	cephalometric machine. 70Kvp, 6m-amp for 1.4 s)					and middle CB)	(THP) passing through point T and perpendicular to TVP. 2. Point T, C & L define a triangle that includes the anterior wall of sella turcica and the anterior and middle CB
Gliddon et al. (2006)	Prospective study on pre-existing radiographs	NA	Randomly selected from the archive	Randomly picked, acceptable or less than perfect Rx	CASSOS prediction tracing software (Soft Enable Technology Ltd, Hong Kong SAR, China)	2 (1 experienced and 1 unexperienc ed)	CB	1. Po and Or, registered at Po 2. S and N, registered at S 3. S, N, Po, Or and Ba 4. Best fit of CB including Ba-S, nasal bone, and occipital bone
Roden-Johnson et al. (2008)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment Rx	2 Rx taken with the same machine, growth between Rx intervals (min 2 years), structures of interest must be visible on Rx	Hand tracing and Quick Ceph 2000 (digital version 3.3, Quick Ceph Systems, Inc, San Diego, Calif)	1	ACB MAX Mandible	1. CB 2. Maxilla 3. Björk's stable structures
Gu et al. (2008)	Prospective study on pre-existing radiographs	NA	Untreated patients from a growth study	Sample of serial headfilms from Mathew and Ware`s implant study	Hand (tracing, superimposition) + Computer program (Dentofacial Planner Plus for digitisation)	2	MAX Mandible	1a. Maxilla: Lingual curvature of the palate registered at internal bony structures 1b. Mandible: Internal cortical outline of the symphysis and the inferior alveolar nerve canals 2. Inferior border of the mandible
Standerwick et al. (2008)	Prospective study on pre-existing radiographs	NA	Untreated patients from a growth study	Adequate quality film	Hand (tracing)	?	ACB Mandible	1. ACB, Sella turcica 2. I-point, ACB
Huja et al. (2009)	Prospective study on pre-existing radiographs	NA	Treated patients from a university archival database.	Good quality Rx. Growing treated patients. Patients who underwent orthognathic surgery and those with congenital syndromes and dental and skeletal asymmetries were excluded	Tracing (hand and digital) + Superimposition (hand and computer program, Dolphin Image and Management Solutions)	NA	ACB MAX Mandible	1. Anterior portion of the sella turcica, the cribriform plate of the ethmoid bone, internal contour of the frontal bone. 2a. Maxilla: Lower border of the palate, internal cortication of the maxilla. 2b. Mandible: Inner contour of the mandibular symphysis, the mandibular canal, apical portion of unerupted 3rd molars

Standerwick et al. (2009)	Prospective study on pre-existing radiographs	NA	Untreated patients from a growth study	Tracings from the Bolton Standards of Dentofacial Developmental Growth. The 8- and 18-year-old tracings were selected for lateral and frontal superimposition	Tracing (hand)?	NA	ACB	1. ACB, Sella turcica 2. I-point, ACB
Arat et al. (2010)	Prospective study on pre-existing radiographs	NA	Treated patients with Class II, division 1 malocclusion	Class II div 1 treated by the same orthodontist	Hand (tracing), template for landmark identification of the Rx of the same patient	1	ACB	1. ACB 2. S-Na line 3. N-Ba line 4. T-W: Reference line
Türköz et al. (2011)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment Rx of patients with (Ex) and without (Non-Ex) premolar extractions	Skeletal Class I subjects with SN/GoGN angle 26° - 38°, treated with fixed standard edgewise appliances and the Roth system, with and without extractions of 4 first premolars	Tracing (hand)	1	ACB	1. Bjork's anterior cranial base structures 2. S-N line, registered at S
Lenza et al. (2015)	Prospective study on pre-existing radiographs	NA	Pre- & post orthodontic treatment Rx of Class I malocclusion including extraction of the 4 first premolars	Good quality Rx. Patients with malocclusion Class I including extraction of the 4 first premolars	Flatbed scanner (Hewlett-Packard Company, Palo Alto, Ca, USA) Data Collection: Radiocef Studio 2.0 cephalometric software (Radiomemory, Belo Horizonte, Brazil	1	CB	1. ACB 2. SN Line 3. N-Ba line at N-point 4. N-Ba line at CC-point

Jabbal et al. (2016)	Prospective study on pre-existing radiographs	Planmeca, Dimax3, Helsinki, Finland.	Pre- & mid-orthodontic treatment cephalograms	1. fixed appliance treatment 2. age 12 - 16 years at treatment start 3. orthodontic treatment for at least a year between the pre- and mid-treatment radiographs 4. no lower incisor extraction Radiographs taken at least 1 year apart (120 weeks; SD = 34.4)	Tracing (hand)	1 AJ	Mandible	1. Outer surface of the symphysis, inner surface of the adjacent cortex, any prominent trabeculae within the symphyseal cortex, internal cortical margin of the inferior dental nerve, lower border of the third molar germ 2. MP: Me–Go 3. MP: Gn–Go 4. MP: Tangent to the lower border of the mandible
Jiang et al. (2020)	Prospective study on pre-existing radiographs	NA	Pre- & post-orthodontic treatment Rx	Pre-treatment (T1) and post-treatment cephalograms taken by the same X-ray machine. Subjects with orthodontic treatment	Tracing (hand) + Superimposition (hand and computer program)	3 senior orthodontic residents	ACB MAX Mandible	1. Areas defined by the following landmarks: i. CB: URP, S, Pt and N ii. Maxilla: Pt, PNS, ANS, A iii. Mandible: LM, Pg, Me and Go 2. ACB: de Coster's basal line MAX: zygomatic process and curvature of the palate Mand.: Facial half of symphysis and mandibular canal or molar tooth germs

ACB: anterior cranial base, CB: cranial base, Rx: radiograph, ANS: anterior nasal spine, NA: Not applicable. IMP: Implant, MAX: Maxilla, MP: Mandibular plane

Table S5. Detailed main and secondary results of the included studies.

Study name	Main outcomes	Secondary outcomes	Main results	Secondary results
Baumrind et al. (1976)	Standard deviations (mm) of horizontal and vertical typical landmark displacements due to errors of tracing superimposition	Standard deviation of angles between levels used to perform superimpositions by four different operators	Sella point. SN: X: 0.36; Y: 0.38; ACB: X: 0.47; Y: 0.38 U1 Edge point. SN: X: 1.10; Y: 0.83; ACB: X: 1.15; Y: 0.68; PP: X: 0.92; Y: 0.97 Menton point. SN: X: 1.52; Y: 0.60; ACB: X: 1.51; Y: 0.45; MB: X:0.44; Y: 0.41	S-N: Registered on sella (Sella-Nasion) 0.80° Anterior cranial base (Sella-Nasion) 0.71° Palatal plane (PP-ANS) 1.44° Mandibular border (Gonion-Menton) 1.23°
Houston and Lee (1985)	Reproducibility of the ACB superimposition with different methods measured through the standard deviations of the displacement (mm) of the midpoints of the registration lines	Relation of the errors made by the two operators	*Method error* 1. Direct: 1.82 mm, 2. Blink: 1.92 mm, 3. Subtraction: 1.62 mm, 4. ACB Tracing: 1.74 mm, 5. Sella Nasion: 1.41 mm	Correlations 1. Direct 0.51, 2. Blink 0.43, 3. Substraction 0.44, 4. ACB Tracing 0.03, 5. Sella nasion 0.33. (Correlations > 0.4 were statistically significant, p< 0.05)
Buschang et al. (1986)	Reproducibility of the vertical and anteroposterior position of the ACB and mandibular reference line points following repeated superimpositions	Change in reproducibility by time (experience)	*Method error* ACB: 0.17 - 0.39 mm. Mandible: 0.45 - 0.93 mm	Reproducibility of ACB reference points did not improve by experience, whereas reproducibility of the mandibular reference points improved
Ghafari et al. (1987)	Agreement of the four methods measured as horizontal and vertical displacement (mm) of the following 6 reference landmarks: ANS, PNS, Point A, Point B, Pogonion, Gonion	Clinically significant differences (> 1 mm) between the 4 methods	M1/M2 (mm): ANS 0.8±0.1, PNS 0.6±0.0, A 0.9±0.1, B 1.0±0.1, POG 1,2±0.1, GON 0.8±0.1. M1/M3 (mm): ANS 1.0±0.1, PNS 0.7±0.1, A 1.0±0.1, B 1.3±0.1, POG 1,4±0.2, GON 1.2±0.2 M1/M4 (mm): ANS 1.2±0.1, PNS 1.1±0.2, A 1.2±0.1, B 1.4±0.2, POG 1,6±0.2, GON 1.4±0.2. M1/M4a (mm): ANS 2.1±0.2*, PNS 2.0±0.2*, A 2.2±0.2*, B 2.3±0.3*, POG 2.4±0.3* GON 2.3±0.3* *p<0.01	M1/M4a (mm): ANS 2.1±0.2*, PNS 2.0±0.2*, A 2.2±0.2*, B 2.3±0.3*, POG 2.4±0.3*, GON 2.3±0.3*. M2/M4a (mm): ANS 2.0±0.2*, PNS 2.0±0.2*, A 2.1±0.2*, B 2.2±0.3*, POG 2.3±0.3*, GON 2.3±0.3*. M3/M4 (mm): ANS 2.1±0.2*, PNS 2.0±0.2*, A 2.1±0.3*, B 2.3±0.3*, POG 2.4±0.4* GON 2.4±0.3* *p<0.01
Baumrind et al. (1987)	Differences between the two superimposition methods in mean vertical and horizontal displacement (mm) of 3	NA	ANS: 2 years X: 0.51±0.81*, Y: 1.22±0.94** / 4 years X: 0.90±0.87**, Y: 1.89±1.60** / 7 years X: 1.45±1.25**, Y: 2.50±2.42** PNS: 2 years X: 0.48±0.78*, Y: -0.87±1.44* / 4 years X: 0.79±0.90**, Y: 1.62±2.03** / 7 years X: 1.29±1.26**,	NA

	reference landmarks: ANS, PNS and A point		Y: 2.66±2.01** Point A: 2 years X: 0.54±0.73**, Y: 1.20±0.89** / 4 years X: 0.91±0.78**, Y: 1.86±1.47** / 7 years X: 1.41±1.29**, Y: 2.50±2.25** *p<0.01, **p<0.001	
Cook et al. (1988)	Reproducibility of the Björk's mandibular structure superimposition (intrarater and interrater) measured as differences in SNB and MMPA at the same Rx traced different times	Operator with vs. operator without experience	*Error of the method (random)* 1: SNB 0.15°, MMPA 0.50°, 2: SNB 0.15°, MMPA 0.96°, 3: SNB 0.35°, MMPA 1.17°, 4: SNB 0.15°, MMPA 0.55°, 5: SNB 0.14°, MMPA 0.44° Vertical errors were higher than horizontal errors	The error between the raters for SNA (1.18°) and SNB (1.04°) angle was significantly different, with the experienced operator showing smaller error
Cook et al. (1989)	Reproducibility of the three techniques measured as the displacement (mm) of a constructed registration point (C), after registering the same tracings at two points, constructed following the 1st and the 2nd superimposition	Reproducibility level achieved by two operators	Mandibular plane: 0.48±0.56 (operator 1), 0.31±0.45 (operator 2) Mandibular outline: 0.55±0.46 (operator 1), 0.39±0.47 (operator 2) Björk's structures:1.30±1.38 (operator 1), 1.21±1.02 (operator 2) Björk's differed significantly from the other two (p < 0.01) in all cases. Mandibular outline did not differ from mandibular plane	No difference between the two operators
Nielsen et al. (1989)	Agreement between different methods of maxillary superimposition measured through the displacement of the maxillary skeletal and dental landmarks: ANS, ANS (Harvold), PNS (Posterior nasal spine), A, U6Cusp, U6Apex, U1Edge, U1Apex (measured on Downs occlusal plane X und perpendicular to this one Y)	Interrater agreement for the structural method	*Implant vs. PP (mm):* PNS x:1.09±1.29*, ANS x: 1.23±1.29**, A x: 1.15±1.09**, PNS y: 1.08±0.96**, ANS y: 1.88±0.89**, A y: 2.24±1.35**, U6C x: 1.53±1.35*, U6C y: 1.33±0.82** Structural vs. PP (mm): PNS x: 0.63±1.02, ANS x: 0.74±1.08, A x: 0.79±1.13*, PNS y: 1.39±0.94**, ANS y: 2.00±1.16**, A y: 1.96±1.09**, U6C x: 0.96±1.33**, U6C y: 1.59±0.77 (p<0.0001). *Implant vs. Structural (mm):* PNS x: 0.53±0.77*, ANS x: 0.56±0.74*, A x: 0.41±0.66, PNS y: -0.33±0.93, ANS y: -0.17±0.73, A y: -0.04±1.19 U6C x: 0.62±0.79 **, U6C y: -0.29±0.76 *p<0.01, **p<0.001	*Interrater agreement for the structural method (mm)* PNS x: 0.17±0.87, ANS x: 0.32±0.94, A x: 0.22±1.49, PNS y: -0.74±2.40, ANS y: 0.88±1.29**, A y: 0.77±1.42*, U6C x: 1.00±2.09, U6A x: 0.32±0.95, U1E x: 1.28±2.67, U1A x: 0.52±1.25, U6C y: -0.22±1.33, U6A y: -0.01±0.97, U1E y: 0.86 ±1.28*, U1A y: 0.52±0.75*** *p<0.05, **p<0.01, ***p<0.001
Iseri et al. (1990)	Mean difference of the maxillary displacement measured as linear and angular displacement of the following landmarks: s-n-ia,st-nt-ia, s-ia hor, s-ia ver, s-ip hor, s-ip ver, RWFcrb/IPLs, ia-ip	NA	No significant difference was identified between mean results of repeated measurements obtained. Mean Difference: s-n-ia 0.04° (Standard Error - SE: 0.10), st-nt-ia 0.01° (SE: 0.08), s-ia hor -0.02 mm (SE: 0.08), s-ia ver -0.08 mm (SE: 0.06), s-ip hor -0.03 mm (SE: 0.05), s-ip ver -0.09mm, (SE: 0.08), REFcrb/IPLs -0.04° (SE: 0.28), ia-ip 0.00 mm (SE: 0.07)	NA

Cook et al. (1994)	1. Mean difference of maxillary superimposition using best fit of the palate and Position 3 2. Reproducibility of MAX superimpositions 3. Mean difference of mandibular superimposition using best fit of the mandible and Position 4 4. Reproducibility of mandible superimpositions	NA	1. [Best fit of palate; Position 3] (mm or °) *p<0.05 U-1 ver. [0.46±1.30; 1.27±1.40]* U-1 hor. [0.19±1.66; 0.67±1.93]* U-6 ver. [0.74±1.47; 0.99±1.74], U-6 hor. [2.24±2.70; 2.62±3.52] U-1 ang [0.36±5.69; 0.46±6.23], U-6 ang. [1.05±9.32; 1.86±10.06] 2. [Best fit of palate; Position 3] (mm or °) (p>0.05) U-1 ver. [0.14; 0.03], U-1 hor. [0.02; 0.21] U-6 ver. [0.10; 0.06], U-6 hor. [0.15; 0.02] U-1 ang. [0.07; 0.03] U-6 ang. [0.64; 0.56] 3. [Björk; Position 4] (mm or °): L-1 ver. [0.48±1.43; 0.23±2.08], L-1 hor. [0.34±1.85; 0.24±2.09] L-6 ver. [0.85±1.87; 0.69±0.80], L-6 hor. [0.43±2.01; 0.16±2.46] L-1 ang. [2.86±5.85; 1.85±5.55], L-6 ang. [5.41±7.13; 6.42±7.23] 4. [Björk; Position 4] (mm or °): (p>0.05): L-1 ver. [0.04; 0.14], L-1 hor. [0.03; 0.17] L-6 ver. [0.02; 0.29], L-6 hor. [0.20; 0.05] L-1 ang. [0.19; 0.19], L-6 ang. [0.03; 0.16]	NA
Baumrind et al. (1996)	Mean vertical and horizontal displacement of 4 reference landmarks (U6C, U6A, U1C, U1A) detected by ACB, IMP_MAX and MAX superimpositions	NA	U6C (mm): ACB X: 7.92±3.30, Y: -10.71±3.02 IMP_MAX X: 4.98±2.67, Y: -7.06±2.30 MAX X: 5.85±2.53, Y: -4.15±1.73 U6A (mm): ACB X: 4.94±2.65, Y: -9.14±2.69 IMP_MAX X: 1.23±1.97, Y: -5.23±1.96 MAX X: 2.15±1.89, Y: -2.45±1.50 U1C (mm): ACB X: 3.90±4.62, Y: -9.79±4.48 IMP_MAX X: 1.10±4.35, Y: 1.86±3.62 MAX X: 1.86±3.62, Y: -2.04±2.83 U1A (mm): ACB X: 3.77±3.51, Y: -9.02±4.36 IMP_MAX X: 0.05±3.01, Y: 1.86±3.62 MAX X: -4.40±3.12, Y: -1.68±3.51	NA
Springate and Jones (1998)	The median difference of Björk's and Rickett's mandibular superimpositions from the implant method, measured through the displacement of six mandibular skeletal and dental landmarks (Me, I, Co, Pog, Mbc, Go)	Inter-operator error.	Median differences from the implant method (mm). *Björk's superimposition* Me-x: -0.01, Me-y: 0.00, Pog-x: -0.01, Pog-y: 0.05, I-x: 0.01, I-y: 0.06, Mbc-x: -0.62, Mbc-y: -0.51, Co-x: -0.94, Co-y: -1.60, Go-x: -0.18, Go-y: -1.14 *Rickett's superimposition* Me-x: 0.82, Me-y: 0.20, Pog-x: -0.28, Pog-y: -0.84, I-x: -2.18 (p<0.01), I-y: -0.38, Mbc-x: -1.29 (p=0.04), Mbc-y: -2.27, Co-x: -1.30 (p<0.01), Co-y: -3.82 (p<0.01),	All superimposition methods. Random error: 0.16 - 0.40 mm Systematic error (p<0.05): Pog-x: 0.12 mm Pog-y: 0.167 mm, Go-x: 0.14 mm

			Go-x: 0.62 (p=0.03), Go-y: -3.17 (p=0.02)	
Efstratiadis et al. (1999)	Mean difference of the ACB & MAX superimpositions measured as linear displacement of pogonion, gnathion and menton	Horizontal and vertical components of displacement vectors of pogonion, gnathion and menton measured through ACB and MAX	*Total displacement (mm) of* Pogonion: ACB 8.75, MAX 6.23, Diff. 2.52, p=0.0001 Gnathion: ACB 8.93, MAX 6.39, Diff. 2.55, p=0.0001 Menton: ACB 9.27, MAX 6.55, Diff. 2.73, p=0.0001	*Vertical displacement (mm)* Pogonion: ACB 7.95, MAX 5.00, Diff. 2.95, p=0.0001 Gnathion: ACB 8.02, MAX 5.36, Diff. 2.66, p=0.0001 Menton: ACB 8.39, MAX 5.43, Diff. 2.95, p=0.0001 *Horizontal displacement (mm)* Non-significant differences
You et al. (1999)	Reproducibility of the 3 superimpositions methods measured through the displacement of the landmarks A, B, Pg, U1, U6 L1, L6 along or vertical to the occlusal plane	Agreement between the 3 superimposition methods (mm) of the maxilla and the mandible	No significant difference was identified between mean results of repeated measurements for each method, at each point (p>0.05).	[Björk; Ricketts; Pancherz] mm A: [0.07; 0.38; 0.35] (p>0.05) B: [3.62; 2.58; 2.48] (p>0.05) Pg: [3.73; 2.30; 2.25] (p>0.05) U1: [2.89; -2.64; -2.37] (p>0.05) U6: [-2.54; -2.75; -2.73] (p>0.05) L1: [1.98; 1.86; 2.65] (p>0.05) L6: [1.06; 1.30; 1.63] (p>0.05)
Arat et al. (2003)	Agreement of the three methods in the displacement in mm of specific landmarks (N, PT, S, Ba) from T1 to T2	1. Intrarater reliability on the main outcome measurements (ICC) 2. Measured displacement at Pg by each method	Horizontal (mm) [Method A; Method B; Method C] N. [1.21; 1.99; 1.33], PT. [-0.10; 0.66; 0], S. [-0.52; 0; -0.45], Ba. [-1.52; -0.66; -1.66] Vertical (mm) [Method A; Method B; Method C] N. [0.86; 0, 0], PT. [0.53; 0.19; 0], S. [0.20; 0; -0.94], Ba. [1.22; 1.43; 0] All landmarks changed their position with growth	1. ICC: 0.93–0.99 2. Horizontal (mm) Pg. Method A: 1.55, Method B: 3.53, Method C: 0.41 Vertical (mm) Pg. Method A: 4.56, Method B: 5.89, Method C: 5.81
Goel et al. (2004)	Reproducibility of 2 superimpositions methods measured through the displacement of the landmarks A, U1, L1, B, Pg along the occlusal plane	Agreement between the two superimposition methods	No significant difference was identified between mean results of repeated measurements with each method at each point (p>0.05)	A: T-method: 0.02; Viazis: 0.50 (P>0.05) U1: T-method: 0.56; Viazis: 1.40 (P>0.05) L1: T-method: 0.44; Viazis: 1.25 (P>0.05) B: T-method: 0.29; Viazis: 1.08 (P>0.05) Pg: T-method: 0.65; Viazis: 1.50 (P>0.05)
Gliddon et al. (2006)	Reproducibility of the four methods (intrarater and interrater) measured as horizontal and vertical displacement of the following reference landmarks: ANS, Point A, Point B, and Pogonion	1. Operator with vs. operator without experience. 2. Differences in reproducibility between the four reference landmarks	*Reproducibility of the different superimposition methods from highest to lowest* Manual geometric method (examiner 1 range: 0.008–0.107 mm; examiner 2 range: 0.036–0.150 mm), LS-5 method, S-N method, F-H method (p<0.05)	1. Reproducibility level differed between the 2 examiners (P<0.001). Reproducibility was consistently higher for the experienced examiner. 2. Reproducibility on the 4 reference landmarks, for both examiners, from highest to lowest: ANS, Point A, Point B, and Pogonion (P<0.001)

Roden-Johnson et al. (2008)	Mean difference between computer and hand superimpositions in the assessment of structural changes between T1 and T2	NA	ACB, Vertical movement of point N (T1-T2) Hand: 0.5 mm, Quick ceph: 0.8 mm (p=0.029) No other significant differences were observed in 42 other variables tested	NA
Gu et al. (2008)	Mean difference of mandibular and maxillary superimposition methods measured through landmark displacement (mm) according to the functional occlusal plane (horizontal – x) and the pterygomaxillary fissure (vertical – y)	NA	*Maxillary landmarks (mm)* Implant superimposition: A: x: 0.6±1.1, y: -3.9±1.7, ANS: x: 0±0.8, y: -3.5±1.8, PNS: x: -5.6±2.2, y: -3.9±1.9. ABO (Maxilla): A: x: 1.7±2, y: -1±2.6 (p<0.05), ANS: x: 1.5±1.7 (p<0.05), y: -0.4±2.5 (p<0.01), PNS: x: -4.1±1.6, y: 0±1.2 (p<0.001). *Mandibular remodeling (mm)* Implant: Condylion: 19.7±3.1, Superior condylion 17.1±2.6, Posterior condylion: 6.6±1.9, Posterior border 5.0±1.7, Antegonial region -2.7±1.7, Menton: 1.4±1.1, Pogonion 0.3±0.3. Mandibular Lower border: 18.4±4, Superior condylion 14.2±3.4, Posterior condylion: 7.8±2.5, Posterior border 6.0±1.9, Antegonial region -1.4±0.7, Menton: -0.2±0.4 (p<0.05), Pogonion: 2.3±1 (p<0.05) ABO (Mandible): 19.2±4.3, Superior condylion 15.2±3.5, Posterior condylion: 7.8±1.8, Posterior border 6.0±1.4, Antegonial region -1.8±2.4, Menton: 1.0±0.9, Pogonion 0.7±1.1	NA
Standerwick et al. (2008)	Agreement of the two methods measured through the displacement of the landmarks between tracings in degrees	NA	Agreement between measurements using I and S as reference: Cond. horiz.: -7.88±9.63°, Cond. perp.: 22.27±33.53°, IAC horiz. -1.79±2.06°, IAC perp. 5.96±5.02°, IS horiz. -1.05±1.24°, IS perp. 4.38±13.26°	NA
Huja et al. (2009)	Mean difference between the best-fit cranial base superimposition and S-N superimpositions using the digital method measured through corresponding T2 landmark distances (mm)	Difference between repeated and between hand and digital superimpositions on the cranial base and S-N measured through corresponding T2 landmark distances (T2 LD)	*Descriptive statistics for T2 LD (mm) for Best-fit vs S-N superimposition* PNS: 0.87mm±0.62, ANS: 1.17mm±1.01 A point: 1.19mm±1.02, B point: 1.53mm±1.33 Pogonion: 1.65mm±1.49, Gonion: 1.30mm±0.98 Gnathion: 1.71mm±1.54, Menton: 1.64mm±1.54	The Friedman test indicated no statistically significant differences for the reproducibility of the hand and digital methods and between superimposition methods (P > .05) for any of the landmarks tested. Mean differences were smaller than 1 mm
Standerwick et al. (2009)	Mean distance between corresponding landmarks (x-axis and y-axis) on serial	Repeatability of each superimposition method (ICC)	Significant differences were identified between measurements using I point and S-ACB as reference. For the ages 8 to 18, sagittal tracing superimpositions displayed an average 7 mm greater cephalad movement of landmarks and 2.4 mm greater	Average ICC values for I-point were nearly identical to SACB; 0.87/ 0.87 for the x-axis and 0.85/0.86 for the y-axis

	tracings superimposed at I-point and S-ACB		ventral movement when superimposition was referenced at I-point as compared to the S-ACB reference	
Arat et al. (2010)	Agreement of the positional changes of ACB landmarks (Nasion, wing, tuberculum sella, sella, basion & pterygomaxillare), detected by 4 ACB superimposition methods. Björk structural superimposition method was used as gold standard	1. Stability of cranial landmarks according to Björk's method during puberal, postpubertal and overall periods. 2. Repeatability of all procedures in 10 patients	As a mean, the T-W method was the most similar to Björk's structural method both in horizontal and vertical directions during all periods	1. Stability of the landmarks in all tested periods with decreasing order: Tuberculum sella (T1-T3, horizontal: 0.13±0.83; vertical: -0.12±0.76) > Wings > Pterygomaxillare > Sella > Nasion > Basion 2. Repeatability of all procedures, Cronbach's alpha: 0.942–0.999
Türköz et al. (2011)	Mean difference of Björk's and Steiner's superimposition methods regarding positional changes of the following landmarks: N, Or, ANS, PNS, A, B, Po, Gn, Me, Go, Co, maxillary rotation and mandibular rotation in extraction and non-extraction cases	Reproducibility level achieved by one operator	No significant difference was found between the groups or methods	The calculated reliability coefficients for the examiner reliability ranged between 0.56 and 1.00
Lenza et al. (2015)	Agreement of the four methods measured as horizontal and vertical displacement (mm) of the following 7 landmarks: ANS, PNS, Gnathion, Gonion, Pogonion, Point A, Point B	NA	No significant difference was identified between mean results of measurements obtained by each method, at each point (p<0.05).	NA
Jabbal et al. (2016)	Agreement of the three commonly used mandibular superimposition methods with Björk's method in the degrees of incisor inclination change induced by orthodontic treatment (ICC)	Agreement between conventional methods	*Mean changes in the lower incisor inclination* Me-Go: 0.31°±7.73, Go-Gn: 0.26°±7.73 Tangent: 0.64°±7.91, Björk: 0.51°±7.69 *ICC values:* Me-Go to Björk: 0.96 Go-Gn to Björk: 0.94 Tangent to Björk: 0.92	*ICC between conventional methods* Me-Go to Go-Gn: 0.98 Me-Go to Tangent: 0.96 Go-Gn to Tangent: 0.97
Jiang et al. (2020)	Mean of T2 landmark distances of paired automated and hand superimposed T1-T2 cephalometric pairs by three different operators	Differences in hand superimposition among the operators	There were no significant differences in interoperator error between hand and automated superimposition (p>0.05)	The T2 landmark differences in hand tracing between the operators ranged from 0.61 mm to 1.65 mm for the three types of superimpositions. ACB: 0.61–1.02 mm

				Maxilla: 0.78–0.82 mm Mandible: 0.62–1.65 mm

ACB: Anterior cranial base, ICC: Intraclass correlation coefficient, ANS: Anterior nasal spine, PNS: Posterior nasal spine, T(x): Timepoint, S: Sella, N: Nasion, NA: Not applicable, MAX: Maxilla, IMP: Implant, Pg: Pogonion, Rx: Radiograph

Supplementary Text 1. Detailed reporting of the results of individual studies.

Superimposition on the cranial base

Seventeen studies superimposed the cephalograms on the cranial base (3, 5, 15, 17, 18, 20, 22–25, 27, 31–36).

Baumrind et al. tested the precision of tracing superimposition, following manual best fit of SN, registered on sella, or of the anterior cranial base (ACB) structures (31). The errors attributed only to the superimposition procedure (no landmark identification error) were relatively large, when considering individual case assessment. Furthermore, different superimpositions provided different outcomes. The main limitation of the study is that it used only good quality radiographs and there was inadequate individual error assessment.

Houston and Lee tested the precision of five ACB manual best fit superimposition methods (directly, blink method, subtraction method, tracing of cranial base, and tracing on S-Na line at sella) and concluded that all superimposition methods have relatively large errors (22). The main limitations of this study are that it did not assess an actual superimposition outcome and that individual differences were not tested.

Buschang et al. tested the reproducibility of Bjork's ACB superimposition (manual best fit) in growing patients (25). The vertical and anteroposterior position of the ACB was assessed following repeated superimpositions. The reproducibility for the ACB ranged from 0.17 to 0.39 mm, which was considered clinically acceptable, and it did not improve by experience. Also, in this study no actual superimposition outcome was tested and group mean values were only considered.

Ghafari et al. tested 5 different methods of superimposition on CB. The best fit superimposition method on ACB, superimposition on SN (S fixed), superimposition on R (Bolton-Nasion plane parallel) and superimposition on Ba-N at CC (Intersection Ba-N and central axis) and on Ba-N at N. The different methods showed varying results on the measured outcomes that can be considered clinically significant (27). The main limitation of the study is that it did not test individual differences and there was no reproducibility assessment.

Iseri et al. tested the reproducibility of Bjork's ACB superimposition method, supplemented by a computerized procedure, on measuring maxilla displacement during growth (32). No significant difference was identified between mean outcomes. The main limitation of the study is that it did not test individual differences and the anatomical stability of the points used as superimposition reference was not verified.

You et al. tested the precision of the Pancherz's CB superimposition method and found it reliable (15). This result should be treated with caution due to the small sample (inadequate statistical power) and the absence of the assessment of individual differences.

Arat et al. compared the displacement of CB and facial landmarks during growth, detected by Steiner's (SN), Ricketts's (NBa) and Björk's (ACB) structural superimposition method (17). Landmarks used for Steiner and Ricketts methods exhibited significant vertical and horizontal displacements according to Björk's structural method. The three methods measured differently changes at Pogonion. This study did not use a gold standard and mean values were only considered.

Goel et al. tested the agreement and reproducibility of T and Viazis cephalometric superimposition methods and concluded that the two methods may show good reproducibility and agreement (24). However, the study has important limitations, such as the assessment of mean values only and the small sample.

Gliddon et al. tested the precision of FH, S-N, LS-5 and geometric method for CB superimposition (23). The reproducibility of the different superimposition methods from the highest to the lowest was: the geometric method, the LS-5 method, the S-N method, and the F-H method ($p<0.05$) and it was consistently higher for the experienced examiner. An important limitation of this study was that identical images were superimposed (no actual clinical data).

Roden-Johnson et al. tested the agreement between hand and digital (Quick ceph) superimposition of the ACB (20). A small statistically significant difference was detected only for Nasion. There might be good agreement between the hand and digital (Quick ceph) superimposition techniques, but this study also considered only differences in mean values.

Standerwick et al. evaluated the agreement of two cephalometric superimposition methods with growth patterns defined by Melsen´s necropsy specimens and the Bjork´s implant studies (33). The authors claimed that I-point shows better agreement to growth patterns than traditional ACB superimposition. However, the study had several limitations, such as the inadequate error evaluation and the absence of assessment of the anatomical stability of I point, which was used as superimposition reference.

Huja et al. evaluated the mean difference between hand and digital superimpositions, as well as between the best-fit CB and S-N superimpositions, using the digital method (34). It was concluded that there might be only small differences between best-fit CB and S-N superimpositions. The limitation of this study is that only good quality radiographs were used, and only mean values were compared.

Standerwick et al. compared the Sella-ACB superimposition with a superimposition referenced at the occipital condyle (I-point) in the demonstration of craniofacial growth and development (35). It was concluded that the S-ACB and I-point superimpositions provide different outcomes. This study also presented only a comparison of mean values.

Arat et al. compared the displacement of cranial landmarks during growth and adulthood through Bjork's structural, Ricketts, Steiner and T-W superimposition methods (18). The study found that the T-W method shows the highest agreement with the Bjork's structural method, which was used as gold standard. An important limitation of this study was that it did not test individual differences.

Lenza et al. tested the agreement of four methods in the displacement of specific landmarks (5). There was no significant difference between mean outcomes obtained by each method, at each point ($p < 0.05$). There might be good agreement between the four methods, but the study had important limitations, such as the assessment of only mean values, and the use of high-quality radiographs.

Türköz et al. tested if there is a difference in cephalometric superimposition outcomes between Björk's ACB structural method and Steiner's method of Sella–Nasion line registered at Sella (36). The study concluded that there are no significant differences between the outcomes of Björk's and Steiner's ACB superimposition. However, there was no testing of individual differences and no adequate reproducibility testing.

Jiang et al. tested the interoperator error of a computer-aided automated method for structural cephalometric superimposition compared with Johnston's free-hand tracing superimposition method (3). It was concluded that the computer-aided superimposition provides comparable interoperator error to that of the traditional hand tracing superimposition. The limitations were that there was no testing of individual differences, no gold standard, and no method error.

Superimposition on the maxilla

Baumrind et al. evaluated the precision of tracing superimposition on the palatal plane and concluded that the mean tracing error, without any landmark identification error, was approximately 1 mm (31). The study assessed individual differences, but it did not report the relevant outcomes adequately.

Baumrind et al. tested the differences between the structural cephalometric superimposition on anatomical structures and that on metallic implants (6). The two methods showed differences in the detected displacement of maxillary ANS, PNS and A point. This study also considered only differences in mean values.

Nielsen et al. compared three different methods and found that the implant method and the structural method showed considerable differences with the palatal plane best fit method (16). The implant method showed limited mean differences with the structural method and the structural method was found to show adequate reproducibility. However, the study did not test individual differences, there was no method error evaluation apart from the structural method, and in some cases only unilateral implants were used. Finally, only good quality radiographs with clearly visible zygomatic processes and no double contours were considered, which is a very rare occasion in actual conditions.

Cook et al. evaluated the mean difference of the palatal best fit superimposition and Ricketts Position 3 superimposition on maxillary tooth positional changes (19). They concluded that the two methods differ significantly in the assessment of changes in the upper central incisors. The limitation of this study was that it did not test the differences between methods in individual cases, but only compared mean values.

Baumrind et al. evaluated the displacement of the teeth using superimposition on maxillary implants and on the best fit of maxillary anatomical structures (37). The methods showed different results in maxillary tooth displacement. The highest differences were evident in the vertical displacement of the maxillary first molar. The best fit method tended to undervalue the vertical growth. Similar to other studies, this one also considered only mean values and did not assess method error.

You et al. tested the precision of Björk's method for the maxilla and concluded that the method was reliable for the assessment of skeletal and dental changes (15). This result should be treated with caution due to the small sample (inadequate statistical power) and the absence of individual difference assessment.

Roden-Johnson et al. tested the agreement between hand and digital Bjork's maxillary superimposition (20). They found good agreement between them, but the study had severe limitations.

Gu et al. tested if there is a difference between the ABO superimposition (lingual curvature of the palate plus internal bony structures of the maxilla) and superimposition on metallic implants in the detected displacements of anatomical landmarks (38). It concluded that Bjork's maxillary superimposition seems to overestimate the forward displacement of point A and underestimate the vertical displacement of points A, ANS and PNS. The study assessed only mean differences on a small sample.

Huja et al. evaluated the difference between hand and digital maxillary regional best-fit superimpositions, performed on the lingual curvature of the palate plus the internal bony structures of the maxilla (34). The study concluded that there were no differences between maxillary superimpositions produced by Dolphin Imaging and those completed by hand. However, only good quality radiographs were used and there was no assessment of individual differences.

Jiang et al. tested the interoperator error of a computer-aided automated method for cephalometric superimposition on ANS, PNS, A, and Pt points compared with Johnston's structural free-hand tracing superimposition (3). The study concluded that computer-aided cephalometric superimposition provided comparable interoperator error to that of traditional hand tracing structural superimposition. However, the study had severe limitations.

Superimposition on the mandible

Baumrind et al. evaluated the precision of tracing superimposition on the mandibular border and found adequate reproducibility, with small mean differences between repeated measurements (approximately 0.4 mm) (31). However, the study used only good quality radiographs and did not report properly on individual differences.

Buschang et al. tested the reproducibility of Bjork's mandibular superimposition (25). The average method error of mandibular reference line points ranged from 0.45 to 0.93 mm and was found to improve by experience. The study had important limitations.

Cook et al. assessed the reproducibility of Bjork's mandibular superimposition (39). Horizontal errors were found to be much less than the vertical. Midline structures were more reliable than bilateral ones and the lower molar tooth germ was more reliable than the inferior nerve canal. The detected precision was considered clinically adequate. However, this study did not assess an actual superimposition outcome; only one time-point was used. Furthermore, only differences in mean values were considered.

Cook et al. evaluated the reproducibility of cephalometric tracing superimposition on the mandibular plane, mandibular outline and Bjork's mandibular superimposition (26). All three techniques showed sizeable errors and Bjork's method showed to be the least reproducible of the three. Mean values were only considered in this study and there was no blinding (the registration points of the first superimposition were present when the second was performed).

Cook et al. tested the reproducibility of Ricketts' position 4 and Bjork's mandibular superimposition methods and concluded that the two methods provided similar outcomes (19). The main limitation of this study was that it did not test the difference between methods in individual cases.

Springate and Jones tested the agreement of Bjork's and Ricketts's superimposition methods with an implant superimposition method in the mandible (21). The results indicated high trueness for the structural method, whereas low for the Rickett's method, assuming the implant method as gold standard. However, no solid conclusion can be drawn, since only mean values were assessed, and only good quality radiographs were included. Furthermore, the results regarding precision (reproducibility) were not reported for each method separately.

You et al. tested the reproducibility and the agreement of Bjork's structural and Ricketts Position 4 superimposition methods for the mandible (15). No significant differences were identified between repeated measurements or between different methods. This result should be treated with caution due to the small sample (inadequate statistical power) and the absence of individual difference assessment.

Roden-Johnson et al. tested the agreement between Bjork's hand and digital (Quick ceph) superimposition of the mandible (20). It was found that there is good agreement between the hand and the digital (Quick ceph) superimposition techniques. Individual differences were also not tested in this study.

Gu et al. tested the differences between the ABO superimposition (internal cortical outline of the symphysis and inferior alveolar nerve canal), the superimposition on the inferior border of the mandible and the superimposition on metallic implants in the detected displacements of anatomical landmarks (38). It was found that the ABO method shows good agreement with the implant superimposition method. However, the study had small sample (inadequate statistical power) and it did not assess individual differences.

Huja et al. compared the hand-traced and computer-based superimpositions on the inner contour of the mandibular symphysis, the mandibular canal, and the apical portion of unerupted third molar (34). No differences were detected between Dolphin Imaging and hand superimposition, although the study considered only mean values and good quality radiographs.

Jabbal et al. evaluated if there is a difference between the change in incisor inclination between Björk's mandibular, Me-Go, Go-Gn, and tangent to the lower border of the mandible superimpositions (4). All compared methods showed a mean change in incisor inclination below 1 degree and good agreement to each other. However, the study did not test the differences between methods in individual cases.

Jiang et al. tested the interoperator error of a computer-aided automated superimposition method on points LM, Pg, Me and Go compared with Johnston's free-hand mandibular tracing superimposition (3). The two methods showed similar amount of error ranging from 0.6 to 1.7 mm. This study also refers only to differences in mean values.

Superimposition on the cranial base compared to the maxilla

Baumrind et al. evaluated maxillary teeth displacement following superimposition on ACB, on maxillary implants and on the best fit of maxillary anatomical structures (37). The methods showed different outcomes in the vertical and horizontal dimension. The best fit method tended to undervalue the vertical growth. Similar to other studies, this one also considered only mean values and did not assess method reproducibility.

Efstratiadis et al. tested differences of the cranial versus the maxillary regional superimposition on measuring mandibular displacement (40). It was found that in growing patients the cranial and maxillary superimpositions led to different results, especially on the vertical dimension. Methodological drawbacks were the lack of blinding in measurements and method reproducibility assessment and that only mean values were tested.

You et al. tested the precision of Bjork's maxillary, Ricketts Position 3 maxillary and Pancherz's CB superimposition methods and their agreement (15). No significant differences were identified between the different methods or between repeated measurements with each method. However, the study sample was small and only differences in mean values were tested.

Superimposition on the cranial base compared to the mandible

You et al. tested the precision of Björk's and Rickett's mandibular methods, as well as of Pancherz's CB superimposition method, to assess mandibular changes (15). The result was that all methods showed good reproducibility and agreement to each other. However, this study presented shortcomings, such as no individual difference assessment and had a high risk of bias. These results should be treated with caution.