Complications, impact and success rate of approaches to treatment of Class II malocclusion in adolescents: A systematic review & meta-analysis

Moaiyad Moussa Pacha, Padhraig S. Fleming, and Ama Johal

Barts and London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

Introduction: We aimed to explore the prevalence and nature of complications associated with Class II correctors in adolescents, and their impact on quality of life (QoL), completion of treatment and success rate. **Methods**: The review was registered in PROSPERO, and a comprehensive electronic search was performed without language or date restrictions. Randomized and non-randomized trials, prospective cohort and cross-sectional studies, case series, and qualitative research were included. The Cochrane Collaboration's Risk of Bias Tool and Newcastle-Ottawa Scale were used to assess the quality of included studies. Data were grouped according to appliances design: removable functional, fixed functional, hybrid functional, headgear and fixed maxillary molar distalization appliances. **Results**: Data from 27 studies were included, of which 11 were deemed eligible for meta-analysis. Overall, 1,676 adolescents were included related to fixed functional (n=682), removable functional (n=682), hybrid functional (n=84), headgear (n=186) and Carriere (n=42) appliances. The mean number of emergencies was 0.8 (95% CI, 1.1, 2.1) and 2 (95% CI, 0.9, 3.0) for removable and fixed designs, respectively. However, the rate of discontinuation was 35% (95% CI, 0.28, 0.42) and just 1% (95% CI, 0.01, 0.1) for removable and fixed designs, respectively. Other QoL dimensions such as eating, sleep, speech and emotional domains were significantly impaired during treatment with removable functional appliances. **Conclusion**: Removable Class II correctors were associated with high rate of treatment discontinuation, most likely due to negative impact on QoL and lack of compliance. More complications were observed with fixed designs, although this did not impact the overall success rates. Further prospective studies are needed to explore patient perceptions and cost-effectiveness to better inform treatment decisions.

Class II malocclusion has a prevalence of approximately 25% of 12-year-olds in the U.K.[1] Affected children are more likely to experience teasing, with resultant psychological harms and distress,[2] and negative connotations for self-esteem and the quality of life (QoL) for both child and family.[3, 4]

Currently, Class II malocclusion in growing adolescents can be treated using a wide variety of appliances and techniques including Class II correction appliances and/or fixed multi-bracket appliance with a combination of selective extraction and/or the use of inter-arch elastics. Removable, fixed, intra- and inter- maxillary Class II correctors exist. Intra-maxillary appliances are routinely fitted to the first maxillary molars aiming to either restrain the forward growth of the maxilla, e.g. the removable headgear, or to distalize the maxillary molars, using fixed devices (e.g. Pendulum, Carriere, and Distal Jet). Functional appliances have an inter-maxillary design and can be grouped into: (i) removable, e.g. Twin Block (TB), Activator and prefabricated appliance, (ii) hybrid, e.g. Dynamax, and (iii) fixed. However, the fixed design can be further sub-classified into fixed rigid, e.g. Herbst, fixed-Twin Block (F-TB) and mandibular protraction appliance (MPA), fixed flexible designs, e.g. Forsus fatigue resistance device (FFRD).The primary difference between fixed and removable Class II correctors is the premium on compliance with removable variants whilst full-time wear is guaranteed with the fixed design.

There is widespread consensus that these appliances and techniques can successfully correct Class II characteristics including the reduction of the overjet to within normal limits by producing a combination of dental and skeletal effects.[5] Moreover, there is also evidence of an association with complications and negative impact on oral health related quality of life (OHRQoL),[6] with successful treatment dependent on patient acceptance and their ability to adapt to the appliance.[7]

It is increasingly accepted that orthodontic research tends to be overly focussed on clinician-centred outcomes, e.g. cephalometric changes rather than those that matter more to patients.[8] Furthermore, a recent systematic review comparing the effectiveness of fixed and removable functional appliances concluded that there was a lack of data relating to patient experiences and potential complications during Class II treatment, and hence emphasized the need for further prospective studies focusing on these outcomes in order to better inform the consent process and treatment decisions. [9]

Whilst several studies have been carried out to assess the impact of Class II malocclusion on the OHRQoL, [2 3] others evaluated the potential role of overjet correction in improving the QoL of affected children.[10] However, there is a little consensus concerning key patient-reported outcomes during Class II treatment. Therefore, the aim of this systematic review was to assess complications and patient experiences during the wear of Class II correctors, and to evaluate their impact on OHRQoL and success rate, in order to evaluate the efficacy and safety of appliances when considering Class II correction.

METHODS AND MATERIALS

Protocol and registration

The protocol for this systematic review was registered with PROSPERO (www.crd.york.ac.uk/PROSPERO, CRD42019121628).

Eligibility criteria:

The following PICO selection criteria where applied:

Participants: Adolescents (under 18 years old) with a Class II malocclusion.

Intervention: Orthodontic treatment with a Class II correction appliance, including any type of functional appliances (fixed or removable), headgear, maxillary molar distalizing device or other Class II correctors.

Comparator: A comparison and/or control group was not essential

Outcomes: (i) Nature, severity (minor, moderate, severe, or treatment failure) and prevalence of emergencies/complications, e.g. pain, harms, breakages and other complications associated with the treatment (ii) Patient experiences and impact of appliance wear on OHRQoL (iii) Impact of associated complications and/or OHRQoL domains on patient compliance and discontinuation of treatment (success rate).

Study design: The following study designs were considered eligible if they incorporated subjective data on complications associated with the Class II treatment and impact on OHRQoL: Randomized controlled trials (RCTs), non-randomized controlled clinical trials (CCTs), observational cohort and cross-sectional studies, and prospective case series (minimum sample size of 10 patients). Qualitative studies exploring patients' views and experiences during Class II treatment were also included. Only data relating to patients'

experiences during Class II correction and treatment discontinuation were extracted, with clinician-centred outcomes omitted from the review.

Exclusion criteria: Studies were excluded if compliance only was measured, if treatment was confined to fixed multiple-bracket appliance alone (e.g. with Class II elastics) and/or simple removable appliance, explored the effects of treated/untreated maloclussion on OHRQoL, involved the effects on the temporomandibular joint, or involved participants with craniofacial growth syndromes.

Information sources, search strategy and study selection

The search strategy included terms related to orthodontic complications, experiences and quality of life during Class II correction involving any type of removable or fixed functional appliance, headgear and other Class II corrector (Appendix 1). Comprehensive searches, without date restrictions, were conducted using the following electronic databases: MEDLINE via PubMed and Ovid, Web of Science, Cochrane, and Embase. A gray literature search was also undertaken using Google Scholar and OpenGrey. Hand searching was performed from the reference lists of the full text articles considered eligible for inclusion and other relevant systematic reviews. Assessments of studies for inclusion in the review were performed independently and in duplicate by two authors (M.M.P; A.J), and any disagreements were resolved by discussion with a third author (P.S.F) who was not involved with the original screening. If further information regarding patient experiences or OHRQoL were required, authors were contacted for clarification.

Risk of bias and quality assessment in individual studies

Two authors independently assessed the risk of bias of the included studies (M.M.P; A.J) and any disagreements were resolved by further discussion and consensus. Due to diversity in the design of included studies, two different tools were used to assess their quality. Randomized controlled trials were assessed using the Cochrane Collaboration's risk of bias tool.[11] The following seven domains were considered: sequence generation, allocation concealment, blinding participants, blinding of outcome assessors, incomplete outcome data, selective reporting and other biases.

An appropriately modified version of the Newcastle-Ottawa scale was used to assess the quality of non-randomized studies including the adapted version of the scale suitable for the assessment of cross-sectional studies.[12] This tool compromise of 9 domains, which are divided

into three broad criteria: patient selection, comparability of study groups, and outcome assessment. Studies are awarded stars according to their methodological quality with a high-quality study achieving the maximum score of 9 stars. Studies achieving between 6 and 8 stars were considered of moderate quality whilst a rating of 5 stars or less indicated low quality. Studies at low or unclear risk of bias, or medium to high quality were planned for inclusion in any subsequent meta-analysis.

Data items and collection

The following characteristics were recorded: study design; sample size; setting; treatment duration; participant details; type of treatment and outcomes relevant to experiences during treatment (e.g. complications and impact of treatment on OHRQoL). Data was extracted and grouped according to their design as removable functional, fixed functional, hybrid functional, headgear and fixed maxillary molar distalization appliance. Measurements or outcomes that related to morphological changes during treatment were omitted. Complications and emergencies during treatment were categorised according to their severity as follows: 1) minor: function of the appliance was not impaired and repair could be undertaken during the same appointment, 2) moderate: function was impaired but emergencies could be resolved at the same appointment at chair-side, 3) severe: function was impaired and repair require any laboratory input, and 4) treatment failure: complication leading to termination of treatment (Table I).

Summary measures and approach to statistical analysis

Data concerning patient experiences of treatment, e.g. number, type and severity of emergency visits were expressed as a number or percentage (prevalence) of all treated patients. OHRQoL scores reporting on the same domain, were combined to obtain pooled mean proportion values, with standard deviation and/or 95% confidence intervals (95% CI) if applicable. Data from qualitative studies was planned for synthesis if the same outcome was reported in more than two studies, followed by integration of quantitative and qualitative results.

Risk of bias across studies

To identify publication bias, standard and contoured enhanced funnel plots were to be drawn if sufficient numbers of studies were identified (>10 studies).

Additional analysis

A meta-analysis was planned for studies with low and/or unclear risk of bias, and moderate to high-quality studies, using similar design and reporting the same outcome, after grouping data according to Class II corrector classification, to estimate the overall impact on the rate/number of emergencies, treatment discontinuation and OHRQoL (patient experiences). However, due to the proportional nature of these outcomes and the use of single-arm data, random-effects specific meta-analysis was undertaken using bespoke software (OpenMeta [Analyst], open source software, Brown University of Public Health, RI, USA). Results were to be presented as forest plots with weighted values and 95% confidence intervals (CIs), with a P-value of less than 0.05 being considered statistically significant. The I^2 statistic test was applied to quantify heterogeneity among studies, with an I^2 value of up to 35% indicating low, 35 - 70% moderate and 70 - 100% high heterogeneity. However, it should be noted, this test reflected particularly the variation in the sample size and the proportional nature of the outcome, e.g. frequency of reported complications, rather than the variability of the comparable outcomes where there was a control group.

RESULTS

Study selection and characteristics of included studies

A total number of 461 studies were initially identified, with a further 19 articles obtained through other sources, of which 51 full-text were evaluated for inclusion (Figure 1). Twenty-two full-text articles were excluded (Appendix 2). Twenty-nine articles met the selection criteria; however, two articles reported on data from the same study,[13,14] with a further two related to the same cohort study.[15,16] Thus, 27 unique datasets were eventually included in the qualitative and quantitative synthesis.[6,13,15,17-40] Of these, nine were RCTs,[6,13,17,18,23,24,30,34,39] six were CCTs,[25,26,33,36-38] one was a case series,[29] eight were prospective observational studies,[15,19,27,28,31,32,35,40] and three were qualitative studies.[20-22] The vast majority of included studies were carried out in a university or hospital setting apart from four studies were undertaken in private practices.[13,24,33,40] The appliances used varied significantly, with some studies evaluating a single treatment and others comparing two or three different types of Class II correction appliances (Appendix 3; Table II).

Overall, 1,676 participants were included, with the majority having received functional appliances. The fixed functional design (n=682) involved Herbst (n=442), FFRD (n=154), MPA (n=54) and F-TB (n=32). The removable functional design (n=682) involved TB

(n=347), Activator (n=153), Bionator (n=79), prefabricated design (n=83) and unknown type (n=20). The hybrid functional design (n=84) involved Dynamax (n=63) or the Herbst appliance with a removable mandibular plate (n=21). The remainder utilized headgear (n=186) and the Carriere appliance (n=42, Tables II & III).

Risk of bias within studies

The methodological quality of the included randomized controlled trials (n=9) is shown in Figure 2. Only one study was considered to be of low risk of bias,[23] with two studies considered to be unclear.[6,18] (Appendix 4). Of the non-randomized studies (n= 15), six studies were judged of low quality,[19,25,27,29,33,38] and nine were of moderate quality.[15,26,28,31,32,35-37,40] (Figure 3; Appendix 5). Finally, of the included studies, three were qualitative in nature involving semistructured interviews,[20-22] and therefore a risk of bias assessment was not indicated. Nevertheless, the participants were derived from ongoing randomized controlled trials on the effect of removable functional appliances.

Overall, 12 included studies were considered eligible for meta-analysis.[6,15,18,23,26,28,31,32,35-37,40] However, one study was excluded because, despite having a moderate risk of bias, there was incomplete reporting of interventions and outcome data precluding combination with the other measurements.[15]

Results of individual studies, meta-analysis, and additional analysis

Complications and emergencies associated with treatment:

Complications rate was found to be high in both hybrid and fixed varieties of functional appliances, with a prevalence of 69% and 34%, respectively. Sixteen percent of these complications were minor, 22% were moderate while only 5% were severe requiring re-make of the appliance (Tables II & III). However, only a limited number of studies (n=6) reported the number of emergencies associated with removable functional appliances, with little data concerning the nature or severity of these complications.[6,18,28,34,35,39] It is noteworthy that in 2 studies the incidence of complications was reported retrospectively by the patient using customized questionnaires,[19,35] and hence the results should be interpreted with caution.

Meta-analysis was performed in relation to the overall mean number of emergencies derived from three removable functional appliances studies involving 141 participants,[6,18,28] and five fixed functional appliances studies involving 157 participants, [6,18,26,36,37] with a mean of 0.8 (95% CI, 1.1, 2.1) and 2 (95% CI, 0.9, 3.0), respectively (Figures 4 & 5). However, although

the included studies were at low or moderate risk of bias and data were collected prospectively from clinical notes, results should be interpreted with caution due to heterogeneity amongst the studies in both removable and fixed categories (I^2 = 88% and 94%, respectively), most likely due to the sample size variation.

Treatment discontinuation:

Treatment discontinuation rate was significantly higher with removable (39%) and hybrid (29%) functional appliances compared to fixed designs (4%; Tables II & III). This result is supported by a meta-analysis involving data from 4 studies with removable functional appliances involving 294 participants,[6,18,28,31] and 7 studies with fixed functional appliances involving 253 participants,[6,18,23,26,36,37,40] in which the rate of discontinuation was found to be 35% (95% CI, 0.28, 0.42) with the removable and just 1% (95% CI, 0.01, 0.1) with the fixed type (Figures 6 & 7). However, the level of heterogeneity amongst studies was low to moderate (I^2= 33.5% - 66%).

Headgear discontinuation was only reported in a single study, in which 50% of participants failed to complete their treatment; however, the overall number of participants was relatively small (n=90).[31]

Patient experiences and OHRQoL:

The OHRQoL during treatment were reported using a wide variety of questionnaires but surprisingly, only two studies used validated questionnaires,[17,32] namely the Child Perception Questionnaire 11-14 (CPQ_11-14) and Oral Health Impact Profile (OHIP), with the remaining studies relying on customized non-validated questionnaires.[6,13,15,19,23-25,27-31,33,35] Therefore, we were unable to combine OHRQoL data in a meta-analysis. However, a qualitative synthesis of data (Tables II & III) showed that participants treated with removable functional appliances were twice (59%) as likely to be concerned about their appearance compared to those receiving the fixed variety (30%). However, similar levels of concern were reported with both removable and fixed designs in relation to oral hygiene and mouth opening issues during treatment, with a rate of 50% and 67%, respectively. Pain prevalence varied according to the type of Class II corrector being reported at 29%, 48%, 54% and 70% with Carriere, fixed functional, removable functional and headgear, respectively. However, the sample size was small in both headgear and Carriere categories and therefore caution is needed when interpreting the results.

Sleep problems were slightly more prevalent with removable functional (26%) compared to fixed (23%) designs. In contrast, eating problems were significantly greater during fixed compared to removable functional appliances, with a range of 45% and 14%, respectively. Similarly, eating problems were reported among 22% of participants during headgear treatment reflecting the removable nature of the headgear. On the contrary, the removable functional was associated with more speech concerns (81%) compared to headgear (54%) and fixed functional appliances (27%).

Qualitative studies:

A limited number (n=3) of qualitative studies involving semi-structured interviews with patients undergoing treatment with removable functional appliances were included.[20-22] Carter *et al.,* (2015), in six patients, explored the impact on eating including process, time and restrictions, during appliance wear. Cirgic *et al*., (2015) interviewed twenty-one patients to investigate the impairment of OHRQoL during TB therapy. El-Huni *et al*., (2019), explored the physical (pain, discomfort, eating, speech, sleep) and psychological (embarrassment, bullying) impacts during TB treatment of twenty-two patients. A thematic synthesis of this data was considered not feasible; however, in general, all included qualitative studies reported a significant negative impact associated with removable functional treatment in relation to eating, pain, discomfort and other psycho-social effects, including embarrassment and being bullied (Table III). Nevertheless, El-Huni *et al*., (2019) reported that the initial negative patient experiences were often followed by a period of adaptation, with comfort levels improving and participants becoming more receptive to treatment over time, particularly as positive treatment induced changes become apparent.

Risk of bias across studies.

Tests for publication bias were not undertaken because no more than 10 studies were included in an individual meta-analysis.

DISCUSSION

A recent systematic review involved evaluation of the prevalence of complications during fixed Class II correctors wear in isolation,[41] concluding that most patients, particularly during Herbst appliance treatment, experienced a significant number of complications such as fracture and dislodgement, requiring additional emergency visits. The results of the present systematic review and meta-analyses support its finding but extends inclusion of all varieties of Class II

correctors, with holistic assessment of patient experiences and the associated impact on OHRQoL and success rate.

It is well known that patient cooperation with removable Class II correctors, such as removable functional and headgear appliances is essential to achieve effective results, with a lack of compliance, leading to a dramatic increase in treatment duration,[42] and an increased risk of failure to complete treatment. A significant rate of treatment discontinuation during treatment with removable extra- and intra-oral Class II correctors was observed in the present study. Similarly, premature termination of treatment with removable functional appliances due to lack of compliance has been observed in previous prospective research, based on either the lack of overjet reduction or frequency of emergencies/breakages during treatment.[43,44] Nevertheless, there has been little emphasis on patient perspective concerning means of Class II correction. However, based on the present systematic review, the success rate with fixed Class II correctors may be considerably higher than with removable alternatives. This may relate to the enforced nature of full-time wear leading to adaptation and acceptance of the appliance sooner than might be the case with removable variants.[22]

In the present review, only a limited number of studies involving treatment with headgear and maxillary molar distalization appliances (intra-maxillary appliances) were included,[24,27,31,32] with a preponderance of inter-maxillary appliances. This may reflect the reported lack of skeletal effects associated with headgear. Furthermore, we identified a higher level of pain with headgear, which may contribute to reticence among clinicians to recommend the appliance in view of the attendant impact on patient compliance.

We found that fixed functional appliances were associated with a significantly greater rate of complications and the need for more emergency visits compared to removable designs. While data describing the nature or severity of complications was briefly reported in the fixed Class II corrector studies, it was, surprisingly, entirely omitted from studies involve removable functional appliances. In addition to the number and prevalence of emergencies, we attempted to quantify the severity and complexity of complications according to their description in the included studies as mild, moderate, severe and failure. However, and unfortunately, reporting in this respect was often incomplete precluding further analysis. Most of the complications associated with fixed functional appliances were mild or moderate in nature being repairable at chair-side, with only 5% requiring laboratory intervention. Nevertheless, modification to

traditional designs of fixed functional appliances including the Herbst, have been introduced, aiming to reduce the frequency of complications and increase patient comfort.[45]

It appears that OHRQoL during Class II correction with removable designs deteriorates immediately, with significant impairment in functional, social, and emotional dimensions, resulting in high rates of treatment discontinuation. Based on the present review, this impairment in OHRQoL is found to be significantly higher with removable compared to fixed designs, although it is noteworthy that Class II characteristics, such as the increase of overjet, have been found to be associated with significant negative impact on emotional and social dimensions. As such, any diminution in OHRQoL is likely temporary in nature. Traditionally, patient experiences and the impairment of QoL during orthodontic treatment are usually measured indirectly, using generic OHRQoL questionnaires or modified versions, by evaluating the improvement of psychosocial and well-being aspects with little attention to the impact of appliances on treatment. Therefore, they may not address directly certain aspects of patient experiences with treatment.[46] In the current review, several unvalidated, customized questionnaires were found to have been used, with considerable heterogeneity in the assessed domains and statistical measurements, making pooling of the data impossible and precluding a meta-analysis. Hence, it is important for further prospective studies to use agreed, valid and reliable condition-specific questionnaires.[47]

More recently, qualitative research methods based on one-to-one interviews have been implemented in orthodontic research, attempting to assess and understand patient perspectives and experiences during treatment more clearly. Qualitative interviewing allows synergistic conversation and may therefore lead to more detailed information regarding treatment experience. Three qualitative studies involving treatment with removable functional appliances were identified, illustrating the facilitators and barriers to diligent wear allied to the experiences of treatment.[20-22] However, the need for further qualitative research, particularly in relation the fixed Class II correctors, is clear.

Limitations

As is often the case with orthodontic systematic reviews, only a limited number of randomized controlled studies were found to be eligible for inclusion. However, because the review has not been carried out to compare the effectiveness of Class II correctors but rather to better understand cross-sectional qualitative information during the treatment, other observational

study designs were also included such as prospective cohort, cross-sectional and qualitative research.

Another limitation was the quality of included studies. While the Cochrane Collaboration tool is considered a robust tool in assessing the internal validity and risk of bias in randomized controlled trials, there is controversy regarding the application of the Newcastle-Ottawa tool in assessing the quality of nonrandomized studies.[48] Similarly, the application of meta-analysis on the observational studies with single arms is debatable due to their potential bias and methodological issues, as well as the diversity in study designs, making the calculation of an estimate of effects potentially problematic.[49] Therefore, the interpretation of the findings should be considered with caution.

We were also unable to assess the cost-effectiveness of Class II correctors. It is well known that the fixed functional appliances, particularly Herbst incur additional costs in terms of materials and laboratory fees. Nevertheless, from the clinical point of view and based on our results, if the increased risk of discontinuation with the removable design were confirmed, it would be essential to evaluate the cost-effectiveness ratio between removable and fixed Class II correctors. For example, assuming that both treatment designs have similar treatment duration and levels of effectiveness, the relative impact of more complications and emergency visits versus fewer failures and lower discontinuation rate with fixed variants, warrants economic analysis.

CONCLUSION

- Removable Class II correctors are associated with a significantly higher rate of treatment discontinuation, most likely as a result of lack of compliance.
- The removable design is associated with significant negative impact on OHRQoL including oral symptoms, functional restrictions and emotional deterioration most likely contributing to the rate of treatment failure.
- Fixed Class II correctors are associated with high rates of complications and the need for further emergency visits, although this did not impact overall success rates.
- Further well-designed prospective studies focusing on patient perceptions and evaluating potential physical and psychological harms associated with Class II

correctors as well as exploring cost-effectiveness, are required in order to better inform treatment decisions.

ACKNOWLEDGMENTS

The authors report no commercial, proprietary, or financial interest in the products or companies described in this article.

REFERENCES

1. Holmes A. The subjective need and demand for orthodontic treatment. Br J Orthod 1992;19:287-297.
2. Seehra J, Fleming P, Newton T, DiBiase AT. Bullying in orthodontic patients and its relationship to malocclusion, self-esteem and oral health-related quality of life. J Orthod 2011;38:247-256.
3. Johal A, Cheung MY, Marcenes W. The impact of two different malocclusion traits on quality of life. Br Dent J 2007;202:E6-E6.
4. Liu Z, McGrath C, Hägg U. The impact of malocclusion/orthodontic treatment need on the quality of life: A systematic review. Angle Orthod 2009;79: 585-591.
5. Thiruvenkatachari B, Harrison J, Worthington H, O'Brien K. Orthodontic treatment for prominent upper front teeth (Class II malocclusion) in children. Cochrane Database of Systematic Reviews 2013; 11.
6. O'Brien K, Wright J, Conboy F, Sanjie ., Mandall N, Chadwick S, et al. Effectiveness of treatment for Class II malocclusion with the Herbst or Twin-block appliances: a randomised, controlled trial. Am J Orthod Dentofacial Orthop 2003;124:128-137.
7. Stewart FN, Kerr W, Taylor PJ. Appliance wear: the patient's point of view. Eur J Orthod 1997; 19:377-382.
8. Tsichlaki A, O'Brien K. Do orthodontic research outcomes reflect patient values? A systematic review of randomised controlled trials involving children. Am J Orthod Dentofacial Orthop 2014; 146, 279-285.
9. Pacha MM, Fleming PS, Johal A. A comparison of the efficacy of fixed versus removable functional appliances in children with Class II malocclusion: A systematic review. Eur J Orthod 2015;38:621-30.
10. O'Brien K, Wright J, Conboy F, Chadwick S, Connolly I, Cook P, et al. Effectiveness of early orthodontic treatment with the Twin-block appliance: a multicenter,

randomised, controlled trial. Part 2: psychosocial effects. Am J Orthod Dentofacial Orthop 2003;124:488-494.

11. Higgins JP, Green S, editors. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons 2011; 24.
12. Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hospital Research Institute 2011.
13. Čirgić E, Kjellberg H, Hansen K. Treatment of large overjet in Angle Class II: division 1 malocclusion with Andresen activators versus prefabricated functional appliances—a multicenter, randomized, controlled trial. Eur J Orthod 2016;1;38:516-24.
14. Čirgić E, Kjellberg H, Hansen K. Discomfort, expectations, and experiences during treatment of large overjet with Andresen Activator or prefabricated functional appliance: a questionnaire survey. Act Odont Scand 2017;3;75:166-72.
15. Sergl HG, Zentner A. A comparative assessment of acceptance of different types of functional appliances. The European Journal of Orthodontics 1998, 20(5), 517-524.
16. Sergl, H. G., Klages, U., and Zentner, A. Functional and social discomfort during orthodontic treatment-effects on compliance and prediction of patients' adaptation by personality variables. Eur J Orthod 2000; 22: 307-315.
17. Alzoubi EE, Hariri R, Mulligan K, Attard N. An evaluation of oral health-related quality of life in orthodontic patients treated with fixed and twin blocks appliances. J Orthod Science 2017;6:65.
18. Baysal A, Uysal T. Soft tissue effects of Twin Block and Herbst appliances in patients with Class II division 1 mandibular retrognathy. Eur J Orthod 2011;35:71-81.
19. Bowman AC, Saltaji H, Flores-Mir C, Preston B, Tabbaa S. Patient experiences with the Forsus fatigue resistant device. Angle Orthod 2012;83:437-46.
20. Carter LA, Geldenhuys M, Moynihan PJ, Slater DR, Exley CE, Rolland SL. The impact of orthodontic appliances on eating—young people's views and experiences. J Orthod 2015;42:114-22.
21. Čirgić E, Kjellberg H, Hansen K, Lepp M. Adolescents' experiences of using removable functional appliances. Orthod Craniofacial Research 2015;18:165-74.
22. El-Huni A, Salazar FB, Sharma PK, Fleming PS. Understanding factors influencing compliance with removable functional appliances: A qualitative study. Am J Orthod Dentofacial Orthop 2019;155:173-81.

23. Elkordy SA, Fayed MM, Abouelezz AM, Attia KH. Comparison of patient acceptance of the Forsus Fatigue Resistant Device with and without mini-implant anchorage: A randomized controlled trial. Am J Orthod Dentofacial Orthop 2015;148:755-64.
24. Feldmann I, List T, Bondemark L. Orthodontic anchoring techniques and its influence on pain, discomfort, and jaw function—a randomized controlled trial. Eur J Orthod 2011;34:102-8.
25. Gandhi P, Meenu G, Puneet B. "Relative comparison and assessment of patient's attitude and discomfort between two different types of fixed functional appliances: A comprehensive survey. J Orthod Research 2013;1:3
26. Hagg U, Tse EL, Rabie AM, Robinson W. A comparison of splinted and banded Herbst appliances: treatment changes and complications. Aus Orthod J 2002;18:76.
27. Hamilton CF, Saltaji H, Preston CB, Flores-Mir C, Tabbaa S. Adolescent patients' experience with the Carriere distalizer appliance. Eur J Paed Dent. 2013;14:219-24.
28. Hedlund C, Feldmann I. Success rate, costs and long-term stability of treatment with activator/headgear combinations. Sw Dent J 2016;40:67-77.
29. Heinig N, Göz G. Clinical Application and Effects of the Forsus™ Spring A Study of a New Herbst Hybrid. J Orofacial Orthop 2001;62:436-50.
30. Idris G, Hajeer MY, Al-Jundi A. Acceptance and discomfort in growing patients during treatment with two functional appliances: a randomised controlled trial. Eur J Paed Dent 2012;13:219-24.
31. Johnson PD, Cohen DA, Wheeler T, Aiosa L, McGorray S. Attitudes and compliance of pre-adolescent children during early treatment of Class II malocclusion. Clin Orthod Research. 1998;1:20-8.
32. Kadkhoda S, Nedjat S, Shirazi M. Comparison of oral-health-related quality of life during treatment with headgear and functional appliances. Int J Paed Dent 2011;2:369-73.
33. Latkauskiene D, Jakobsone G, McNamara JA. A prospective study on the clinical effectiveness of the stainless steel crown Herbst appliance. Progress Orthod 2012;13:100-8.
34. Lee RT, Kyi CS, Mack GJ. A controlled clinical trial of the effects of the Twin Block and Dynamax appliances on the hard and soft tissues. Eur J Orthod 2007;29:272-82.
35. Lena Y, Bozkurt AP, Yetkiner E. Patients' and Parents' Perception of Functional Appliances: A Survey Study. Tur J Orthod 2017;30:33.

36. Moro A, Janson G, Moresca R, Freitas MR, Henriques JF. Comparative study of complications during Herbst treatment with Cantilever Bite Jumper and removable mandibular acrylic splint. Dent Press J Orthod 2011;16:E1-0.
37. Read MJ, Deacon S, O'Brien K. A prospective cohort study of a clip-on fixed functional appliance. Am J Orthod Dentofacial Orthop 2004;125:444-9.
38. Schiöth T, von Bremen J, Pancherz H, Ruf S. Complications during Herbst appliance treatment with reduced mandibular cast splints. J Orofacial Orthop 2007;68:321-7.
39. Thiruvenkatachari B, Sandler J, Murray A, Walsh T, O'Brien K. Comparison of Twin-block and Dynamax appliances for the treatment of Class II malocclusion in adolescents: A randomized controlled trial. Am J Orthod Dentofacial Orthop 2010;138:144-E1.
40. Wiechmann D, Vu J, Schwestka-Polly R, Helms HJ, Knösel M. Clinical complications during treatment with a modified Herbst appliance in combination with a lingual appliance. Head Face Med 2015;11:31.
41. Phuong A, Fagundes NC, Abtahi S, Roberts MR, Major PW, Flores-Mir C. Additional appointments and discomfort associated with compliance-free fixed Class II corrector treatment: a systematic review. Eur J Orthod 2018;41:404-14.
42. Al-Moghrabi D, Salazar FC, Pandis N, Fleming PS. Compliance with removable orthodontic appliances and adjuncts: A systematic review and meta-analysis. Am J Orthod Dentofacial Orthop 2017;152:17-32.
43. Banks P, Wright J, O'Brien K. Incremental versus maximum bite advancement during twin-block therapy: a randomized controlled clinical trial. Am J Orthod Dentofacial Orthop 2004;126:583-8.
44. Illing HM, Morris DO, Lee RT. A prospective evaluation of Bass, Bionator and Twin Block appliances. Part I-The hard tissues. Eur J Orthod 1998;20:501-16.
45. Miles P. Rigid fixed functional appliances. Orthodontic Functional Appliances: Theory and Practice 2016;7:70-92.
46. Yassir YA, McIntyre GT, Bearn DR. Three questionnaires to assess the perception of fixed orthodontic therapy before, during and after treatment: validity and reliability. Eur J Orthod 2016;39:402-10.
47. Benson PE, Cunningham SJ, Shah N, Gilchrist F, Baker SR, Hodges SJ, Marshman Z. Development of the Malocclusion Impact Questionnaire (MIQ) to measure the oral health-related quality of life of young people with malocclusion: part 2–cross-sectional validation. J Orthod 2016;43:14-23.

48. Hartling L, Milne A, Hamm MP, Vandermeer B, Ansari M, Tsertsvadze A, Dryden DM. Testing the Newcastle Ottawa Scale showed low reliability between individual reviewers. J Clin Epi. 2013;66:982-93.
49. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Jama. 2000;283:2008-12.

Figures Captions:

Figure 1. PRISMA flow chart of article identification and selection

Figure 2. Risk of bias for the included randomized controlled trials: low (green); unclear risk (white); high (red).

Figure 3. Newcastle-Ottawa scores for nonrandomized studies (n= 15)

Figure 4. Forest plot for estimated number of emergencies during removable functional treatment

Figure 5. Forest plot for estimated number of emergencies during fixed functional treatment.

Figure 6. Forest plot for estimated average of treatment discontinuation with removable functional appliances

Figure 7. Forest plot for estimated average of treatment discontinuation with fixed functional appliances.

Table I. Classification of potential complications and adverse events associated with Class II correction using fixed or removable appliances

Severity of incidence	Removable appliance	Fixed appliance
Minor	• Loosening or crib fracture, where retention and stability of the appliance was still acceptable • Soft tissue irritation due to rubbing or sharp edges	• Band/crown decementation • Soft tissue irritation due to rubbing or sharp edges
Moderate	• Loss of appliance retention and stability compromising the use of the appliance	• Detachment, distortion or loss of the appliance which can be repaired or replaced at the same appointment
Severe	• Fracture of acrylic component that requires laboratory repair • Loss of the appliance	• Component impinging/embedded on the mucosal tissue to the degree that removal followed by replacement after healing or re-fabrication is required • Fracture of key components
Failure	• Severe damage/harm to the teeth or the mucosal tissue • Very poor oral hygiene to the degree that treatment become harmful	• Severe damage/harm to the teeth or the mucosal tissue • Very poor oral hygiene to the degree that treatment become harmful

Table II. Baseline characteristics and outcome measurements of the included studies

Study/ design	Appliance/ Participants /duration/setting	Outcomes/Measurement method	Prevalence/Number/Severity of complications n (% amongst participants) OR M(SD)	OHRQoL/Patient experiences n (% amongst respondents)	Rate of discontinuation n (% amongst respondents)
-Alzoubi *et al*,[17] 2017 -RCT	-TB: (n= 49) 20 M, 29 F -Age range: 10-16 years -University of Malta, Malta	-OHRQoL & patients experience -OHIP-14 questionnaire (at baseline, 6-week, 12-week, 6-month)	-	(n= 29) -Speech problems: -1.3 ± 1.3 -Pain: -0.8 ± 0.9 -Eating Problems: -1.1 ± 1.1 -Embarrassment: -1 ± 1.1 -Works and jobs: -0.6 ± 0.9	-
-Bysal & Usal,[18] 2011 -RCT	-Herbst (n= 23): (9 M, 11 F) -Duration: 15.8 ± 6 months -TB (n= 24): (9 M, 11 F), 16.2 ± 7.5 months -Mean age: 12.9 ± 1.1 years -Erciyes University, Turkey	-Number of complications -Clinical notes	-TB (n= 20): 0.4 ± 0.6 -Herbst (n= 20): 0.95 ± 1.1	-	-TB (n= 24): 4 (16.7%) -Herbst (n=23): 3 (13%)
-Bowman *et al*,[19] 2013 -Cross-sectional	-FFRD (n= 70): (40 F, 30 M) -Duration: ≥2 months -Mean age: (14.5±1.5 years) -University of Buffalo & private clinic, USA	-OHRQoL, patients experiences, & prevalence/number of complications -Modified Smiles Better questionnaire & clinical notes	(n= 67) 25 (37.3%)	(n= 67) -Look scary: 5 (7.1%) -Speech problems: 2 (3%) -Eating problems: 5 (7.5%) -Sleep problems: 8 (12%) -Teasing: 0 -Pain: 5 (7.5%) -Schoolwork: 0	-
-Carter *et a*l,[20] 2015 -Qualitative	-Removable functional (n= 6): (5 F, 1 M) -Age range: 11-14 years -Newcastle Dental Hospital, UK	-Impacts on eating -Semistructured interviews	-	(n= 6) -Eating problems: (limitation of food choices, distress eating with appliance, longer time to eat, being messy, embarrassment, difficulties of chewing, alteration of taste & lack of adaptation)	-
-Cirgic *et al*,[21] 2015 -Qualitative	-Myobrace (n= 9): (5 F, 4 M), Duration: 6 months -Activator (n= 12): (6 F, 6 M), Duration: 6 months -Mean age: 13.2 ± 1.25 years -University of Gothenburg, Sweden	-OHRQoL & patients experience -Semistructured interviews		(n= 21) -Pain & discomfort: (painful at start, difficulty keeping appliance in mouth) -Teasing and being bullied and embarrassment	

Study/ design	Appliance/ Participants /duration/setting	Outcomes/Measurement method	Prevalence/Number/Severity of complications n (% amongst participants) OR M(SD)	OHRQoL/Patient experiences n (% amongst respondents)	Rate of discontinuation n (% amongst respondents)
-Cirgic *et al*[13, 14] 2016; 2017 -RCT	- Activator (n= 40): (16 F, 24 M) -Myobrace (n= 57): (28 F, 29 M) -Mean age: 10.3 ± 1.6 years -12 dental practices, Sweden	-Treatment discontinuation and patients experiences -Clinical notes & customized questionnaire (13-item)	-	Activator (n= 20); Myobrace (n= 24) -Falling out during sleep: Activator: 6 (30%); Myobrace: 17 (70%) -Sleep problems: Activator: 3 (17%); Myobrace: 11 (45%) -Pain: Activator: 7 (33%); Myobrace: 14 (60%) -Forget to wear: Activator: 6 (28%); Myobrace: 12 (50%) -Social discomfort: Activator: 4 (18%); Myobrace: 3 (12%)	-Activator (n= 40): 21(53%) -Myobrace (n= 57): 40 (70%)
-El-Huni *et al*,[22] 2019 -Qualitative	-TB (n= 22): (7 F, 15 M) -Duration: ≥ 3 months -Mean age 12.5 years -Royal London Hospital, U. K	-OHRQoL & patients experiences -Semistructured interview	-	-Physical impairment: discomfort and difficulty to speak and eat -Psychologic impairment: history of teasing or receiving negative comments while wearing the appliance -Recall issues: forgetfulness to wear the appliance and follow instructions -Daily activities: interference with social and educational activities -Adaptability: initial negative impact followed by adaptation	-
-Elkordy *et al*,[23] 2015 -RCT	-FFRD (n= 16 F) -Duration: 5 ± 1.5 months -Age range: 11 - 14 years -Cairo University, Egypt	-OHRQoL & patients experience, prevalence/severity of complications -Clinical notes & customized questionnaire	-Mild (swelling of cheeks, gum bleeding): 6 (37.5%) -Moderate (separation of parts): 4 (25%) -Severe (breakages): 3 (19%)	-Noticeability: 9 (56.5%) -Speech problems: 1 (6.5%) -Eating problems: 7 (44%) -Sleep problems: 4 (25.5%) -OH problems: 0 -Pain: 7 (43.5%)	0
-Feldmann *et al*,[24] 2011 -RCT	-HG (n= 30): (15 F, 15 M) -Mean age: 14 ± 1.7 years -Private clinic, Sweden	-Patient experiences (pain, discomfort and jaw function) -3 customized customised questionnaires	-	-At 1-day: Median (IQR) Pain-incisors: 1 (0 - 3) Pain-molars: 0.8 (0 - 3.4) Jaw function limitation: low - moderate -At 6-week: Median (IQR) Pain-incisors: 0 (0 - 2) Pain-molars: 0 (0 – 2)	-

Study/ design	Appliance/ Participants /duration/setting	Outcomes/Measurement method	Prevalence/Number/Severity of complications n (% amongst participants) OR M(SD)	OHRQoL/Patient experiences n (% amongst respondents)	Rate of discontinuation n (% amongst respondents)
-Gandhi *et al*,[25] 2013 -CCT	-FFRD (n= 12) -MPA (n= 12) -Duration: ≥ 2 months -Mean age: 14.5 ± 1.5 years -Institute of Dentistry, India	-OHRQoL & patients experiences -Modified Smiles Better questionnaire (at 1-day, 7-day, 14-day and 30-day)	-	-Speech problems: FFRD: 4.5 (37.5%); MPA: 7.5 (62.5%) -Eating problems: FFRD: 6 (50%); MPA: 7.5 (62.5%) -Sleep problems: FFRD: 0 (0%); MPA: 0 (0%) -Teasing: FFRD: 4.5 (37.5%); MPA: 12 (100%) -Pain-teeth: FFRD: 6 (50%); MPA: 1.5 (12.5%) -Pain-jaw: FFRD: 1.5 (12.5%); MPA: 1.5 (12.5)	-
-Hagg *et al*,[26] 2002 -CCT	-Banded-Herbst (n= 14): (6 F, 8 M), Duration: 6.4 ± 0.7 months -Casted-Herbst (n= 14): (8 F, 6 M), Duration: 7.1 ± 0.8 months -Mean age 13 ± 1 year -University of Hong Kong	-Number/Severity of complications -Clinical notes	Banded-Herbst & Casted-Herbst-H (n= 24) -Moderate (dislodged): 1.8 ± 2.4 -Severe (fractured): 1.7 ± 2.3	-	0
-Hamilton *et al*,[27] 2013 -Cross-sectional	-Carriere (n= 42): (26 F, 15 M, 1 unknown) -Age range: 11 - 17 years -Mean age: 13.9 ± 1.3 years -University of Buffalo & 2 private clinics, USA	-OHRQoL & patient experiences -Modified Smiles Better questionnaire	-	-Speech problems: 1 (2.5%) -Eating problems: 4 (10%) -Sleep problems: 2 (5%) -Appearance: 3 (7.5%) -Teasing: 1 (2.5%) -Teeth pain: 15 (36.5%) -Jaws pain: 9 (22%)	-
-Hedlund & Feldmann,[28] 2016 -Cohort	-Activator (n= 85): (33 F, 52 M) -Duration: 25.8 ± 12.7 months -Mean age: 10.9 ± 1.4 years -Public Dental Service, Sweden	-Treatment discontinuation, prevalence/number of complications, OHRQoL & patient experiences -Clinical notes & customized questionnaire	(n= 85) 0.53 ± 0.76	(n= 38) Median (IQR) -Pain/discomfort: 42 (22 - 66) -Sleep problems: 46 (10 - 59) -Soreness: 26 (11 - 43) -Mood: 6 (20 - 18) -Teasing: 1 (0 - 5)	(n= 35) 14 (41.2%)
-Heinig *et al*,[29] 2001 -Case series	-FFRD (n= 13): (5 F, 8 M) -Duration: 4 months -Mean age: 14.2 years -University of Tübingen, Germany	-OHRQoL & patient experience	-	-Eating problems: 1 (8%) -Speech problems: 1 (8%) -Pain-teeth: 0 -Pain-Jaws: 0 -Mouth opening: 5 (38%) -Teeth cleaning: 6 (46%) -Sleep problems: 0 -Appearance: 9 (69%)	-

Study/ design	Appliance/ Participants /duration/setting	Outcomes/Measurement method	Prevalence/Number/Severity of complications n (% amongst participants) OR M(SD)	OHRQoL/Patient experiences n (% amongst respondents)	Rate of discontinuation n (% amongst respondents)
-Idris *et al*,[30] 2012 -RCT	-T4K (n= 26): (12 F, 14 M) -Activator (n=28): (14 F, 14 M) -Mean age: 10.5 years -Hama University, Syria	-OHRQoL & patient experiences -Sergl et al questionnaire (at 7-day, 14-day, 3-month and 6-month)	-	(Mean value) -Pain: Activator: (1.2); T4K: (1.2) -Pressure: Activator: (1.3); T4K: (1.5) -Teeth sensitivity: Activator: (1.7); T4K: (2.3) -Speech problems: Activator: (1.4); T4K: (3.2) -Lack of confidence: Activator: (1.3); T4K: (1.8)	-
-Johnson et al,[31] 1998 -Cross-sectional	-Bionator (n= 79): (30 F, 49 M) -HG (n= 89): (37 F, 52 M) -Duration: ≤ 24 months -Mean age: 9.5 years -University of Florida, USA	-OHRQoL & patient experiences -Customized questionnaire (28-item)	-	-Pain: Bionator: 43 (54%); HG: 62 (70%) -Eating problems: Bionator: 9 (11%); HG: 22 (25%) -Speech problems: Bionator: 65 (82%); HG: 48 (54%) -Chewing problems: Bionator: 25 (31%); HG: 37 (41%) -Embarrassment: Bionator: 29 (37%); HG: 29 (33%)	- Bionator: 30 (38%) - HG: 45 (50%)
-Kadkhhoda et al,[32] 2011 -Cross-sectional	-TB (n=67) -Headgear (n=67) -Duration: ≥ 3 months -Mean age: 12.5 ± 1.3 years -Location: Tehran University, Iran	-OHRQoL & patient experiences -CPQ (11-14), at 3-month		-Pain: TB: (0.4 ± 0.8); HG: (0.6±0.9) -Bad breath: TB: (1.3 ± 1.1); HG: (1.2 ± 1) -Speech problems: TB: (1.5 ± 1.3); HG (0.6 ± 1.1) -Sleep problems: TB: (0.8 ± 1.1); HG: (1.1 ± 1.3) -Teasing: TB: (0.6 ± 1.1); HG: (0.4 ± 0.8) -Upset: TB: (1.2 ± 1.2); HG: (1 ± 1.2)	
-Latkauskiene et al,[33] 2011 -CCT	-Crowned-Herbst (n= 180) -Duration: 12 months -Gender: not clear -Private clinic, Lithuania	-Treatment discontinuation, prevalence/number/severity of complications, OHRQoL & patient experiences -Clinical notes & customized questionnaire (at 6-month of appliance removal)	(n= 175) -Prevalence: 48 (27.4%) -Minor (loosing crown, bending rods): 27 (15.4%) -Moderate (unscrewing screw, damage attachment): 19 (10.9%) -Severe (fractures): 2 (1.5%)	(n=87) -Function problems: 0 -Discomfort: 16 (14%) -Noticeability:17 (19.5%)	(n=180) 5 (2.8%)

Study/ design	Appliance/ Participants /duration/setting	Outcomes/Measurement method	Prevalence/Number/Severity of complications n (% amongst participants) OR M(SD)	OHRQoL/Patient experiences n (% amongst respondents)	Rate of discontinuation n (% amongst respondents)
-Lee et al,[34] 2007 -RCT	-TB (n= 31): (17 F, 14 M) -Dynamax (n= 31): (17 F, 14 M) -Duration: 9 months -Age range: 10.6 - 14.7 years -Royal London Hospital, U. K	-Treatment discontinuation, prevalence/number/severity of complications -Clinical notes	-TB (n= 28): Minor (Adams clasp breakages), 10 (35%) -Dynamax (n= 28): Severe (vertical components breakages), 15.5 (55%)	-	-TB (n= 31): 3 (10%) -Dynamax (n= 31): 3 (10%)
-Lena et al,[35] 2017 -Cross-sectional	-FFRD (n= 43) -MPA (n= 42) -TB (n= 39) -Duration: 6 months -Mean age: 13.3 years -Ege University, Turkey	-OHRQoL & patient experiences, prevalence of complications -Customized questionnaire (31-item at 6-month)	-TB: 27 (69.3%) -FRD: 35 (81.4%) -MPA: 41 (97.6%)	-Appearance: TB: 23 (59%); FFRD: 27 (62.8%); MPA: 41 (97.6%) -Eating problems: TB: 7 (18%); FFRD: 32 (74.4%); MPA: 33 (78.6%) -Speech problems: TB: 31 (79.6%); FFRD: 15 (34.9%); MPA: 23 (54.7%) -Sleep problems: TB: 24 (61.6%); FFRD: 9 (21%); MPA: 27 (64.3%) -Schoolwork: TB: 15 (38.5%); FFRD: 4 (9.3%); MPA: 14 (33.3%) -Pain-teeth: TB: 24 (61.6%); FFRD 29 (67.4%); MPA: 31 (73.9%) -Pain-jaw: TB: 20 (51.3%); FFRD 26 (60.5%); MPA: 29 (63.5%) -Opening limitation: TB: 26 (66.7%); FFRD: 30 (69.7%); MPA: 31 (73.9%) -OH: TB: 18 (46.1%); FFRD: 24 (55.8%); MPA: 29 (69%)	-
-Moro et al,[36] 2011 -CCT	-Crowned-Herbst (n= 21): (6 F, 16 M) -Acrylic-Herbst (n= 21): (10 F, 11 M) -Duration: 12 months -Mean age: 11.3 - 12.3 years -Bauru Dental School & private clinic, Brazil	-Prevalence/number/severity of complications -Clinical notes	-Prevalence: Crowned-Herbst: 14 (66.7%); Acrylic-Herbst: 18 (85.7%) -Number: Crowned-Herbst: 24 (1.1 ± 1); Acrylic-Herbst: 53 (2.5 ± 1.8) -Minor (e.g. lesions in soft tissue & crown debond): Crwoned-Herbst: 15 (62.5%); Acrylic-Herbst: 8 (15%) -Moderate (e.g. screw loosening, rod distortion): Crowned-Herbst: 10 (41.6%); Acrylic-Herbst: 27 (51%) -Severe (e.g. fracture of appliance parts): Crowned-Herbst: 1 (4.2%); Acrylic-Herbst: 18 (34%)	-	0 (0%)

Study/ design	Appliance/ Participants /duration/setting	Outcomes/Measurement method	Prevalence/Number/Severity of complications n (% amongst participants) OR M(SD)	OHRQoL/Patient experiences n (% amongst respondents)	Rate of discontinuation n (% amongst respondents)
-O'Brien et al,[6] 2003 -RCT	-Casted-Herbst (n= 105): (55 F, 50 M), Duration: 5.8 months -TB (n= 110): (62 F, 48 M), Duration: 11.2 months -Mean age: 12.5 years -13 NHS Hospital, United Kingdom	-Treatment discontinuation, number of emergencies, OHRQoL & patients experiences -Clinical notes & Smile Better questionnaire at 4-month	-TB (n= 36): 1.6 ± 1.6 -Casted-Herbst (n= 60): 4.3 ± 2.9	-	−TB (n= 110): 37 (33.6%) -Casted-Herbst (n= 105): 18 (17%)
-Read et al,[37] 2004 -CCT	-F-TB (n= 32) -Duration: 5.1 ± 2 months -Age range: less than 15 years -University of Manchester, United Kingdom	-Treatment discontinuation, number/severity of complications -Clinical notes	Severe (Replacement of loose blocks and repair fractured bands): 1.7 ± 1.6	-	2 (6.3%)
-Schioth et al,[38] 2007 -CCT	-Casted-Herbst (n= 50) -Duration: 8 months -Mean age: 14.5 - 15.5 years -University of GieBen, Germany & Berne, Switzerland	-Treatment discontinuation, prevalence/severity of complications -Clinical notes	-Prevalence: 29 (58%) -Moderate (loosening splint): 26 (89) % -Severe (telescope & splint breakages): 3 (11%)	-	0
-Sergl & Zenter,[15] 1998 -Cohort	-Removable functional (n= 14) -Duration: 6 months -Mean age: 12.8 ± 4 years -University of Mainz, Germany	-OHRQoL & patient experience -Customized questionnaire (at 14-day, 3-month, 6-month)	-	(Mean value) - Pain: (1.5) - Teeth sensitivity: (1.5) -Speech problems: (2) -Swallowing problems: (1.7) -Lack of confidence (1.7)	-
-Thiruvenkatachari et al,[39] 2010 -RCT	-TB (n= 32): (16 F, 16 M) -Dynamax (n= 32): (16 F, 16 M) -Duration: 9 months -Age range: 10 - 14 years. -University of Manchester, U. K	-Treatment discontinuation, prevalence/number/severity of complications -Clinical notes	-Breakages: TB: 11 (34.4%); Dynamax: 18 (56.3%) -Adverse events: TB: 5 (17%); Dynamax: 26 (81%) -Complications: TB: 8 (25.7%); Dynamax: 22 (68.7%) -Nature of Dynamax complications (vertical spur escaping, vertical spur imbedded in soft tissue LA buried in mucosa, fractures of vertical spurs, fracture LA, fracture of maxillary plate)	-	TB: 8 (25%) Dynamax: 21 (65.6%)
-Wiechmann et al,[40] 2015 -Case control	-WIN-Herbst (n= 35): (23 F, 12 M) -Duration: 10.5 months -Mean age: 16.9 years -Private clinic, Germany	-Prevalence/number/severity of complications -Clinical notes	-Number of Complications: 13 -Prevalence: 10 (28.6%) -Mild: 0 -Moderate (loosening attachments): 7 (20%) -Severe (fracture of l-pin): 5 (14.3)		(n=35): 0 (0%)

M (SD), mean (standard deviation); RCT, randomized controlled trial; F, female; M, male; TB, Twin-Block; OHRQoL, oral health related quality of life; FFRD, Forsus Fatigue Resistance Device; CCT, controlled clinical trial; MPA, mandibular protraction appliance; HG, headgear; T4K, Trainer four Kids appliance; F-TB, Fixed Twin-Block; OHIPQ, oral health impact profile questionnaire; IQR, interquartile range; CPQ, child perception questionnaire; TPA, trans-palatal arch; LA, lingual a

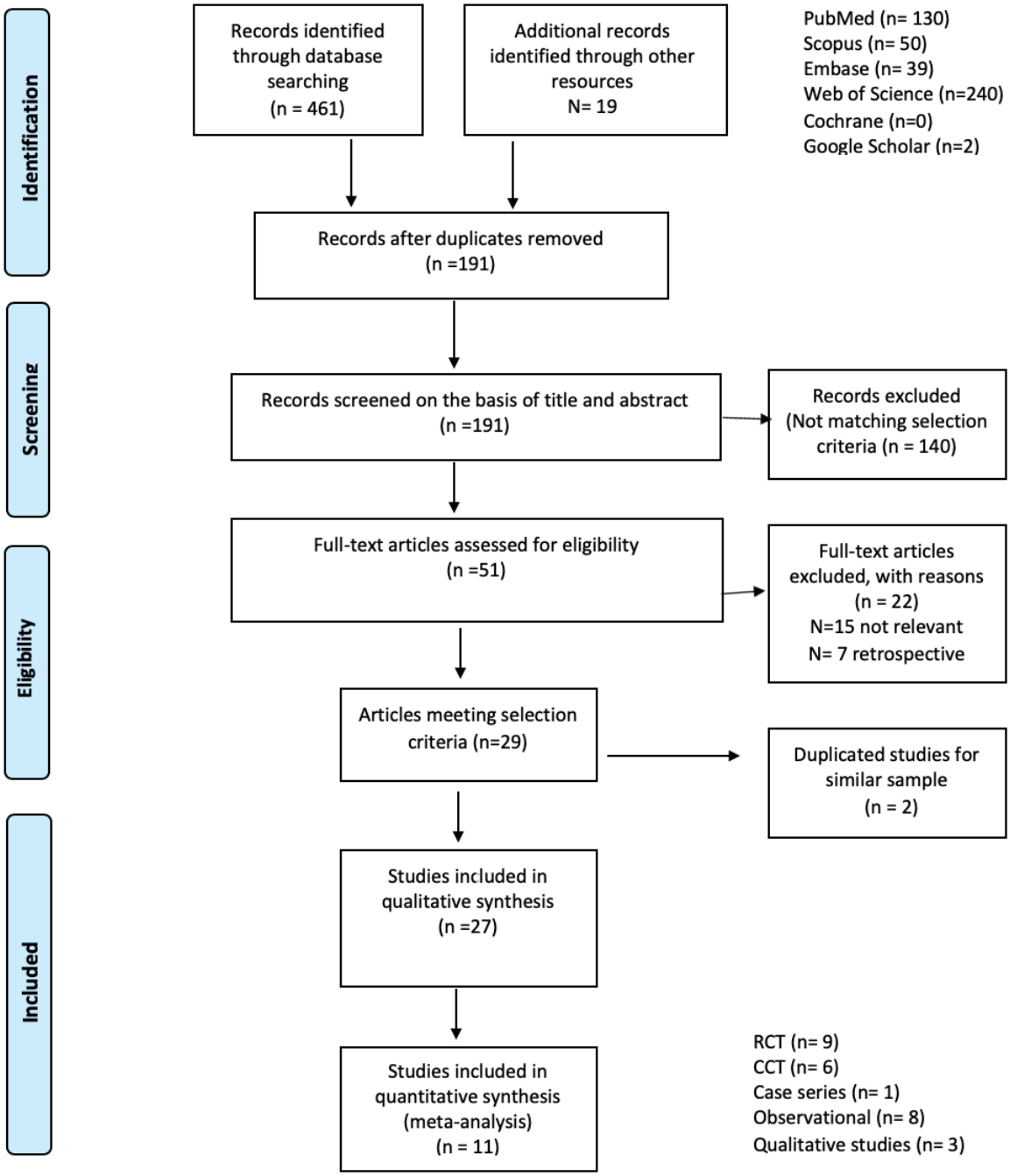

Fig 1. PRISMA flow chart of article identification and selection.

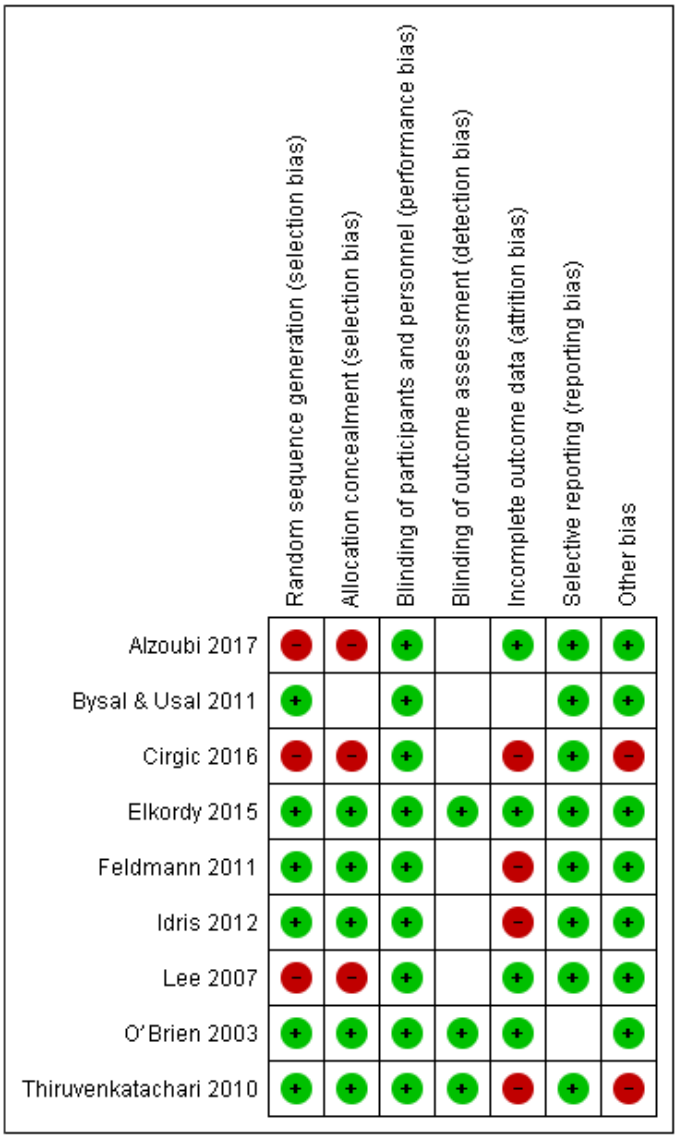

Figure 2. Risk of bias for the included randomized controlled trials: low risk of bias (green); unclear risk of bias (white); high risk of bias (red).

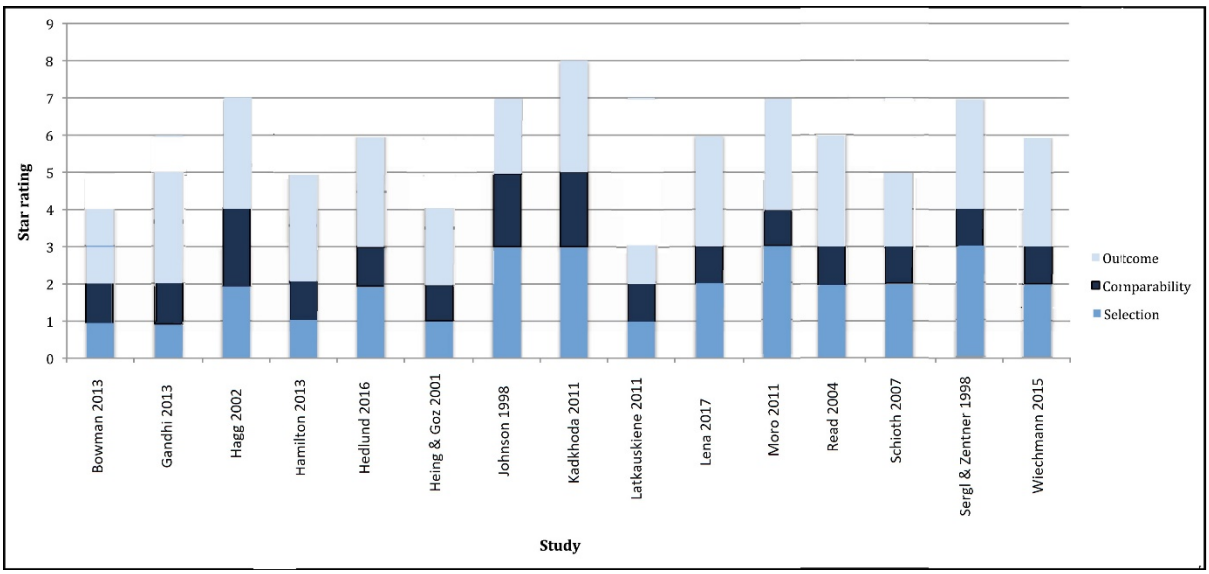

Fig 3. Newcastle-Ottawa scores for nonrandomized studies (n= 15).

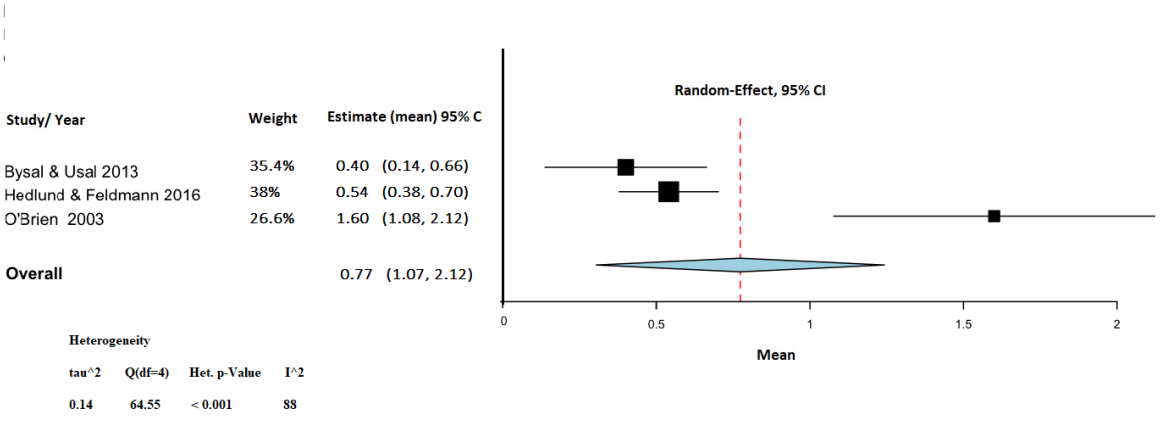

Figure 4. Forest plot for estimated number of emergencies during removable functional treatment.

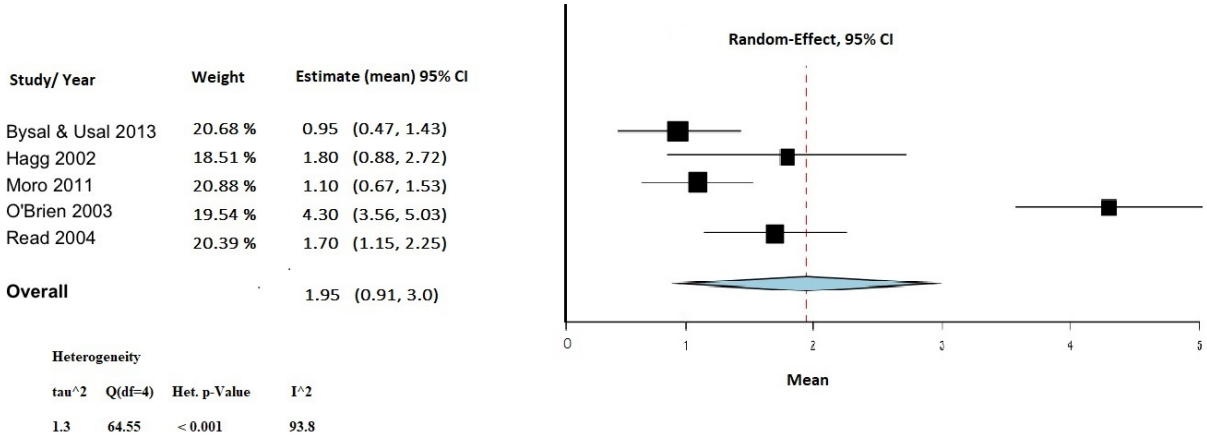

Figure 5. Forest plot for estimated number of emergencies during fixed functional treatment.

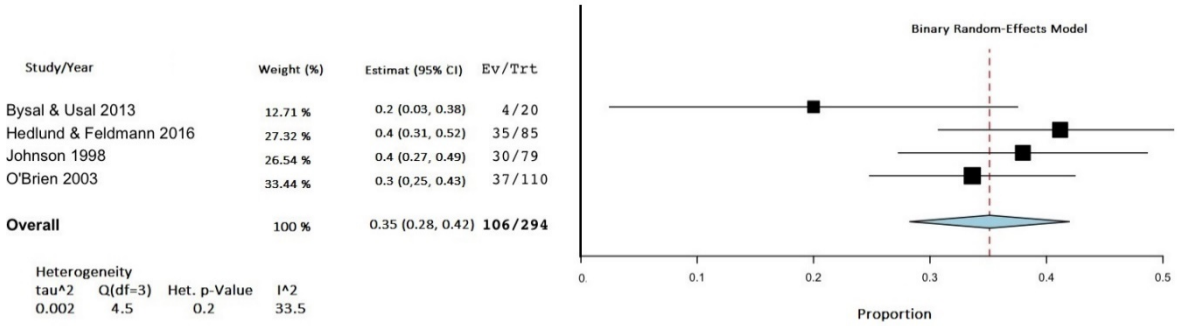

Figure 6. Forest plot for estimated average of treatment discontinuation with removable functional appliances.

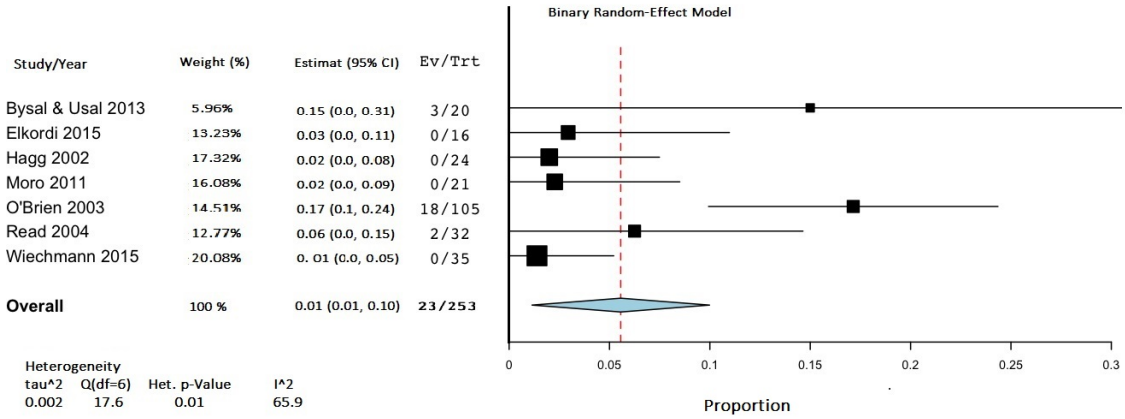

Figure 7. Forest plot for estimated average of treatment discontinuation with fixed functional appliances.

Table III. Pooled results

Outcomes	Fixed Functional Appliances (n= 682)	Removable Functional Appliances (n= 682)	Hybrid Functional Appliances (n=84)	Headgear (n=186)	Molar Destalizer (n=42)
EMG Prevalence	n= 327: 112 (34%)	n=60: 34 (56%)	n=81: 55.5 (69%)	-	-
EMG number	n= 178: 2.5 ± 2.5	n=141: 0.9 ± 1.2	n=53: 150	-	-
Minor EMG	n= 318: 48 (16%)	-	-	-	-
Moderate EMG	n= 297: 66 (22%)	-	-	-	-
Severe EMG	n= 297: 14 (5%)	-	-	-	-
Failure Prevalence	n= 667: 28 (4%)	n= 458: 178 (39%)	n= 84: 24 (29%)	n= 90: 45 (50%)	-
Eating problems	n= 205: 92 (45%)	-n= 118: 16 (14%) -n= 49: 2.2 ± 1.1	-	n= 90: 22 (25%)	n= 4: 12 (30%)
Speech problems	n= 205: 55 (27%)	-n= 118: 96 (81%) -n= 116: 1.9 ± 1.3 -n= 54: M (2.3)	-	n= 90: 48 (54%)	n= 41: 4 (10%)
Pain	-Pain-teeth (n= 380): 96 (25%) -Pain-jaw (n= 122): 59 (48%)	-n= 162: 88 (54%) -n= 116: 1 ± 1.1 -n= 54: M (1.2)	-	-n= 90: 62 (70%) -n= 61: 0.6±0.9 -n= 30: median (IQR) 0 (0-2)	n= 41: 12 (29%)
Sleep problem	n= 205: 48 (23%)	-n= 368: 100 (26%) -n= 67: 0.8 ± 1.1	-	n= 69: 1.1 ± 1.3	n= 41: 2 (5%)
Appearance	n= 356: 108 (30%)	n= 39: 23 (59%)	-	-	n= 41: 3 (8%)
Teasing	n= 91: 17 (19%)	n= 67: 0.6 ± 1.1	-	n= 61: 0.4 ± 0.8	n= 41: 1 (3%)
Embarrassment	-	-n= 79: 29 (37%) -n=49: 2.1±1.1	-	n= 90: 29 (33%)	-
Doing works	-	-n= 39: 15 (59%) -n= 49: 1.7 ± 0.9	-	-	-
OH problems	n= 114: 59 (52%)	n= 39: 18 (46%)	-	-	-
Opening limitation	n=98: 66 (67%)	n= 39: 26 (67%)	-	-	-

EMG, emergencies; OH, oral hygiene; M, mean; IQR, (Interquartile range)

APPENDIX

Appendix 1. Search strategy as used in databases

Database	Search Strategy	Results
Medline via PubMed	Search orthodontics AND (((((((functional appliance) OR ((orthodontics) OR ((orthodontics) AND headgear)) OR ((orthodontics) AND molar destalizer)) OR trans-palatal arch) OR Herbst)))) AND ((orthodontics) AND ((((((quality of life) OR complications) OR breakages) OR patient-experiences) OR patient-concerns) OR expectations))	130
Web of Science	Search (orthodontics) AND ((((((functional appliances) OR complications) OR breakages) OR patient-experiences) OR patient-concerns) OR expectations)	240
Embase:	Same as PubMed	39
Scopus	(orthodontics) AND (((functional AND appliances)) AND (patient AND experiences)) AND (complications) AND (LIMIT-TO (DOCTYPE," ar")) AND (LIMIT-TO (SUBJAREA," DENT"))	50
Cochrane		0
Google Scholar		2

Appendix 2. Excluded studies with reasons

Agar et al., 2005	Studied the compliance rate with Headgear treatment and associated psychosocial factors in improving compliance
Dann ate al., 1995	Studied the psychological effects of overjet on children and did not report patient experiences during twin block treatment
Dowsing et al., 2015	Personal opinion concerning management of emergencies of functional appliances
Egolf et al 1990	Studied factors that impact patient's compliance with headgear and intraoral elastic wear
Ghafari et al., 1998	Studied compliance and successful rate during Class II treatment and did not report patient experiences or complications.
Gill and Lee 2005	Reported discontinuation rate during twin blocks treatment but did not report patient experiences
Harradine 2000	Retrospective study reported compliance and discontinuation rate with twin block appliance
Kavaliauskiene et al., 2012	Included sample size of patients over 18 years old. Data was very difficult to interpret and outcomes from different types of orthodontic appliances were mixed up together.
Prove et al., 1997	Comparison of fixed and removable plate. Not a Class II malocclusion
Rawji 2008	Studied impact of orthodontic appliances wear on sleep quality in sleep laboratory
Sergl et al., 1998	Laboratory experiment on functional appliances on adults
Stewart et al., 1997	Comparison of fixed and removable plate. Not a Class II malocclusion
Sergl et al., 2000	Duplicated study using the similar sample size of another included study
Wiedel., 2016	Removable appliances to treat anterior cross bite (Class III malocclusion)
O'Brien., 2003	Studied effect of malocclusion and corrected overjet on children's self-esteem and did not reported patient experiences from functional appliances
Franzen, 2011	Studied effects of corrected overjet on patient's self-esteem and did not reported patient experiences from functional appliance
Duterloo et al., 1998	Retrospective study on complications in the treatment of angle class II div 1
Kanuru et al., 2017	Retrospective study of complications in removable acrylic and splint Herbst for Class II
Manni et al., 2014	Retrospective study of emergencies and failure in acrylic splinted and Hanks Herbst
Manni et al., 2018	Retrospective study of complications in conventional and Manni Herbst
Nilson et al., 2016	Retrospective comparative study between Twin Block and Activator-Headgear
Sanden et al., 2004	Retrospective study compared complications in casted and banded Herbst
Silva et al., 2015	Retrospective study of complications in removable acrylic and cantilever Herbst

Appendix 3. Class II correctors as identified from the included studies according to their classification

Appliance type		Author/year
Fixed functional appliances	Forsus Fatigue Resistant Device (FFRD)	Heinig and Goz 2001 Bowman *et al.,* 2013 Gandhi, et al., 2013 Elkordy *et al.,* 2015 Lena et al., 2017
	Banded Herbst	Hagg et al., 2002
	Casted splinted Herbst with or without Hyrax	O'Brien et al., 2003 Hagg et al., 2003 Baysal & Usal 2011 Schioth *et al.,* 2007
	Crowned Herbst cantilever design with anchorage adjuncts	Moro *et al.,* 2011
	Crowned Herbst bite jumper without anchorage	Latkauskiene *et al.,* 2004
	WIN-Herbst used with a lingual system	Wiechmann *et al.,* 2015
	Clip-on fixed Twin-Block	Read et al., 2004
	Mandibular Protraction Appliance (MPA)	Gandhi et al., 2013 Lena et al., 2017
Removable Functional appliances	Twin-Block Appliance (TB)	O'Brien et al., 2003 Lee et al., 2007 Theruvenkatchari et al., 2010 Baysal & Usal 2011 Kadkhoda et al., 2011 Alzoubi et al., 2017 Lena et al., 2017 El-Huni et al., 2019
	Activator Appliance	Idris et al., 2012 Cirgic et al., 2015, 2016, 2017 Hedlund & Feldmann 2016
	Bionator Appliance	Johnson et al., 1998
	Prefabricated Functional Appliance (Myobrace)	Idris et al., 2012 Cirgic et al., 2015, 2016, 2017
	Unknown removable types	Sergl & Zentner 1998 Carter et al., 2015
Hybrid Functional appliance	Dynamax Appliance (maxillary removable and mandibular fixed parts)	Lee et al., 2007 Theruvenkatchari et al., 2010
	Acrylic Herbst (maxillary fixed and mandibular removable acrylic parts)	Moro *et al.,* 2011
Extraoral appliances	Headgear	Johnson et al., 1998 Feldmann et al., 2011 Kadkhoda et al., 2011
Maxillary molar destalizer	Carrier Destalizer Appliance (CDA)	Hamilton et al., 2013

Appendix 4. Risk of Bias assessment for Randomized Clinical Trials.

Alzoubi 2017	Risk of bias	Explanation
Random sequence generation	High	Inadequate randomization, and sequence generation not described. "A total of 98 patients.... were selected randomly..."
Allocation concealment	High	Lack of randomization and allocation concealment not described.
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding outcome assessment	Unclear	Not described.
Incomplete outcome data	Low	No dropouts or losses to follow up mentioned
Selective reporting	Low	The published report includes all the study's pre-specified outcomes
Other bias	Low	Imbalance in gender and age distribution at baseline
Baysal and Usal 2011	Risk of bias	Explanation
Random number generation	Low	"Randomization was made at the start of the study with prepared random number tables with block stratification on gender"
Allocation concealment	Unclear	Not described
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding assessor	Unclear	Not described.
Free of incomplete data	low	Balanced dropout between groups (13% in Herbst vs 16% drop out of Twin Block)
Selective outcome reporting	low	The published report includes all the study's pre-specified outcomes
other	Low	The study appears to be free of other sources of bias.
Cirgic 2016, 2017	Risk of bias	Explanation
Random number generation	High	Inadequate randomization, and sequence generation not described. "Randomization was performed by lottery"
Allocation concealment	High	Description not adequate and allocation was not concealed. "At each clinic two envelops were available one for girls and one for boys with 5AA and 5PFA notes for each gender…"
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding

Blinding assessor	Unclear	Not described
Free of incomplete data	high	Large number of dropouts in questionnaire outcome at -month (n=15 of each group) and then at 6-month (5 of Activator and 18 of Myobrace)
Selective outcome reporting	Low	The published report includes all the study's pre-specified outcomes
other	High	Imbalanced sample size among groups (40 versus 57) suggest inadequate randomization. Imbalanced gender distribution suggest lack of stratification. Questionnaires completed at home and retrieved by mail resulting large number of no response
Elkordy 2015	Risk of bias	Explanation
Random number generation	Low	"Random sequence generation was done with a computer-generated list of random numbers obtained from an Excel spreadsheet"
Allocation concealment	Low	"Allocation concealment was achieved with sequential numbered and sealed opaque envelopes that were concealed..."
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding assessor	Low	", date were extracted by uninvolved person. The person responsible for the statistical analysis was not informed about the nature of the trial"
Free of incomplete data	Low	No dropouts or losses to follow up mentioned, 32 randomized and 32 analyzed.
Selective outcome reporting	Low	The published report includes all the study's pre-specified outcomes
Other	Low	Appears to be free of other sources of bias.
Feldmann 2011	Risk of bias	Explanation
Random number generation	Low	" ...the patients were randomized in blocks and stratified by gender..." The allocation sequence was computer generated..."
Allocation concealment	Low	"...and concealed in envelops until randomization"
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding assessor	Unclear	Not described
Free of incomplete data	High	Imbalanced dropout among groups (6 vs 1) and analysis on per-protocol base was applied
Selective outcome reporting	Low	The published report includes all the study's pre-specified outcomes
Other	Low	Appears to be free of other sources of bias
Idris 2012, 2019	Risk of bias	Explanation

Random number generation	Low	"...using a computer-generated list of random numbers..."
Allocation concealment	Low	"...performed using opaque sealed envelopes..."
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding assessor	Unclear	Not described
Free of incomplete data	High	Imbalanced dropout among groups (Activator= 6%; T4K= 13%) and intention-to-treat base was not applied
Selective outcome reporting	Low	The published report includes all the study's pre-specified outcomes
Other	Low	Appears to be free of other sources of bias.
Lee 2007	Risk of bias	Explanation
Random number generation	High	Inadequate randomization, and sequence generation not described "...and then randomly allocated to an appliance group by a non-clinician"
Allocation concealment	High	Lack of randomization and allocation concealment not described
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding assessor	Unclear	Not described
Free of incomplete data	Low	Balanced dropout among groups (3 from each group). Intention-to-treat base applied and data of all randomized patients was analyzed.
Selective outcome reporting	Low	The published report includes all the study's pre-specified outcomes
Other	Low	Appears to be free of other sources of bias.
O'Brien 2003	Risk of bias	Explanation
Random number generation	Low	"At the beginning of study, random number tables were used to prepare randomization lists, stratified"
Allocation concealment	Low	"Randomization performed using a central telephone line"
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding assessor	Low	"Cephalogram and study casts were both scored with the examiner unaware of the group to which the patient had been allocated"
Free of incomplete data	low	Unbalanced dropout among groups (TB=37; Herbst=18) but intention-to-treat base applied and data of all randomized patients was analyzed.

Selective outcome reporting	unclear	The published report missed some pre-specified outcome concerning patients' experiences during treatment
Other	Low	Appears to be free of other sources of bias.
Thiruvenkatachari 2010	Risk of bias	Explanation
Random number generation	Low	Randomization and sequence generation using Minim software
Allocation concealment	Low	"Treatment allocation was performed centrally by independent research assistants using minimization to one of treatments"
Blinding participants and personnel	Low	Blinding treatment is not feasible, but outcomes and measurements are not likely to be influenced by lack of blinding
Blinding assessor	Low	"The DMC assessors and the trial statistician were blinded to treatment allocation"
Free of incomplete data	High	Imbalanced dropout among groups (TB=7; Dynamax=3) and study terminated early
Selective outcome reporting	Low	The published report includes all the study's pre-specified outcomes
Other	High	The RCT was terminated early and patients in Dynamax group were moved to different treatment

Appendix 5. Quality assessment according to the Newcastle-Ottawa Scale adapted for prospective non-randomised and cross-sectional studies.

Study ID	Selection (Max 4 stars)				Comparability (Max 2 stars)		Outcomes assessment (Max 3 stars)			Total scores (Max 9)	High quality: 9 Moderate quality: 6-8 Low quality: 1-5
	i	ii	iii	iv	v	vi	vii	viii	ix		
Bowman 2013	✵	0	0	0	✵	0	✵	✵	0	4	Low quality
Gandhi 2013	✵	0	0	0	✵	0	✵	✵	✵	5	Low quality
Hagg 2002	✵	0	✵	0	✵	✵	✵	✵	✵	7	Moderate quality
Hamilton 2013	✵	0	0	0	✵	0	✵	✵	✵	5	Low quality
Hedlund 2016	✵	✵	0	0	✵	0	✵	✵	✵	6	Moderate quality
Heing & Goz 2001	✵	0	0	0	✵	0	✵	✵	0	4	Low quality
Johnson 1998	✵	✵	0	✵	✵	✵	✵	✵	0	7	Moderate quality
Kadkhoda 2011	✵	✵	✵	0	✵	✵	✵	✵	✵	8	Moderate quality
Latkauskiene 2011	✵	0	0	0	✵	0	✵	✵	0	4	Low quality
Lena 2017	✵	✵	0	0	✵	0	✵	✵	✵	6	Moderate quality
Moro 2011	✵	✵	✵	0	✵	0	✵	✵	✵	7	Moderate quality
Read 2004	✵	0	✵	0	✵	0	✵	✵	✵	6	Moderate quality
Schioth 2007	✵	0	✵	0	✵	0	✵	✵	0	5	Low quality
Sergl & Zentner 1998	✵	✵	✵	0	✵	0	✵	✵	✵	7	Moderate quality
Wiechmann 2015	✵	0	✵	0	✵	0	✵	✵	✵	6	Moderate quality

✵: 1, 0: no✵, i: Representativeness of the sample; ii: Non-respondents or Selection of controls; iii: Ascertainment of exposure (validity of measurement tool, secure records); iv: Justification of study sample size; v: Study controls for Class II treatment. vi: Study controls for additional confounding factor (e.g. gender, age); vii: Assessment of outcome (independent blind assessment, self-report, no description); viii: Was follow-up long enough for outcome to occur; ix: Statistical test (was appropriate and described?)