Assessment reliability of infraposition of implant supported maxillary incisors

A study using three different methods.

Lovisa Andersson
Karin Loqvist

Supervisor:
Lars-Åke Johansson
Christina Lindh

Master Thesis in Odontology (30 ECTS)
Program of dentistry
February 2019

Malmö University
Faculty of Odontology
205 06 Malmö
Tillförlitlighet vid bedömning av infraposition hos implantatstödda överkäksincisiver

En metodstudie

Lovisa Andersson
Karin Loqvist

Handledare:
Lars-Åke Johansson
Christina Lindh

Examensarbete (30 hp)
Tandläkarprogrammet
Februari, 2019

Malmö universitet
Odontologiska fakulteten
205 06 Malmö
Content

Content ....................................................................................................................................... 1
Abstract ...................................................................................................................................... 3
Sammanfattning ......................................................................................................................... 3
Introduction ................................................................................................................................ 4
Infraposition ............................................................................................................................ 4
Craniofacial growth ................................................................................................................ 5
Adult growth pattern ............................................................................................................... 6
Implants and osseointegration ................................................................................................. 6
Recommendations for dental implant insertion ...................................................................... 6
Earlier described methods to evaluate infraposition ............................................................... 7
Radiological evaluation ............................................................................................................. 7
Aesthetic indices ..................................................................................................................... 7
Intraoral scanning .................................................................................................................... 9
Relevance of the study .......................................................................................................... 10
Aim of study ......................................................................................................................... 10
Hypothesis .............................................................................................................................. 10
Material and methods ............................................................................................................... 11
Subjects .............................................................................................................................. 11
Raters ................................................................................................................................ 11
Evaluation .............................................................................................................................. 12
Method 1: Radiographic method ...................................................................................... 12
Method 2: Aesthetic index ................................................................................................ 13
Method 3: A new method based on clinical photographs and study cast ......................... 13
Calculations ....................................................................................................................... 14
Statistics ............................................................................................................................... 14
Ethical consideration ........................................................................................................... 15
Results ...................................................................................................................................... 15
Method 1: Radiographic method ...................................................................................... 15
Method 2: Aesthetic index ................................................................................................ 17
Method 3: New method based on clinical photographs and study casts ....................... 18
Limitations of the study ......................................................................................................... 23
**Abstract**

**Aim:** To evaluate reliability of two previously described methods and one newly developed method in assessment of infraposition of implant supported maxillary incisors.

**Material and method:** A retrospective material consisting of periapical radiographs, study casts and intraoral photographs from baseline and follow-up from 10 patients with 11 implant-supported crowns in maxillary incisor position, was used. Six raters: one oral radiologist, one prosthodontist/periodontist and four dental students, independently assessed infraposition with three methods (radiographic method, aesthetic index and a new method based on clinical photographs and study casts). Reevaluation was performed after at least two weeks. Inter-and intrarater agreement was calculated with Cohen’s kappa, weighted kappa, IRR and ICC.

**Results:** The highest intrarater agreement was achieved using the new method (ICC<sub>min-max</sub> 0.760-0.910) and the least with the radiographic method (ICC<sub>min-max</sub> 0.073-0.830). Highest interrater agreement was achieved using the new method (ICC 0.67) and the least with the aesthetic index (weighted kappa 0.234). According to our findings specialists had higher agreement than students in the radiographic method whereas students had higher agreement in the aesthetic index. In the new method this comparison was not possible due to missing data.

**Conclusions:** Our findings indicate a low reliability for the two previously described methods. However, assessments with the new method are reproducible which imply a very good reliability. Therefore the new method may be suitable for evaluating infraposition of implant supported maxillary incisors. Whether observed infraposition is close to the truth cannot be evaluated until a reference method has been proposed.

**Sammanfattning**

**Syfte:** Att utvärdera reliabiliteten för bedömning av infraposition av implantatförankrade incisiver i överkäken för två tidigare beskrivna metoder och en ny metod.

**Material och metod:** I studien ingick 11 implantatstödda singelkronor i överkäkens framtandsområde från 10 patienter. Studiematerialet omfattade studiemodeller, intraorala kliniska fotografier periapikala röntgenbilder från baseline och uppföljning. En radiologisk metod, ett estetiskt index och en ny metod som använde kliniska fotografier och studiemodeller, användes för att utvärdera infraposition. Sex observatörer, en oral radiolog, en parodontolog/protektiker och fyra tandläkarstudenter utförde bedömningar med samtliga metoder vid två tillfällen. Inter- och intraobservatörs-överensstämmelsen analyserades med hjälp av Cohen’s kappa, viktat kappa, IRR och ICC.

**Resultat:** Högst intra-observatörsöverensstämmelse erhölls då nya metoden användes (ICC<sub>min-max</sub> 0,760-0,910) och lägst med radiologiska metoden (ICC<sub>min-max</sub> 0,073-0,830). Interrobservatörsöverensstämmelsen var högst då den den nya metoden användes (ICC 0,67) och lägst avseende estetiskt index (viktat kappa 0,234). Utifrån vårt resultat hade specialisterna högre överensstämmelse med varandra jämfört med studenterna för den radiologiska metoden medan studenterna hade högre överensstämmelse med det estetiska indexet. För den nya metoden var jämförelse mellan observatörsgrupperna ej möjlig pga att data saknades.

**Slutsats:** Vårt resultat indikerar att de tidigare beskrivna metoderna har låg reliabilitet, framförallt mellan observatörerna. Dock visar den nya metoden hög reproducerbarhet och därmed är reliabiliteten mycket god. Därför kan nya metoden passa för att mäta graden av infraposition hos implantatförankrade maxillära incisiver. Huruvida uppmätt infraposition är sannolik kan inte säkerställas med mindre än att en referensmetod finns att tillgå.
Introduction

Since Brånemark 1981 introduced osseointegration as a direct and lasting connection between vital bone and screw-shaped titanium implants, dental implant treatment has become a common choice in dental therapies (1). In as many as 20 percent of dental implant treatments, aesthetics is a primary motivation for the patient, especially in the anterior maxilla (2). Survival has been reported to be high in implant-supported prosthetics but many factors affect a successful treatment outcome (3,5). One example is infraposition ascribed to continuous eruption of adjacent teeth (4,5). Due to self-perceived unsatisfied aesthetics, infraposition is an indication for replacement of the implant-supported prosthetic restoration. Many studies have tried to clarify why some patients are at risk of infraposition. To be able to answer this question, to diagnose correctly and to compare the infraposition as a function of time, an accurate measurement of the degree of infraposition is needed. Methods to evaluate infraposition are described in the literature, but are described methods reliable? Is there a better method to evaluate infraposition? By finding a valid and reliable method, correlations between patients and the degree of infraposition should be possible. Such a method could also be useful in the future in identifying patients at risk for infraposition.

Infraposition

Infraposition of one or several teeth can be found in individuals with ankylosed teeth but also in patients treated with dental implants. Infraposition is synonymous with infraocclusion that is described as one or more teeth that fail to extend to the occlusion plane (6). Symmetry and harmony are central parameters in the aesthetic zone and that includes symmetric position of incisal edges in relation to adjacent and lateral teeth (7,8). Since the dental implants behave as an ankylotic tooth it is unable to adjust its position according to craniofacial growth and the eruption processes. These processes were thought to cease when skeletal growth was completed (9,10), and thus assumed to be a problem in growing children and adolescent patients. However, Bernard and co-workers followed both young adult (15.5 to 21 years old) and adult patients (40 to 55 years old) with maxillary incisor and canine implants. They found infraposition of the implant supported crowns ranging from 0.1-1.65 mm in all patients in a young adult group and infraposition ranging from 0.12-1.86 mm in all patients in a mature adult group after a mean of 4 years follow up (10). Similar results have been found in adults by other authors (11, 12). In an 18-year follow up, Andersson et al. found that 40% of the adult patients presented infraposition of at least 0.5 mm on anterior implants (11) compared to 65% of the adult patients and 57% in young adults in a study by Bernard et al (10). Infraposition of implant supported crowns in younger patients (mean age 15 years old, 3 years follow up) was found to be proportional to the residual growth in a previous study (13). The mean changes were 0.46 mm in the anterior positions in young patients (13).

Different factors are suggested to affect infraposition of implant supported maxillary incisors. In a study by Thilander et al (1999), it was found that missing tooth contacts in the anterior region were associated with a higher degree of infraposition in anterior implants. Patients with good incisor stability tended to be less affected by infraposition (14,15). Missing tooth contact due to implants failing to reach the occlusal plane, allowed vertical eruption and/or lingual displacement of the opposing tooth (9).

There are contradicting results of the rate of infraposition of implants in males and females. Bernard et al (2004) found no difference in tooth movements between the sexes (10). On the
other hand, a study by Andersson et al found that the presence of infraposition is higher amongst females. The same study showed weak association between infraposition and “long faces” (11). The term “long-face” is correlated to increased vertical dimensions of the face due to adult anterior face growth. It is seen in both sexes but mostly in women (16). The same association between infraposition and “long-face” is described by Jemt et al 2007 (15). To be able to understand the processes leading to infraposition in implant-supported prosthetics it is important to understand the craniofacial growth processes.

Craniofacial growth

Most individuals follow a similar growth rate pattern. Different phases in this pattern shows individual timing and each phase begin and terminate different among every individual. Growth rate is high in infancy but reaches a minimum during prepuberty and increase again during puberty, called pubertal growth spurt. This pubertal growth spurt varies between sexes as well as among individuals with different duration (17). Termination of the pubertal growth spurt are not the termination of craniofacial growth as it continues even in adults (16, 18, 19).

Postnatal craniofacial growth depends on bone remodelling and bone displacement together with the rotation of the jaws. These processes characterize the development of the face. Bone remodelling is a mechanism of deposition and resorption of bone leading to possible change in size and shape of the maxillary and mandibular bones. Bone displacement is the movement of bone due to expansion of the bone or bones in adjacent areas leading to relocation (20).

The rotation affects the eruption path of the teeth and thus influences the occlusion (21). Both jaws have the ability to rotate either anteriorly or posteriorly, which results in different facial types and various post-eruptive patterns of the dentition (20,21). An anterior rotation is most common in both jaws. Maxillary rotation is characterized by remodeling in anterior or posterior maxilla that results in anterior or posterior rotation (21). The transversal rotation is more pronounced in the maxilla than in the mandible, due to apposition at the suture (22). The mandible exhibits more vertical rotation than the maxilla. The mandibular rotation and displacement are influenced by midfacial displacement, rotation and soft tissue growth (22). The displacement of the mandible due to developmental growth can be caused by translation only or a combination of translation and rotation. Translation alone occur when the displacement follows an equal lowering of the posterior and anterior part of the mandible. Mandibular rotation occurs as a consequence of an uneven lowering of the mandible (20). The rotation is also correlated with the intensiveness of the condylar growth.

Post-eruptive movement of the dentition will compensate for the rotation. This mechanism to maintain interarch relations and adapt to changed relationship between the jaws is called the dentoalveolar compensatory mechanism. This mechanism is important during adolescents and continues in adults. Factors involved in the dentoalveolar compensatory mechanism are the eruption of teeth, forces exerted by soft tissues on the dental- and alveolar arches and the effect on tooth position by neighboring teeth and inclination plane of antagonizing teeth (20). The alveolar ridge is continuously remodeled throughout life and its height and volume is locally maintained by the stimulation of a tooth and its periodontal ligament. When a tooth is absent, ankylosed or replaced by an implant, the stimulation is somewhat altered due to the lack of PDL and changed internal forces. Thus, theoretically, the implant will not be able to adapt with eruption, nor with alveolar compensation, to changed inter-arch or interdental relationships. The dentoalveolar compensatory mechanism will then be dysfunctional.
Adult growth pattern

The fact that the adult face undergoes dimensional changes was stated by Forsberg, 1979 (18). Studies (16,19) show that the craniofacial growth continues even after the second decade. The growth rate seems to decrease with time, which means that changes are minor with increased age (16, 19). In men, the horizontal changes are greater than the vertical, whereas in women the vertical growth dominates (19,23). Mean changes in male and female faces from late teens to the early forties show comparable changes in anterior face height between the sexes whereas an increased posterior face height is seen in males (24).

In a 20-year follow-up study, it was shown that the anterior facial height increased by 1.6 mm on average with an approximate 1 mm contribution from the eruption of the maxillary incisors (25). The mandibular growth rate is higher for adult men. The rate declines after the third decade, which could indicate that changes due to growth do not continue throughout lifetime. The mandibular inclination increases in women during the beginning of the third and the end of the fourth decade, which results in posterior rotation of the mandible. Thus the pattern of vertical growth differs between the sexes due to the posterior rotation of the mandible in women resulting in an increased anterior height and a decreased posterior height in contrast to men that display a stable mandibular inclination (16).

Implants and osseointegration

The use of dental implants to support prosthetic restorations provides aesthetical benefits and preservation of dental tissues by avoiding preparation of intact teeth (13). In 2011, total 63 856 dental implants were inserted in 24 373 Swedish patients, 49 % of these patients were over 65 years according to the Swedish Health Insurance Office (26). The use of single-tooth implants is most common in the anterior maxilla, where the aesthetic demands are high and limitations of occlusion, space and bone volume are common (27). Indications for implant treatment could be congenitally missing lateral incisors, trauma, decay and congenital and acquired jaw defects in the alveolar process (28).

Healing of the alveolar bone around dental implant is due to osseointegration, a physiological reaction where osteoblasts eventually will form an osteoid in direct contact to the outer porous metal and hence integrate the dental implant surface to the living bone (29). Mineralization occurs within the following months, which will stabilize the dental implant. The difference between the dental implant and a tooth is hence the fixed position in the alveolar bone, a lack of periodontal ligament and hence a disability to respond to normal tooth movements (29,30).

Recommendations for dental implant insertion

There is no fixed chronological age recommended for dental implant insertion according to consensus statements from a conference on oral implants in young patients in Jönköping 1995 (31). Chronological age cannot be used for planning dental implant insertion since chronological age does not correlate to skeletal and dental maturation (32). The biological age is on the other hand a common factor for deciding appropriate time for insertion of osseointegrated implants. The general recommendation to avoid infraposition is to wait with insertion until the dentition is fully erupted and terminated skeletal growth accomplished (31).
There are different methods to evaluate skeletal maturation. An increase in body height not more than 0.5 cm/year indicates a nearly terminated growth and thus skeletal maturation. The dental implant treatment should wait for at least 2-year, with height increase not surpassing 0.5 cm/year. This is to ensure growth termination, measured by following the growth curve (31). By using hand-wrist radiographs it is also possible to estimate skeletal maturation (31, 33). These methods are according to consensus statements from 1995 the way to ensure skeletal maturation (31). The additional radiation received from hand-wrist radiographs together with its limited capacity of predicting further craniofacial growth makes this method questionable (34). Therefore, stature is a useful way to evaluate maturation without side effects received from radiographs (35).

The biological ankylosis process associated with dental implant installation will arrest both dental eruption and alveolar bone formation around the dental implant. This process might also negatively affect neighboring tooth germs. This can lead to locally disturbed tooth formation and growth patterns in the jaw and alveolar bone. In some cases, the dental implant becomes buried in the alveolar bone and the apical point might even become exposed as the nasal and antral floor is remodeled in the maxilla. (28)

**Earlier described methods to evaluate infraposition**

**Radiological evaluation**

The vertical relation of the dental implant to an adjacent tooth and vertical bone height was assessed with a radiological method by Thilander et al (1999) and Jemt (2008) (4,14). The method was also used and modified by Andersson et al (2013) (11). Andersson measured the distance between fixture-abutment junction (FAJ) and cement-enamel junction (CEJ) at the adjacent mesial tooth in periapical radiographs. If the mesial tooth was missing or reconstructed, CEJ at the distal tooth was used. At follow up a measurement of the vertical position of adjacent tooth (CEJ) was compared to the same reference point at the implant FAJ (11).

Andersson and colleagues found that observed infraposition by this method, after 15-19 years of function, were usually less than 1 mm. The study found no clear relationship of the measurement between the FAJ and the CEJ to the degree of infraposition (11).

**Aesthetic indices**

In a successful implant reconstruction an optimum aesthetic outcome with subjective and objective pleasing prosthetics and healthy, well adapted soft tissues are important factors. Symmetry and harmony are central parameters in the aesthetic zone that includes symmetric position of incisal edges in relation to adjacent and lateral teeth (7,8,10). Different esthetic indices have been proposed for assessing the outcome of implant supported crowns in the aesthetic zone. Evaluation is mainly performed by various qualitative assessments of peri-implant mucosa and crown parameters. There is yet no consensus on a single standardized method and the reproducibility of existing methods has been questioned (36-39).

Of the aesthetic indices found in the literature there are two recurrent methods used, and often combined, in studies: the ‘Pink Esthetic Score’ (PES) (40) evaluating the mucosa surrounding the implant crown such as the papilla, the colour and the contour of the labial surface; and the
‘White Esthetic score’ (WES) (41). WES describes the form, texture and outline of the implant-supported crown. None of these indexes evaluated infraposition as a single parameter but might include infraposition in the criteria: ‘outline/volume’ (41).

Later studies have suggested improved methods to assess the aesthetics of implant crowns and mucosa. Meijer et al. (2005) developed the ‘Implant Crown Aesthetic Index’, ICAI, with the complement of five implant crown related parameters to PES/WES. One of the parameters were “the position of the incisal edge of the crown: the position must be in harmony with the adjacent and collateral tooth on a five-point rating scale”. In their pilot study, they found good intra-observer agreement and moderate interobserver agreement. (36)

In a study by Tettamanti et al (2016), PES/WES and ICAI was evaluated and compared to a new index “Peri-Implant and Crown Index” (PICI) (39). The new index, PICI, evaluated three pink, three white and tree overall subjective esthetic characteristics in comparison with the appearance of the contralateral tooth, see table 1. For the three subjective overall categories, the observer rated the implant on a 100 mm line ranging from “not aesthetic at all” to “very aesthetic”. They found that PES/WES and PICI had a higher intra-observer agreement than ICAI. ICAI also showed the lowest clinical acceptance when four groups of raters with different background assessed cases with the included indices (39).

Table 1. Criteria of three esthetic indexes: PICI, ICAI and PES/WES. From Tettamanti et al 2016 (39).

<table>
<thead>
<tr>
<th>Criteria of the peri-implant mucosa (pink esthetic)</th>
<th>PICI (New Index)</th>
<th>ICAI</th>
<th>PES/WES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papillae</td>
<td>Labial margin</td>
<td>Mesial papilla</td>
<td></td>
</tr>
<tr>
<td>Zenith</td>
<td>Papillae</td>
<td>Distal papilla</td>
<td></td>
</tr>
<tr>
<td>Root Convexity</td>
<td>Contour of the labial surface</td>
<td>Facial curvature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Colour and surface</td>
<td>Level of facial mucosa</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Root convexity and color</td>
<td>Tooth form</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Outline/volume</td>
<td></td>
</tr>
<tr>
<td>Criteria of the implant crown (white esthetic)</td>
<td>Shape</td>
<td>Color (hue/value)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Width</td>
<td>surface texture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>translucency and characterisation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Labial convexity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Color/translucency</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective overall criteria</td>
<td>Crown</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mucoса</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Reference tooth</td>
<td>Contralateral tooth</td>
<td>Contralateral tooth</td>
<td></td>
</tr>
<tr>
<td>Scores per criteria</td>
<td>100-mm visual analogue scale</td>
<td>2 (No deviation)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0–45 Points</td>
<td>1 (Small deviation)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6–20 Points</td>
<td>5 (Large deviation)</td>
<td></td>
</tr>
<tr>
<td>Overall score</td>
<td>≥360 points</td>
<td>6 Points = 0%</td>
<td></td>
</tr>
<tr>
<td>Threshold of clinical acceptability</td>
<td>4 Points = 50%</td>
<td>10 Points = 50%</td>
<td></td>
</tr>
<tr>
<td>Calculation to percentage scale</td>
<td>600 points = 100%</td>
<td>20 Points = 100%</td>
<td></td>
</tr>
</tbody>
</table>

At Copenhagen University, the Copenhagen Index Score (CIS) was invented to be used by students. CIS rated crown morphology, crown color, symmetry/harmony, mucosal discoloration and a papilla index. The parameters were rated with a four-point rating scale. The crown morphology score was a modification of the California Dental Association, CDA (1997) index criteria for anatomic form with several sub-parameters, including crown length in relation to neighboring teeth. Other parameters were similar to PES/WES. The validity of the CIS was tested for several observers and found to have a substantial intra-observer
agreement for 2/3 of the assessments and moderate (>50%) taking all assessments into account. In the same study inter-observed agreement was found to be moderate (>50%) in 4/5 of all observations. The crown morphology and symmetry/harmony scores had the lowest agreement, showing significant difficulties in evaluation of these features. The CIS was also compared with a VAS score (visual analog scale 1-10) of evaluated overall aesthetics and found that all CIS scores correlated significantly to the corresponding VAS scores. (38)

In contrast to other studies, Jemt et al (2007) focused on infraposition as a single parameter (15). The position and extent of infraposition of the implant supported crown was evaluated in relation to adjacent teeth, by visually evaluating photographs and then categorizing according to the following score:

- Score A: No obvious clinical vertical implant infraposition was identified in relation to the incisal edges of the adjacent teeth.
- Score B: Small, insignificant, clinical vertical implant infraposition assessed to be less than half a millimeter (< 0.5 mm) was observed in relation to the adjacent teeth.
- Score C: Obvious clinical vertical implant infraposition assessed not to exceed 1 mm (≤ 1.0 mm) in relation to the adjacent teeth was observed.
- Score D: Clinically significant vertical implant infraposition assessed to be more than 1 mm (> 1.0 mm) was present.

This method has been modified and used in several studies (11,14,39,42). In a few studies the score has been modified, using smaller intervals (11,42). To be able to quantify infraposition to a ratio-scale Thilander et al (1999) measured infraposition of the implant compared to adjacent teeth on study casts with a digital caliper. The measurements to the closest 0.1 mm were recorded (14). Tettamanti et al (2009) also included study casts to visually evaluate the same parameters (39).

**Intraoral scanning**

A few studies evaluated infraposition using intraoral scanning. In a review, Benic and colleagues (2012) stressed the importance of stability of reference structures and found that several indices assessing implant aesthetics have questionable reference structures for repeated measurements (43). This was also addressed by Ghislanzoni et al (2017) who used 3D scanned (Trios 3Shape, Copenhagen, DK) plaster models to assess vertical continuous eruption of teeth in a 10-year longitudinal study in 10 adult women. Two different methods were used with different anatomical structures as reference points. In the first method the palate at baseline and after 10 years were manually superimposed in a software program (Viewbox dHAL Software, Kifissia, Greece) using palatal rugae and an area posteriorly extended along the palatal raphe on the hard palate, in a “mushroom shape”, as reference in scanned models. The second method marked the functional occlusal plane, on first molar and premolar cusps, thus being able to compare canine and incisive edge displacement relative to the occlusal plane upon superimposition of markings. The methods showed no statistical difference when evaluating results of vertical migration. Intra-variation was low and the superimposition’s random method error of the vertical displacement was calculated to 0.13 mm. However, the standard deviation in the different measurements was high, indicating a heterogeneous material. (44)

Brahem et al (2017) suggested using three stable reference areas, preferable three single-crown implants, and if not possible, single-crown implants combined with palatal medial and
lateral point of third rugae as reference markers. The extent of vertical and horizontal tooth displacement relative to the incisal edges of the central incisors was categorized with 0.25 mm distances from A (<0.25 mm) to E (≥1 mm), the index modified from Jemt et al (2007). Movements in mesio-distal directions could not be analyzed in the scanned models with this technique and Brahem and co-workers found difficulties in aligning the models when using medial and lateral rugae for superimposition of scanned dental casts (42). In a case study, 3D-scanned plaster models were used to evaluate infraposition and tooth movements. The calculated three-dimensional accuracy was 0.1 to 0.2 mm and the author concluded that "Only small vertical tooth movements could be observed in the upper jaw during 16 years of follow-up when using the palate as a reference. This is to some extent in contrast to the clinical observations, in which tooth movements seem to be greater when using the osseointegrated implant as the reference". The accuracy of this technique has been questioned (9).

Relevance of the study

A majority of earlier studies have used aesthetic indices that assess the overall aesthetic of the implant-supported crown and surrounding mucosa, and not infraposition exclusively. Although there are some methods focusing exclusively on infraposition (radiographic and intraoral scanning methods) their validity and reliability have been questioned (36-39). Methods using intraoral scanners have been suggested with varying results and more studies are needed (11,42). Both for scientific and clinical reasons, a reproducible and accurate method to measure the degree of infraposition is needed. This could in the future lead to increased knowledge on infraposition, correlation of risk factors in adult patients, prediction of infraposition and the prevalence of infraposition. The clinical benefit would be to evaluate whether the patient is qualified for an implant-supported crown or if other treatment options are more suitable. The dentist could also evaluate the success of the implant treatment with a standardized and reliable method. It would benefit the patient in having the best choice of therapy and benefit both patients and society, reducing costs for failed or remade implant-supported crowns.

Aim of study

The aim of this study was to evaluate the agreement between raters when assessing infraposition using three different methods based on intraoral radiography, clinical photographs and study casts. Scientific issue was “is there a change in reliability using different methods in assessing infraposition of implant supported maxillary incisors?”

Hypothesis

-The hypothesis is that the two previously described methods have a low reliability since they are constructed to evaluate infraposition either in a subjective way or without a standardized equipment. Thus, a new method that compensates for these drawbacks would have a higher reliability.

-There will be no difference in the reliability from experienced and non-experienced raters when using the new method due to a low effect of decision-making errors in this method.
Material and methods

This study was based on a retrospective material consisting of eleven implants inserted in the maxillary incisor regions of ten patients.

Several methods already established in the literature where found to be insufficiently described, which excluded them from this study. Two described methods and one new developed method, inspired by an earlier described method, were chosen to be included in the study. Method 1 was described by Thilander et al (1999) and modified by Jemt et al (2008). In this radiographic method, infraposition is evaluated from periapical radiographs. Method 2, an aesthetic index, was described by Brahem et al (2017), who modified and extended an index described by Jemt et al (2007). Method 3, a new method based on clinical photographs and study casts, was inspired by a method described by Thilander et al (1999) but re-designed to measure infraposition using clinical photographs and study casts.

Subjects

Patients, 18-52 years old at baseline, were selected by inclusion- and exclusion criteria (table 2) from a material consisting of 31 patients with 39 implants. Ten patients with altogether eleven implants were included in the study. The patients had been treated with implant-supported crowns in central or lateral maxillary incisor position between 2008-2011 (T0). Follow-up time (T1) was 3.5-7 years. Inserted implants were all Straumann Bone Level implants. The material from each case (implant) consisted of study casts from baseline, intraoral photographs and periapical radiographs from baseline and follow-up. Radiography and clinical photography were performed during treatment process according to standard procedures. Intraoral clinical photographs had a centric projection and radiographs had an as near as possible isometric projection, although no standardized equipment was used. All implant treatments were performed by the same, well experienced periodontist/prosthodontist, working at the Public Dental Health Service in Halmstad. Radiographs and clinical photographs were taken by the same operator.

Table 2. Inclusion- and exclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion:</th>
<th>Exclusion:</th>
</tr>
</thead>
<tbody>
<tr>
<td>· At least one maxillary incisors implant.</td>
<td>· Absence of one or more neighbouring teeth.</td>
</tr>
<tr>
<td>· Study cast at baseline.</td>
<td>· Neighbouring teeth that have received prosthetic treatment or remaking of a previous prosthetic treatment during the follow-up period.</td>
</tr>
<tr>
<td>· Periapical radiographs from baseline and at follow-up.</td>
<td>· Non-isometric periapical radiograph.</td>
</tr>
<tr>
<td>· Intraoral photographs from baseline and at follow-up.</td>
<td>· Non-centric intraoral photograph.</td>
</tr>
</tbody>
</table>

Raters

Six raters with different experience in dentistry performed assessments and evaluations. Two dentists, one specialist in periodontology/prosthodontics and one specialist in oral- and maxillofacial radiology, both with more than 40 years of experience within their fields, and four 4-year dental students at the School of Dentistry, Malmö University, attended as raters. The raters measured the possible infraposition of each study case at two different occasions,
following a protocol of described methods. Differences in the individual measurements and differences between raters were calculated to estimate the error in the biometric assessments and to evaluate the reliability. When numerical data were included in the method, measurements were calculated to the nearest 0.01 mm. The raters were blinded to patient personal data.

Evaluation

All raters were concurrently instructed on the methods and data collection procedures before the first occasion of evaluation. The three described methods were exemplified. In method 2, calibration of the extent of infraposition according to the method’s index was conducted by raters’ consensus in five photographs. Calibration photographs were not included in the study. The evaluation was performed individually in a dark room on a computer with a screen specialized for radiographic examination. Each rater evaluated infraposition using the three methods at two occasions with at least 2 weeks interval (observation 1 and observation 2). By following a protocol, the raters were guided through the procedure and asked to fill in their measured data. All cases were presented and evaluated in the software program Powerpoint, a common, “easy to use” software. The data was saved as a Powerpoint file on the computer. The study was blinded. The authors were nor informed of the identity of the patients nor of the rates’ identity. Clinical photographs and periapical radiographs from each case were randomly arranged, independent of baseline and follow-up, to minimize the effect of memorization and thus reduce bias.

Method 1: Radiographic method

Periapical radiographs (n=22) from baseline (T0) and follow-up (T1) were used to assess the distance in mm between the fixture-abutment junction (FAJ) and cemento-enamel junction (CEJ) at the adjacent mesial tooth. If the mesial tooth was missing or CEJ was not possible to recognize, CEJ at the distal tooth was used. After identifying FAJ, a line was drawn along the FAJ border and the same was done at the border of CEJ. These two lines constituted two parallel lines, see figure 5.

![Figure 5](image)

**Figure 5.** The fixture-abutment junction (FAJ) and cemento-enamel junction (CEJ) at the adjacent tooth was marked with parallel lines as well as the 1st and 4th thread crest of the implant. The vertical distances (dotted lines) were measured in pixel by the authors and converted to mm.
A calibration was made from the inter-thread distance of the implants (0.8 mm), similar to Gjevold et al 2017 (45). The first and fourth distinct crest were marked by two parallel lines, see figure 1, to minimize the consequence of the projection errors. This in an attempt to evaluate a magnification most likely representing that of the crown.

**Method 2: Aesthetic index**

In the second method clinical photographs (n=22) from baseline and follow-up were used to assess the extent of infraposition with an esthetic index. This was done by evaluating the implant-supported crown in comparison with the collateral tooth. All clinical photographs were graded from A-F by using a method described by Jemt and colleagues (2007) and a modified and extended version of the index as described by Brahem et al (2017) (15,42). The index by Brahem used intervals of 0.25 mm instead of Jemt’s index with 0.5 mm intervals. This would give a better precision in measuring the extent of infraposition with a risk to compromise the reliability. The Brahem index scale was modified to add a sixth group, group F, to include dental implant crowns that were perceived to be more prominent than the collateral tooth. This was indicated since it was found in a few baseline clinical photographs.

- **group A**: no difference seen between the incisal edge at the implant and the collateral tooth incisal edge (0-0.25 mm).
- **group B**: a little, insignificant vertical difference between the incisal edge of the implant and the collateral tooth incisal edge seen (0.25-0.49 mm). Implant in infraposition.
- **group C**: a significant vertical difference between the incisal edge of the implant and the collateral tooth incisal edge seen (0.50-0.74 mm). Implant in infraposition.
- **group D**: a significant vertical difference between the incisal edge of the implant and the collateral tooth incisal edge seen (0.75-0.99 mm). Implant in infraposition.
- **group E**: a significant vertical difference between the incisal edge of the implant and the collateral tooth incisal edge seen (> 1 mm). Implant in infraposition.
- **group F**: implant crown incisal edge more inferior than the collateral tooth incisal edge.

**Method 3: A new method based on clinical photographs and study cast**

Clinical photographs (n=22) from baseline and follow-up were used to assess possible infraposition. The method was designed to compensate for discrepancies in the horizontal projection in the photographs. In a software program (Powerpoint) the rater aligned a pre-drawn, standardized perpendicular cross to the vertical and horizontal plane of the bite, with the mesial central incisor as a reference tooth, see figure 6. The raters were asked to mark the maximal width of the reference tooth with a line. Maximal width of the reference tooth was also measured in the study cast with an orthodontic manual slide caliper, to the closest 0.1 mm. These measurements were used to compensate for different magnification of the pictures at baseline and follow-up.
Figure 6. Raters were asked to align a cross to the longitudinal axis of central incisors and to the incisal edge of the reference tooth, mesial to the implant (i). A line was also drawn at the maximal width (w) of the reference tooth.

Calculations

All data, radiographs and photographs, was saved in Powerpoint and then copied to Adobe Photoshop CS6 to be analyzed. This was done by one of the authors and enabled measurements of distances in pixels in radiographs and photographs.

In the radiological method, the CEJ-FAJ distance and inter-thread distance were measured in pixels and converted to mm by using the known inter-thread distance of the implants (0.8 mm). Infraposition was calculated from measurements in T0 compared to T1 radiographs. Positive values (in mm) indicated an increased infraposition from T0-T1.

In the new method based on clinical photographs and study casts, the distance between the horizontal line of the cross and the most prominent incisal edge of the implant crown was measured in pixels. The values from baseline (T0) and follow up (T1) were compared to assess the infraposition over time. An increased distance suggested infraposition. To evaluate the magnification in the photograph, the maximal width of the reference tooth (in pixels) was compared to the maximal width of the corresponding tooth in the study cast (in millimeters).

Statistics

Intraclass correlation coefficient (ICC) with 95% confidence interval (CI) was calculated to estimate reproducibility in methods 1 and 3. To assess the reproducibility in method 2 intra-rater reliability (IRR), Cohen’s kappa (κ) and weighted Cohen’s kappa (κw) was generated for intra-rater agreement and Fleiss kappa (κ) was generated for inter-raters agreement with 95% CI. Validation of included methods are not possible in this study since there is no “true value” (reference standard) for infraposition. Confidence intervals was calculated to estimate statistical significance. κ, κw, and ICC was interpreted according to Landis & Koch (1977)(46), see table 3. The statistical analysis were performed in collaboration with a statistician at Malmö University.
Table 3. From Landis & Koch (1977) (46).

<table>
<thead>
<tr>
<th>Kappa Statistic</th>
<th>Strength of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.00</td>
<td>Poor</td>
</tr>
<tr>
<td>0.00-0.20</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21-0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41-0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61-0.80</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81-1.00</td>
<td>Almost Perfect</td>
</tr>
</tbody>
</table>

Ethical consideration

The material was earlier ethically approved by the Regional Ethical Review Board, Lund, Sweden, (dnr 2014/655).

Results

Method 1: Radiographic method

There was a large variation in the raters’ assessments and the reliability was generally fair to moderate in method 1. Negative values of infraposition were received in a few cases indicating that the implant crown had become longer than the reference tooth rather than shorter. In a few cases, the same rater presented reversed values between the two observations, one negative followed by one positive value of infraposition. The mean infraposition for each case where generally much smaller than the standard deviation in this method, see table 9.

The intra-rater agreement expressed as intra-class correlation (ICC) was moderate in three out of six raters, see table 4. The lowest ICC was 0.073 (CI= -0.495-0.616) and the highest ICC was 0.830 (CI= 0.499-0.951). All ICC indicated a high uncertainty due to wide intervals of confidence (CI).

Table 4. Intra-rater agreement of infraposition assessment expressed as intra-class correlation (ICC) with 95% confidence interval (CI).

<table>
<thead>
<tr>
<th>Rater</th>
<th>ICC</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.487</td>
<td>-0.222-0.846</td>
</tr>
<tr>
<td>2</td>
<td>0.073</td>
<td>-0.495-0.616</td>
</tr>
<tr>
<td>3</td>
<td>0.830</td>
<td>0.499-0.951</td>
</tr>
<tr>
<td>4</td>
<td>0.238</td>
<td>-0.457-0.725</td>
</tr>
<tr>
<td>5</td>
<td>0.569</td>
<td>0.046-0.858</td>
</tr>
<tr>
<td>6</td>
<td>0.437</td>
<td>-0.231-0.814</td>
</tr>
</tbody>
</table>
Inter-raters agreement was calculated for all raters together as well as separately for the two specialists and for the four students in the first and second observation. The agreement between specialists was higher than for students in both observations, see table 5. The specialists presented a substantial inter-rater agreement whilst ICC in all raters and students had low inter-rater agreement.

### Table 5. Inter-rater agreement of infraposition assessment at first and second observation expressed as intra-class correlation (ICC) with 95% confidence interval (CI). ICC= intra-class correlation.

<table>
<thead>
<tr>
<th>Rater</th>
<th>ICC</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>All raters, observation 1</td>
<td>0.301</td>
<td>0.078-0.641</td>
</tr>
<tr>
<td>Specialists observation 1</td>
<td>0.558</td>
<td>-0.061-0.861</td>
</tr>
<tr>
<td>Students observation 1</td>
<td>0.212</td>
<td>-0.057-0.601</td>
</tr>
<tr>
<td>All raters observation 2</td>
<td>0.222</td>
<td>0.024-0.583</td>
</tr>
<tr>
<td>Specialists observation 2</td>
<td>0.604</td>
<td>-0.023-0.886</td>
</tr>
<tr>
<td>Students observation 2</td>
<td>0.102</td>
<td>-0.098-0.472</td>
</tr>
</tbody>
</table>

**Figure 7.** Measurement in the same case by two different raters, visualizing the problem observed with identifying the cemento-enamel junction (CEJ) of the mesial reference tooth.

The measurements of CEJ and FAJ in the same radiograph showed a wide range of measured values both between raters and within the same rater. Especially the line identifying CEJ varied in position, see figure 7. Measurements of CEJ-FAJ in the 11 cases at T0 (radiograph 1a-11a) and T1(radiograph 1b-11b), in first observation, showed that in a few cases there was a wide range between lowest-highest value, as demonstrated in figure 8. In other cases, the homogenous values of CEJ-FAJ indicated a more distinct CEJ.
Method 2: Aesthetic index

Regarding visual assessment of infraposition according to an esthetic index (group A-F), most cases were categorized in group A (0-0.24mm) and B (0.25-0.49mm). One case was categorized in group D (0.50-0.74 mm) by one rater and no case were classified as group E (0.75-1.0 mm). Intra-rater reliability (IRR) varied from 50% to 82% (table 6). According to $\kappa$ this method showed a moderate to almost perfect strength of agreement. The interval of confidence was generally high.

Table 6: Method 2. Intra-rater agreement expressed as Cohen’s kappa ($\kappa$) and weighted linear kappa ($\kappa_w$). CI=confidence interval. Interpretation of $\kappa$, $\kappa_w$: $<0$=poor, $<0.2$=slight, 0.21–0.4=fair, 0.41–0.6=moderate, 0.61–0.8=substantial, 0.81–1=almost perfect.

<table>
<thead>
<tr>
<th>Raters</th>
<th>IRR</th>
<th>$\kappa$</th>
<th>CI</th>
<th>Interpretation</th>
<th>$\kappa_w$</th>
<th>CI</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50%</td>
<td>0.39</td>
<td>0.14-0.65</td>
<td>Fair</td>
<td>0.58</td>
<td>0.35-0.81</td>
<td>Moderate</td>
</tr>
<tr>
<td>2</td>
<td>82%</td>
<td>0.65</td>
<td>0.34-0.96</td>
<td>Substantial</td>
<td>0.87</td>
<td>0.73-1</td>
<td>Almost perfect</td>
</tr>
<tr>
<td>3</td>
<td>77%</td>
<td>0.66</td>
<td>0.40-0.92</td>
<td>Substantial</td>
<td>0.75</td>
<td>0.49-1</td>
<td>Substantial</td>
</tr>
<tr>
<td>4</td>
<td>82%</td>
<td>0.67</td>
<td>0.37-0.96</td>
<td>Substantial</td>
<td>0.69</td>
<td>0.46-0.92</td>
<td>Substantial</td>
</tr>
<tr>
<td>5</td>
<td>77%</td>
<td>0.61</td>
<td>0.31-0.91</td>
<td>Substantial</td>
<td>0.87</td>
<td>0.74-1</td>
<td>Almost perfect</td>
</tr>
<tr>
<td>6</td>
<td>50%</td>
<td>0.31</td>
<td>0.02-0.60</td>
<td>Fair</td>
<td>0.52</td>
<td>0.28-0.76</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

The results from the different raters were compared and it was found that the inter-rater agreement according to Fleiss kappa ($\kappa$) was low, except for students $\kappa$, that was fair. The $\kappa$ was higher for students (0.44/0.38) than for specialists (0.19/0.02) in both observations (1 and
2), see table 7. Generally the κ showed slightly higher values than the κ values. The second method thus showed a low reliability between raters but a substantial reliability within the same rater.

**Table 7. Method 2. Inter-rater agreement, Fleiss kappa (κ). CI = 95% Confidence interval.**

<table>
<thead>
<tr>
<th></th>
<th>Observation 1</th>
<th>Observation 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>κ</td>
<td>CI</td>
</tr>
<tr>
<td>All raters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>κ</td>
<td>0.197</td>
<td>0.10-0.29</td>
</tr>
<tr>
<td>κ_w</td>
<td>0.297</td>
<td>0.14-0.39</td>
</tr>
<tr>
<td>Specialists</td>
<td></td>
<td></td>
</tr>
<tr>
<td>κ</td>
<td>0.189</td>
<td>0.02-0.38</td>
</tr>
<tr>
<td>κ_w</td>
<td>0.185</td>
<td>0.07-0.38</td>
</tr>
<tr>
<td>Students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>κ</td>
<td>0.211</td>
<td>0.08-0.33</td>
</tr>
<tr>
<td>κ_w</td>
<td>0.437</td>
<td>0.18-0.60</td>
</tr>
</tbody>
</table>

**Method 3: New method based on clinical photographs and study casts**

In this method there was a good consistency of the individual measurements of maximal tooth width in the study casts and in the photographs. Assessment of the infraposition in the new method showed substantial to almost perfect intra-class correlation coefficient (ICC) for all raters (table 8). The result indicates excellent intra-rater agreement. Results from one specialist were excluded in the second observation due to inconsistently used methodology. Thus, comparison between students and specialists was not possible. Calculated ICC between all raters in the first and second observation was 0.67 (first observation) and 0.64 (second observation), see table 8. The results indicate good reliability and better concordance compared to method 1 and 2.

**Table 8. Method 3. Intra-rater agreement for rater 2-6 and inter-rater agreement for all observers calculated as ICC. CI = 95% Confidence interval.**

<table>
<thead>
<tr>
<th>Rater</th>
<th>ICC</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.91</td>
<td>0.71-0.98</td>
</tr>
<tr>
<td>3</td>
<td>0.76</td>
<td>0.28-0.94</td>
</tr>
<tr>
<td>4</td>
<td>0.75</td>
<td>0.27-0.94</td>
</tr>
<tr>
<td>5</td>
<td>0.83</td>
<td>0.50-0.95</td>
</tr>
<tr>
<td>6</td>
<td>0.86</td>
<td>0.59-0.96</td>
</tr>
<tr>
<td>Interrater correlation, observation 1</td>
<td>0.67</td>
<td>0.42-0.89</td>
</tr>
<tr>
<td>Interrater correlation, observation 2</td>
<td>0.64</td>
<td>0.36-0.88</td>
</tr>
</tbody>
</table>
Table 9. Descriptive statistics of method 1, 2 and 3. Mean infraposition (mm) and standard deviation in method 1 and 3 (n=11). Majority score of the esthetic index (A-F) and correlated median interval (mm) in method 2. SD=standard deviation. (n=11).

<table>
<thead>
<tr>
<th>Case</th>
<th>Method 1, observation 1 Mean infraposition ±SD (mm)</th>
<th>Method 1, observation 2 Mean infraposition ±SD (mm)</th>
<th>Method 2 Aesthetic index Majority score (corresponding mm)</th>
<th>Method 3, observation 1 Mean infraposition ±SD (mm)</th>
<th>Method 3, observation 2 Mean infraposition ±SD (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.09±0.44</td>
<td>0.25±0.66</td>
<td>A 0-0.24</td>
<td>-0.04 ±0.30</td>
<td>-0.03±0.19</td>
</tr>
<tr>
<td>2</td>
<td>-0.32±0.40</td>
<td>-0.37±0.43</td>
<td>A 0-0.24</td>
<td>0.10 ±0.18</td>
<td>0.05±0.28</td>
</tr>
<tr>
<td>3</td>
<td>0.61±0.27</td>
<td>0.24±0.48</td>
<td>B 0.25-0.5</td>
<td>0.74 ±0.13</td>
<td>0.19±0.16</td>
</tr>
<tr>
<td>4</td>
<td>0.49±0.61</td>
<td>0.23±0.66</td>
<td>C 0.5-0.74</td>
<td>0.37 ±0.19</td>
<td>0.40±0.15</td>
</tr>
<tr>
<td>5</td>
<td>0.55±1.01</td>
<td>0.97±0.93</td>
<td>B-C 0.25-0.74</td>
<td>0.16 ±0.21</td>
<td>0.26±0.17</td>
</tr>
<tr>
<td>6</td>
<td>-0.25±0.30</td>
<td>0.01±0.50</td>
<td>A 0-0.24</td>
<td>0.90 ±0.34</td>
<td>0.75±0.05</td>
</tr>
<tr>
<td>7</td>
<td>0.35±0.53</td>
<td>0.15±0.68</td>
<td>B 0.25-0.49</td>
<td>-0.25 ±0.23</td>
<td>-0.34±0.78</td>
</tr>
<tr>
<td>8</td>
<td>-0.22±0.40</td>
<td>-0.29±0.59</td>
<td>F supraposition</td>
<td>-0.45 ±0.71</td>
<td>-0.39±0.75</td>
</tr>
<tr>
<td>9</td>
<td>0.81±0.41</td>
<td>0.52±0.35</td>
<td>B 0.25-0.49</td>
<td>-0.04 ±0.06</td>
<td>-0.12±0.43</td>
</tr>
<tr>
<td>10</td>
<td>0.73±0.90</td>
<td>0.53±0.61</td>
<td>A 0-0.24</td>
<td>-0.03 ±0.16</td>
<td>-0.04±0.04</td>
</tr>
<tr>
<td>11</td>
<td>0.66±0.25</td>
<td>0.51±0.63</td>
<td>A 0-0.24</td>
<td>-0.04 ±0.30</td>
<td>-0.03±0.19</td>
</tr>
</tbody>
</table>
Discussion

In this study we proposed a new method to assess infraposition of implant supported maxillary incisors. The aim was to evaluate the agreement between raters when assessing infraposition using the new method based on clinical photographs and study casts and two previously described methods from the literature. Main issue was to answer the question if any of the methods were more reliable. The results indicate that infraposition is difficult to evaluate and that reliability is limited for the two previously described methods. However, very good intra- and interobserver agreements were found for the new method, proposed in this study. For the new method, the inter-rater agreements were substantial (ICC= 0.64-0.67) and the intra-rater agreement was substantial to almost perfect (ICC=0.75-0.91) (table 8). Both inter- and intrarater agreement was generally higher than for the other two methods used in this study. This supports our hypotheses and suggests that the new method is more reliable as the concordance is good between raters. The new method may therefore be suitable for evaluating infraposition of implant supported maxillary incisors, though a larger study is needed to draw significant conclusions.

According to Swets and Pickett (, the performance of an observer could be considered as “a measure of the extent to which the inaccuracy of a system is the result of decision-making errors” (47). Based on the variation among dentists when detecting the extension and depth of caries lesions (48), it can be expected that this variation may also exist for identifying for example CEJ on radiographs and infraposition on clinical photographs. Sennerby and colleagues (2016) also discussed the problem of evaluating the accuracy of a method when no reference standard is available and cited Wenzel et al “when no valid expression of the true state of the disease can be obtained a study of precision ...is appropriate” (48,49). Thus, in this study the reliability was tested.

To be able to perform a method study on infraposition, the material should present the required feature to some degree. The 11 cases (implants) in this study were randomly selected from 31 patients with 39 implants. Prevalence of infraposition (>0.5mm) has been reported in as many as 65% of the adult patients (40-55 years) and 57% in young adults (15.5-21 years) in a study by Bernard et al (10) and thus our material could be expected to include cases presenting infraposition. A correlation of all the results in our study found two cases with mean infraposition (mean obs1 and obs2) of 0.24 mm or more in all three methods, which may indicate that infraposition existed in at least two out of eleven (18%) cases of the study group, see table 9. The new method generally presented lower values for infraposition compared to the aesthetic index and radiographic method, and only two cases presented values higher than 0.24 mm in both observations, see table 9. This could be compared to infraposition recorded in five cases with the aesthetic index and seven (obs 1) and five (obs 2) cases with the radiographic method, see table 9. These results are in line with earlier studies (10,11). Prevalence in the literature indicate a surprisingly high number of adult patients affected by an aesthetical problem that dentists so far have paid little attention to. Though, according to our study, using earlier described methods might result in a high prevalence and magnitude of infraposition of implant supported incisors in a population. Should the assessment of infraposition prove to be overestimated in earlier scientific studies using the earlier described methods, this might explain why there is small clinical attention to infraposition in adults. On the other hand, it could be hypothesised that infraposition is an unrecorded issue due to lack of awareness and inconsistent diagnostic methods.
Assessment of infraposition in periapical radiographs has been used in several studies (10,11,14,45). The results from the radiographic method generally showed fair-moderate intra-class correlation coefficients (ICC). ICC varied both within each operator (ICC= 0.073-0.83) and between the operators (ICC=0.10-0.60), see table 4 and 5. The limited reliability of the method could be explained by difficulties in identifying the cemento-enamel junction (CEJ) in the radiographs. It has been shown that CEJ is an anatomic structure difficult to distinguish in radiographs (50). Interpretation of CEJ thus can be expected to vary between raters, especially if radiographs are not standardized and with an optimal quality. The grade of distortion will also differ both within and between radiographs when a complete parallel technique is not possible to obtain (37, 39). In a clinical situation, limited time or non-standardized projections could result in distortion and unsharp radiographs which might affect the ability to distinguish CEJ. According to figure 8, in some cases the recorded distance between CEJ-FAJ differed about 100 pixels from highest to lowest value by different raters. There was a tendency that CEJ was less apparent in radiographs with large variations in CEJ-FAJ measurements. In cases where CEJ were more apparent the lowest and highest measurements of CEJ-FAJ only differed up to 25 pixels (figure 8).

In the radiographic method, specialists had better agreement than the students, see table 5. This may be explained by the fact that the specialists, one radiologist and one prosthodontist/periodontist, were more trained and well experienced in evaluating radiographs compared to students. Due to experience, the clinicians are probably able to make a more reliable evaluation than students if CEJ are difficult to distinguish in radiographs. The better ICC among specialists could also be a coincidence, since there are two specialists compared to four students. Considering all raters the intra-class correlation was 0.30 and 0.22 in first and second observation, respectively, indicating poor agreement between raters. It was also found that in a few patient cases, raters could present both negative or positive values, resulting in a large standard deviation, as seen in table 9. Since the majority of implants are followed up by general dentists, this method would risk to give an inconsistent assessment of infraposition according to our findings.

The second method, using an aesthetic index, showed a low to limited reliability between raters but a good reliability within the same rater, see table 6-7. This indicated that the individual rater was consistent in interpreting the position of the implant compared to the collateral tooth at two different observations. Inter-rater agreement considering all raters was low. Students had a better agreement (obs1 \( \kappa = 0.44 \), obs2 \( \kappa = 0.38 \)) compared to specialists (obs1 \( \kappa = 0.19 \), obs2 \( \kappa = 0.02 \)). The weighted linear kappa, \( \kappa \), take into account the rate of disagreement between raters which is important when there is a order of the rating scores as in the aesthetic index. Generally the \( \kappa \) indicate a higher agreement than \( \kappa \) but the \( \kappa \) may discredit the results due to the fact that this statistical method is designed for normally distributed material and our results were predominantly categorized in group A and B. The students presented better agreement than experienced dentist which could be explained by the fact that students rated most of the cases as having no, or a small infraposition (group A and B). The experienced dentists tended to give infraposition higher and more varied scores with the aesthetic index. Similar results have been found in a study by Cho et al 2010 periodontists gave higher scores with an aesthetic index, compared to students (51). Could it be a result of low experience in students to evaluate implants, leading to underestimation of infraposition or could experienced dentists generally overestimate, and “search for” infraocclusion?
It has also been suggested that the professional background, i.e. examiners from different specialities, may have an impact on the result of aesthetic evaluations (11,36,38,39,52). The poor agreement of the specialists in our study could thus be explained by the fact that they were from different specialities. It could be hypothesized that the prosthodontist have increased experience in evaluating prosthetic outcomes in contrast to the radiologist. No earlier studies on aesthetic evaluations include radiologists as observers. Results for other aesthetic assessments of implants found contradicting results. With the ICAI-index, it was found that two prosthodontics had higher inter-observer agreement compared to two oral surgeons (36). In contrast, interobserver agreement was low (4-28%) between two prosthodontists and periodontists had the highest agreement compared to orthodontists, oral surgeons, dental technicians and dental assistants in another study (52). Since both these studies had two observers of each expertise and our study only included one of each, there is a risk that the results depend on individual factors more than disciplinary.

Further, the low inter-rater agreement for all observers using the aesthetic index in this study implied that raters disagreed about the extent of infraposition in spite of calibration. Thus, the aesthetic index could be reliable when used by the same dentist/rater assessing infraposition at baseline and follow-up examination. It could be questioned to use the method in scientific research if results from different raters or studies are compared. Also, in this study the first and second observation were accomplished within a few weeks. Since follow-up examinations and longitudinal studies are conducted after years rather than weeks, future studies should evaluate intra-rater agreement over a longer period. An unpublished study by Nilsson, Johansson and Ekfeldt (2019) using the method described by Jemt et al 2007 found that re-evaluation of infraposition after 18 months gave a kappa-value of 0.83 indicating a very good intra-rater agreement.

In this study we found that using the aesthetic index earlier described by Jemt et al (2007) in method 2, a majority of cases were placed in the first category, group A: “no difference seen between the incisal edge at the implant and the collateral tooth incisal edge (infraposition 0-0.25 mm)”. Other studies rated more cases to higher categories (15,42). This could, of course, depend on the population included in the study. Although we used a small index-interval of 0.25mm, introduced by Brahem et al. 2017, we found a limitation in using the index since it did not distinguish a small infraposition from no infraposition (42). This might prevent observations of small tooth movements and to follow a slow progression of infraposition. On the other hand, the question is whether it is possible to register differences smaller than 0.25 mm and if tooth movements less than 0.25 mm is of clinical significance. Several studies compared patient satisfaction of implant aesthetics compared to dentists (11,38,39,45,53). Patients generally rated the aesthetics significantly higher than professionals and were less critical to small imperfections (11,38,39,45).

The mean infraposition in our study varied from -0.32-0.97mm in all cases (n=11) with the radiographic method, <0-0.74mm with the aesthetic index and -0.45-0.90mm with the new method. These values seem to be comparable to other studies that found infraposition of -1.28 to 2.06 mm, (mean=-0.13 mm after 7.5 years) and 0.10-1.86 mm (mean 0.68 mm after 4.2 years) in a mature adult group and a mean of 0.46 mm in young patients (mean age 15 years old, 3 years follow up) (10,13,45). The recorded negative values indicate that implants would present a negative infraposition/ “supraposition” due to the ankylosic behavior of the implant and thus the validity of all three methods could be questioned. In 50% of the cases the mean infraposition in observation 1 differed from the mean infraposition in observation 2 more than 0.25 mm in the radiographic method, see table 3. Comparing to the esthetic index used in
method 2, such a difference between observations would result in a shift of category and thus might lead to a different clinical “diagnose” by the same rater at follow up examination.

There are several studies evaluating infraposition and the aesthetic outcome of implant-supported crowns (11,38,39,45,52,53). However, few studies have been able to identify a significant correlation of patients or groups at risk for infraposition and to be able to evaluate the best possible treatment in individual cases. This has also been addressed by Jemt and colleagues (2007) who suggested that focus would rather be to identify risk factors for infraposition instead of describing the problem (15). A reliable method to ensure the extent of infraposition must be of interest to be able to evaluate individuals with risk of infraposition.

Limitations of the study

The small sample size in this study is a limiting factor as the material is too small to draw significant conclusions, which is supported by the wide intervals of CI for most results. Also, a limited number of raters, for example only two specialist dentists compared to four students could affect the inter-rater agreement results in all methods. According to Swets and Pickets (1982), six or seven raters is considered optimal when including a reasonable large material. In this case the material might be small but as a methodological study on rater performance must be considered sufficient.

In the radiographic method we used a modification of Jemt’s method. The distance between the four crests located closest to the crown was measured. This in an attempt to evaluate a magnification index most likely representing that of the crown where infraposition was evaluated. An orthoradial and isometric projection is of importance to distinguish the implant threads. Due to the lack of standardized equipment, image distortion in the radiographs is a source of error in our study but may be representative to the reality in many clinical situations. Crown and root/implant will be subjected to different grade of distortion resulting from complete parallel technique not being possible to obtain (37,39).

Although the results are promising, improvements of the new method could be suggested. In this study the software program Powerpoint was used. This may be a shortcoming as the projection of the photographs was not standardized. This was compensated by aligning a standardized reference “cross” to the vertical and horizontal plane of the incisors. The software only permitted rotations sometimes larger than desired. This on the other hand impacted the horizontal line that touched the incisal edge of the reference tooth. This could have had an impact of the measurements of infraocclusion in some cases, either increased or decreased values depending on the case. As the method is unvalidated our results were not able to tell the extent of infraposition, which was, however, not our aim. Whether this affected our results is hard to say but there is a risk that the raters avoided adjusting the perpendicular cross due to the apparent deviation of the cross at rotation. This may be a source of error that could be improved by using another software program, for example Adobe Photoshop, where the cross can be rotated gradually.

We also suggest that dentists, rather than students, could participate in a similar study since students have lower experience of interpreting the structures evaluated in this study. Students were chosen due to the limited number of available voluntary dentists at the time of assessment.
Ethical considerations

The aesthetic outcome of an implant-supported crown has shown discrepancies between the dentists’ and the patients’ perception. Studies support the fact that the patients are more satisfied with the results according to an evaluation of several aesthetic criteria, among these infraposition (9,43). There is a risk that studies evaluating infraposition increases the patient's consciousness of infraposition and negatively influence patient satisfaction. That risk has to be considered. In a long-term perspective, a good method to assess the implant and the implant-supported crown over time could benefit individual patients, for example in cases where it can be questioned whether to remake the prosthetic due to infraposition or not.

Dentists should also consider whether provided dental care and evaluations are cost beneficial for the patient and for society. Even though the prevalence of an infraposition more than 0.25mm in implant-supported crowns are reported in as many as 40-87% of an adult population (10,11, unpublished data Nilsson, Johansson et al 2019), anterior infraposition may be considered mainly an aesthetical problem and the cost for the evaluation and time spent might overcome the benefit in most cases.

Future

A prospective aim is to validate a method to assess infraposition. Based on the results on this rather small study the new method based on clinical photographs and study casts could be evaluated further to assess the accuracy rather than only the reliability.

According to Bernard et al (2004), the future trust in digital technique to find a valid method to measure posteruptive movements and infraposition. Though, the digital technique with intraoral scanning, have shortcomings. Earlier studies have concretized problems with intraoral scanning (42,44). They found difficulties finding stable reference points not affected by changes over time and difficulties to diminish the size of measured errors as the magnitude of errors sometimes exceed the expected infraposition. (44) These problems need to be solved. Until then, scanning is as uncertain as any other method described.

Conclusion

The results of this study indicate that infraposition is difficult to evaluate and that reliability is limited in the methods previously described in the literature, due to fair inter-rater agreement.

The best intra- and interobserver agreements were found for the new method based on clinical photographs and study casts. The results from the new method show good/excellent reliability both between raters and within the same rater. This indicate that the new method may be useful to assess infraposition in anterior maxillary implants, as the method makes it possible to record measurement with better concurrence. Considering the limited material we are unable to draw significant conclusions.

The new method suggests better reliability, objectiveness and is possible a better method to evaluate infraposition of implant supported maxillary incisors over time.

Whether these measurements are close to the true infraposition cannot be evaluated until a valid method has been proposed.
References


