Systematic Review

Clinical outcomes of lingual orthodontic treatment: a systematic review

Ilias Mistakidis*, Hattan Katib**, Georgios Vasilakos***, Dimitrios Kloukos**, Nikolaos Gkantidis**

*Department of Orthodontics, Faculty of Dentistry, Aristotle University of Thessaloniki GR-54124, Greece, **Department of Orthodontics and Dentofacial Orthopedics, University of Bern, Freiburgstrasse 7 CH-3010, Switzerland and ***Private Practice, Frankfurter Strasse 610, DE-51145, Cologne, Germany

Correspondence to: Nikolaos Gkantidis, Department of Orthodontics and Dentofacial Orthopedics, University of Bern, CH-3010, Freiburgstrasse 7, Bern, Switzerland. E-mail: nikosgant@yahoo.gr

Summary

Objectives: To assess the available evidence on the effectiveness of lingual orthodontic treatment and related clinical parameters through a systematic review of relevant studies.

Materials and methods: Eligible clinical studies published from January 2000 to March 2015 were identified through electronic (five major databases) and hand searches. Risk of bias was assessed using the Cochrane risk of bias tool for prospective studies and a specially designed tool for retrospective studies.

Results: From the 3734 articles identified by the search, after application of specific inclusion and exclusion criteria, 16 papers were included in the study. Eleven studies were retrospective, four were prospective, and only one was a RCT. In detail, six studies evaluated differences of the treatment outcome from the pre-treatment set-up prediction, two studies evaluated the effect of treatment on periodontal and microbial parameters, and 10 studies assessed various clinical treatment related parameters. Despite several promising findings, the quality of evidence supporting them was found to be low in most cases.

Conclusions: This systematic review showed encouraging results on the clinical outcome of lingual orthodontic treatment, especially in regards to the achievement of individualized treatment goals and the reduction of decalcifications on the bonded surfaces of the teeth. However, additional well-designed prospective clinical trials with larger samples are needed to confirm those findings. Several aspects of lingual orthodontic treatment were difficult to be conclusively evaluated due to the study design, the heterogeneity, the small samples sizes, and the high risk of bias seen in the majority of the included studies.

Introduction

The aesthetic demands of patients have been increased during the years regarding not only the treatment objectives, but also the influence of orthodontic appliances in patient's aesthetic appearance. Conventional orthodontic treatment has been shown to compromise facial appearance (1) and this is a major concern of patients seeking orthodontic treatment (2). As a result, aesthetic materials and techniques have been introduced in clinical practice to overcome these limitations (3).

Lingual orthodontics comprise a fundamental expression of this necessity (4). Since its introduction in 1980s, several systems and techniques have been introduced, revealing the increased interest of patients and doctors for this treatment approach (5). The major advantage of lingual orthodontics is that the appliances are not visible. However, lingual orthodontics comprises a relatively new treatment option that has certain differences to the well-established labial orthodontic treatment, mainly attributed to the position of the appliances (6, 7). A recent systematic review of six studies that compared certain adverse effects of
lingual and labial orthodontic treatment (8) provided further support to this viewpoint. Differences that may exist in various aspects of treatment, and can be related to the patient, the practitioner, or the appliance itself, might be responsible for the limited, although increasing use of lingual orthodontics in everyday practice (5, 7, 9, 10).

During the years, several case reports, reviews, and clinical research papers have discussed various aspects of lingual orthodontic treatment. However, a thorough systematic evaluation of clinical studies related to clinical outcomes of lingual orthodontic treatment is lacking from the literature. Such a review will allow gather the supporting evidence on this expanding treatment approach, will help clinicians to provide evidence-based treatment, and will provide recommendations for targeted future research in the field. Thus, the purpose of the present study is to assess the available evidence on the effectiveness of lingual orthodontic treatment and related clinical parameters through a systematic review of recent relevant studies (published later than 1 January 2000).

**Material and methods**

The present systematic review was conducted and reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines for reporting studies that evaluate health care interventions (11, 12), to the Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines for reporting observational studies (13, 14), and to the Cochrane Non-Randomized Studies Methods Group (15). The Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach (16) was used for grading the quality of evidence. A pilot Pubmed search followed by systematic evaluation of 10 potentially eligible and randomly selected studies were performed in order to draft the review protocol. Methods of search and analysis, eligibility assessment, data extraction forms, and all other methods and tools used were defined in advance of the study based on the results of the pilot study.

The main type of intervention, e.g. buccal or lingual-fixed orthodontic appliances is unlikely to be randomized, since patients usually decide about this issue based on individual needs and preferences. Furthermore, this is a relatively new and unexplored treatment, especially in terms of systematic evaluations. Thus, we decided to evaluate for eligibility also non-randomized and retrospective studies following a quality assessment protocol (see details below).

**Search methods for identification of studies**

Electronic search was conducted by two authors (G.V. and N.G.) in five major databases, Medline, EMBASE, Google scholar beta, all Cochrane Databases, and Conference Paper Index with the last update on March 2015. A complex Medline search (via Pubmed) was performed to identify any relevant study, based upon various combinations of key words including: ‘lingual orthodontics’; ‘lingual patients’, ‘lingual brackets’, in all fields (Limits: Humans, English, German, French, Italian). A detailed description of the Medline electronic search strategy is provided in supplemental Appendix 1. Similar searches were performed in all the electronic databases used for the study.

All papers assessed through full text reading, all included papers, and all reviews identified were subjected to hand search of their references for relevant papers that might have been missed during the electronic search.

Unpublished literature was searched through Cochrane Central Register of Controlled Trials, Conference Paper Index, and Google scholar.

**Selection of studies**

Eligibility assessment was performed in a standardized manner, independently, and in duplicate by two reviewers (H.K. and G.V.) who were not blinded to the identity of the authors, their institution, or the results of the research. Any disagreement was resolved by consensus and through discussion among all authors. Titles and abstracts were screened first and afterwards full text review of any relevant and potential for inclusion article was conducted.

A positive exclusion method was used, whereby only those publications that did not meet one or more of the inclusion criteria were excluded. An independent reviewer (I.M.) checked a random selection (20 per cent) of filtered articles for consistency. Inter-rater agreement on study eligibility was assessed by Cohen’s kappa.

**Eligibility criteria**

The following inclusion criteria were applied:

1. Randomized clinical trials (RCTs), prospective controlled clinical trials (CCTs), or retrospective studies with a control group reporting on results or treatment parameters related to orthodontic treatment performed with fixed lingual brackets bonded at least to all teeth between first molars, in one or both arches.
2. English, German, French, and Italian languages.

The exclusion criteria were:

1. Studies reporting results emerging from questionnaires.
2. In vitro and animal studies.
3. Case reports/case series (sample size ≤10).
4. Editorials, opinions, reviews, and technique description articles, without reported sample.
5. Studies referring to lingual treatment performed in selected teeth of an arch or combined with other appliances, such as labial brackets, in a non-defined manner and not in order to investigate a specific hypothesis under a split-mouth design.
6. Studies published prior to January 2000 were excluded to control for confounding derived from the large differences of bracket types and application techniques of that time compared with the current practice.

**Types of participants**

Healthy participants who require orthodontic treatment with fixed appliances with no age limit. Studies including patients receiving any kind of medication, which can influence orthodontic treatment or patients receiving orthognathic surgery, syndromic patients, patients with cleft lip and palate, or any systemic disease were excluded.

**Types of interventions**

Any treatment approach utilizing orthodontic fixed lingual appliances placed in one or both dental arches, used to correct any type of malocclusion, were included.

**Control**

Conventional orthodontic treatment with fixed labial appliances or no treatment. The planned treatment outcome and comparable patients receiving fixed appliance treatment with different lingual bracket systems or application techniques (i.e. different bonding protocols) were also considered.

**Types of outcome measures**

Any measure of performance or effectiveness of the intervention tested including its accuracy and any parameter regarding the quality
of the treatment outcome, periodontal, or dental parameters (plaque index, gingival index, root resorption, tooth necrosis/loss of vitality etc.), or various clinical aspects of treatment (duration of treatment, stability of final outcome, bond failures etc.).

Data extraction process
Data extraction was performed by two authors (I.M. and H.K.), independently in the pre-determined data extraction forms that were also used for quality assessment of the included studies. In cases of inconsistencies re-examination of the original studies by the two reviewers and discussion resolved any disagreements. N.G. was responsible for checking the data extraction forms. Inter-rater agreement on data extraction was assessed by Cohen’s kappa.

Quality assessment of individual studies
The quality assessment of the eligible studies was performed by two investigators, independently (D.K. and N.G.). In areas of disagreements, a joint decision was obtained after discussion between all authors in order to reach a consensus. Quality assessment of randomized studies was performed using the Cochrane Risk of bias tool (17). The same tool was also used for non-randomized prospective studies in the applicable domains.

To assess the quality of the retrospective studies, two investigators working independently (D.K. and N.G.) evaluated several items according to a modified quality assessment tool designed based on Newcastle–Ottawa scale (18), relevant guidelines of the Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (15), and the GRADE approach (16).

We categorized overall risk of bias in any included study according to the following.

1. Low risk of bias (plausible bias unlikely to seriously alter the results) if all key domains were assessed as at low risk of bias.
2. Unclear risk of bias (plausible bias that raises some doubt about the results) if one or more key domains were assessed as at unclear risk of bias.
3. High risk of bias (plausible bias that seriously weakens confidence in the results) if one or more key domains were assessed as at high risk of bias.

Dealing with missing data
The analysis was performed using only the available data (ignoring missing data); methods for estimating missing SDs as described in Section 7.7.3 of the Cochrane Handbook for Systematic Reviews of Interventions (15), were used where applicable.

Data synthesis
A meta-analysis was planned only if there were at least two studies of similar comparisons, reporting similar outcome at similar time points. For qualitative synthesis, the included studies were to be assigned to general groups according to their subject category.

Determination of available evidence supporting clinical recommendations
Following the quality assessment of individual articles, each one was assigned to a group according to the studied subject. For each subject, the overall quality of the body of evidence was determined after considering the quality (assessment of individual studies), quantity (magnitude of treatment effect, number of studies, sample size across studies), and consistency (the extent of similarity of different studies in their findings) of the available studies and their findings on the subject. Clinical recommendations were formulated based on these considerations and by balancing the desirable and undesirable consequences of each intervention.

Results

Literature flow
The initial electronic search yielded 3734 studies. After reading titles, abstracts, and application of inclusion and exclusion criteria, 87 studies were examined in their full-text forms. From these, 16 studies were deemed eligible for inclusion in the review and for qualitative and/or quantitative synthesis (6, 18–32). The flow chart of study selection together with reasons for exclusion is provided in Figure 1. The kappa scores for the initial selection and data extraction procedures were 0.86 and 0.91, respectively.

Characteristics of included studies
The 16 included studies were assigned to three general groups according to the main outcome tested: 1. accuracy of treatment prediction, 2. periodontal parameters, and 3. other clinical parameters. Two included studies were assigned in two groups.

Eleven studies were retrospective, four were prospective CCTs and one was RCT. Six retrospective studies tested the accuracy of lingual treatment through the comparison of set-ups with actual treatment outcome (19–23, 29). One of them also tested the occlusal outcome of lingual treatment compared with labial treatment (23). Ten studies tested various clinical aspects of treatment, such as effectiveness, bond failures, formation of white spot lesions, anchorage loss, and treatment duration (6, 18, 22–25, 28, 30–32). Finally, two prospective CCTs assessed treatment effects on various periodontal parameters, such as the plaque index and the gingival index (26, 27). An overview of the characteristics of the included studies is provided in Table 1.

Publication bias
Statistical analysis of publication bias was not indicated, since no quantitative synthesis was undertaken.

Risk of bias of included studies
For non-randomized prospective CCT studies, the items of random sequence generation and allocation concealment were not applicable and were set by default as unclear. This decision was based on the specific characteristics of the clinical question studied; primarily the inability of personnel to assign patients in groups and/or predict favourable versus unfavourable outcomes. Initial inter-rater disagreement existed in 3 out of 16 cases (agreement greater than 80 per cent) and these were all between unclear and high risk ratings. Disagreements were resolved through discussion by all authors until consensus was reached.

RCTs
Only one RCT was included and was assessed as unclear overall risk of bias (6). Figure 2 shows the summary of risk of bias assessment for the RCT study according to the Cochrane Risk of bias tool.

Prospective CCTs
From the four CCTs included, two were assessed as unclear (26, 27) and two as high overall risk of bias (25, 30). Figure 2 shows the summary of risk of bias assessment for non-randomized prospective CCTs.
Retrospective studies

From the 11 retrospective studies included, 10 were assessed as high (18, 20–24, 28, 29, 31, 32) and 1 as unclear overall risk of bias (19). Table 3 shows the assessment of risk of bias for retrospective studies.

Qualitative synthesis of included studies

A meta-analysis was not feasible since there were no studies of similar comparisons, reporting similar outcomes at similar time points. In order to proceed to the qualitative analysis, the 16 studies were divided into three main groups according to the aspect of lingual treatment tested. The aforementioned categorization did not derive from any established methodological background, but was implemented based on clinical interest. Furthermore, there is a variety of outcome measures reported in each study that does not allow in most cases for standardized reporting of results. An overview of the set-up, the findings, and the overall risk of bias assessment of individual studies is provided in Tables 1 and 2.

Effects of interventions

Accuracy testing

In total, six studies tested the accuracy of lingual orthodontic treatment. All utilized a retrospective study design and tested the Incognito System (3M Unitek, Monrovia, California, USA) (19–23, 29). Five of them had high (20–23, 29) and one unclear risk of bias (19) (Table 3). Two studies (19, 20) assessed the deviations in tooth position between the planned (setup) and the actual treatment outcome measured on three-dimensional (3D) digital dental casts. Both agreed that in most cases anterior teeth were positioned within 1 mm and within 4 degrees of their planned positions, whereas differences increased from anterior to posterior, with the second molars showing the poorest results. Another study tested the deviations in lower incisor inclination between the predicted (setup) and the actual outcome measured on 3D digital dental casts and reported a difference of 0.44 (SD: 0.29) mm (21). A study that tested the deviations in lower incisor inclination between the planned (setup) and the actual outcome in Class III patients reported a difference of 3.75 degrees (SD: 3.06 degrees) (29). This difference was increasing in extraction cases and when undersized finishing wires were placed.

Clinical parameters testing

In total, 10 studies tested various clinical aspects of treatment, such as effectiveness, bond failures, formation of white spot lesions, anchorage loss, or treatment duration. Nine of them had high (18, 22–25, 28, 30–32) and one unclear risk of bias (6) (Table 3; Figure 2). A retrospective study on the effect of the Incognito lingual technique + Herbst reported a change in mandibular incisor inclination of −16.5 degrees to 5.8 degrees that was according to the planned movement in each case (22). Another retrospective study compared the occlusal outcome and treatment duration with lingual (Incognito) and labial appliances and did not report any significant difference (P > 0.05), although there was a tendency for higher improvement in weighted PAR index through labial treatment (23). One retrospective study tested potential differences in anchorage loss in premolar extraction cases between lingual bidimensional preadjusted brackets (Ormco Corp, Glendora, California, USA) and labial appliances and reported almost double amount of anchorage loss in labial treatment (24). The same issue
was tested in a prospective CCT study that used another bracket system (STb, Ormco, Orange, California, USA) and yielded similar results (30). A prospective study tested changes in the lower dental arch of Class I patients with crowding following orthodontic therapy with labial or lingual two-dimensional (2D) technique (Forestadent, St Louis, Missouri, USA). The only significant difference regarded the position of incisors relative to A–Po (lingual: 1.0, SD: 0.5 mm; labial: 0.0, SD: 1.5 mm, \( P < 0.05 \)) (25). The only included RCT study tested the development of white spot lesions during lingual (Incognito) and labial appliance treatment. This study utilized a split mouth design and was characterized as unclear risk of bias mainly because binding was not possible and this might have resulted to performance bias. The study concluded that the number of buccal caries that formed or progressed during treatment was 4.8 times higher than the number of lingual lesions (\( P = 0.01 \)). Furthermore, there was an overall increase in carries extent/total lesion area that was more for buccal surfaces in a subject than for lingual surfaces (\( P < 0.05 \)) (6). Another retrospective study from the same research group tested the effect of the addition of an extra layer of a hydrophilic resin, during bonding in the incidence of demineralization beneath the base of lingual brackets (Incognito). The addition of hydrophilic resin induced a 3-fold decrease in demineralization incidence after treatment (28). A retrospective study compared treatment with lingual (STb, Ormco) and labial appliances and found similar PAR reduction, treatment duration and root resorption with both techniques. The only difference was detected on post-treatment interincisal angle which was increased more by lingual treatment (T0–T1, labial: 11.2 degrees, SD: 11.6 degrees; lingual, 17.9 degrees, SD: 8.9 degrees, \( P < 0.05 \)) (31). In another retrospective study, the number of lost brackets during the first year of treatment was found to be similar in lingual (Incognito) and labial treatment (32). Finally, a recent retrospective study showed that treatment duration was significantly smaller with one lingual bracket system (WIN; S1: 17.96 m, S2: 20.49 m) compared to another (Incognito; S1: 22.7 m, S2: 29.79 m) (18).

**Periodontal parameters testing**

Two prospective CCTs with unclear risk of bias assessed treatment effects on various periodontal parameters (26, 27). One split-mouth study tested clinical periodontal and microbial indices before and 4 weeks after bonding of Incognito lingual appliances. Plaque index and bleeding on probing significantly increased in this period in the bonded sites, while no difference was detected for probing depth. Periodontopathic bacteria did not change significantly (26). The other study assessed clinical periodontal parameters after at least 6 months on treatment, at three consecutive appointments (3–4 week interval), in patients treated with Ormco seventh generation lingual (Ormco Europe, Amersfoort, The Netherlands) or InvisibleAlign appliances. All the indices apart from probing depth improved significantly between the first and third assessments in the InvisibleAlign patients, while they were significantly worse in lingual treatment in all time points (27).

**Discussion**

The growing interest of the scientific community for lingual orthodontics over the years is evident by the increasing number of relevant publications in the literature. The search of the present systematic review covered a period of 15 years (2000–2015) and 13 out of the 16 included studies were published in the last 5 years. This underlines the need for orthodontics to meet the increased aesthetic demands expressed by patients (33). However, lingual orthodontic treatment is still not considered a conventional option with widespread use in current orthodontic practice. Possible explanations for this could be the relatively unexplored clinical performance of the technique, the additional knowledge required to be applied, the fact that lingual orthodontics are not taught in the majority of postgraduate programs at the same extent as conventional buccal techniques, as well as the higher costs that are usually involved.

We decided to include only recent studies (from January 2000 and on) by applying time restriction to the search strategy in an attempt to be comparable with current clinical conditions. Otherwise, if older studies were included, the evolution of lingual appliances and techniques over the years (34, 35) could give results not applicable to the present clinical practice. The earliest included study was published in 2003 and the remaining studies were published after 2007. The 16 included comparative studies were divided in three major groups according to the study subject. These regarded 1. accuracy, 2. clinical parameters, and 3. periodontal parameters testing.

In the first group, six retrospective studies tested the accuracy of lingual orthodontic treatment by comparing the planned treatment goal in the setup with the actual outcome obtained after treatment (19–23, 29). The results of these studies were encouraging, indicating that modern lingual orthodontic systems can achieve to a significant extent the individualized treatment goals established by the setup. The retrospective nature of all these studies, however, does not allow for elucidating the parameters that influence the achievement of the planned treatment outcome for the individual patient. For this, well-designed prospective clinical trials are necessary.

The second subject group included 10 studies that investigated various clinical aspects of lingual orthodontic treatment. The only available RCT showed that the number of buccal caries lesions that formed or progressed during buccal fixed appliance treatment was approximately five times higher than the number of lingual caries lesions in lingual treatment (6). However, since bonding was not possible because of the split-mouth study design, performance bias during the study period cannot be excluded. Because of the high importance of the results of this study for the dental health, additional RCTs with large samples are needed to confirm this positive effect of lingual appliance treatment. In another retrospective study, it was found that the incidence of demineralization under the lingual bracket base was considerably reduced in children and adolescents through the addition of hydrophilic resin during the bonding procedure (28).

The other studies in the second subject group investigated various issues including effectiveness, treatment duration, anchorage loss, and bond failures. A retrospective study tested in a small group of patients the effect of lingual technique + Herbst on mandibular incisor proclination and concluded that the incisors are positioned according to the planned movement (22). Another retrospective study reported similar occlusal outcome and treatment duration with lingual (Incognito) and labial appliances (23). Two studies tested differences in anchorage loss in premolar extraction cases between lingual (STb or bidimensional, Ormco) and labial appliances and reported almost double amount of anchorage loss in labial treatment (24, 30). One prospective study found that labial and lingual (Forestadent) 2D treatment induces in most cases similar changes to the lower dental arch in Class I patients with crowding (25). A retrospective study concluded that treatment with lingual (STb) and labial appliances result in similar PAR reduction, treatment duration, and root resorption (31). In another study, the number of lost brackets during the first year of treatment was found to be similar for lingual and labial appliances (32). Finally, a recent retrospective study compared two different lingual systems and showed that treatment...
Table 1. Characteristics of included studies grouped according to their main subject: (a) accuracy testing, (b) various clinical aspects testing, and (c) periodontal parameters testing.

<table>
<thead>
<tr>
<th>Study</th>
<th>Subject group</th>
<th>Main objective</th>
<th>Place of sample collection</th>
<th>Study design</th>
<th>Treatments tested and sample size</th>
<th>Method of sample recruitment</th>
<th>Main type of collected data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grauer and Proffit, Am J Orthod Dentofacial Orthop 2011 (19)</td>
<td>Accuracy</td>
<td>Accuracy of Incognito lingual orthodontic technique</td>
<td>Bad Essen, Germany</td>
<td>Retrospective case series</td>
<td>Incognito lingual treatment (n = 94)</td>
<td>Consecutive patients debonded between January 2008 and January 2009</td>
<td>Tooth position in the setup and actual outcome measured at scanned dental casts</td>
</tr>
<tr>
<td>Pauls AH, J Orofac Orthop 2010 (20)</td>
<td>Accuracy</td>
<td>Accuracy of Incognito lingual orthodontic technique</td>
<td>Bad Essen, Germany</td>
<td>Retrospective case series</td>
<td>Incognito lingual treatment (n = 25)</td>
<td>Selection based on specific criteria</td>
<td>As above</td>
</tr>
<tr>
<td>Thalheim and Swestka-Polly, Inf Orthod Kieferorthop 2008 (21)</td>
<td>Accuracy</td>
<td>Accuracy of Incognito lingual orthodontic technique</td>
<td>Bad Essen, Germany</td>
<td>Retrospective case series</td>
<td>Incognito lingual treatment (n = 20)</td>
<td>Selection based on specific criteria</td>
<td>Lower intercanine distance in the setup and actual outcome measured at scanned dental casts</td>
</tr>
<tr>
<td>Lossdörfer et al., J Orofac Orthop 2013 (29)</td>
<td>Accuracy</td>
<td>Accuracy of Incognito lingual orthodontic technique</td>
<td>Bad Essen, Germany</td>
<td>Retrospective case series</td>
<td>Incognito lingual treatment (n = 34)</td>
<td>Selection based on specific criteria</td>
<td>Lower incisor inclination in the setup and actual outcome measured at scanned dental casts superimposed into cephalograms</td>
</tr>
<tr>
<td>Wiechmann et al., Head Face Med 2010 (22)</td>
<td>Accuracy and clinical</td>
<td>Effect of the Incognito lingual technique + Herbst on mandibular incisor proclination</td>
<td>Bad Essen, Germany</td>
<td>Retrospective case series</td>
<td>lingual treatment + Herbst (n = 12)</td>
<td>Selection based on specific criteria</td>
<td>Lower incisor position before treatment, in the setup and after treatment, measured at scanned dental casts PAR scores measured in the setup and actual outcome (final) casts for lingual treatment and between initial and final casts for both groups</td>
</tr>
<tr>
<td>Smith C, dissertation 2010 (23)</td>
<td>Accuracy and clinical</td>
<td>Accuracy of Incognito lingual orthodontic technique; occlusal outcome and treatment duration with lingual and labial appliances</td>
<td>Illinois, USA</td>
<td>Retrospective case series</td>
<td>Incognito lingual (n = 21) versus various labial (n = 22)</td>
<td>Consecutive selection based on specific criteria</td>
<td>Anchorage loss (amount of mesial movement of the upper first molar) measured in cephalograms and dental casts</td>
</tr>
<tr>
<td>Geron et al., Angle Orthod 2003 (24)</td>
<td>Clinical</td>
<td>Anchorage loss relative to extraction site (first versus second premolar), appliance (lingual versus labial), age (adolescents versus adults), overjet and crowding</td>
<td>Tel Aviv, Israel</td>
<td>Retrospective case series</td>
<td>Bidimensional lingual (Ormco; n = 13) versus labial (Victory System, 3M Unitek; n = 15)</td>
<td>Selection based on specific criteria</td>
<td>Anchorage loss (amount of mesial movement of the upper first molar) measured in cephalograms and dental casts</td>
</tr>
<tr>
<td>Venkatesh et al., APOS Trends Orthod 2015 (30)</td>
<td>Clinical</td>
<td>Anchorage loss during first premolar extraction space closure with labial and lingual appliances</td>
<td>Karnataka, India</td>
<td>Prospective CCT</td>
<td>STb lingual (Ormco; n = 10) versus labial (Victory System, 3M Unitek; n = 10)</td>
<td>Selection based on specific criteria</td>
<td>Anchorage loss (amount of mesial movement of the upper first molar) measured in cephalograms</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Subject group</th>
<th>Main objective</th>
<th>Place of sample collection</th>
<th>Study design</th>
<th>Treatments tested and sample size</th>
<th>Method of sample recruitment</th>
<th>Main type of collected data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soldanova et al., Eur J Orthod 2011 (25)</td>
<td>Clinical</td>
<td>Changes in the lower dental arch with the labial or the lingual 2D technique</td>
<td>Prague, Czech republic</td>
<td>Prospective CCT</td>
<td>lingual 2D (Forestadent; ( n = 25 )) versus labial (Mintrim Roth, Dentaurum; ( n = 25 ))</td>
<td>Selection based on specific criteria</td>
<td>Cephalometric measurements and images of dental casts obtained pre and post-treatment</td>
</tr>
<tr>
<td>van der Veen et al., Eur J Oral Sci 2010 (6)</td>
<td>Clinical</td>
<td>Caries incidence after treatment with lingual or labial appliances</td>
<td>Bad Essen, Germany</td>
<td>RCT (split mouth)</td>
<td>Incognito lingual ( (n = 14) ) versus labial (Orthos, Ormco; ( n = 14 ))</td>
<td>Consecutive selection based on specific criteria</td>
<td>White spot caries lesions measured at white light photographs and quantitative light-induced fluorescence images pre and post-treatment</td>
</tr>
<tr>
<td>Beyling et al., Head Face Med 2013 (28)</td>
<td>Clinical</td>
<td>To test the effect of the addition of an extra layer of a hydrophilic resin, during bonding, in the demineralization beneath the lingual bracket base</td>
<td>Bad Essen, Germany</td>
<td>Retrospective case series</td>
<td>Incognito lingual bonded conventionally ( (n = 20) ) versus bonded with an extra resin layer ( (n = 20) )</td>
<td>Consecutive selection based on specific criteria</td>
<td>Sub bracket lesions measured at standardized photographs, pre and post-treatment, in the six maxillary anterior teeth</td>
</tr>
<tr>
<td>Deguchi et al., Angle Orthod 2015 (31)</td>
<td>Clinical</td>
<td>Differences in cephalometric values, root resorption, occlusal indices, and functional aspects, between labial and lingual treatment</td>
<td>Japan</td>
<td>Retrospective case series</td>
<td>STb lingual (Ormco) ( (n = 24) ) versus labial ( (n = 25) )</td>
<td>Consecutive selection based on specific criteria</td>
<td>Cephalometric values, degree of root resorption, PAR, objective grading system (OGS), and functional evaluation pre and post-treatment</td>
</tr>
<tr>
<td>Ziebura et al., Am J Orthod Dentofacial Orthop 2014 (32)</td>
<td>Clinical</td>
<td>To compare frequency and position of bond failures in patients treated with lingual or buccal appliances</td>
<td>Munster, Germany</td>
<td>Retrospective case series</td>
<td>Incognito lingual ( (n = 59) ) versus Mini Diamond Brackets/Accent Molar Tubes (Ormco) labial ( (n = 44) )</td>
<td>Selection based on specific criteria</td>
<td>Number and position of lost brackets in the first year of treatment</td>
</tr>
<tr>
<td>Knösel et al., Head Face Med 2014 (18)</td>
<td>Clinical</td>
<td>To compare treatment duration with two types of customized lingual appliances</td>
<td>Bad Essen, Germany</td>
<td>Retrospective case series</td>
<td>Incognito lingual (3M Unitek) ( (n = 220) ) versus WIN lingual (DW Lingual Systems) ( (n = 136) )</td>
<td>Selection based on specific criteria</td>
<td>Treatment duration</td>
</tr>
<tr>
<td>Demling et al., Angle Orthod 2010 (26)</td>
<td>Periodontal health</td>
<td>Short-term effect of lingual appliances on periodontal and microbial parameters</td>
<td>Hannover, Germany</td>
<td>Prospective CCT (control: maxilla and buccal sites)</td>
<td>Incognito lingual versus no appliances ( (n = 20) ); split mouth</td>
<td>Consecutive selection based on specific criteria</td>
<td>Clinical periodontal evaluation and microbial analysis before and 4 weeks after bonding</td>
</tr>
</tbody>
</table>
### Table 1. Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Subject group</th>
<th>Main objective</th>
<th>Place of sample collection</th>
<th>Study design</th>
<th>Treatments tested and sample size</th>
<th>Method of sample recruitment</th>
<th>Main type of collected data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miethke and Brauner, J Orofac Orthop 2007 (27)</td>
<td>Periodontal health</td>
<td>Periodontal health of patients treated with Invisalign or lingual appliances</td>
<td>Berlin, Germany</td>
<td>Prospective CCT (control: second, fourth quadrant)</td>
<td>Ormco 7gen lingual (n = 30) versus Invisalign (n = 30)</td>
<td>Consecutive selection based on specific criteria</td>
<td>Clinical periodontal evaluation after at least 6 months during treatment at three consecutive appointments (3–4 weeks interval)</td>
</tr>
</tbody>
</table>

### Table 2. Results of individual studies, quality assessment, and synthesis of results.

<table>
<thead>
<tr>
<th>Study</th>
<th>Subject group</th>
<th>Definition of pre-specified main outcome</th>
<th>Summary outcome data</th>
<th>Additional outcomes</th>
<th>Quality assessment*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grauer and Profitt, Am J Orthod Dentofacial Orthop 2011 (19)</td>
<td>Accuracy</td>
<td>Level of accuracy and effect of age, gender, crowding, overbite, overjet, ANB, days in treatment, days in slot-filling wire, Class II elastics, vertical elastics, interproximal reduction, rebondings, jaw, and tooth type</td>
<td>In most cases, anterior teeth were positioned within 1 mm and within 4° of their planned positions; differences increased from anterior to posterior; highly significant differences in all discrepancies except tooth long-axial rotation for the maxilla versus the mandible, and in all parameters for tooth type</td>
<td>Various parameters were statistically related to the amount of rotational and translational discrepancy, but each of these factors explained only a small amount of the total discrepancy</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>Pauls AH, J Orofac Orthop 2010 (20)</td>
<td>Accuracy</td>
<td>Level of accuracy and effect of age, jaw, various finishing wires, bracket generation, original front tooth axial position, change in the front tooth inclination during lingual treatment, type of tooth, type of scanner</td>
<td>Mean deviations of set-up from end result were: tip (UJ/upper jaw 5.53°; LJ/lower jaw 5.72°), torque (UJ 5.16°; LJ 5.23°), supra/infraposition (UJ 0.89 mm; LJ 0.97 mm), rotation (UJ 5.82°; LJ 5.40°), in/out (UJ 1.19 mm; LJ 1.07 mm) and anterior/posterior position (UJ 1.04 mm; LJ 1.00 mm); significant difference between jaws in the antero-posterior position only</td>
<td>Significant differences for final archwires and different bracket generations and tooth types; the upper central incisors yielded the lowest values after the lower front teeth with deviations in rotation less than 4.5° and in translation less than 0.40 mm; from anterior to posterior, feasibility decreased by approx. 2° and 1 mm, with the second molars showing the poorest values</td>
<td>High risk</td>
</tr>
<tr>
<td>Thalheim and Swezka-Polly, Inf Orthod Kieferorthop 2008 (21)</td>
<td>Accuracy</td>
<td>Effect of lingual treatment on lower intercanine distance</td>
<td>Difference in intercanine distance of set-up versus final result 0.44 ± 0.29 mm Difference of clinically achieved versus planned inclinations was 5.89 ± 5.43°, after correcting for changes in occlusal plane inclination. It was 3.75 ± 3.06° versus 8.59 ± 6.58° for full-size versus undersized finishing archwires and 7.88 ± 6.36° versus 4.65 ± 4.48° for extraction versus non-extraction cases</td>
<td></td>
<td>High risk</td>
</tr>
<tr>
<td>Lossdörfer et al., J Orofac Orthop 2013 (29)</td>
<td>Accuracy</td>
<td>Changes in lower incisor inclination during treatment and differences between set-up and treatment result</td>
<td></td>
<td></td>
<td>High risk</td>
</tr>
<tr>
<td>Study</td>
<td>Subject group</td>
<td>Definition of pre-specified main outcome</td>
<td>Summary outcome data</td>
<td>Additional outcomes</td>
<td>Quality assessment*</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Wiechmann et al., Head Face Med 2010 (22)</td>
<td>Accuracy and clinical</td>
<td>Changes in lower incisor inclination during lingual + Herbst treatment and differences between set-up and treatment result</td>
<td>Planned incisor inclination: $-16.5^\circ$ to $5.8^\circ$; difference between planned and final incisor rotations was $2.2^\circ \pm 1.0^\circ$ ($P &gt; 0.05$)</td>
<td>-</td>
<td>High risk</td>
</tr>
<tr>
<td>Smith C, dissertation 2010 (23)</td>
<td>Accuracy and clinical</td>
<td>PAR score in lingual and labial treatment</td>
<td>treatment time: lingual $22.35 \pm 5.19$ m, labial $22.35 \pm 5.15$ m (ns); total reduction in PAR score: lingual $15.24 \pm 7.71$, labial $18.27 \pm 11.15$ (ns); percentage of PAR reduction: lingual $77.51 \pm 23.11%$, labial $79.83 \pm 19.34%$ (ns); diagnostic set-up (0.4) versus post-tx (4.5) total weighted PAR ($P &lt; 0.05$)</td>
<td>There was no significant difference between any of the individual PAR components and in total weighted PAR scores in the two groups pre-treatment; the only difference post-treatment was in overbite ($P &lt; 0.05$; labial mean PAR = 0; lingual = 0.29)</td>
<td>High risk</td>
</tr>
<tr>
<td>Geron et al., Angle Orthod 2003 (24)</td>
<td>Clinical</td>
<td>Anchorage loss (amount of mesial movement of the upper first molar) during lingual treatment</td>
<td>Cephalograms: lingual $1.84 \pm 1.4$ mm, labial $3.0 \pm 1.4$ mm ($P &lt; 0.05$); casts: lingual $2.4 \pm 1.9$, labial $3.9 \pm 2.7$ (ns)</td>
<td>—</td>
<td>High risk</td>
</tr>
<tr>
<td>Venkatesh et al., APOS Trends Orthod 2015 (30)</td>
<td>Clinical</td>
<td>Anchorage loss in first premolar extraction cases (mesial movement of the upper first molar)</td>
<td>Lингual: $1.24 \pm 0.17$ mm, labial: $2.26 \pm 0.39$ mm ($P = 0.001$)</td>
<td>—</td>
<td>High risk</td>
</tr>
<tr>
<td>Soldanova et al., Eur J Orthod 2011 (25)</td>
<td>Clinical</td>
<td>Cephalometric and cast measurements in lingual and labial treatment</td>
<td>Significant difference in the position of incisors relative to A-Po (lingual: $1.6 \pm 1.5$ mm; labial: $0.0 \pm 1.5$ mm $P &lt; 0.05$); no other differences in position of incisors, canines or molars</td>
<td>Active treatment time: lingual: $14.6 \pm 6.9$ m; labial: $15.44 \pm 7.3$ m</td>
<td>High risk</td>
</tr>
<tr>
<td>van der Veen et al., Eur J Oral Sci 2010 (6)</td>
<td>Clinical</td>
<td>Presence/absence or progress of lesions in lingual and labial treatment</td>
<td>The number of caries that formed (buccal: $n = 16$, lingual: $n = 1$) or progressed (buccal: $n = 21$, lingual: $n = 7$) was higher in buccal compared to lingual lesions ($P = 0.01$); the overall increase in caries extent/total lesion area was $10.6$ times higher on buccal surfaces in a subject than on lingual surfaces ($P &lt; 0.05$)</td>
<td>The number of new lesions developing or progressing on bracketed buccal surfaces was $4.8$ times higher and the caries extent, or integrated fluorescence loss, was $10.6$ times higher than in lingual surfaces</td>
<td>High risk</td>
</tr>
<tr>
<td>Beyling et al., Head Face Med 2013 (28)</td>
<td>Clinical</td>
<td>Incidence of demineralization beneath the lingual bracket base at the end of treatment with two bonding protocols</td>
<td>With standard indirect chemical cure bonding, 18 lesions occurred in 120 surfaces (15%). With the addition of an extra layer of a hydrophilic resin, 6 lesions occurred in 120 surfaces (5%), $P = 0.01$</td>
<td>Gender and age were not significantly related to the development of new lesions</td>
<td>High risk</td>
</tr>
</tbody>
</table>
### Table 2. Continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Subject group</th>
<th>Definition of pre-specified main outcome</th>
<th>Summary outcome data</th>
<th>Additional outcomes</th>
<th>Quality assessment*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deguchi et al., Angle Orthod 2015 (31)</td>
<td>Clinical</td>
<td>Cephalometric values, degree of root resorption, PAR, OGS, and functional evaluation in lingual and labial treatment</td>
<td>Pre-treatment PAR, labial: 25.0±9.6, lingual: 28.0±7.2 (ns); post-treatment PAR, labial: 4.0±2.2, lingual: 4.2±2.2 (ns). Pre-treatment cephalometric measurements (ns). Post-treatment only interincisal angle (T0–T1, labial: 11.2±11.6; lingual, 17.9±8.9, <em>P</em> &lt; 0.05). Root resorption, labial: 1.1±0.5 mm, lingual: 1.3±0.6 mm (ns). Jaw movements increased in the same way in both groups.</td>
<td>Average treatment time, labial: 29.4±5.6 m, lingual: 32.5±6.7 m (ns)</td>
<td>High risk</td>
</tr>
<tr>
<td>Zaebura et al., Am J Orthod Dentofacial Orthop 2014 (32)</td>
<td>Clinical</td>
<td>Number and positions of lost brackets in the first year of lingual and labial treatment</td>
<td>Bond failures per patient lingual: 2.63±2.77, buccal: 2.61±3.41 (<em>P</em> = 0.428)</td>
<td>Molar brackets failed more than premolar (<em>P</em> &lt; 0.05); premolar brackets more than canine and incisor (<em>P</em> &lt; 0.05); Tendency for more failures in the mandible (lingual: <em>P</em> = 0.069; buccal: <em>P</em> = 0.036); Occlusal pads in lingual lowered failures in molars, but not in premolars; Significant influence of 'appliance type', and 'case severity' on treatment duration, not dependent on each other; no difference in frequencies of bracket losses</td>
<td>High risk</td>
</tr>
<tr>
<td>Knösel et al., Head Face Med 2014 (18)</td>
<td>Clinical</td>
<td>Treatment duration with two types of customized lingual appliances</td>
<td>Treatment duration WIN (Severity1: 17.96 m; S2: 20.49 m; Incognito (S1: 22.7 m; S2: 29.79 m) <em>P</em> &lt; 0.001</td>
<td>No difference in control sites in any of the tested variables between t0 and t1</td>
<td>High risk</td>
</tr>
<tr>
<td>Demling et al., Angle Orthod 2010 (26)</td>
<td>Periodontal health</td>
<td>Changes in periodontal and microbial parameters after 1 month of lingual treatment on the lower arch</td>
<td>Mandible lingual side Pf: t0 = 0.3±0.3, t1 = 1.0±0.7 (<em>P</em> = 0.001); BOP: t0 = 23.4±22.5, t1 = 46.2±23.5 (<em>P</em> = 0.001); PPD: t0 = 2.1±0.4, t1 = 2.2±0.3 (ns); prevalence of Aa: t0 = 25%, t1 = 35%; prevalence of Pg: t0 = 5%, t1 = 5%</td>
<td>All the indices improved significantly between the first and third assessments in the Invisalign patients, while they were significantly worse in lingual treatment</td>
<td>Unclear risk</td>
</tr>
<tr>
<td>Miethke and Brauner, J Orofac Orthop 2007 (27)</td>
<td>Periodontal health</td>
<td>Difference in gingival index (GI), papillary bleeding index (PBI), Plaque index (PI), sulcus probing depth (SPD) between lingual and Invisalign treatment at three consecutive appointments (T1, T2, T3)</td>
<td>The SPD of approximately 2.5 mm was similar in both groups in all time points</td>
<td>All the indices improved significantly between the first and third assessments in the Invisalign patients, while they were significantly worse in lingual treatment</td>
<td>Unclear risk</td>
</tr>
</tbody>
</table>

*Low risk, unclear risk, or high risk of bias study; ns, non-significant.
duration was significantly smaller with WIN compared to Incognito system (18). The results of these investigations, however, have to be interpreted taken into account the possibility of the influence of the various biases on the study outcomes.

The third subject group included two prospective CCTs that investigated the effect of lingual orthodontic treatment on various periodontal parameters. The first study tested, in a split-mouth design, the short-term influence of Incognito lingual appliances on various clinical periodontal and microbial indices and showed a worsening of plaque index and bleeding on probing in the bonded sites, while no significant difference was detected on probing depth and periodontopathogenic bacteria (26). The other study tested the effect of Ormco seventh generation lingual treatment as well as Invisalign treatment on clinical periodontal parameters. All the indices apart from probing depth improved significantly between assessments in the Invisalign patients, while they were significantly worse in lingual treatment in all time points (27).

Concerning the interpretation of the findings of these two studies, however, it must be taken into consideration that both are single non-randomized studies with unclear risk of bias.

Limitations
There is a shortage of large, high-quality prospective clinical trials investigating clinical outcomes and effects of lingual orthodontic treatment (only one RCT was available). Thus, the majority of the included studies had a retrospective design and the risk of bias was high. An assessment of high risk of bias does not necessarily mean that the study results are biased; it means that the suspicion of biased results is higher compared to those of a study classified at low risk of bias.

In most cases, no synthesis of studies was feasible since there was only a single study on each subject. In the presence of more than one study on a subject, grouping and/or synthesis of results and conclusions was not feasible due to the clinical heterogeneity of the included studies (e.g. different bracket systems are tested).

Language restrictions might be an additional limitation in the present systematic review.

Conclusions
This systematic review showed encouraging results on the clinical outcome of lingual orthodontic treatment, especially in regards to the achievement of individualized treatment goals and the decreased risk of decalcifications on the bonded surfaces of the teeth. However, additional well-designed prospective clinical trials with large samples are needed to confirm those findings. Several aspects of lingual orthodontic treatment were difficult to be conclusively evaluated due to the study design, the heterogeneity, the small samples, and the high risk of bias seen in the majority of the included studies.

Supplementary material
Supplementary material is available at European Journal of Orthodontics online.
References


