Performance of Hawley-type appliances: a systematic review of randomized clinical trials

By

Wafa Jaber Al Rahma

DDS, Ajman University of Science and Technology, 2005

MSc, Royal College of Surgeon Ireland, 2009

Principal Supervisor:

Professor Athanasios E. Athanasiou

Co-supervisor:
Assistant Professor Eleftherios G. Kaklamanos

Presented to the Hamdan Bin Mohammed College of Dental Medicine of
Mohammed Bin Rashid University of Medicine and Health Sciences
in Partial Fulfilment of the Requirements for the Degree of
Master of Science in Orthodontics
January 2017

ABSTRACT

Performance of Hawley-type appliances: a systematic review of randomized clinical trials

Wafa Jaber Al Rahma, DDS, MSc

Supervisors:

Professor Athanasios E. Athanasiou

Assistant Professor Eleftherios G. Kaklamanos

BACKGROUND: Although post-treatment changes are almost inevitable, and retention has long been recognized as one of the most critical and routine problems faced by orthodontists, there remains a lack of certainty regarding the parameters of any definitive retention protocol following orthodontic treatment.

AIM: To compare the performance of the Hawely-type appliances to that of the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) immediately used after completion of orthodontic treatment, together with a comparison of the effectiveness of different wearing schedules.

MATERIALS AND METHODS: An electronic search without restrictions for published and unpublished literature, together with hand searching, was carried out. We reviewed randomized clinical trials (RCTs) investigating the performance of the Hawley-type appliances. The risk of bias was assessed using the Cochrane Collaboration's Risk of Bias

assessment tool for RCTs and the quality of evidence assessed according the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach.

RESULTS: The initially identified 1174 records were finally reduced to 10 full-text reports analyzing various outcomes in 854 patients who had been followed for a maximum period of one year after the removal of the fixed orthodontic appliances. Eight of the eligible publications investigated groups of subjects using Hawley removable appliances and clear thermoplastic retainers, another study compared the Hawley appliance to positioner use during the retention period and, finally, one study involved patients allocated to groups using different wearing schedules for Hawley retention appliances. Three studies were considered as being of low risk of bias, four of unclear and three of high risk of bias. In general, few differences were observed between the comparative performance of the Hawely-type appliances and the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment regarding outcomes relevant to maxillary and mandibular dental arch measurements, dental arch relationships and occlusal contacts, speech evaluation, patient reported outcomes, adverse effects and problems related to the appliances, in addition to economic evaluation related outcomes. Moreover, no differences were found between different appliance wearing schedules and protocols. Overall, the quality of the available evidence was considered low.

CONCLUSIONS: Based on the available data there are few differences were observed between the comparative performance of the Hawley-type appliances to the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment. Moreover, no differences were found between different appliance wearing schedules and protocols.

DEDICATION

I would like to dedicate this thesis to my husband, parents and my kids for their support and motivation; without their support I would never have completed this research.

DECLARATION

I declare that all the contents of the thesis is my own work. There is no conflict of interest with any other entity or organization.

Name: Wafa Jaber Al Rahma

Signature:

ACKNOWLEDGEMENTS

I would like to express my deep sense of gratitude to my mentor, Dean, Professor and Program Director in Orthodontics Athanasios E. Athanasiou, for his valuable guidance, keen interest and encouragement at various stages of my program period.

I would like also to express my deep sense of gratitude to my supervisor, Assistant Professor in Orthodontics Eleftherios G. Kaklamanos; the completion of this dissertation could not have been possible without his endless support, encouragement and motivation.

To my friend and sister Dr. Shaima Al Naqbi, my love and thanks for her assistance and support during my program period.

To all relatives and friends who, in one way or another, gave their support, thank you!

Table of contents

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

GCF: Gingival Crevicular Fluid

GRADE: Grades of Recommendation, Assessment, Development and Evaluation

LILACS: Literatura Latino-Americana e do Caribe em Ciências da Saúde

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROSPERO: International prospective register of systematic reviews

VOT: Voice onset time

1. Introduction

Orthodontic treatment involves the application of forces which create a cellular response resulting in tooth movement (Parker, 1972;Thilander, 2011; Maltha et al., 2015). Once the teeth have been moved, the supporting tissues needs to remodel to maintain the tooth in its new position. This is why changes in gingival and periodontal fibers constitute one of several post-treatment factors affecting teeth after orthodontic treatment, in addition to occlusal factors, soft tissues, growth and physiologic dento-alveolar adaptation (Parker,1972; Thilander, 2011; Maltha et al., 2015). All of the above-mentioned parameters interact to move teeth back toward their pre-treatment position, a phenomenon that is observed in the vast majority of orthodontic patients (Joondeph, 2011; Blake and Garvey, 1998; Blake and Bibby, 1998; Melrose and Millett, 1998). It has been reported that up to 90% of orthodontic patients may have an unacceptable dental alignment 10 years after orthodontic treatment (Thilander, 2000).

As post-treatment changes are almost inevitable, retention has long been recognized as one of the most critical and routine problems faced by orthodontists in treating patients (Oppenheim, 1934). Since the 1950's reports have appeared in the orthodontic literature of studies on the importance of maintaining teeth in their required ideal positions following orthodontic treatment (Johnston and Littlewood, 2015). Various types of retainers are currently used to assist this process, including acrylic and wire retainers, such as the Hawley retainer, vacuum-formed retainers, and fixed retainers (Wong and Freer, 2004; Keim et al., 2008; Renkema et al., 2009).

Nevertheless, there remains a lack of certainty regarding the parameters of any definitive retention protocol in orthodontic treatment. Much of current practices seem to be contradictory regarding the type, daily regimen and total duration of orthodontic retention, indicating the need to develop evidence-based practice guidelines (Wong and Freer 2004; Renkema et al., 2009;

Singh et al., 2009; Valiathan and Hughes, 2010; Pratt et al., 2011; Van-devska-Radunovic et al., 2013).

The aim of this systematic review is to compare the performance of the Hawely-type appliances to the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment.

2. REVIEW OF THE LITERATURE

Retention has been defined as holding teeth in their "treated position after orthodontic treatment for the period of time necessary to maintain the results" (Moyers, 1973), or as holding them "in ideal esthetic and functional position" (Riedel, 1976). While relapse is defined as "tendency of the teeth to return to their former position" (Reitan, 1959).

For many years there was no general agreement among orthodontists on the need for retention. As far back as the 1800s, it was believed that the occlusion of the teeth was the most potent factor in determining their stability in a new position (Kingsley, 1880). Later, in 1925, Lundström suggested that the osseous apical base was the most important factor in maintaining retention (Lundström, 1925) while McCauley believed that the transverse widths of the canines and molars played a major role in sustaining stability (McCauley, 1944). Eventually, the Tweed philosophy gained credence, where the incisors were considered less prone to relapse if they were upright, and the extraction of premolars was promoted (Tweed, 1944). Lastly, it was thought that active retention needed to be lifelong following orthodontic treatment, as there would always be a tendency for relapse to occur (Little et al., 1988).

2.1. Factors involved in post-treatment changes

During orthodontic treatment, the force applied to the tooth crown that creates the cellular response resulting in tooth movement will also cause changes in the surrounding tissues. If the area is not allowed to remodel in the post-movement condition, teeth will show the tendency to return to their original position (Reitan, 1969; Ericsson and Thilander, 1980). There are several possible factors implicated in these changes; these can be summarized as follows: gingival and

periodontal factors, occlusal factors, soft tissue factors, growth factors, and physiologic dento-alveolar adaptation (Blake and Bibby, 1998; Littlewood, 2013).

2.1.1. Periodontal factors

Throughout active tooth movement, the periodontal ligament fibers stretch along the tension side because of the application of orthodontic force, and then become embedded in the newly deposited bone (Maltha et al., 2015). Reitan (1960; 1967) found that these fibers are elongated, and that their reorganization, as well as that of the alveolar bone after force removal could lead to relapse. Hence, holding the teeth in their treated position after orthodontic therapy would allow for periodontal and gingival reorganization, minimize changes from growth, permit neuromuscular adaptation to the corrected tooth position, and maintain a potentially unstable tooth position, if such positioning is required for compromise or esthetics (Blake and Bibby, 1998).

The re-organization of the supporting tissues is thought to be completed within one year after the removal of the appliances, but each element of the tissue structure has its own mechanism and time span for remodeling. The periodontal ligament reorganizes over a 3 to 4 month period, the gingival collagen fiber network takes 4 to 6 months to remodel (Reitan, 1959; 1967; 1969), while the elastic supracrestal fibers need more than 7 months (Reitan, 1967). For these reasons, relapse during the initial post-debond period is more likely to occur due to the rebound of the periodontal fibers. After this period, the most significant cause of relapse is likely to be the gingival fibers (Boese, 1969), which, as already mentioned, exhibit a slower rate of change (Roberts and Chase 1981) and can be observed stretched and unremodeled up to 232 days after experimental tooth rotation (Reitan, 1959).

A rotated tooth could be the best example to demonstrate the above-mentioned phenomena. During the correction of this type of problem, the fibers in the supporting tissues become stretched (Maltha et al., 2015). This stretching causes tension in the fiber system (Reitan, 1967), which then becomes extensively displaced and deformed (Edwards, 1968). While, periodontal fibers reorganize during and after rotation, the gingival collagen fiber network takes up to six months to remodel, which explains the protracted period lasting several months of the tendency for such a corrective movement towards relapse (Reitan, 1967; Edwards, 1968; Boese, 1969).

In the time period subsequent to the re-organization of the supporting tissues, teeth are subjected to variable muscular forces acting on the periodontium together with the associated modeling and remodeling (Horowitz and Hixon, 1969; Littlewood, 2013), as well as the other factors that have been implicated in post-treatment changes that will be analyzed later. Adaptations that happen during this phase cannot usually be distinguished from normal aging developments, which occur naturally in any person regardless of whether they are under orthodontic treatment or not (Thilander, 2011).

2.1.2. Occlusal factors

To ensure stable results after orthodontic treatment, the posterior teeth should be in interdigitation, while the anterior teeth should be in an edge centroid relationship (Houston, 1989; Littlewood, 2013), in which the edges of the lower incisors lie 0-2 mm anterior to the mid-point of the root axis of the upper incisors (Littlewood, 2013). In addition, the inter-incisal angle should be around 135 degrees to produce an occlusal stop serving to prevent continued eruption of lower incisors (Littlewood, 2013). Any of the above mentioned criteria, if not properly satisfied, might constitute factors leading to post-treatment changes.

2.1.3. Soft tissues

Soft tissues are also considered to be crucial factors affecting the stability and alignment of teeth in the alveolar bone. This explains why it is important to achieve a balanced equilibrium between oral soft tissues, including equilibrium between lips and tongue (Littlewood, 2013), and an equilibrium between the intraoral and extra-oral soft tissues (Mills, 1966; Horowitz and Hixon,1969; Melsen and Athanasiou, 1978).

Function is considered to be critical for the final position of teeth on the alveolar base and it plays a role in affecting the retention of the achieved occlusion (Nanda and Burstone, 1993). Hence, post-orthodontic stability is facilitated if the occlusion obtained is in harmony with the functions of the oral cavity and, in particular, with the activity of the masticatory muscles (Dahlquist et al., 1996; Konstantonis et al., 2013). This highlights the importance of making a correct diagnosis and always defining a treatment plan based on the clinical and instrumental data obtained, taking into account the functional aspects of the oral cavity (Ciger et al., 2005; Raucci et al., 2014).

2.1.4. Growth

Continuing growth may be an additional reason for instability in the occlusion presenting a considerable time after treatment, especially in patients treated at pre-adolescence and adolescence (Nanda and Burstone, 1993). This is why cases showing greater degrees of growth during treatment tend to show less relapse (Joondeph, 2011). Moreover, the increase in the degree of lower incisor crowding defined as the maturation of dentition, might also be

considered in this category (Little et al,. 1988; Little, 1990; Nanda and Zernick, 1993; Littlewood, 2013; Johnston and Littlewood, 2015). Consequently, growth might be a support in the correction of orthodontic problems, but is also a potential source of changes in treated cases (Riedel, 1960; Joondeph, 2011).

2.1.5. Physiologic dento-alveolar adaptation

During normal dento-alveolar adaptation it is expected that there will be a moderate increase in intercanine width until eruption of the permanent canines (Moorrees, 1959; DeKock, 1972), followed by a reduction in this dimension (Moorrees, 1969; Sinclair and Little, 1985). While the inter-molar width tends to be stable for 13 to 20 years (Moorrees, 1959; DeKock, 1972; Moyers, 1973; Sinclair and Little, 1983;1985), and there will be a reduction in the anteroposterior dimension of the mandibular arch with time (Moorrees, 1959; Sinclair and Little, 1983; Allred, 1986). All these phenomena constitute physiologic changes occurring independently of orthodontic treatment (Nanda and Zernick, 1993). This is why the lack of retention in the short term may undesirably affect tooth alignment and interproximal contacts, thus confirming the necessity for immediate retention after orthodontic treatment (Lyotard et al., 2010).

2.2. Types of orthodontic retainers

Fixed wire retainers and removable appliances are usually used to counteract any tendency for relapse (Kaklamanos et al., 2016)

2.2.1. Fixed retainers

Fixed retainers are typically formed from a segment of wire (or less often fiber-reinforced composite or polyethylene strips) bonded to the lingual /palatal surface of the anterior teeth and can be fabricated in different forms. Retainers bonded only to the canines use rigid 0.025" stainless steel wire, which facilitates maintaining oral hygiene in the incisor region. Retainers bonded to all the anterior teeth usually entail the use of 0.0175" multistrand wire (Zachrisson, 1977; 1995; Littlewood, 2013; Joondeph, 2011). Using this approach, the wire should be flexible enough to allow physiologic tooth movement (Bearn, 1995; Joondeph, 2011), while still maintaining the teeth in the intended position (Littlewood, 2013; Joondeph, 2011). Multistrand wire is considered to be the gold standard for a fixed retainer (Joondeph, 2011). A recent systematic review could not locate any differences in terms of survival rates between conventional multistrand stainless steel wire retainers and the newer types of polyethylene retainers (Littlewood et al., 2016).

Fixed retention has the advantage of not relying on patient compliance as the patient does not need to remember to wear it (Littlewood, 2013). However, there is general agreement that the prolonged use of such types of retainers makes maintaining oral hygiene more challenging (Lew, 1989; Lima et al., 2012). The difficulty may arise from the wire being crossed in the interdental regions; thus creating an area that is difficult to clean; especially in the interproximal and areas gingival to the wire (Keim et al., 2002; Butler, 2005; Pandis et al., 2007). It is possible that compromised oral biofilm removal could lead to periodontal decline and changes such as recession, increased probing depths and bleeding on probing, as well as enamel demineralization (Pandis et al., 2007; Levin et al., 2008; Corbett et al., 2015; Al-Kuwari et al., 2015). The possibility of iatrogenic damage to the teeth and their supporting structures may occur

irrespective of the type or size of wire used and appears to be more correlated with the length of time the bonded retainer is in place (Artun, 1984). Hence, prolonged use fixed retainers could possibly lead to greater adverse effects compared to the use of removable retainers (Pandis et al., 2007; Al-Kuwari et al., 2015; Rody et al., 2016). Indeed, it has been observed that the presence of mandibular fixed retainers may influence GCF biomarker and lead to increased plaque accumulation and gingivitis levels (Rody et al., 2011; Dietrich et al., 2015). However, it does not seem be associated with severe clinical damage to the periodontal supporting tissues (Heier et al., 1997; Booth et al., 2008; Corbett et al., 2015; Rody et al., 2016; Dietrich et al., 2015) nor increased demineralization of the enamel (Al-Kuwari et al., 2015).

Another recently recognized problem with flexible spiral wire retainers bonded to all 6 mandibular anterior teeth is that they might prompt undesirable movement of the anterior teeth such as torque differences between the adjacent mandibular incisors and increased buccal inclination and the movement of mandibular canines (Katsaros et al., 2007; Renkema et al., 2011). Although rare, such complications might be extensive enough to warrant retreatment (Katsaros et al., 2007; Pazera et al., 2012).

As the clinical impact of the abovementioned data has not been fully clarified, it has been suggested that retention protocols be cautiously applied after a thorough consideration of anatomical, oral hygiene and social factors, as well as, the close monitoring of patients through frequent recalls (Pandis et al., 2007). Thus, some clinicians may opt to prescribe removable appliances, especially in patients unable to follow a recall program or patients that have presented problems with oral hygiene.

2.2.2. Removable retainers

Removable retainers are appliances that patients can insert, wear and remove by themselves. The advantage of this type are: they offer easy cleaning, leading to good oral hygiene maintenance; socially acceptability; part time wear is possible, as can be removed if required, and inexpensive. On the other hand, the disadvantages are: patient cooperation is paramount; speech difficulties may occur in the first few days; and they suffer high rates of breakage and loss (Blake and Garvey, 1998).

Various types of removable retainers are used in orthodontics; acrylic and wire appliances like the Hawely retainer (Hawley, 1919) and the Begg retainer, clear thermoplastic retainer like the Essix retainer and the positioner (Littlewood, 2013).

The **Hawley retainer** is the original removable type retainer. It was used as an active removable appliance and then found usable as a retainer as well (Littlewood, 2013; Joondeph, 2011). It has many advantages including being easy to fabricate, effective and can be worn while eating (Littlewood, 2013; Johnston et al., 2013; Johnston and Littlewood, 2015). It maintains the transverse correction, since it is rigid, easy to add pontics to in order to provide good esthetics and retain a missing tooth space (Joondeph, 2011; Littlewood, 2013). It has the following components: (1) The palatal or lingual part made of acrylic, covering the whole or part of the palatal mucosa using a horseshoe form in contact with part of the mucosa and lingual surface of the teeth (Joondeph, 2011); (2) A labial bow fabricated of stainless steel wire, 0.020 to 0.036 inches which sits in contact with the labial surface of the anterior teeth; either four or six maxillary anterior teeth (Joondeph, 2011); (3) Retention clasp; and (4) An anterior bite plane, which can be easily added for deep bite cases (Johnston and Littlewood, 2015). These components should not interfere with either tooth movement or the occlusion (Joondeph, 2011).

The **Essix retainer** was introduced in 1993 by Sheridan et al. it is made of plastic 0.025" thick and provides full coverage of the teeth of the upper and lower dentition. It has the advantages of being an esthetic, comfortable, inexpensive modern alternative to traditional retainers (Littlewood, 2013). It is easy to manufacture and can be delivered to the patient on the same day as debonding the orthodontic appliance (McNamara et al., 1985; Sheridan et al., 1993; Johnston et al., 2013). On other hand, disadvantages have been mentioned such as a short life span, tendency to fracture, and the development of unwanted occlusal contacts, which may affect the treatment result in cases where premature occlusal contacts are found in the posterior teeth potentially leading to an anterior open bite (Lindauer and Shoff, 1998; Sheridan et al., 2001). Moreover, when this retainer is worn full time, it could increase the risk of the pathological decalcification of enamel due to disruption of the salivary flow, particularly in conjunction with the acidity of certain type of diet, such as beverages (Sheridan et al., 2001). Thus, it is contraindicated for patient with poor oral hygiene despite being removable (Sheridan et al.,1993; Littlewood, 2013).

The positioner appliance is rarely used because of its expense and the need for patient compliance (Blake and Garvey, 1998). The patient cannot eat or talk with it in place; which results in limiting the wearing time. In addition, it is contraindicated for any patient with an airway obstruction and mouth breathing because the patient will encounter difficulty in respiration with this appliance (Joondeph, 2011). Nevertheless, it has been suggested to be good for cases with gingival hyperplasia which occur during orthodontic treatment and reestablishing normal tissue tone (Joondeph, 2011).

2.3. Retention regimens

There are, as yet, no definite indications for choosing the optimal period of retention or type of retainer for specific cases (Melrose and Millett, 1998; Johnston et al., 2013) and protocols for ideal orthodontic retention practices are still undetermined (Pratt et al., 2011).

In present orthodontic practice, significant variation exists in the duration of the retention period used (Johnston and Littlewood, 2015). This is based on a number of factors including the preference of the individual orthodontist, the variety of occlusal, skeletal and soft tissue relationships, as well as the paucity of well-controlled scientific studies (Johnston and Littlewood, 2015). According to a survey conducted in the United Kingdom during the 1990s, the greatest retention period used was found to be 12 months (Clark et al., 1997). Nevertheless, it is believed that variations in the duration of removable retainer wear in particular are clinically acceptable (Johnston and Littlewood, 2015). It has been suggested that the duration of retention must be decided for each case specifically, in combination with the informed patient, taking into consideration future growth (Nanda and Nanda, 1992; Haines and Williams, 1995).

Variations in the type of orthodontic retainer used have also been found among various countries or even within the same country (Wong and Freer, 2004; Singh et al., 2009; Lai et al., 2014). There are some similarities, such as the frequent use of fixed retainers in the mandibular dental arch or the growing popularity of removable vacuum-formed retainers among orthodontists practicing in countries such as Australia and New Zealand, Netherlands, United Kingdom, United States, Norway and Malaysia (Wong and Freer, 2004; Renkema et al., 2009; Singh et al., 2009; Valiathan & Hughes, 2010; Pratt et al., 2011; Vandevska-Radunovic et al., 2013; Ab Rahman et al., 2016). However, in Norway, they use both fixed and removable retainer for the maxillary dental arch (Vandevska-Radunovic et al., 2013). In Switzerland they have found that

bonded retainers are the most popular type of retention in the upper and lower dental arches; but these may be supplemented with removable retention in cases such as extraction treatment, or in patients who have had maxillary expansion (Lai et al., 2014). Moreover, in Saudi Arabia the Hawley retainer is preferred for the maxilla with 90.3% of them recommended for full-time maxillary removable retainer wear (Al-Jewair et al., 2016). Removable retainers, in particular, are often used despite concerns about poor compliance, especially in adolescents (Ackerman and Thornton, 2011).

From the abovementioned data it seems clear that these differences are based on personal preference, experience and other nonscientific criteria, and not on a solid scientific base (Renkema et al., 2009), although, a high percentage of orthodontists agree about retention and post-treatment changes being problems in daily practice and a scientific area where guidelines are necessary (Renkema et al., 2009). Despite the large number of studies on stability after orthodontic treatment, particularly involving the use of different removable retainers, evidence-based conclusions are few (Lai et al., 2014). Thus, proper retention guidelines need to be developed based on a well established, firm scientific base (Melrose and Millett, 1998; Renkema et al., 2009).

3. AIM

3.1. Aim of the systematic review

To compare the performance of the Hawley-type appliances with the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment.

3.2. Objectives of the systematic review

a. To compare the performance of the Hawley-type appliances with the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment regarding outcomes relevant to maxillary and mandibular dental arch measurements, dental arch relationships and occlusal contacts, speech evaluation, patient reported outcomes, adverse effects and problems related to the appliances, as well as, economic evaluation related outcomes.

b. To compare the performance of different appliance wearing schedules and protocols of Hawley-type appliances used after the completion of orthodontic treatment regarding outcomes relevant to maxillary and mandibular dental arch measurements, dental arch relationships and occlusal contacts, speech evaluation, patient reported outcomes, adverse effects and problems related to the appliances, as well as, economic evaluation related outcomes.

3.3. Null hypotheses

a. There is no difference between the comparative performances of the Hawley-type appliances with the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment regarding outcomes relevant to maxillary and mandibular dental arch measurements, dental arch relationships and occlusal contacts, speech evaluation, patient reported outcomes, adverse effects and problems related to the appliances, as well as, economic evaluation related outcomes.

b. There is no difference in the comparative the performance of different appliance wearing schedules and protocols of Hawley-type appliances used after the completion of orthodontic treatment regarding outcomes relevant to maxillary and mandibular dental arch measurements, dental arch relationships and occlusal contacts, speech evaluation, patient reported outcomes, adverse effects and problems related to the appliances, as well as, economic evaluation related outcomes.

4. MATERIALS AND METHODS

4.1. Protocol development

The present systematic review was based on a specific protocol developed following the guidelines outlined in the PRISMA statement (Moher et al., 2001) and the Cochrane Handbook for Systematic Reviews of Interventions (version 5.1.0) (Higgins and Green, 2011).

The protocol was registered with PROSPERO - International Prospective Register of Systematic Reviews, produced by the Centre for Reviews and Dissemination, University of York, United Kingdom (see Appendix I, Wafa Al Rahma, Eleftherios G. Kaklamanos, Athanasios E. Athanasiou. Performance of acrylic-based retention appliances: a systematic review of randomized controlled trials. PROSPERO 2016:CRD42015029279; Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015029279).

4.2. Selection criteria applied for the review

The selection criteria for the domains of study design, participant characteristics, intervention characteristics and principal outcome measures applied for the present review were as follows:

4.2.1. Types of study design

Studies included in the present thesis had to be Randomized Clinical Trials (RCTs) evaluating the performance of Hawley- type in maintaining orthodontic treatment result.

Animal studies, non-comparative studies (case reports and case series), systematic reviews and meta-analyses were excluded.

The type of study design was assessed using the algorithm available from SIGN (Scottish Intercollegiate Guidelines Network) available from http://www.sign.ac.uk (Appendix II).

4.2.2. Types of participants

The included studies could involve patients of any age referred for retention after a full course of orthodontic treatment.

Studies involving patients at the end of an initial phase of removable or functional appliance treatment, after combined orthodontic and surgical treatment, exhibiting congenital anomalies of the craniofacial region, as well as, physical or mental handicaps that could compromise dexterity or compliance with the applied retention protocol were excluded.

4.2.3. Types of interventions

The included studies could involve patients allocated to groups using any Hawley-type removable retention appliances (i.e. typical Hawley retainers or Begg retainers) compared to groups wearing other removable retention appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.). In addition, the included studies could involve patients

allocated to groups of different wearing schedules of Hawley-type removable retention appliances.

As Hawley-type removable retention appliances were considered to be the appliances consisting of an acrylic base-plate covering the palate or the lingual cervical portion of the mandibular teeth combined with clasps (usually Adam's clasps) on the first molars and a labial bow on the anterior teeth (like the Hawley retainers). Moreover, appliances similar to the aforementioned, but characterized by the absence of clasps crossing the occlusal surfaces and the presence of a labial bow wire, running on the labial surfaces of all teeth (like the Begg retainers), were considered.

Clear thermoplastic appliances were considered to be the removable retainers made from a clear material and molded with the aid of heat and/or pressure over a patient's cast, separately for the upper and lower dental arch, after the completion of active orthodontic treatment. Positioners were regarded the removable appliances that are usually made from a single piece plastic material and fit simultaneously in the patient's upper and lower dental arch, guiding the teeth into ideal position and interdigitation.

Studies involving the additional use of any type of fixed retainers or any kind of adjunctive techniques like inter-proximal reduction or pericision were excluded.

4.2.4. Types of outcome measures

The studies included in the present review had to primarily provide quantified measurements on any outcome reflecting stability of treatment result, i.e. changes in teeth alignment, arch form and occlusion such as the Irregularity Index (Little, 1975), intercanine and intermolar width, arch length, overbite and overjet.

Secondary outcomes considered in the present review were patient reported outcomes, speech assessment, data on compliance, data on the condition of retainers (e.g. integrity and failure), economic evaluation data, as well as hard and soft oral tissue health and possible adverse effects.

4.3. Search strategy for identification of studies

The principal investigator (WJA) developed detailed search strategies for each database involved. These were based on the strategy developed for MEDLINE, but revised appropriately for each database to take account of the differences in controlled vocabulary and syntax rules. The following electronic databases were searched (Appendix III): MEDLINE via PubMed (http://www.ncbi.nlm.nih.gov/pubmed), the Cochrane Central Register of Controlled Trials (CENTRAL) (http://onlinelibrary.wiley.com/cochranelibrary), Cochrane Database of Systematic Reviews (http://0-ovidsp.tx.ovid.com.amclb.iii.com/sp-3.16.0b),Scopus (www.scopus.com), Web of Science™ Core Collection (http://apps.webofknowledge.com/), Latin-American and Caribbean System on Health Sciences Information (LILACS) (http://lilacs.bvsalud.org/en/), Scielo (http://www.scielo.org/php/index.php?lang=en),National Databases of Indian Medical Journals (IndMed) (http://indmed.nic.in/indmed.html), Deutsche Zentralbibliotek fuer Medizin (https://www.livivo.de) and Arab World Research Source (http://0-web.a.ebscohost.com.amclb.iii.com). Unpublished literature was accessed electronically using Google Scholar (https://scholar.google.com), ClinicalTrials.gov (http://clinicaltrials.gov), International Standard Randomised Controlled Trial Number (ISRCTN) registry (http://www.isrctn.com) and OpenGrey (http://www.opengrey.eu). In addition, Pro-Quest Dissertation and Theses Global database (http://search.proquest.com) was searched.

No restriction was placed on the language, date or status of publication. Together with these searches, efforts were made to obtain conference proceedings and abstracts where possible and

the reference lists of all eligible studies for additional records were searched. Inter-reviewer agreement was not assessed because it is not recommended as slandered procedure (Higgins and Green, 2011).

4.4. Selection of studies and data extraction

The principal investigator (WJA) and the thesis co-supervisor (EGK) independently assessed the retrieved records for inclusion. They were not blinded to the identity of the authors, their institution, or the results of the research. They obtained and assessed, again independently, the full report of records considered by either reviewer to meet the inclusion criteria. Disagreements were resolved by discussion or consultation with the thesis supervisor (AEA). A record of all decisions on study identification was kept.

The same two persons performed data extraction independently and any disagreements were again resolved by discussion or consultation with the thesis supervisor (AEA). Data collection forms were used to record the desired information.

a. Bibliographic details of the study.

b. Details on study design and verification of study eligibility.

c. Participant characteristics.

d. Intervention characteristics.

e. Details on outcomes assessed and assessment procedures. When a study recorded outcomes at various time points, then only the latest measurements were extracted. Where needed, numerical data where transformed into the desired formats and tested statistically using MedCalc (©2016 MedCalc, Belgium) and QuickCalcs (©2016 GraphPad Software, Inc, USA.

f. Additional information: a prior sample size calculation, methodology reliability assessment and data on compliance.

The outcomes retrieved from the studies included in the present review were categorized as follows:

a. Maxillary and mandibular dental arch measurements.

b. Dental arch relationships and occlusal contacts.

c. Speech evaluation.

d. Patient reported outcomes.

e. Adverse effects and problems related to the appliances.

f. Economic evaluation related outcomes.

Inter-reviewer agreement was not assessed because it is not recommended as slandered procedure (Higgins and Green, 2011).

4.5. Data synthesis and assessment of publication bias

In situations where the retrieved data used different indices measuring the same concept on different scales with a high degree of correlation, the effects of the interventions were planned to be expressed as standardized values (i.e. the Standardized Mean Difference (SMD) together with the relevant 95% Confidence Interval (CI)), in order to enable quantitative synthesis (Deeks et al., 2001). In a case where, in a particular comparison, only one index was recorded, the intervention effect was planned to be expressed as the Weighted Mean Difference (WMD) together with the 95% CI.

The random effects method for meta-analysis was to be used to combine data from studies that reported similar measurements in appropriate statistical forms (Der Simonian and Laird, 1986, Borenstein et al., 2009), since they were expected to differ across studies due to clinical diversity, in terms of participant and intervention characteristics.

To identify the presence and extent of between-study heterogeneity, the overlap of the 95% CI for the results of individual studies was to be inspected graphically, and Cochrane's test for homogeneity and the I^2 statistic were to be calculated (Higgins and Green, 2011). The results of the I^2 statistic were to be interpreted as follows (Higgins and Greene, 2011):

- I^2 from 0% to 40%: heterogeneity might not be important;
- I^2 from 30% to 60%: may represent moderate heterogeneity;
- I^2 from 50% to 90%: may represent substantial heterogeneity;
- I^2 from 75% to 100%: considerable heterogeneity.

If deemed possible, exploratory subgroup analyses were planned according to participant and intervention characteristics.

In addition, if a sufficient number of trials were identified, analyses were planned for "small-study effects" and publication bias (Higgins and Green, 2011).

All analyses were to be carried out with Comprehensive Meta-analysis software 2.2.046 (©2007 Biostat Inc.). Significance (a) was set at 0.05, except for 0.10 used for the heterogeneity tests (Ioannidis, 2008).

4.6. Risk of bias assessment and determination of the quality of evidence

The principal investigator (WJA) and the thesis co-supervisor (EGK) were to assess the risk of bias in the included studies, independently and in duplicate, during the data extraction process, using The Cochrane Collaboration's Risk of Bias assessment tool for RCTs (Higgins and Green, 2011). Any disagreements were to be resolved by discussion or consultation with the thesis supervisor (AEA). The Risk of Bias assessment tool includes the following domains.

a. Random sequence generation (selection bias).

b. Allocation concealment (selection bias).

c. Blinding of participants and personnel (performance bias).

d. Blinding of outcome assessors (detection bias).

e. Incomplete outcome data (attrition bias).

f. Selective outcome reporting (reporting bias).

g. Other sources of bias.

After entering in the data extraction form the information reported in each study, every domain would receive a judgment of low, high or unclear risk of bias (indicating either lack of sufficient information to make a judgment or uncertainty over the risk of bias) (Higgins and Green, 2011). Subsequently, studies were to be judged as being of low, unclear or high risk of bias (Higgins and Green, 2011).

a. Low risk of bias (plausible bias unlikely to seriously alter the results)

b. Unclear risk of bias (bias that raises some doubt about the results)

c. High risk of bias (bias that seriously weakens confidence in the results)

The quality of evidence and strength of recommendations at the longest follow up available for key outcomes of the systematic review were ultimately to be assessed based on the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) approach (Guyatt et al., 2011).The GRADE profiler (GRADEpro) software (available www.gradepro.org; © 2015, McMaster University and Evidence Prime Inc.) was to be used to facilitate the summary regarding the quality of evidence using the GRADE approach. The principal investigator (WJA) and the thesis co-supervisor (EGK) were to assess the quality of evidence independently and in duplicate. Any disagreements were to be resolved by discussion or consultation with the thesis supervisor (AEA). Inter-reviewer agreement was not assessed because it is not recommended as slandered procedure (Higgins and Green, 2011).

During the GRADE assessment and for the purpose of summarizing risk of bias across studies, where possible, relevant information was to be judged as being of low, unclear or high risk of bias.

a. Low risk of bias: most information is from studies at low risk of bias.

b. Unclear risk of bias: most information is from studies at low or unclear risk of bias.

c. High risk of bias: information from studies at high risk of bias could have an effect on the interpretation of the results.

5. RESULTS

5.1. Results of the search

The flow of records through the reviewing process is shown in Figure 1. We initially identified 1146 records through database search and 28 records through other sources, but excluded 630 as duplicates, and 516 more on the basis of their title and abstract. Finally, 10 full-text reports were included in the systematic review (Zhang and Wang, 2003; Rohaya et al., 2006; Hichens et al., 2007; Rowland et al., 2007; Liao, 2010; Shawesh et al., 2010; Tsai, 2010; Sun et al., 2011; Atik et al., 2016; Wan et al., 2016).

5.2. Study characteristics

The characteristics of the studies included in the present systematic review are presented in Tables 1 and 2. The papers, which were published between 2003 and 2016, randomized in different groups 1024 subjects referred for retention after a full course of orthodontic treatment, and analyzed various outcomes in 854 of them that were followed for a maximum period of one year after the removal of the fixed orthodontic appliances.

Eight of the eligible publications investigated groups of subjects using Hawley removable appliances and clear thermoplastic retainers (Rohaya et al., 2006; Hichens et al., 2007; Rowland et al., 2007; Liao, 2010; Tsai, 2010; Sun et al., 2011; Atik et al., 2016; Wan et al., 2016). Two of the aforementioned publications comprised part of a larger trial. Another study (Zhang and Wang, 2003) compared the Hawley appliance to positioner use during the retention period.

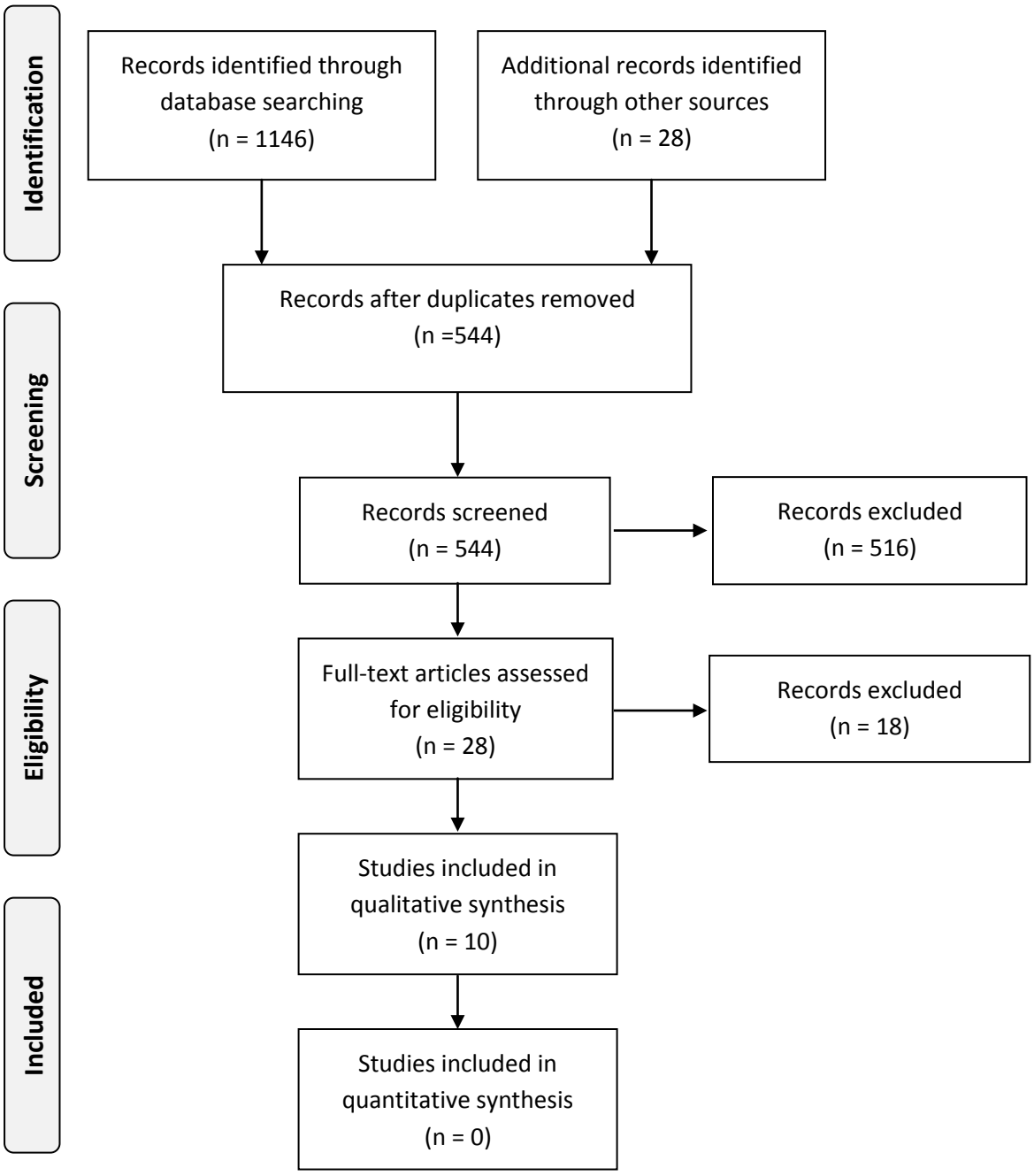


Figure 1. Flow of records through the reviewing process

Finally, one study (Shawesh et al., 2010) involved patients allocated to groups using different wearing schedules for Hawley retention appliances.

Table 1. General characteristics of the studies included in the systematic review.

Study	Intervention characteristics	Included outcomes	Additional information
Atik et al., 2016 **Turkey**	**Group 1**: **Hawley retainers** Made with Adams clasps, labial wire with vertical loops, and lingual acrylic. The acrylic part of the Hawley retainer had a uniform thickness of 2–3 mm, and was trimmed into a horseshoe shape. **Group 2**:**Clear thermoplastic retainers** Constructed from plastic, copolyester 0.75 mm Essix sheet materials according to the manufacturer's instructions, and the maxillary retainers trimmed into a horseshoe shape. Patients were instructed to wear the retainers 24 hours a day for 6 months, including while eating, but to remove them when brushing their teeth. It was recommended that the patients not read paragraphs out loud to expedite speech adaptation, in order to allow spontaneous flow during speech adaptation.	**Formant frequencies of sustained vowels** [a, e, i, u] and of vowel [a] in combination with consonants [b, d, g, t, ş, ç, l, z, c, m, n] **Voice onset time (VOT) values** of voiceless [t] and voiced [d] stop consonants in combination with vowel [a] The speech sound assessments were performed by a speech–language pathologist. The recording procedure was repeated for each participant at four different time points: 1. on the first day, 2. 1 week later, 3. 4weeks later, and 4. 3 months later. On the first day, the assessments were conducted prior to inserting the retainers, immediately after maxillary and mandibular retainer application, individually, and with both retainers applied. The later assessments were conducted while both retainers were worn. The recording order was the same for every patient.	**A priori sample calculation:** Thirty patients were selected at the beginning of the study to provide a power of 80 per cent with a 5 per cent significance level to detect a true difference. For the most important five variables (ç,g, m, u, a), the sample sizes were calculated separately based on the difference in the groups between the different time points; maximum sample size was taken into account. The power analysis was done via the two-way repeated measures. **Information on compliance:** No **Reliability of measurements:** NR
Hichens et al., 2007 **United Kingdom**	**See Rowland et al. (2007)**	**Patient satisfaction questionnaire:** completed at 3 and 6 months post-debonding **Costs to the National Health Service** **Costs to the orthodontic practice** **Costs to the patient** Unit costs in euros were used at 2003 prices to value the resources included in the analysis. No adjustment or discount of the costs was carried out. All direct costs were included. The type of costs included in the analysis depended on the perspective. Total clinical time cost was calculated by multiplying the gross clinical time cost per minute by the total clinical time spent on retainer appointments, over the 6-monthperiod, for each retainer group.	**A priori sample calculation:** Existing, but not based on these outcomes. **Information on compliance:** Self-reported **Reliability of measurements:** NR

NR: Not Reported

Table 1. [Continued]

Study	Intervention characteristics	Included outcomes	Additional information
Liao, 2010 **China**	**Group 1**: **Hawley retainers** Made with circumferential clasps on the molars (0.9 mm), labial wire with vertical loops (0.8 mm), and lingual acrylic. **Group 2**:**Clear thermoplastic retainers** 0.15 mm (??) hard transparent sheet (Thermo-Forming Materials, Netherlands). The edges exceeded 1 mm from the gingival margin on the vestibular side and 2-3 mm from the palatal/lingual side. Patients were instructed to wear the retainers 24 hours a day, except meals, for 6 months	**Intercanine and Intermolar width** **Arch length [canines and molars]** Measurements on dental casts taken at the removal of the appliances and 6 months after.	**A priori sample calculation:** NR **Information on compliance:** No **Reliability of measurements:** NR
Rohaya et al., 2006 **Malaysia**	**Group 1**: **Hawley retainers** A standard upper Hawley retainer, with a plain labial bow, Adam's clasps on the first molars and cold-cured acrylic. Patients were asked to wear the retainer full time for the first 3 months, followed by 6months nighttime wear only. **Group 2**:**Clear thermoplastic retainers** A vacuum-formed retainer made fromEssix™ type C thermoplastic copolyester material (RaintreeEssix Inc., Ortho-Care (UK) limited) with a 0.5mm thickness after thermoformed, full-arch tooth coverage and the edges extended about 2-3mm into labial gingival. Patients were asked to wear the retainer full-time for one week, followed by evening and night-wear for three months and 6 months for night time wear only.	**Rotations of the upper teeth (labial segment and premolars)** were measured using a protractor on patients' casts at the end of the follow up period (9 months). Rotations were either scored as present (i.e. teeth that were rotated 15^{o} or more from normal arch alignment) or absent. For the upper central incisors especially with V shaped arch, they were considered rotated disto-palatally if the angle between them was less than 90^{o}.	**A priori sample calculation:** The intra and inter-examiner agreement analyzed using Kappa statistics showed very good agreement of intra-examiner (k= 0.82) and moderate agreement of inter-examiner(k= 0.47). Most of the disagreement involved the upper labial segment, which can be difficult to assess due to the shape of their arch form. The disagreements were discussed and a standardized protocol agreed upon**.** **Information on compliance:** No **Reliability of measurements:** Yes.

NR: Not Reported

Table 1. [Continued]

Study	Intervention characteristics	Included outcomes	Additional information
Rowland et al., 2007 **United Kingdom**	**Group 1**: **Hawley retainers** The Hawley retainers were constructed with an acrylic base plate and Adams clasps fabricated of0.7-mm diameter stainless steel wire on the first standing molars. A Hawley bow (open looped short labial bow) was also made from 0.7-mm stainless steel wire; it extended from canine to canine. The Hawley bow was then contoured with acrylic resin to contact the labial surfaces of the incisors. The patients were instructed to wear the maxillary and mandibular Hawley retainers 24 hours a day for 3months, including while eating, but to remove them when brushing their teeth. After 3 months, wear time was reduced to 12 hours a day. **Group 2**:**Clear thermoplastic retainers** Constructed from an Erkodur blank (Erkodent, Erich Kopp, GmbH, Pfalzgrafenweiler, Germany)1.5 mm in thickness. The retainer was trimmed to provide 1 to 2 mm buccal and 3 to 4 mm lingual extensions past the gingival margin. All occlusal surfaces were covered up to and including the most distal tooth. Patients were instructed to wear the VFRs 24 hours a day for the first week and remove them only for eating and brushing their teeth. After the first week, wear time was reduced to 12 hours a day.	**Tooth rotations:** The rotation of the incisors and the canines was determined by constructing a line that bisected2 points per tooth that best marked its rotational angulation (Fig 5). The angle formed by the intersection of this line with the line forming arch depth gave the measurement of rotation of the tooth. Arch depth was defined as the length of a line perpendicular to the intermolar width that passed through the midpoint of the contact points of the central incisors. The measurement of rotation for the premolars was calculated by constructing a line that bisected the buccal and the palatal/lingual cusp tips.In premolars with 2 lingual/palatal cusps, the mesiopalatal/mesiolingual cusp was bisected to form the line. The angle formed by the intersection of this line with the line forming arch depth gave the rotation of the tooth. **Intercanine width** **Intermolar width** **Little's Irregularity Index:** Displacement between the midpoint of the incisor edges. Measurements on digitized images of study casts taken at debonding, 3 and 6 months after.	**A priori sample calculation:** Total sample size of 388subjects would give a power of 80% with a 5%significance level to detect a true difference in contact point displacement of greater than 0.2 mm. **Information on compliance:** Self-reported **Reliability of measurements:** The intra-observer reliability coefficients ranged from 0.96 - 1.0 for linear measurements and from 0.93to 1.0 for angular, demonstrating that the method had good reliability.

NR: Not Reported

Table 1. [Continued]

Study	Intervention characteristics	Included outcomes	Additional information
Shawesh et al., 2010 **United Kingdom**	**Group 1**: **Hawley retainers full-time** 6 months full-time followed by 6 months night-only **Group 2**:**Hawley retainers part-time** 12 months night-only Upper Hawley retainer. Adams' cribs were placed on both upper first permanent molars and a long labial bow was taped and soldered to, and extended from, the bridges of these cribs. The anterior part of the labial bow was covered in acrylic to engage the embrasures between the incisors. The base plate was manufactured with acrylic that contacted the palatal surface of all teeth around the entire arch. Lower Hawley retainer. Adams' cribs were placed on both lower first permanent molars. The labial bow was constructed to extend from the lower permanent canine to the canine and the labial aspect of the bow was covered in acrylic that engaged the embrasures between the incisors. The base plate was manufactured with acrylic that contacted the lingual surface of the teeth all the way round the arch.	**Little's Irregularity Index:** The sum of the distances between the anatomic contact points from the mesial of the left canine to the mesial of the right canine in each labial segment **Incisor crowding:** The difference between the sum, in mm, of the canine-to-canine tooth widths and the space in the labial segment from canine to canine. The available space in the labial segment was measured by dividing the labial segment into two straight-line segments, extending from the distal contact point to the canine on each side to the midpoint between the central incisors. Measurements on study casts taken at debonding and 12 months after.	**A priori sample calculation:** Power of 90 per cent to detect a clinically significant difference of 2 mm in labial segment alignment, assuming that the common standard deviation is 2 mm using a two-group t-test with a 0.05, two-sided significance level. **Information on compliance:** NR **Reliability of measurements:** Little's index/lower arch, 0.96; Little's index/upper arch, 0.95; labial segment crowding/lower arch, 0.90; and labial segment crowding/upper arch, 0.81.
Sun et al., 2011 **China**	**Group 1**: **Hawley retainers** Composed of a 2-mm-thickacrylic resin base plate, one-arm clasps (0.9 mm stainless steel (SS) wire on the first molars, and a Hawley bow (0.7mm SS). The base plate covered most of the hard palate and was meticulously adapted to the lingual surfaces of the teeth. **Group 2**:**Clear thermoplastic retainers** 0.75-mm-thick thermoplastic material (Biolon, Dreve Dentamid GmbH, Unna, Germany). The buccal edge paralleled the gingival margin, and the lingual portion extended 4 to 5 mm beyond the lingual gingival margin. The distal portion extended as far as the second molars, and all occlusal surfaces were tightly covered. Patients in both groups were required to wear the retainers full-time, except during meals	**Retainer failure as judged by the professional:** Retainer fracture, retainer loss, retainer no longer fitting (because of retainer deformation) or tooth relapse, retainer with serious abrasion causing penetration. Retainers which had slight cracks on the surface were not considered as a breakage unless the retainers could not be worn because of crack expansion. The numbers of each type of failure were totalled at the end of the 12 month follow-up period.	**A priori sample calculation:** NR **Information on compliance:** NR **Reliability of measurements:** Not applicable

NR: Not Reported

Table 1. [Continued]

Study	Intervention characteristics	Included outcomes	Additional information
Tsai, 2010 **United States of America**	**Group 1**: **Hawley retainers** Fabricated with ball clasps mesial to the first molar. **Group 2**:**Clear thermoplastic retainers** Fabricated using Essix C+ and were trimmed 2 mm below the gingival margins. Both groups were instructed to wear their retainers full-time, except when eating and brushing.	**Occlusal contacts:** Quantified in the posterior area with an imaging software aiming at investigating the thickness of bilateral bite registrations taken in maximum intercuspation using a silicon bite registration material. **Subjective assessment of occlusion questionnaire** Assessments at debonding, one and three months after the removal of fixed orthodontic appliances.	**A priori sample calculation:** NR **Information on compliance:** NR **Reliability of measurements:** NR
Wan et al., 2016 **China**	**Group 1**: **Hawley retainers** constructed with a 2-mm-thick U-shaped acrylic base-plate, one-arm clasps with an 0.8-mm diameter stainless steel wire on the first standing molars, and a Hawley bow with 0.8-mm diameter stainless steel wire. **Group 2**:**Clear thermoplastic retainers** 0.8-mm-thick thermoplastic material, tightly covering all occlusal surfaces and trimmed to provide approximately2-mm lingual extensions past the gingival margin. Patients instructed to wear the maxillary and mandibular retainers all the time, except while eating or brushing their teeth.	**Articulation of four long vowels and five voiceless fricatives** from the International Phonetic Alphabet: Patients asked to pronounce 36 words (four typical words for each symbol selected from a dictionary as a speech stimulus) at their comfort levels of pitch and loudness in a quiet, soundproof room. The recorded samples were analyzed on the basis of their acoustic characteristics using specific software (Praat Software, version 5.4.21; Amsterdam, The Netherlands). **Formant frequencies of vowels** **Upper boundary frequency (UBF) of voiceless fricatives** Speech was evaluated immediately after wearing both upper and lower retainer, 24 hours later, 1 week later, 1 month later and 3 months later.	**A priori sample calculation:** G*Power software Version 3.1 (Heinrich-Heine-Universitat Dusseldorf, Dusseldorf, Germany), with an alpha value of 0.05 and a power of 80%; revealed the need for 10 subjects per group. **Information on compliance:** NR **Reliability of measurements:** NR
Zhang and Wang, 2003 **China**	**Group 1**: **Hawley retainers** No other details available. **Group 2**:**Positioners** No other details available. No details available on wearing instructions.	**Plaque Index, Gingival Index, Probing Depth at 16, 13, 21, 36, 33, 41** [mesiobuccal angle, labial side (or buccal side),distobuccal angle and right in the middle of palatal (lingual) side] **Examination was done 1 week, 1 month and 3 month after retainer placement**	**A priori sample calculation:** NR **Information on compliance:** NR **Reliability of measurements:** NR

NR: Not Reported

Table 2. Participant characteristics of the studies included in the systematic review.

Study	Eligibility criteria	Number of patients randomized and analyzed
Atik et al., 2016	**Inclusion criteria:** Native Turkish speakers raised in a monolingual environment. At the beginning of the observation period, none of the patients had known cognitive deficits, definite dysmorphology such as cleft lip and/or palate, neurological disorders, phonological problems, articulation problems, or hearing loss. Patients were treated with non-extraction treatment protocol and had Class I and II malocclusion.	**Randomized:**30subjects **Group 1**: **Hawley retainers**: 15 **Group 2**:**Clear thermoplasticretainers**15 **Analyzed:**25subjects **Group 1**: **Hawley retainers**: 12 (mean age 16.3 ±2.56 years) **Group 2**:**Clear thermoplasticretainers**13 (mean age 15.3 ±2.4 years)
Hichens et al., 2007	**See Rowland et al., 2007**	**Randomized:**397subjects **Group 1**: **Hawley retainers**: 196 **Group 2**:**Clear thermoplastic retainers**201 **Analyzed [Questionnaires]:**355subjects **Group 1**: **Hawley retainers**: 168 **Group 2**:**Clear thermoplastic retainers**182 **Analyzed [Cost interview]:**60subjects **Group 1**: **Hawley retainers**: 41 **Group 2**:**Clear thermoplastic retainers**19
Liao, 2010	**Inclusion criteria:** Patients who just finished fixed orthodontic treatment.	**Randomized and analyzed:**50subjects **Group 1**: **Hawley retainers**: 25 **Group 2**:**Clear thermoplastic retainers**25
Rohaya et al., 2006	**Inclusion criteria:**1) Consecutively treated patients who had either upper fixedor upper and lower fixed appliances; with and without aprior phase of functional appliance therapy and with or without extraction of teeth. 2) Patients prepared to wear either a Hawley or a Vacuum-formed retainer. **Exclusion criteria:** 1) Patients who had undergone surgical repositioning of the jaws as part of their treatment.2) Hypodontia treated by reopening spaces for prosthetic replacement of teeth. 3) Patients with marked alveolar bone loss related to treated periodontal disease (now stabilized), who usually need permanent bonded retainers.4) Cases that required substantial expansion of the upper arch, which necessitated the use of either rapid maxillary expansion or a Quad-helix. 5) Cleft lip and palate patients or patients who presented with any craniofacial syndrome.	**Randomized and analyzed:**139subjects **Group 1**: **Hawley retainers**: 68 **Group 2**:**Clear thermoplastic retainers**71

NR: Not Reported

Table 2. [Continued]

Study	Eligibility criteria	Number of patients randomized and analyzed
Rowland et al., 2007	**Inclusion criteria:** Fixed appliance treatment involving both arches; pre-adjusted edgewise appliances; Pretreatment records, treatment plan, and study models available; Willing to wear maxillary and mandibular retainers. **Exclusion criteria:** Single-arch or sectional fixed appliance treatment, hypodontia requiring tooth replacement on the retainer a sa temporary measure, rapid maxillary expansion, bonded retainers, poor periodontal status, early debonding, transfer patients, learning difficulties, or cleft lip or palate.	**Randomized:**397subjects **Group 1**: **Hawley retainers**: 196 **Group 2**:**Clear thermoplastic retainers**201 **Analyzed:**310subjects **Group 1**: **Hawley retainers**: 155 **Group 2**:**Clear thermoplastic retainers**155
Shawesh et al., 2010	**Inclusion criteria:** 10–16 years of age, labial segment crowding or tooth contact point displacement at the start of orthodontic treatment, clinically acceptable labial segment alignment at the end of active treatment, and good oral hygiene. **Exclusion criteria:** lack of consent, severe rotations or midline diastema suggesting the need for a bonded retainer, and patients with a restorative need in the labial segment, e.g. implant, bridges, or missing teeth.	**Randomized:**67subjects **Group 1**: **Hawley retainers full-time**: 34 **Group 2**:**Hawley retainers part-time**:33 **Analyzed:**52subjects **Group 1**: **Hawley retainers full-time**: 28 **Group 2**:**Hawley retainers part-time**: 24
Sun et al., 2011	**Inclusion criteria:** (1) Age the 18 yrs. (2) All second molars erupted and in occlusal contact. (3) Agreement to the research procedures and signing of an informed consent, accompanied by a parent or legal guardian. **Exclusion criteria:** allergic reaction to acrylic resin, rejection of either of the 2 types of retainers, individuals unable to comply with follow-up appointments during 12 months.	**Randomized:**120subjects **Group 1**: **Hawley retainers**: 61 **Group 2**:**Clear thermoplastic retainers**59 **Analyzed:**111subjects **Group 1**: **Hawley retainers**: 56 **Group 2**:**Clear thermoplastic retainers**55
Tsai, 2010	**Inclusion criteria:** Completed full orthodontic treatment to Class I molar and canine relationships. **Exclusion criteria:** History of temporomandibular disorder, large restorations on posterior teeth, allergies to any materials used this study and periodontal disease and /or muscular dysfunction. Subjects were also withdrawn from the study if they were non-compliant with regards to retainer wear, lost their retainers, or did not have longitudinal records.	**Randomized:**60subjects **Group 1**: **Hawley retainers**: 30 **Group 2**:**Clear thermoplastic retainers**30 **Analyzed:**40subjects **Group 1**: **Hawley retainers**: 20 **Group 2**:**Clear thermoplastic retainers**20
Wan et al., 2016	**Inclusion criteria:** Native Chinese who completed active orthodontic treatment. **Exclusion criteria:**(1) cleft lip or cleft palate, (2) surgical correction of the jaws, (3) dialects,(4) hearing and speech disorders, (5) temporomandibularjoint dysfunction syndrome, (6) younger than 18years, or (7) suffering from serious periodontitis.	**Randomized and analyzed:**20subjects **Group 1**: **Hawley retainers**: 10(mean age 24.6 ±2.6 years) **Group 2**:**Clear thermoplastic retainers**10(mean age 24.1 ±3.1 years)

NR: Not Reported

Table 2. [Continued]

Study	Eligibility criteria	Number of patients randomized and analyzed
Zhang and Wang, 2003	**Inclusion criteria:** Finished fixed orthodontic treatment, brushing carefully 3-5 min, 3 times per day. **Exclusion criteria:** Use of antibiotics. Before the start of the study all patients received periodontal treatment and their gingival conditions were good with normal color and normal probing depths.	**Randomized and analyzed:**62subjects(mean age 17.7; range: 9 - 26 years) **Group 1**: **Hawley retainers**: 31(**Group 2**:**Clear thermoplastic retainers**31

NR: Not Reported

5.3. Results of risk of bias assessment

Table 3 presents the summary findings of the risk of bias assessment for the included studies; more details can be found in Appendix IV. Three studies were considered as being of low risk of bias (Rowland et al., 2007; Shawesh et al., 2010; Wan et al., 2016), four of unclear (Hichens et al., 2007; Liao, 2010; Atik et al., 2016; Zhang and Wang, 2003) and three (Rohaya et al., 2006; Tsai, 2010; Sun et al., 2011) of high risk of bias.

In general, most studies included in the present review were considered to present a low risk of bias regarding the domains of random sequence generation and allocation concealment, although some were categorized as unclear regarding this domain because of insufficient information to form a definite judgment. Blinding of the participants, caregivers and the personnel providing the instructions was not feasible or easily achievable in most cases. However, in the context of the present research design, there was no reason to suggest that bias could be introduced because of absence of blinding in these cases. Moreover, the review authors did not believe that bias could be introduced by the methods described in most publications included in the present review regarding blinding of outcome assessment, although many studies were considered to be at unclear risk of bias. In addition, most studies exhibited low risk of bias due to the observed dropouts, whereas, more than half of them where considered to be at unclear or high risk of reporting bias. Finally, the risk of the included studies of presenting other potential threats to validity was considered unclear in most cases because it was impossible to determine how the subjects' compliance to the recommended retention protocol could have influenced the results, as the relative information was sparse and inadequately reported.

Table 3. Summary of risk of bias assessment [Domains examined: 1: Random sequence generation; 2: Allocation concealment; 3: Blinding of participants and personnel; 4: Blinding of outcome assessment; 5: Incomplete outcome data; 6: Selective outcome reporting; and 7: Other potential threats to validity]

	Domain							
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Summary assessment**
Atik et al., 2016	Unclear	Low	Low	Low	Low	Low	Unclear	**Unclear**
Hichens et al., 2007	Low	Low	Low	Low	Low	Unclear	Unclear	**Unclear**
Liao, 2010	Unclear	Unclear	Low	Unclear	Low	High	Unclear	**Unclear**
Rohaya et al., 2006	Low	Low	Low	Low	High	High	Unclear	**High**
Rowland et al., 2007	Low	Low	Low	Low	Low	High	Unclear	**Low**
Shawesh et al., 2010	Low	Low	Low	Low	Low	Low	Unclear	**Low**
Sun et al., 2011	Low	Low	Low	High	Low	High	Unclear	**High**
Tsai, 2010	Unclear	Unclear	Low	Unclear	High	Unclear	Unclear	**High**
Wan et al., 2016	Low	Low	Low	Low	Low	Unclear	Low	**Low**
Zhang and Wang, 2003	Unclear	Unclear	Low	Unclear	Unclear	Low	Unclear	**Unclear**

5.4. Comparison of Hawley retainers to clear thermoplastic retainers

The results of the studies included in the present review are presented below. Because of the lack of extensive relevant data, differences in the methodology used and the statistical elaboration of the obtained results, quantitative data synthesis was not possible. Moreover, we were not able to conduct analyses for "small-study effects" and publication bias (Higgins and Green, 2011).

5.4.1. Maxillary and mandibular dental arch measurements

Three studies investigated the comparative effects of Hawley and clear thermoplastic retainers on different maxillary and mandibular dental arch measurements at different follow-ups (Rohaya, 2006; Rowland et al., 2007; Liao, 2010) (Table 4).

Table 4. Comparative effects of Hawley and clear thermoplastic retainers on different maxillary and mandibular dental arch measurements at different follow-ups.

	Observation - 6 months		Observation - 9 months
	Rowland et al., 2007	**Liao, 2010**	**Rohaya et al., 2006**
Maxillary measurements			
Little's Irregularity Index	**0.013 [HR>CR]**		
Intercanine width	0.12	0.56	
Intermolar width	0.53	0.40	
Arch length [canines]		0.48	
Arch length [molars]		0.37	
Teeth rotations	>0.05		**0.04 [HR>CR]**
Mandibular measurements			
Little's Irregularity Index	**<0.001 [HR>CR]**		
Intercanine width	0.09		
Intermolar width	0.17		
Teeth rotations	>0.05		

HR: Hawley retainers; CR: Clear thermoplastic retainers

Rowland and co-workers (2007) showed significantly greater changes in ***Little's Irregularity Index*** in the Hawley retainer group than in the subjects wearing the clear thermoplastic retainers

at 6 months. At 9 months, a significantly greater proportion of ***maxillary teeth rotations*** had relapsed in subjects wearing Hawley retainers (Rohaya et al., 2006). Otherwise, no statistically significant differences were observed.

The quality of available evidence, using the GRADE approach (Guyatt et al., 2011), for selected variables was considered as low (Table 5).

Table 5. Quality of available evidence for selected maxillary and mandibular dental arch measurements

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**HR**	**CR**	**Absolute**	
Maxillary Little's Irregularity Index [follow up: 6 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	155	155	Median **0.56mm higher** $p<0.001$	⊕⊕◯◯ **LOW**
MandibularLittle's Irregularity Index [follow up: 6 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	155	155	Median **0.25mm higher** $p=0.013$	⊕⊕◯◯ **LOW**

HR: Hawley retainers; CR: Clear thermoplastic retainers

[1] Results were based on specific appliances and retention protocols. [2] The results are based only in one study.

5.4.2. Dental arch relationships and occlusal contacts

The comparative effects of Hawley and clear thermoplastic retainers on dental arch relationships and occlusal contacts were investigated in two papers at different follow-ups (Rowland et al., 2007; Tsai, 2010) (Table 6).

Overall, there was no statistically significant difference observed in ***overjet*** and ***overbite*** measurements between the subjects belonging to the two retainer groups at the 6-month evaluation. Regarding the ***occlusal contacts***, they were quantified in the posterior area by investigating the thickness of bilateral bite registrations taken in maximum intercuspation using a silicon bite registration material, three months after the removal of fixed orthodontic appliances.

The only statistically significant difference observed was an increase in the Hawley retainer subjects of the area of the silicon bite registration with thickness between 100-150 μm.

Table 6. Comparative effects of Hawley and clear thermoplastic retainerson dental arch relationships and occlusal contacts at different follow-ups.

	Observation - 3 months	**Observation - 6 months**
	Tsai, 2010	**Rowland et al., 2007**
Dental arch relationships		
Overjet		0.24
Overbite		0.32
Area of the bite registration		
with ≤ 50 μm thickness	>0.05	
with50-100 μm thickness	>0.05	
with100-150 μm thickness	**<0.05 [HR>CR]**	
with150-200 μm thickness	>0.05	
with200-250 μm thickness	>0.05	
with250-300 μm thickness	>0.05	
with300-350 μm thickness	>0.05	
Area of the bite registration [cumulative]		
with ≤ 50 μm thickness	>0.05	
with ≤ 100 μm thickness	>0.05	
with ≤ 150 μm thickness	>0.05	
with ≤ 200 μm thickness	>0.05	
with ≤ 250 μm thickness	>0.05	
with ≤ 300 μm thickness	>0.05	
with ≤ 350 μm thickness	>0.05	

HR: Hawley retainers; CR: Clear thermoplastic retainers

The quality of available evidence, using the GRADE approach (Guyatt et al., 2011), for selected variables was considered as very low (Table 7).

Table 7. Quality of available evidence for selected dental arch relationship measurements.

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**HR**	**CR**	**Absolute**	
Overjet[follow up: 6 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	Serious[3]	155	155	*p*=0.24	⨁◯◯◯ **VERY LOW**
Overbite[follow up: 6 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	Serious[3]	155	155	*p*=0.32	⨁◯◯◯ **VERY LOW**

HR: Hawley retainers; CR: Clear thermoplastic retainers

[1] Results were based on specific appliances and retention protocols. [2]The results are based only in one study.[3]The results are not adequately reported.

5.4.3. Speech evaluation

Evaluation of different aspects of speech was carried out in two studies following patients up to three months after the removal of fixed orthodontic appliances (Atik et al., 2016; Wan et al., 2016).

Wan et al. (2016) evaluated the ***articulation*** of four long vowels and five voiceless fricatives from the International Phonetic Alphabet, by instructing native Chinese patients to pronounce 36 words (four typical words for each symbol selected from a dictionary as a speech stimulus) at their comfort levels of pitch and loudness in a quiet, soundproof room. The recorded samples were analyzed on the basis of their acoustic characteristics using specific software (Praat Software, version 5.4.21; Amsterdam, The Netherlands). No statistically significant differences were observed between the two groups the number of sound distortions in the investigated sounds in the 3-month evaluation.

Atik et al. (2016) investigated the ***formant frequencies*** of four sustained vowels [a, e, i, u] and of vowel [a] in combination with consonants [b, d, g, t, ş, ç, l, z, c, m, n] in Turkish native speakers raised in a monolingual environment. Moreover, ***Voice onset time (VOT) values*** of

voiceless[t] and voiced [d] stop consonants in combination with vowel [a] were measured using waveforms and spectrograms generated from their speech samples in order to study articulatory–phonatory timing co-ordination. At the end of the 3-month evaluation period the most apparent and statistically significant changes were for [za], [ca], [ma], [na] and [şa] in the Hawley group. The sounds [a] and [ga] were affected in the clear retainer group and the vowel [e] in both groups. VOT values did not differ between the two groups of subjects.

The quality of the available evidence, using the GRADE approach (Guyatt et al., 2011),regarding the articulation of four long vowels and five voiceless fricatives was considered as low (Table 8).

Table 8. Quality of available evidence for articulation.

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**HR**	**CR**		
Articulation[follow up: 3 months; assessed with: number of sound distortions]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	10	10	$p>0.05$	⊕⊕◯◯ **LOW**

HR: Hawley retainers; CR: Clear thermoplastic retainers
[1] Results were based on specific populations and sounds evaluated. [2]The number of patients analyzed was limited.

5.4.4. Patient reported outcomes

Evaluation of different kind of patient reported outcomes was carried out in two studies following patients up to 6 months after the removal of fixed orthodontic appliances (Hichens et al., 2007; Tsai et al., 2016) (Table 9).

Patient ***self-reported compliance with the instructed wear regimen*** was not found to differ at three months after the removal of the appliances between the two retainer groups (Tsai et al., 2016). However, an evaluation at 6 months showed that subjects wearing clear retainers were

self-reporting to be more compliant with the instructions they received from the orthodontist if they belonged to the clear retainer group (Hichens et al., 2007).

Table 9. Comparative effects of Hawley and clear thermoplastic retainers on various patient reported outcomes.

	Observation - 3 months Tsai, 2010	**Observation - 6 months Hichens et al., 2007**
Patient compliance		
How often do you wear your retainer as instructed by your orthodontist? [VAS – Very often/ Never]	>0.05	
Wear retainer as instructed? [Yes/No]		**0.02 [HR<CR]**
Patient satisfaction		
Wear retainers away from home? [Always/ Nearly always/ Sometimes/ Never]		0.103
Embarrassed to wear retainer? [Yes/No]		**0.005[HR>CR]**
Amount of discomfort [Never/ On the occasion/ Most of the time]		0.271
Overall rate compared with fixed appliances [Much better/ Better/ Same/ Worse/ Much worse]		**<0.001 [HR<CR]**
Subjective assessment of the occlusion		
How well do your back teeth fit together when you bite down hard? [VAS – Very well/ Very poor]	>0.05	
Do your back teeth contact each other evenly when you bite down hard? [VAS – Very well/ Very poor]	>0.05	
How well can you chew tough meats, such as steak or chops, with your back teeth? [VAS – Very well/ Very poor]	>0.05	
How much pain do you feel when you bite down hard on your back teeth? [VAS – None/ Very much]	>0.05	

HR: Hawley retainers; CR: Clear thermoplastic retainers

The Hichens and co-workers study (2007) also reported on the frequency of responses to the most relevant questions in a ***patient satisfaction questionnaire*** they administered 6 months after the removal of fixed appliances to subjects participating in a broader trial investigating the performance of Hawley retainers compared to the clear ones. Although participants in the latter group reported less embarrassment than those wearing the Hawley appliances and a better overall assessment regarding their retainers compared with fixed appliances, no statistically significant difference in the frequency of wearing the appliance away from home was observed nor in the amount of discomfort felt.

Tsai (2010) reported data on the ***subjective assessment of occlusion*** following the use of Hawley and clear thermoplastic retainers. No statistically significant difference between the subjects belonging to the two groups was observed three months after the removal of the fixed appliances.

The quality of the available evidence, using the GRADE approach (Guyatt et al., 2011),for patient reported outcomes was considered as very low (Table 10).

Table 10. Quality of available evidence for patient reported outcomes.

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**HR**	**CR**	**Relative& Absolute (95% CI)**	
Self-reported compliance [follow up: 6 months; assessed with: patients not wearing their retainers as instructed]									
1	Serious[1]	Not serious	Serious[2]	Serious[3]	None	25/166 [15.1%]	9/181 [5.0%]	**RR0.33** (0.15 to 0.68) **3 fewer per 100** (from 2 fewer to 4 fewer)	⊕◯◯◯ **VERY LOW**
Embarrassment to wear the retainer[follow up: 6 months; assessed with: patients feeling embarrassed]									
1	Serious[1]	Not serious	Serious[2]	Serious[3]	None	29/167 [17.4%]	13/181 [7.2%]	**RR 0.41** (0.22 to 0.76) **4 fewer per 100** (from2 fewer to 6 fewer)	⊕◯◯◯ **VERY LOW**

HR: Hawley retainers; CR: Clear thermoplastic retainers; CI: Confidence interval; RR: Relative risk

[1]The study was considered at high risk of bias for these outcomes because of the self-reported nature and the fact that only the most relevant questions in the patient satisfaction questionnaire were reported.[2]Results were based on specific populations and treatment protocols. [3]Results are based only on one study.

5.4.5. Adverse effects and problems related to the appliances

Possible adverse effects related to the retainers were not investigated in the studies included in the present review. Two papers eligible for inclusion in the present thesis gave comparison data on retainer failure (Hichens et al., 2007; Sun et al., 2011).

Hichens and co-workers (2007) in their 6-month investigation compared the self-reported overall ***retainer failure*** of Hawley and clear thermoplastic appliances. Although more participants reported that they had broken their retainers in the first group, no difference was observed in the

proportions of patients reporting having lost their appliances. In more long-term study, Sun et al. (2011) also compared the failure of retainers as assessed by professionals, separately for the maxillary and the mandibular dental arches. Only in the mandibular dental arch was the proportion of clear retainers presenting with fractures or exhibiting overall failure greater than the proportion in the Hawley retainer group (Table 11).

Table 11. Comparative retainer failure for the Hawley and clear thermoplastic retainer groups.

	Observation - 6 months Hichens et al., 2007	**Observation - 12 months Sun et al., 2011**
Self-reported retainer failure		
Broken retainer?[Yes/No] [Overall]	**<0.001[HR>CR]**	
Lost retainer? [Yes/No] [Overall]	0.681	
Professional assessment of retainer failure		
Overall failure [Yes/No] [Maxillary]		0.443
Retainer fracture [Yes/No] [Maxillary]		0.688
Retainer loss [Yes/No] [Maxillary]		0.105
Retainer no longer fitting [Yes/No] [Maxillary]		1.000
Retainer with serious abrasion [Yes/No] [Maxillary]		1.000
Overall failure [Yes/No] [Mandibular]		**0.001[HR<CR]**
Retainer fracture [Yes/No] [Mandibular]		**0.010 [HR<CR]**
Retainer loss [Yes/No] [Mandibular]		0.216
Retainer no longer fitting [Yes/No] [Mandibular]		0.174
Retainer with serious abrasion [Yes/No] [Mandibular]		0.495

HR: Hawley retainers; CR: Clear thermoplastic retainers

The quality of the available evidence, using the GRADE approach (Guyatt et al., 2011), for retainer failure was considered to be very low (Table 12).

Table 12. Quality of available evidence for retainer failure.

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**HR**	**CR**	**Relative& Absolute (95% CI)**	
Self-reported retainer failure [follow up: 6 months; assessed with: patients reporting a broken retainer]									
1	Serious[1]	Not serious	Serious[2]	Serious[3]	None	32/165 [19.4%]	12/182 [6.6%]	**RR 0.34** (0.18 to 0.63) **4 fewer per 100** (from 2 fewer to 5 fewer)	⊕◯◯◯ **VERY LOW**
Professional assessment of retainer failure [follow up: 12 months; assessed with: fractured mandibular retainer]									
1	Serious[4]	Not serious	Serious[2]	Serious[3]	None	14/56 [25.0%]	27/55 [49.1%]	**RR 1.96** (1.15 to 3.32) **47more per 100** (from 7 fewer to 100 fewer)	⊕◯◯◯ **VERY LOW**

HR: Hawley retainers; CR: Clear thermoplastic retainers; CI: Confidence interval; RR: Relative risk
[1]The study was considered at high risk of bias for this outcome because of the self-reported nature. [1]Results were based on specific populations and treatment protocols. [2]Results are based only on one study.[4]The study was considered at high risk of bias because blinding of the assessor was not possible.

Sun et al. (2011) investigated the **overall survival** with the Kaplan-Meier method and used the log-rank test to calculate p values for time to breakage of maxillary and mandibular retainers during the 12 month follow-up. No statistically significant differences were observed between the 2 groups in maxillary (p=0.254) and the mandibular dental arch (p=0.188).

Regarding the ***visits to the orthodontist because of problems with the retainers***, a statistically significantly greater proportion of subjects returned from the Hawley group (41 out of 172) than the clear retainer group (21 out of 183) for extra appointments (p=0.001).

5.4.6. Economic evaluation related outcomes

Two papers eligible for inclusion in the present thesis gave comparison data on economic evaluation related outcomes (Hichens et al., 2007; Sun et al., 2011).

Regarding analysis of the ***associated costs and calculations on cost-effectiveness***, the mean bootstrapped cost to the United Kingdom *National Health Service* (NHS) per subject was €152

for the Hawley appliance group (95% Confidence Interval (CI): €150.86 to €153.15) and €122.02 for the clear retainer group (95% CI: €120.84 to €123.21). The difference in bootstrapped mean cost to the NHS per subject between the two retention groups was €31.35 (95% CI: €28.06 – €34.68). From the perspective of the *orthodontic practice,* a bootstrapped mean profit (as indicated by the negative sign) of − €1.00 (95% CI: − €1.78 to − €0.22) and− €34.00 (95% CI: − €34.57 to € 33.34) per patient, was calculated respectively for the Hawley and the clear retainer group. The difference in bootstrapped mean cost from the perspective of the orthodontic practice per participant between the two intervention groups was €32.60 (95% CI: €30.58 to €34.67). Regarding *cost to the individual*, the bootstrapped means were €11.63 (95% CI: €9.67 to €13.59) for patients wearing Hawley retainers and €6.92 (95% CI: €5.29 to €8.53) for subjects wearing clear retainers. The difference in bootstrapped mean cost to the individual per participant between retainer groups was €2.15 (95% CI: € 2.90 to €7.57). Overall, based on the fact that the in Little's Irregularity Index was statistically significantly greater in the Hawley group compared with the group using thermoplastic retainers, over 6months (Rowland et al., 2007), the latter were considered more *cost-effective* from all three perspectives, although for the comparison concerning the perspective of the patient the evidence was deemed to be weak.

The quality of the available evidence for the cost analysis is presented in Table 13.

Table 13. Quality of available evidence for the cost analysis.

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**HR**	**CR**	**Absolute (95% CI)**	
Cost analysisfrom the perspective of the health system [follow up: 6 months; assessed with:**€**]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	172	183	BMD **€31.35** (28.06higher to 34.68 higher)	⊕⊕◯◯ **LOW**
Cost analysis from the perspective of the orthodontic practice[follow up: 6 months; assessed with:**€**]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	172	183	BMD **€32.60** (30.58higher to 34.67 higher)	⊕⊕◯◯ **LOW**
Cost analysisrom the perspective of the individual [follow up: 6 months; assessed with:**€**]									
1	Serious[3]	Not serious	Serious[1]	Serious[2]	None	41	19	BMD **€2.15** (2.90higher to 7.57 higher	⊕⊕⊕◯ **VERY LOW**

CI: Confidence interval; BMD: Bootstrapped Mean difference
[1]Results were based on specific populations and treatment protocols. [2]Results are based only on one study[3]The study was considered of high risk of bias because of limited number of patients analyzed for this outcome.

5.5. Comparison of Hawley retainers to positioners

Only one study (Zhang and Wang, 2003) compared the effects between Hawley retainers and positioners. The results are presented below.

5.5.1. Adverse effects and problems related to the appliances

Zhang and Wang(2003) investigated ***Plaque Index (PI)***, ***Gingival Index (GI)*** and ***Probing Depth (PD)*** in 62 patients randomly allocated to group using Hawley appliances and positioners for retention, up to 3 months. Only the PI was observed to be statistically higher in the positioner group compared to the Hawley group. The quality of the available evidence, using the GRADE approach (Guyatt et al., 2011),was considered to be very low (Table 14).

Table 14. Quality of available evidence for selected maxillary and mandibular dental arch measurements.

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**HR**	**P**	**Absolute (95% CI)**	
Plaque Index[follow up: 3 months]									
1	Serious[1]	Not serious	Serious[2]	Serious[3]	None	31	31	MD**0.13mm higher** (0.06 higher to 0.20higher) *p*<0.001	⊕◯◯◯ **VERY LOW**
Gingival Index[follow up: 3 months]									
1	Serious[1]	Not serious	Serious[2]	Serious[3]	None	31	31	MD**0.16mm higher** (0.002higher to 0.232higher) *p*=0.0535	⊕◯◯◯ **VERY LOW**
Probing Depth[follow up: 3 months; assessed with:mm]									
1	Serious[1]	Not serious	Serious[2]	Serious[3]	None	31	31	MD**0.08mm higher** (0.08higher to 0.24higher) *p*=0.326	⊕◯◯◯ **VERY LOW**

HR: Hawley retainers; P: Positioner; CI: Confidence Interval; MD: Mean Difference
[1] The confidence in the retrieved estimates was diminished because of the many unclear domains during risk of bias assessment.[2]Results were based on specific appliances and retention protocols. [2] The results are based only in one study and limited number of patients.

5.6. Comparison of different wearing schedules of Hawley retention appliances

Shawesh et al. (2010) compared full-time wear (6 months 24 hours per day followed by 6 months night-only) with part-time (1 year night-only) wear of Hawley retainers. The results are presented below.

5.6.1. Maxillary and mandibular dental arch measurements

No statistically significant differences were observed between the two retention regimens regarding maxillary and mandibular Little's Irregularity Index or labial segment crowding. The quality of the available evidence was considered to be low (Table 15).

Table 15. Quality of available evidence for selected maxillary and mandibular dental arch measurements.

Quality assessment						№ of patients		Effect	Quality
Studies	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other**	**H-FT**	**H-PT**	**Absolute (95% CI)**	
Maxillary Little's Irregularity Index [follow up: 12 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	28	24	MD**0.02mm lower** (0.57lower to 0.53higher) *p*=0.94	⊕⊕◯◯ **LOW**
Maxillary Crowding[follow up: 12 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	28	24	MD **0.03mm higher** (0.26 lower to 0.33 higher) *p*=0.82	⊕⊕◯◯ **LOW**
Mandibular Little's Irregularity Index [follow up: 12 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	28	24	MD **0.26mm lower** (0.88 lower to 0.36 higher) *p*=0.39	⊕⊕◯◯ **LOW**
Mandibular Crowding[follow up: 12 months; assessed with: mm]									
1	Not serious	Not serious	Serious[1]	Serious[2]	None	28	24	MD **0.16mm higher** (0.17 lower to 0.50 higher) *p*=0.33	⊕⊕◯◯ **LOW**

H-FT: Hawley retainer full-time (6 months 24 hours per day + 6 months night time only); H-PT: Hawley retainer part-time (12 months night time only); CI: Confidence Interval; MD: Mean Difference

[1]Results were based on specific appliances and retention protocols.[2]The results are based only in one study and limited number of patients.

6. DISCUSSION

During orthodontic treatment, the force that produces tooth movement will also alter the surrounding tissues. If this area is not allowed to remodel during the post-movement period, the teeth will show the tendency to return to their original position (Reitan, 1969; Ericsson and Thilander, 1980) under the influence of various factors such as gingival and periodontal factors, occlusal factors, soft tissue factors, growth factors, and physiologic dento-alveolar adaptation (Blake and Bibby, 1998; Littlewood, 2013). The process of orthodontic retention will reduce the changes caused by growth, reorganize the tissue and will allow neuromuscular adaptation to the corrected tooth position (Moorrees, 1959). It is generally advised that almost all patients and all type of malocclusion treated with orthodontic therapy should be provided with some type of retainers (Renkema et al., 2009; Johnston et al., 2013). Removable retainers, in particular, are often used, despite reservations about poor compliance, especially in adolescents (Ackerman and Thornton, 2011). However, There are, as yet, no definite indications for choosing the optimal period of retention, or the type of retainer for specific cases (Melrose and Millett, 1998; Johnston et al., 2013) and protocols for ideal orthodontic retention practices remain undetermined (Pratt et al., 2011).

6.1. Summary of available evidence

From the initially identified records, ten full-text randomized clinical trials reports comparing the performance of the Hawely-type appliances with the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.), or different wearing schedules and protocols, used after the completion of orthodontic treatment were included in this

systematic review, reflecting the scarcity of relevant research at the top of a widely accepted hierarchy of scientific evidence. Considerable research has been done in trials that were not randomized despite the generally accepted fact that well-designed and properly executed RCTs provide the best evidence on the efficacy of health care interventions (Altman et al., 2001; Oxford Centre for Evidence-based Medicine, 2009). The consequent lack of extensive data of high evidence based potential is rather surprising bearing in mind post-treatment changes may occur in a significant proportion of the orthodontically treated population (Thilander, 2000) and that a high percentage of orthodontists consider them as constituting an important area requiring guidelines (Renkema et al., 2009). Thus, proper retention guidelines should be developed based on a well established, firm scientific base (Melrose and Millett, 1998; Renkema et al., 2009), in order to support and consolidate the care provided.

6.1.1. Comparison of Hawley retainers to clear thermoplastic retainers

Eight of the eligible publications investigated groups of subjects using Hawley removable appliances and clear thermoplastic retainers. Possible adverse effects related to the retainers were not investigated in the studies included in the present review.

According to the results of the studies examining **maxillary and mandibular dental arch measurements** at different follow-ups, in general, no statistically significant differences were observed (Rohaya et al., 2006; Rowland et al., 2007; Liao, 2010). However, subjects wearing Hawley retainers had significantly greater proportion of *maxillary teeth rotation* (Rohaya et al., 2006) and significantly greater changes in *Little's irregularity index* (Rowland et al., 2007).

Regarding **dental arch relationships and occlusal contacts**, no statistically significant differences were observed in *overjet* and *overbite* measurements (Rowland et al., 2007), and almost no differences regarding *occlusal contacts* (Tsai, 2010).

Evaluation of **different aspects of speech** was carried out in two studies (Atik et al., 2016; Wan et al., 2016). No statistically significant differences were observed between the two groups regarding articulation assessed by the number of sound distortions (Wan et al., 2016), whereas some differences were noted regarding the *formant frequencies* and *Voice onset time (VOT) values* in some sounds (Atik et al., 2016).

Furthermore, in the category of **patient reported outcomes**, patient *self-reported compliance with the instructed wear regimen* was observed to be greater in the clear retainer group at the six month evaluation. However, there was no statistically significant difference in the frequency of wearing the appliance away from home, the amount of discomfort felt (Hichens et al., 2007) or the *subjective assessment of occlusion*.

In addition, the proportion of broken retainers was greater for the *self-reported* overall **retainer failure** (Hichens et al. 2007), although, when *professionals* assessed failure it was found more for the clear retainers in the mandibular arch (Sun et al., 2011).

Finally, regarding **economic evaluation,** clear thermoplastic retainers were considered more *cost-effective* from the perspectives of the NHS, the orthodontic practice and the individual patient.

6.1.2. Comparison of Hawley retainers to positioners

Only one study compared the various periodontal parameters (**Plaque Index**, **Gingival Index** and **Probing Depth**) between patients wearing Hawley retainers and positioners (Zhang and Wang, 2003). Only, the Plaque Index was observed to be statistically higher in the positioner group compared to the Hawley group.

6.1.3. Comparison of different wearing schedules of Hawley retention appliances

Only one study involved patients allocated to groups of different wearing schedules of Hawley retention appliances. No statistically significant differences were observed between the two retention regimens regarding maxillary and mandibular **Little's Irregularity Index** or **labial segment crowding** (Shawesh et al., 2010).

6.2. Quality of the available evidence

In general, the quality of evidence (confidence in the observed estimates) assessed with the GRADE approach (Guyatt et al., 2011) was considered at best as low, indicating caution regarding the strength of the relevant recommendations.

Because of the lack of extensive relevant information and differences in the methodology used the data for the evaluated comparisons were based only on one study. In this context, **inconsistency** during the GRADE assessment was not to be considered serious. However, the

numbers of patients analyzed were limited, creating serious problems regarding the **precision** of the results obtained.

In all the cases considered, the overall quality of evidence was downgraded because of problems related to serious **indirectness** of the evidence retrieved. The results obtained were derived from specific populations, appliances and retention protocols, hence even this limited set of data cannot be applied with certainty in clinical settings characterized by a different patient mix or variable type and frequency of retention modalities.

Moreover, an important factor leading to downgrading the overall quality of evidence originated from the **risk of bias assessment** for the outcomes considered in the included studies. Overall, three studies were considered as being of low risk of bias (Rowland et al., 2007; Shawesh et al., 2010; Wan et al., 2016), four of unclear (Hichens et al., 2007; Liao, 2010; Atik et al., 2016; Zhang and Wang, 2003) and three of high risk of bias (Rohaya et al., 2006; Tsai, 2010; Sun et al., 2011). In general, most studies included in the present review were considered to present a low risk of bias regarding the domains of random sequence generation and allocation concealment. Moreover, the review authors did not believe that bias could be introduced by the methods regarding blinding of outcome assessment, although many studies were considered to be at unclear risk of bias. In addition, most studies exhibited low risk of bias due to the observed dropouts, whereas, more than half of them where considered to be at unclear or high risk of reporting bias. Finally, the risk of the included studies of presenting other potential threats to validity was considered unclear in most cases because it was impossible to determine how the subjects' compliance to the recommended retention protocol could have influenced the results, as the relative information was sparse and inadequately reported. However, the latter observation did not result in the downgrading of our general level of confidence in the presented results.

6.3. Strengths and limitations of the present review

The strengths of the present review include using a methodology following well-established guidelines and the fact that it focused exclusively on randomized clinical trials, as it is widely accepted that well-designed and properly executed RCTs provide the best evidence, with decreased risk of bias, on the efficacy of health care interventions (Altman et al., 2001; Oxford Centre for Evidence-based Medicine, 2009). The available empirical evidence suggests that intervention effects in orthodontic research seem to differ in non-RCTs compared to RCTs (Papageorgiou et al., 2015).

Moreover, the search strategy employed in the present review was exhaustive, covering electronic, manual, and gray literature material up to December 2016. In addition the search was comprehensive including every available randomized clinical trial evaluating the performance of Hawley-type appliances in maintaining orthodontic treatment result, irrespective of language, date and status of publication; an approach that led to the retrieval of more relevant information compared to a recent systematic review (Littlewood et al., 2016). Every effort to decrease bias in the methodology employed was made. Screening, verification of eligibility, abstraction of information, assessment of risk of bias and of the quality of evidence were performed in duplicate, and any disagreement was resolved by discussion or consultation with the thesis co-supervisor until a final consensus was achieved.

There are also some limitations to the present review, arising mainly from the nature and the characteristics of the data retrieved during the review process, which resulted in the assessment of the level of available evidence as, at best, low. The scarcity of relevant high quality hierarchically evidence based information from RCTs, precluded meta-analytic procedures for all outcomes. Another limitation of the data retrieved in this study stems from the small number

of patients analyzed resulting in problems regarding the precision of the effect estimates. Also the data on adverse effects and problems related to the employed appliances were very sparse. In addition, the results obtained were derived from specific populations, appliances and retention protocols, thus curtailing the generalizability of the retrieved information to other clinical settings or situations.

Finally, the generalizability of the data found is limited by two additional factors:

Firstly, the lack of long term data. Investigations up to now have been short term in nature and show there is little difference between various retention protocols. It is unclear, however, which retention strategy is most effective in the long term. "What retainer(s) and for how long should I use them?" are questions yet to be answered based on methodologically rigorous randomized clinical trials (Lai et al., 2014). Patients who spent months or years in treatment and the orthodontist who provided it, would be very much interested in having an insight into the effectiveness of various retention approaches over periods exceeding the one year forming the longest follow up of patients in the studies included in the present systematic review.

Secondly, as one of the most important factors in retention is patient co-operation, it would be useful to have more details on patients' compliance. Current data on this topic from studies where compliance was assessed as objectively as possible suggest that patients tend to wear their retainers less than the prescribed daily schedule (Kourakou, 2016; Ackerman and Thornton, 2011; Pauls et al., 2013; Schott et al., 2013; Tsomos et al., 2014; Hyun et al., 2015; Kourakou et al., 2016).

6.4. Recommendations for future research

As the overall quality of the relevant available evidence was considered low, and appropriate guidelines and protocols for ideal orthodontic retention practice are still not determined (Pratt et al., 2011), further research is recommended. Proper retention guidelines should be developed based on a well established, firm scientific base (Melrose and Millett, 1998; Renkema et al., 2009).

Well-designed and properly executed RCTs provide the best evidence with decreased risk of bias on the efficacy of health care interventions (Altman et al., 2001; Oxford Centre for Evidence-based Medicine, 2009). In addition, there is a need to determine which approach is most effective in the long term through adequate follow-up periods. Retention studies are not easy to undertake, as long-term follow-up of participants is difficult in practice and financially demanding. However, given that the vast majority of people requiring orthodontic treatment undergo a phase of retention, this vital area of orthodontic research should be given priority.

Furthermore, as the evidence-based data up to the present have involved specific appliance protocols used in a particular populations, future research should also be directed to alternative protocols, as well as, different populations according to type of malocclusion, age and growth pattern, so as better understanding of the different aspects of retention can be achieved. In addition, every effort should be made to increase the number of patients analyzed in such studies. Increasing the number of patients analyzed will increase the precision of effect estimates (Ellis, 2010).

Finally, future studies including objective assessment of patients' compliance with retainer wearing, and correlating patients' compliance with tooth stabilization could be important. Such studies could provide an answer regarding the ''optimal'' wear time for maintaining the final

orthodontic outcome, so that treatment protocols could be readjusted accordingly and support in a more meaningful way the care provided.

7. CONCLUSIONS

Based on the findings of the present systematic review, conducted following well-established guidelines few differences were observed between the comparative performance of the Hawley-type appliances to the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment.

Regarding the comparative performance of Hawley and clear thermoplastic retainers the following were observed:

1. Greater changes in Little's Irregularity Index were observed with Hawley retainers.
2. Greater proportion of maxillary anterior teeth rotation was observed with Hawley retainers.
3. Patient self-reported compliance with the instructed wear regimen was greater in the clear retainer group.
4. Less embarrassment in wearing the retainer was observed in clear thermoplastic retainer group.
5. Mandibular clear thermoplastic retainers seem to exhibit failure more often.
6. Clear thermoplastic retainers are considered more cost-effective.

Regarding the comparative performance of Hawley retainers and positioners the following were observed:

1. Plaque Index is higher in the positioner group.

Moreover, no differences were found between different appliance wearing schedules and protocols. Given the overall, low level of confidence in the observed estimates and the multitude of parameters which may have affected the results of the included trials, good practice would

suggest further research in the respective field to seriously augment the body of high quality information.

REFERENCES

Ab Rahman A, Low TF, Idris NS. A survey on retention practice among orthodontists in Malaysia. Korean J Orthod 2016;46:36-41.

Ackerman MB, Thornton B. Posttreatment compliance with removable maxillary retention in a teenage population: a short-term randomized clinical trial. Orthod (Chic.) 2011;12:22–7.

Al-Jewair TS, Hamidaddin MA, Alotaib HA, Alqahtani ND, Albarakati SF, Alkofide EA, Al-Moammar KA. Retention practices and factors affecting retainer choice among orthodontists in Saudi Arabia. Saudi Med J 2016;37:895-901.

Al-Kuwari HM, Al Balbeesi HO, Al Thobiani S, Sogaian MB. Caries incidence in lower anterior teeth bonded with fixed orthodontic retainer. Saudi J Dent Res 2015;6:3-8.

Allred K. Stability and relapse of maxillary anterior alignment: a post-retention study of premolar extraction cases treated by traditional edgewise orthodontics [Masters Thesis]. Seattle: University of Washington, School of Dentistry; 1986.

Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663-94.

Artun J. Caries and periodontal reactions associated with long-term use of different types of bonded lingual retainers. Am J Orthod 1984;86:112-18.

Assumpção WK, Ota GKB, Ferreira RI, Cotrim-Ferreira FA. Orthodontic retainers: Analysis of prescriptions sent to laboratories. Dental Press J Orthod 2012;17:36.e1-6.

Atik E, Esen Aydınlı F, Kulak Kayıkçı ME, Ciğer S. Comparing the effects ofEssix and Hawley retainers on the acoustics of speech. Eur J Orthod 2016 Aug 9.pii: cjw050. [Epub ahead of print]

Bearn DR. Bonded orthodontic retainers: a review. Am J Orthod Dentofac Orthop 1995;108:207-13.

Blake M, Bibby K. Retention and stability: a review of the literature. Am J Orthod Dentofac Orthop 1998;114:299-306.

Blake M, Garvey MT. Rationale for retention following orthodontic treatment. J Can Dent Assoc 1998;64:640-1.

Boese LR. Increased stability of orthodontically rotated teeth following gingivectomy in Macaca nemestrina. Am J Orthod 1969;56:273-90.

Booth FA, Edelman JM, Proffit WR. Twenty-year follow-up of patients with permanently bonded mandibular canine-to-canine retainers. Am J Orthod Dentofac Orthop 2008;133:70-6.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to metaanalysis. Chichester: Wiley, 2009.

Butler J, Dowling P. Orthodontic bonded retainers. J Ir Dent Assoc 2005;51:29-32.

Ciger S, Aksu M, Germeç D. Evaluation of posttreatment changes in Class II Division 1 patients after nonextraction orthodontic treatment: cephalometric and model analysis. Am J Orthod Dentofac Orthop 2005;127:219-23.

Clark JD, Kerr WJ, Davis MH. CASES—clinical audit; scenarios for evaluation and study. Br Dent J 1997;183:108-11.

Corbett AI, Leggitt VL, Angelov N, Olson G, Caruso JM. Periodontal health of anterior teeth with two types of fixed retainers. Angle Orthod 2015;85:699-705.

Dahlquist A, Gebauer U, Ingervall B. The effect of a transpalatal arch for the correction of first molar rotation. Eur J Orthod 1996;18:257-67.

DeKock WH. Dental arch depth and width studies longitudinally 12 years of age to adulthood. Am J Orthod 1972;62:56-66.

Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, eds. Systematic Reviews in Health Care, 2nd Edition, BMJ Books, London. 2001,pp 285–312.

Der Simonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.

Dietrich P, Patcas R, Pandis N, Eliades T. Long-term follow-up of maxillary fixed retention: survival rate and periodontal health. Eur J Orthod 2015;37:37-42.

Edman-Tynelius G, Bondemark L, Lilja-Karlander E. A randomized controlled trial of three orthodontic retention methods in Class I four premolar extraction cases—stability after two years in retention. Orthod Craniofac Res 2013;16:105-15.

Edwards JG. A study of the periodontium after orthodontic rotation of teeth. Am J Orthod 1968;54:441-59.

Ellis PD. The Essential Guide to Effect Sizes: An Introduction to Statistical Power, Meta-Analysis and the Interpretation of Research Results. Cambridge: Cambridge University Press, 2010.

Ericsson I, Thilander B. Orthodontic relapse in dentitions with reduced periodontal support: An experimental study in dogs. Eur J Orthod 1980;2:51-57.

Fukui T, Yamane A, Komatsu K, Chiba M. Restoration of mechanical strength and morphological features of the periodontal ligament following orthodontic retention in the rat mandibular first molar. Eur J Orthod 2003;25:167-74.

Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol 2011;64:380–2.

Haines WF, Williams DW. Consent and orthodontic treatment. Br J Orthod 1995;22:101-5.

Hawley CA. A removable retainer. Int J Orthod Oral Surg 1919;5:291-305.

Heier EE, De Smit AA, Wijgaerts IA, Adriaens PA. Periodontal implications of bonded versus removable retainers. Am J Orthod Dentofac Orthop 1997;112:607-16.

Hichens L, Rowland H, Williams A, Hollinghurst S, Ewings P, Clark S, Ireland A, Sandy J. Cost-effectiveness and patient satisfaction: Hawley and vacuum-formed retainers. Eur J Orthod 2007;29:372–8.

Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions, version 5.1.0. The Cochrane Collaboration. 2011.

Horowitz SI, Hixon EH. Physiologic recovery following orthodontic treatment. Am J Orthod 1969;55:1-4.

Horton JK, Bushang PH, Oliver DR, Behrents RG. Comparison of the effects of Hawley and perfector/spring aligner retainers on postorthodontic occlusion. Am J Orthod Dentofac Orthop 2009;135:729-38.

Houston W. Incisor edge-centroid relationships and overbite depth. Eur J Orthod 1989;11:139-43.

Houston WJB, Isaacson KG. Orthodontic treatment with removable appliances. Bristol: John Wright & Sons Ltd, 1980:123.

Houston WJB, Stephens CD, Tulley WJ. Stability and retention. In: Houston WJB, ed. A Textbook of Orthodontics. Oxford: Wright, 1992:346-56.

Hyun P, Preston CB, Al-Jewair TS, Park-Hyun E, Tabbaa S. Patient compliance with Hawley retainers fitted with the SMART® sensor: A prospective clinical pilot study. Angle Orthod 2015;85:263-9.

Ioannidis JP. Interpretation of tests of heterogeneity and bias in meta analysis. J Eval Clin Pract 2008:14: 951e957.

Johnston C, Burden D, Morris D. Clinical guidelines: orthodontic retention (Revised by Parvizi F, Morris D, Atack N). London: British Orthodontic Society, 2013.

Johnston CD, Littlewood SJ. Retention in orthodontics. Br Dent J 2015;218:119-22.

Joondeph DR. Stability, retention and relapse. In: Graber TM, Vanarsdall RL, Vig KWL, eds. Orthodontics. Current principles and techniques. Philadelphia: Elsevier, 2011.

Kaklamanos EG, Kourakou M, Kloukos D, Doulis I, Kavvadia S. Performance of clear vacuum-formed thermoplastic retainers depending on retention protocol: a systematic review. Odontology 2016 Jun 6. [Epub ahead of print]

Kaplan H. The logic of modern retention procedures. Am J Orthod Dentofac Orthop 1988;93:325-40.

Katsaros C, Livas C, Renkema AM. Unexpected complications of bonded mandibular lingual retainers. Am J Orthod Dentofacial Orthop 2007;132:838-41.

Keim RG, Gottlieb EL, Nelson AH, Vogels DS 3rd. JCO study of orthodontic diagnosis and treatment procedures. Part 1. Results and trends. J Clin Orthod 2002;36:553–68.

Keim RG, Gottlieb EL, Nelson AH, Vogels DS 3rd. JCO study of orthodontic diagnosis and treatment procedures. Part 1: Results and trends. J Clin Orthod 2008;42:625-40.

Kingsley NW. A treatise on oral deformities as a branch of mechanical surgery: New York: D. Appleton, 1888.

Konstantonis D, Anthopoulou C, Makou M. Extraction decision and identification of treatment predictors in Class I malocclusions. Prog Orthod 2013;14:47.

Kourakou M, Kaklamanos EG, Kloukos D, Papadopoulos MA, Kavvadia S. Objective assessment of subjects' compliance with thermoplastic vacuum-formed retainers following fixed appliances: a prospective two-centre study. Eur J Orthod 2016;38:e2-e229.

Kourakou M. Objective assessment of patients' compliance with thermoplastic vacuum-formed retainers following fixed appliances, (MSc Thesis), Thessaloniki: Aristotle University of Thessaloniki, 2016.

Lai CS, Grossen JM, Renkema AM, Bronkhorst E, Fudalej PS, Katsaros C. Orthodontic retention procedures in Switzerland. Swiss Dent J 2014;124:655-61.

Levin L, Samorodnitzky-Naveh GR, Machtei EE. The association of orthodontic treatment and fixed retainers with gingival health. J Period 2008;79:2087-92.

Lew K. Direct-bonded lingual retainer. J Clin Orthod 1989;23:490-91.

Liao Y. The research of clinic application of Hawley retainer and transparent wraparound retainer. J Luzhou Med Col 2010; 4: 416-17.

Lima VSA, Carvalho FAR, Almeida RCC, Capelli Júnior J. Different strategies used in the retention phase of orthodontic treatment. Dental Press J Orthod 2012;17:115-21.

Lin F, Sun H, Ni Z, Zheng M, Yao L. A feasible method to improve adherence of Hawley retainer in adolescent orthodontic patients: a randomized controlled trial. Patient Prefer Adherence 2015;9:1525-30.

Lindauer SJ, Shoff RC. Comparison of Essix and Hawley retainers. J Clin Orthod 1998;32:95-7.

Litowitz R. A study of the movement of certain teeth during and following orthodontic treatment. Angle Orthod 1948;18:113-32.

Littelwood SJ. Retention. In: Michell L, ed. An Introduction to Orthodontics. Oxford: Oxford University Press, 2013:193-205.

Little RM, Riedel RA, Artun J. An evaluation of changes in mandibular anterior alignment from 10 to 20 years postretention. Am J Orthod Dentofac Orthop 1988;93:423-28.

Little R M. Stability and relapse of dental arch alignment. Br J Orthod 1990;17**:**235-41.

Little RM. The irregularity index: a quantitative score of mandibular anterior alignment. Am J Orthod 1975;68:554–63.

Littlewood SJ, Millett DT, Doubleday B, Bearn DR, Worthington HV. Orthodontic retention: a systematic review. J Orthod 2006;33:205-12.

Littlewood SJ, Millett DT, Doubleday B, Bearn DR, Worthington HV. Retention procedures for stabilising tooth position after treatment with orthodontic braces. Cochrane Database Syst Rev 2016:CD002283.

Lundström AF. Malocclusion of the teeth regarded as a problem in connection with the apical base. Int J Orthod Oral Surg 1925;9:591–602.

Lyotard, N, Hans M, Nelson S, Valiathan, M. Short-term postorthodontic changes in absence of retention. Angle Orthod 2010;80:1045-50.

Maltha JC, Vandevska-Radunovic V, Kuijpers-Jagtman AM. The biological background of relapse of orthodontic tooth movement. In: Krishnan V, Davidovitch Z, eds. Biological Mechanisms of Tooth Movement. Chichester: John Wiley & Sons, 2015:248-59.

McCauley DR. The cuspid and its function in retention. Am J Orthod and Oral Surg 1944;30:196-205.

McNamara JA, Kramer KL, Juenker JP. Invisible retainers. J Clin Orthod 1985;19:570-8.

Melrose C, Millett DT. Toward a perspective on orthodontic retention? Am J Orthod Dentofac Orthop 1998;113:507-14.

Melsen B, Athanasiou AE. Soft Tissue Influence in the Development of Malocclusion. Aarhus: The Royal Dental college , 1987.

Mills RJE. The long term results of proclination of lower incisors. Br Dent J 1966;120:355-63.

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Open Medicine. 2001;3:123–30.

Moorrees CFA, Gron A, Lebret LML, Yen BKJ, Frohlich FJ. Growth studies of the dentition: a reiew. Am J Orthod 1969;55:600-16.

Moorrees CFA. The Dentition of the Growing Child. Cambridge: Harvard University Press, 1959.

Moyers RE. Handbook of Orthodontics for the Student and General Practitioner. Chicago: Year Book, 1973:442.

Nanda R, Burstone CJ. Retention and Stability in Orthodontics. Philadelphia: WB Saunders, 1993.

Nanda R, Zernick J. Retention and Stability - An Overview. In: Nanda R, Burstone CJ, eds. Retention and Stability in Orthodontics. Philadelphia: WB Saunders, 1993:1-8.

Nanda RS, Nanda SK. Considerations of dentofacial growth in long term retention and stability: is active retention needed? Am J Orthod Dentofac Orthop 1992;101:297-302.

Oppenheim A. The crisis in orthodontia. Part I. 2. Tissue changes during retention. Skogsborg's septotomy. Int J Orthod 1934;20:639-44.

Oxford Centre for Evidence-based Medicine – Levels of Evidence; 2009. Accessed at 31st of July 2016. Available at http://www.cebm.net/index.aspx?o=1025.

Pandis N, Vlahopoulos K, Madianos P, Eliades T. Long-term periodontal status of patients with mandibular lingual fixed retention. Eur J Orthod 2007;29:471-6.

Papageorgiou SN, Xavier GM, Cobourne MT. Basic study design influences the results of orthodontic clinical investigations. J Clin Epidemiol 2015;68:1512-22.

Parker G. Transeptal fibers and relapse following bodily retraction of teeth. Am J Orthod 1972;61:331-44.

Pauls A, Nienkemper M, Panayotidis A, Wilmes B, Drescher D. Effects of wear time recording on the patient's compliance. Angle Orthod 2013;83:1002-8.

Pazera P, Fudalej P, Katsaros C. Severe complication of a bonded mandibular lingual retainer. Am J Orthod Dentofac Orthop 2012;142:406-9.

Pratt MC, Kluemper GT, Hartsfield JK Jr, Fardo D, Nash DA. Evaluation of retention protocols among members of the American Association of Orthodontists in the United States. Am J Orthod Dentofac Orthop 2011;140:520-6.

Raucci G, Pachêco-Pereira C, Grassia V, d'Apuzzo F, Flores-Mir C, Perillo L. Maxillary arch changes with transpalatal arch treatment followed by full fixed appliances. Angle Orthod 2014;85:683-9.

Redlich M, Rahamim E, Gaft A, Shoshan S. The response of supraalveolar gingival collagen to orthodontic rotation movement in dogs. Am J Orthod Dentofac Orthop 1996 ;110:247-55.

Redlich M, Shoshan S, Palmon A. Gingival response to orthodontic force. Am J Orthod Dentofac Orthop 1999;116:152-58.

Reitan K. Clinical and histologic observations on tooth movement during and after orthodontic treatment. Am J Orthod 1967;53:721-45.

Reitan K. Principles of retention and avoidance of posttreatment relapse. Am J Orthod 1969;55:776-90.

Reitan K. Tissue rearrangement during retention of orthodontically rotated teeth. Angle Orthod 1959;29:105-13.

Reitan K. Tissue behavior during orthodontic tooth movement.Am J Orthod 1960;46:881-900.

Renkema AM, Renkema A, Bronkhorst E, Katsaros C. Long-term effectiveness of canine-to-canine bonded flexible spiral wire lingual retainers. Am J Orthod Dentofac Orthop 2011;139:614-21.

Renkema AM, Sips ET, Bronkhorst E, Kuijpers-Jagtman AM. A survey on orthodontic retention procedures in The Netherlands. Eur J Orthod 2009;31:432-7.

Riedel RA. A review of the retention problem. Angle Orthod 1960;30:179-94.

Riedel RA. Retention and relapse. J Clin Orthod 1976;10:454-72.

Riedel RA. Retention. In: Graber TM, Swain BF, eds. Current Orthodontic Concepts and Techniques. Philadelphia: WB Saunders, 1969:875-918.

Roberts W, Chase D. Kinetics of cell proliferation and migration associated with orthodontically induced osteogenesis. J Dent Res 1981;60:174-81.

Rody Jr WJ, Elmaraghy S, McNeight AM, Chamberlain CA, Antal D, Dolce C, Wheeler TT, McGorray SP, Shaddox LM. Effects of different orthodontic retention protocols on the periodontal health of mandibular incisors. Orthod Craniofac Res 2016;19:198-208.

Rody WJ Jr, Akhlaghi H, Akyalcin S, Wiltshire WA, Wijegunasinghe M, Filho GN. Impact of orthodontic retainers on periodontal health status assessed by biomarkers in gingival crevicular fluid. Angle Orthod 2011;81:1083-9.

Rohaya MAW, Shahrul Hisham ZA, Doubleday B. Randomised clinical trial: comparing the efficacy of vacuum-formed and Hawley retainers in retaining corrected tooth rotations. Malays Dent J 2006;27:38–44.

Rowland H, Hichens L, Williams A, Hills D, Killingback N, Ewings P, Clark S, Ireland AJ, Sandy JR. The effectiveness of Hawley and vacuum-formed retainers: a single-center randomized controlled trial. Am J Orthod Dentofac Orthop 2007;132:730-7.

Schott TC, Schlipf C, Glasl B, Schwarzer CL, Weber J, Ludwig B. Quantification of patient compliance with Hawley retainers and removable functional appliances during the retention phase. Am J Orthod Dentofac Orthop 2013;144:533-40.

Shawesh M, Bhatti B, Usmani T, Mandall N. Hawley retainers full or part time? A randomized clinical trial.Eur J Orthod 2010;32:165–70.

Sheridan J, LeDoux W, McMinn R. Essix retainers: fabrication and supervision for permanent retention. J Clin Orthod 1993;27:37-45.

Sheridan JJ, Armbruster P, Moskowitz E, Nguyen P. Avoiding demineralization and bite alteration from full-coverage plastic appliances. J Clin Orthod 2001;35:444-8.

Sinclair PM, Little RM. Dentofacial maturation of untreated normals. Am J Orthod Dentofac Orthop 1985;88:146-56.

Sinclair PM, Little RM. Maturation of untreated normal occlusions. Am J Orthod. 1983;83:114–23.

Singh P, Grammati S, Kirschen R. Orthodontic retention patterns in the United Kingdom. J Orthod 2009;36:115-21.

Sun J, Yu MY, Liu L, Chen L, Li HW, Zhang L, ZhouY, Ao D, Tao R, Lai WL. Survival time comparison between Hawley and clear overlayretainers: a randomized trial. J Dent Res 2011;90:1197–1201.

Thilander B. Orthodontic relapse versus natural development. Am J Orthod Dentofac Orthop 2000;117**:**562-63.

Thilander B. Tissue reactions in orthodontics. In: Graber TM, Vanarsdall RL, Vig KWL, eds. Orthodontics. Current Principles and Techniques. Philadelphia: Elsevier, 2011.

Tsai C. Comparison of the effects of an Essix and Hawley retainer on post-orthodontic occlusion. (MSc Thesis). Saint Louis: Saint Louis University, 2010.

Tsomos G, Ludwig B, Grossen J, Pazera P, Gkantidis N. Objective assessment of patient compliance with removable orthodontic appliances: A cross-sectional cohort study. Angle Orthod 2014;84:56-61.

Tweed CH. Indications for the extraction of teeth in orthodontic procedure. Am J Orthod Oral Surg 1944;30:405-28.

Valiathan M, Hughes E. Results of a survey-based study to identify common retention practices in the United States. Am J Orthod Dentofac Orthop 2010;137:170-7.

Vandevska-Radunovic V, Espeland L, Stenvik A. Retention: type, duration and need for common guidelines. A survey of Norwegian orthodontists. Orthodontics (Chic.) 2013;14:e110-7.

Wan J, Wang T, Pei X, Wan Q, Feng W, Chen J. Speech effects of Hawley andvacuum-formed retainers by acoustic analysis: A single-center randomized controlled trial. Angle Orthod 2016 Aug 24. [Epub ahead of print]

Wasserman I, Barberá A, Conte F, Zajia E. Essix Retainer as an alternative in orthodontics: A systematic review. Rev Salud Bosque 2014;4:37-50.

Wong PM, Freer TJ. Acomprehensive survey of retention procedures in Australia and New Zealand. Aust Orthod J 2004;20:99-106.

Zachrisson BU. Clinical experience with direct-bonded orthodontic retainers. Am J Orthod 1977;71:440-8.

Zachrisson BU. Third-generation mandibular bonded lingual 3-3 retainer. J Clin Orthod 1995;29:39-48.

Zhang B, Wang Q. Periodontal implication of positioner versus removable retainer. Beijing J Stomatol 2003;3:146-7.

APPENDICES

Appendix I. Systematic review protocol used for registration with international prospective register of systematic reviews (PROSPERO).

Review question(s)

The aim of this study is to compare the performance of the Hawley-type appliances to the other removable appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.) used after the completion of orthodontic treatment.

Searches

Comprehensive electronic database searches will be undertaken (up to November 2015) without language restriction in the following databases:

MEDLINE via PubMed, the Cochrane Central Register of Controlled Trials (CENTRAL), the Cochrane Database of Systematic Reviews, Scopus, Web of Science, LILACS, IndMed, Scielo, Arab World Research Source and Deutsche Zentralbibliotek fuer Medizin. Unpublished literature will be accessed electronically using Google Scholar (https://scholar.google.com), ClinicalTrials.gov (http://clinicaltrials.gov), International Standard Randomised Controlled Trial Number (ISRCTN) registry (http://www.isrctn.com) and OpenGrey (http://www.opengrey.eu). In addition, Pro-Quest Dissertation and Theses Global database will be searched. Efforts will be made to obtain conference proceedings and abstracts where possible. Authors will be contacted to identify unpublished or ongoing clinical trials and to clarify methodology and data as necessary. Reference lists of included studies will be screened for additional relevant research.

Types of study to be included

The trials to be included should be RCTs

Condition or domain being studied

Retention after orthodontic treatment

Participants/ population

Patients of any age referred for retention after orthodontic treatment of any type.

Intervention(s), exposure(s)

Hawley-type appliances.

Comparator(s)/ control

Other removable retention appliances (clear thermoplastic appliances, tooth positioners and modifications, etc.).

Outcome(s)

Primary outcomes

Changes in teeth alignment and arch form: Irregularity Index, intercanine and intermolar width, arch length, etc.

Changes in occlusion: overbite, overjet, PAR score, etc.

Secondary outcomes

Patient reported outcomes, compliance, data on retainers' condition (such as thickness and integrity), retainer longevity, hard and soft oral tissue health and possible adverse effects, economic evaluation data.

Data extraction, (selection and coding)

All assessments including titles and/or abstract screening, full text evaluation, and extraction of data will be performed independently and in duplicate by two investigators (WAR and EGK). The investigators will not be blinded to the authors or the results of the research. Disagreements will be resolved by discussion and consultation with a third author where necessary (AEA).

Risk of bias (quality) assessment

Assessment of risk of bias will be performed independently and in duplicate by two investigators (WAR and EGK) using the Cochrane Collaboration risk of bias tool that considers seven domains: random sequence generation; allocation concealment; blinding of participants and personnel; blinding of assessors; incomplete outcome data; selective reporting of outcomes; and other potential sources of bias.

Each domain will receive an assessment of low, high or unclear risk of bias (indicating either lack of sufficient information to make an assessment or uncertainty over the risk of bias). Studies will be finally grouped into the following categories:

• low risk of bias (plausible bias unlikely to seriously alter the results): if all key domains of the study are at low risk of bias,

• unclear risk of bias (bias that raises some doubt about the results): if one or more key domains of the study are unclear, and,

• high risk of bias (bias that seriously weakens confidence in the results): if one or more key domains are at high risk of bias.

Disagreements will be resolved by discussion and consultation with a third author where necessary (AEA).

Strategy for data synthesis

A meta-analysis will be undertaken if studies present adequate clinical and methodological homogeneity. We will pool the results using a random-effects meta-analysis analysis in view of the likely variation in population groups and settings. Depending on the variation of the indices used to quantify primary or secondary outcomes we will use weighted or standardized mean differences for continuous outcomes and risk ratios for binary outcomes, and calculate 95% confidence intervals and two sided *p* values for each outcome. Heterogeneity will be assessed using both the Chi-squared test and the I-squared statistic. If an adequate number of trials are identified, we will carry out analyses for "small-study effects" and publication bias.

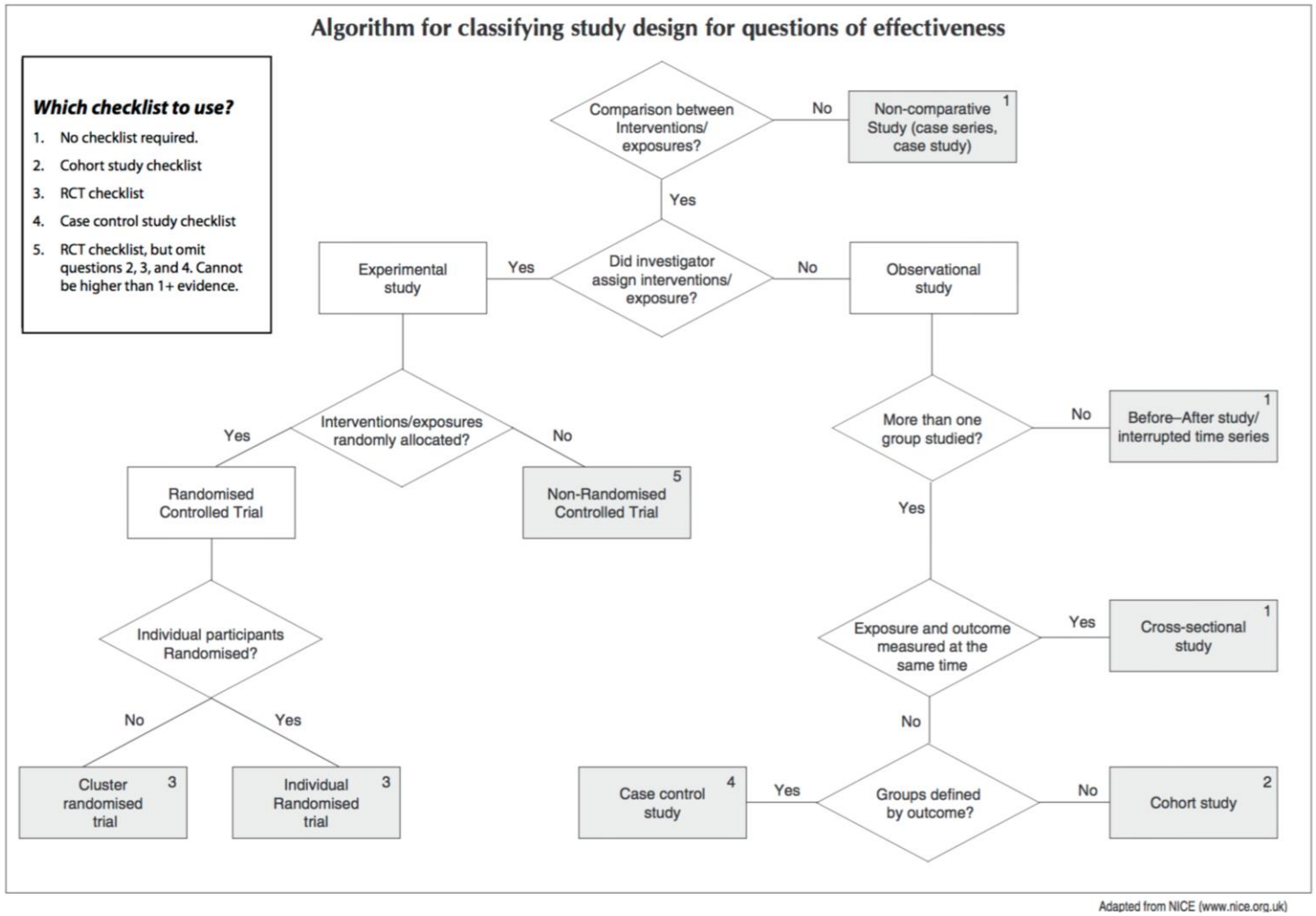
Analysis of subgroups or subsets

If the necessary data are available, subgroup analysis will be performed for each type of the performed orthodontic treatment or appliance type comparison.

Dissemination plans

Peer-reviewed orthodontic journal

Appendix II. Scottish Intercollegiate Guidelines Network (SIGN) algorithm for classifying study design for questions of effectiveness.



Appendix III. Strategy for database search [until December 15, 2016].

Database	Search strategy	Hits
General Sources		
PubMed http://www.ncbi.nlm.nih.gov/pubmed	((((((((((((randomized controlled trial[pt]) OR controlled clinical trial[pt]) OR randomized[tiab]) OR placebo[tiab]) OR drug therapy[sh]) OR randomly[tiab]) OR trial[tiab]) OR groups[tiab])) NOT ((animals[mh] NOT humans[mh]))))) AND orthodon*) AND (((((retain*) OR retent*) OR stability)) OR relaps*)) AND ((Hawley*) OR removable)	**96**
Cochrane Central Register of Controlled Trials http://onlinelibrary.wiley.com/cochranelibrary/search	(orthodon*) AND (retain* OR retent* OR stability OR relaps*) AND (Hawley* OR removable) in Title, Abstract, Keywords in Trials'	**39**
Cochrane Database of Systematic Reviews http://0-ovidsp.tx.ovid.com.libopac.mbru.ac.ae/sp-3.23.1b/ovidweb.cgi	orthodon* {Including Limited Related Terms}	**42**
Scopus https://www.scopus.com/search/form.url?zone=TopNavBar&origin=searchbasic	(TITLE-ABS-KEY (orthodon*)) AND ((TITLE-ABS-KEY (retain*) OR TITLE-ABS-KEY (retent*) OR TITLE-ABS-KEY (stability) OR TITLE-ABS-KEY (relaps*))) AND ((TITLE-ABS-KEY (hawley) OR TITLE-ABS-KEY (removable))) AND (LIMIT-TO (SUBJAREA , "DENT"))	**279**
Web of Science™ Core Collection http://apps.webofknowledge.com/	TOPIC: ((“randomized controlled trial” OR “controlled clinical trial” OR randomized OR placebo OR drug therapy OR randomly OR trial OR groups) AND (orthodon*) AND (retain* OR retent* OR stability OR relaps*) AND (Hawley* OR removable)) Refined by: WEB OF SCIENCE CATEGORIES: (DENTISTRY ORAL SURGERY MEDICINE) DocType=All document types; Language=All languages;	**136**
Regional sources		
LILACS http://lilacs.bvsalud.org/en/	orthodon$ Type of study: Controlled clinical trial	**135**
IndMed http://indmed.nic.in/indmed.html	orthodontic AND (Hawley OR removable)	**5**
Scielo http://www.scielo.org/php/index.php?lang=en	(orthodon*) AND (retain* OR retent* OR stability OR relaps*) AND (Hawley* OR removable)	**12**
Arab World Research Source http://0-web.a.ebscohost.com.amclb.iii.com	orthodon*	**81**
Deutsche Zentralbibliotek fuer Medizin https://www.livivo.de	(orthodon*) AND (retain* OR retent* OR stability OR relaps*) AND (Hawley* OR removable) in Catalogue ZB MED	**83**
Grey literature sources		
Google Scholar https://scholar.google.com	allintitle: orthodontic randomized Excluding patents & citations	**164**
ClinicalTrials.gov http://clinicaltrials.gov/	orthodontic AND (Hawley OR removable)	**15**
ISRCTN registry http://www.isrctn.com	orthodontic AND (Hawley OR removable)	**17**
OpenGrey http://www.opengrey.eu/	orthodon* AND (Hawley OR removable)	**2**
UMI PROQUEST Dissertations Express http://dissexpress.umi.com/dxweb/search.html	orthodontic AND (Hawley OR removable)	**40**

Appendix IV. Details of risk of bias assessment [Domains examined: 1: Random sequence generation; 2: Allocation concealment; 3: Blinding of participants and personnel; 4: Blinding of outcome assessment; 5: Incomplete outcome data; 6: Selective outcome reporting; and 7: Other potential threats to validity]

Study	Rating		Reasons for rating
Atik et al., 2016	**1.**	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk'.
	2.	Low	"Numbered and closed envelopes were prepared before the trial including the treatment allocation card. And a secretary out of the study was responsible for opening the envelope in sequence.
	3.	Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4.	Low	Outcome assessor blinded to the type of retainer used ["The speech sound assessments were performed by a speech–language pathologist (FE) who was blinded to the nature of the study and did not have a thorough knowledge of the potential effects of the retainers."]
	5.	Low	Dropouts are described and explained.
	6.	Low	All measured outcomes were reported.
	7.	Unclear	It is unclear how compliance may have influenced the results of the study.
Hichens et al., 2007	**1.**	Low	The exact method of randomization is mentioned. ["A blocked randomization method (taken fromhttp://www.randomisation.com) was chosen, based on equal numbers of both types of retainers allocated per block of 20 patients."]
	2.	Low	No statement about envelopes, however there is reason to infer that the investigator could influence group allocation. ["Concealment of allocation wasachieved by ensuring that randomization was undertaken after obtaining patient consent."]
	3.	Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4.	Low	Outcome assessor blinded to the type ofretainer used. However, with regards to retainer breakages or other problems the assessor could not be blinded and it is not clear how this fact could have influenced this specific part of the study.
	5.	Low	Dropouts are described and explained. However, the costs to the patient were calculated based on interviews of a small subset of the participants and it is not clear how this fact could have influenced this specific part of the study.
	6.	Unclear	Regarding cost-effectiveness calculations there is no reason to infer that there might be risk of bias. However, the authors reported the "most relevant questions in the patient satisfaction questionnaires", and a significant risk of bias may exist regarding this part.
	7.	Unclear	It is unclear how compliance may have influenced the results of the study. Moreover, although the authors mention that "questionnaires had been previously piloted on a separate sample of patients on two occasions, 1 week apart, and were found to have good readability and reproducibility", the exact questionnaire and its detailed psychometric properties are not available.
Liao, 2010	**1.**	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk'.
	2.	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk'.
	3.	Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4.	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk';
	5.	Low	No dropouts occurred.
	6.	High	Many important and relevant outcomes were not considered.
	7.	Unclear	It is unclear how compliance may have influenced the results of the study.

Appendix IV. [Continued]

Study	Rating	Reasons for rating
Rohaya et al., 2006	**1.** Low	The exact method of randomization is mentioned. ["Stratified randomization system was used to ensure that the patients in the two retainer groups were equivalent according to their original malocclusion and its rotation severity."]
	2. Low	No statement on the envelopes being sequentially administered. No other reason to infer that the investigator could influence group allocation. ["Once stratified, the patients were then randomly allocated one of the retainer types using an opaque envelope system containing either the letter H or V that selected by the third person who was not involved in the study."]
	3. Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4. Low	"The investigator who analyzed the study models was blind to the type of retainer that had been fitted."
	5. High	Very high number of drop-outs(79 subjects out of 218 initially randomized)that was not taken into account in the analysis.
	6. High	Many important and relevant outcomes were not considered.
	7. Unclear	It is unclear how compliance may have influenced the results of the study.
Rowland et al., 2007	**1.** Low	The exact method of randomization is mentioned. ["A blocked randomization method (taken fromhttp://www.randomisation.com) was chosen, based on equal numbers of both types of retainers allocated per block of 20 patients."]
	2. Low	No statement about envelopes, however there is reason to infer that the investigator could influence group allocation. ["Concealment of allocation was achieved by ensuring that randomization was undertaken after obtaining patient consent."]
	3. Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4. Low	Outcome assessor blinded to the type of retainer used
	5. Low	Dropouts are described and explained.
	6. High	Some outcomes (tooth rotations mesial to the first permanent molars, overjet and overbite) were measured but not reported adequately.
	7. Unclear	It is unclear how compliance may have influenced the results of the study.
Shawesh et al., 2010	**1.** Low	The exact method of randomization is mentioned. ["The subjects were randomly allocated to one of the two retention regimen groups using a restricted randomization technique, in blocks of 12, to ensure that equal numbers were allocated to each group. The allocation was decided by throwing an unweighted die where throws of 1, 2, or 3 = group 1 and 4, 5, or 6 = group 2."]
	2. Low	No statement on the envelopes being sequentially administered. No other reason to infer that the investigator could influence group allocation ["the random allocation was sealed in numbered opaque envelopes and held in a central place"].
	3. Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4. Low	"the study model assessor (MS) was blind to the retention regimen used"
	5. Low	Dropouts are described and explained.
	6. Low	All measured outcomes were reported.
	7. Unclear	It is unclear how compliance may have influenced the results of the study.

Appendix IV. [Continued]

Study	Rating		Reasons for rating
Sun et al., 2011	**1.**	Low	The exact method of randomization is mentioned. ["Eligible, consenting individuals were randomly assigned, according to computer-generated random sequence, to receive either the HR or the COR.]
	2.	Low	No statement about envelopes, however there is reason to infer that the investigator could influence group allocation. ["We concealed allocation to avoid selective bias, which means that both patient and researcher were blinded to randomization unless the patient agreed to participate in the study and signed informed consent.]
	3.	Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4.	High	Blinding of the assessor was not possible in this type of study, as the assessorhad to see the retainer used in order to note the type of problem encountered.
	5.	Low	Dropouts are described and explained.
	6.	High	The registered protocol reported that plaque index as the primary outcome.
	7.	Unclear	It is unclear how compliance may have influenced the results of the study.
Tsai, 2010	**1.**	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk'.
	2.	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk'.
	3.	Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4.	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk';
	5.	High	Very high number of dropouts (20 subjects out of 60 initially randomized)that was not taken into account in the analysis.
	6.	Unclear	All measured outcomes reported but not adequately.
	7.	Unclear	It is unclear how compliance may have influenced the results of the study.
Wan et al., 2016	**1.**	Low	The exact method of randomization is mentioned. ["Twenty eligible, consenting individuals were randomized according to a sortation randomization method]
	2.	Low	No statement on the envelopes being sequentially administered. No other reason to infer that the investigator could influence group allocation ["Allocation concealment was achieved with opaque sealed envelopes.]
	3.	Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4.	Low	Outcome assessor blinded to the type of retainer used ["evaluators for data collection or analysis were blinded."]
	5.	Low	No dropouts occurred.
	6.	Unclear	All measured outcomes were reported but not adequately
	7.	Low	It is unclear how compliance may have influenced the results of the study, but there is no reason to believe that it could have influenced the results of the present study.

Appendix IV. [Continued]

Study	Rating		Reasons for rating
Zhang and Wang, 2003	**1.**	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk'.
	2.	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk'.
	3.	Low	Blinding of the participants, caregivers and personnel was not possible. However, the review authors believe that the outcome is not likely to be influenced by lack of blinding.
	4.	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk';
	5.	Unclear	Insufficient information to permit judgment of 'Low risk' or 'High risk';
	6.	Low	All measured outcomes were reported.
	7.	Unclear	It is unclear how compliance may have influenced the results of the study.