European Journal of Orthodontics 30 (2008) 418–424
doi:10.1093/ejo/cjn021
Advance Access publication 16 July 2008

A randomized controlled study of early headgear treatment on occlusal stability—a 13 year follow-up

Viktorija Krušinskienė*, Päivi Kiuttu**, Johanna Julku**, Anna-Sofia Silvola**, Tuomo Kantomaa** and Pertti Pirttiniemi**

*Clinic of Orthodontics, Faculty of Odontology, Kaunas University of Medicine, Lithuania and **Department of Oral Development and Orthodontics, Institute of Dentristry, University of Oulu, Finland

SUMMARY The purpose of this investigation was to assess the long-term occlusal stability in a group treated early with headgear (HG) compared with a control group. The total study group comprised 68 children (40 males and 28 females) aged 7.6 years (standard deviation 0.3), randomly divided into two groups of equal size. In the first group, HG treatment was initiated immediately, while in the control group only minor interceptive procedures were performed during the follow-up period. Fixed appliance treatment, if needed, including extraction of permanent teeth due to crowding, was undertaken after the completion of early treatment. The records were available from the start of the early treatment and at follow-up after 2, 4, 8, and 13 years. The US-weighted Peer Assessment Rating (PAR) Index, graded according to the severity of malocclusion, was used to evaluate occlusal stability. Little's Irregularity Index (LII)and intercanine distance in the lower arch were measured at all time periods. The Aesthetic Component (AC) of the Index of Orthodontic Treatment Need (IOTN) scores was used for evaluation of dental aesthetics at the last follow-up. Parametric tests were applied for statistical analyses, except for the evaluation of aesthetics, where a non-parametric test was used.

No significant differences were found when long-term stability between the HG and control groups was evaluated at the 13 year follow-up. Lower PAR scores were observed in patients treated without extraction of teeth. A greater irregularity in lower incisor alignment before treatment was found in subjects later treated with extractions. The findings of this study seem to suggest that treatment timing has only a minor influence on stability.

Introduction

There are only a few reports describing mixed dentition treatment for Class II malocclusion subjects, based on randomized controlled trials (RCTs; Ghafari *et al.*, 1998; King *et al.*, 2003; O'Brien *et al.*, 2003; Tulloch *et al.*, 2004), all of which have used different orthopaedic appliances and treatment mechanics. In addition, the timing of the treatment and inclusion criteria for RCT has been different.

Early treatment still remains controversial. Much of the discussion concerning the effectiveness of early treatment has been concentrated on timing and suitable methods of intervention in Class II malocclusion subjects. In the majority of the studies, the examined groups received a second phase of fixed appliance treatment, and it seems that much of the gained orthopaedic effects are lost during later growth (King *et al.*, 2003; Tulloch *et al.*, 2004).

One of the main goals of orthodontic treatment is to achieve long-term stability of the occlusion. There is, however, only limited data available concerning long-term stability after early treatment. Most post-retention studies have generally centred on the mandibular incisor segment because of its tendency to exhibit relapse after active treatment and retention or even with ageing (Sinclair and Little, 1983; Little, 1990; Kahl-Nieke *et al.*, 1995; Rossouw *et al.*, 1999; Ormiston *et al.*, 2005). The difference in long-term stability of orthodontically induced changes of arch alignment in patients treated early and those treated late is not well established. Haruki and Little (1998), in their study of early versus late treatment of crowded first premolar extraction cases, found that the late treatment group had greater mandibular anterior irregularity at the post-retention stage. It has been reported that lower anterior crowding continues to increase for up to 20 years after retention, with only 10 per cent of the treated cases having clinically acceptable mandibular anterior alignment (Little *et al.*, 1988).

To obtain objective data on treatment efficacy, it is necessary to use valid and reliable measures of treatment outcome. The Peer Assessment Rating (PAR) Index, which measures the severity of dental malocclusion, was designed to effectively evaluate treatment outcomes (Richmond *et al.*, 1992). The index has been shown to have a good intra- and interexaminer reliability and generally demonstrates a high correlation with the opinions of dentists on the severity of malocclusion. Its reliability and validity have been established in the United Kingdom (UK; Richmond *et al.*, 1992) and, with different weightings and by eliminating the mandibular anterior alignment component, in the United States of America (USA; DeGuzman *et al.*, 1995).

Aesthetic dental appearance has increasingly become more important in orthodontics. Subjects with a normal dental appearance have been judged to be more attractive and aesthetics is generally considered to be the main reason to seek orthodontic treatment (Kiekens *et al.*, 2006). Therefore, the assessment of the long-term outcome of orthodontic treatment should also include evaluation of aesthetics. Nowadays, the Aesthetic Component (AC) of the Index of Orthodontic Treatment Need (IOTN), as described by Evans and Shaw (1987), is widely accepted for the evaluation of dental aesthetics.

The purpose of this study was to assess long-term occlusal stability in adulthood in a group treated early with headgear (HG) and in a control group. A further aim was to assess the influence of different treatment modalities on dental aesthetics.

The hypothesis of the study was that there are no significant differences between the groups at long-term follow-up in occlusal stability and aesthetic outcome.

Subjects and methods

The experimental design was a prospective, longitudinal RCT of early HG treatment. The study group comprised 68 children of both genders (40 males and 28 females) aged 7.6 years [standard deviation (SD) 0.3 years]. The complete details of this trial have been reported previously (Mäntysaari *et al.*, 2004; Pirttiniemi *et al.*, 2005). In brief, the inclusion criteria were a need for orthodontic treatment due to moderate crowding and a Class II tendency. The crowding was clinically diagnosed as moderate, based on the degree of space deficiency in the anterior regions of the dental arches. Twenty per cent of the children had an Angle Class II molar relationship and 80 per cent a bilateral or a unilateral cusp to cusp relationship.

The children were randomly divided into two groups of equal size using random numbers. In the HG group, treatment was initiated immediately. In the second group, which served as the control, only minor interceptive procedures were performed during the follow-up period. The criterion for providing interceptive treatment in the control group was to achieve improved alignment of the anterior teeth during the early mixed dentition. The interceptive procedures in the control group, to permit the eruption of the lateral incisors, were extraction of the upper primary canines in 38 per cent and lower primary canines in 35 per cent. In addition, in 19 per cent of the subjects in the control group, interdental stripping was carried out.

The dental casts included in the present study were taken before (T0) and after follow-up periods of 2 (T1), 4 (T2), 8 (T3), and 13 (T4) years.

Between T1 and T3, there was no difference between the groups in the treatment protocol. Orthodontic treatment, if needed during this phase, comprised fixed appliance therapy, including extraction of permanent teeth due to crowding. In 27 per cent of the HG and in 57 per cent of the control group, fixed appliance therapy, after the first phase of treatment, was undertaken. In 16 per cent of the HG and in 34 per cent of the control group, extraction of the permanent teeth was carried out (Pirttiniemi *et al.*, 2005). Fifty-three patients (83 per cent of the total study group) who continued to the second phase of treatment at T2 completed the follow-up at T3 and full records were available. Thirty-four subjects (53 per cent of the total study group) attended a recall appointment at T4 at the mean age of 20.6 years.

Measurements on study casts

At T2, T3 and T4 dental cast analysis was carried out using the US-weighted PAR Index according to the malocclusion severity (DeGuzman *et al.*, 1995). At T0 and T1, there were difficulties in using the index in the early mixed dentition. The scores were weighted for the separate components, and summed to obtain a total score (PAR Index), expressing the severity of malocclusion.

Little's Irregularity Index (LII; Little, 1975) was measured as the sum of the linear displacements of the anatomical contact points of each mandibular incisor from the adjacent tooth anatomical points for all time periods.

Intercanine distance in the lower arch was measured by marking the tips of the cusps of the canines at a tangent to the occlusal plane with a sharp pencil. The cusp tips were consistently marked at the same point: the intersection of the cusp occlusal contour with the buccal vertical axis of the cusp.

All measurements were made by one author (VK) directly on dental casts using a digital calliper with an accuracy of 0.01 mm.

Dental aesthetics was evaluated in the patients by two calibrated observers (PK and ASS) using the AC of IOTN scores (Evans and Shaw, 1987) at the last follow-up (T4).

Statistical analysis

Normality of the sample was assessed before the analyses and, as there were only minor deviations, the use of parametric tests was preferred when applicable. To analyze the PAR and LII indices between the groups, an independent samples *t*-test was used. A one-sample *t*-test was performed for analysis of the differences between the different time points of PAR scores and LII. For intercanine distance comparisons, and also for comparison between PAR and LII, univariate analysis of variance was used.

A random sample of 20 sets of study casts were evaluated and measured by the same author (VK) twice for all parameters with an interval of 1 week. The intraobserver error of the method for dental cast analysis was measured using intraclass correlation (ICC). Non-parametric tests (kappa) were carried out for evaluation of observer agreement in IOTN scores and for comparison of AC/IOTN scores between the observers.

The Statistical Package for Social Sciences, version 14.0 (SPSS Inc., Chicago, Illinois, USA) was used for statistical analyses.

Results

The ICC scores for intraobserver error of the method in dental cast analysis ranged from 0.988 to 0.998. Kappa for observer agreement in scores of AC/IOTN between the two dental specialists was 0.78. The results of the two examiners were pooled and the mean score of the two observers was used in the analysis.

No differences were found between the HG and control groups at any time points in PAR score but the increase in PAR scores from T3 to T4 was significant for the control group [mean increase 3.75 (SD 4.94; $P = 0.003$); Table 1, Figure 1].

When the HG and control groups were analyzed separately according to whether or not extractions had been carried out, the mean weighted PAR score was lower for the control group without extractions than for the control group with extractions at T3 ($P = 0.026$) and at T4 ($P = 0.015$). There was a significant increase in PAR scores for the controls with extractions ($P = 0.022$) also for the controls without extractions ($P = 0.044$) from T3 to T4. The PAR score for the HG group without extractions was lower than for the HG group with extractions ($P = 0.050$) at T4 (Table 1, Figure 2). When comparing the patients with and without extractions, in general, there was a tendency for the extraction group to exhibit larger PAR scores and the differences were significant at T3 ($P = 0.032$) and T4 ($P = 0.001$). The increase in the PAR scores was significant for patients treated with ($P = 0.004$) and without ($P = 0.016$) extractions from T3 to T4 (Table 1, Figure 3).

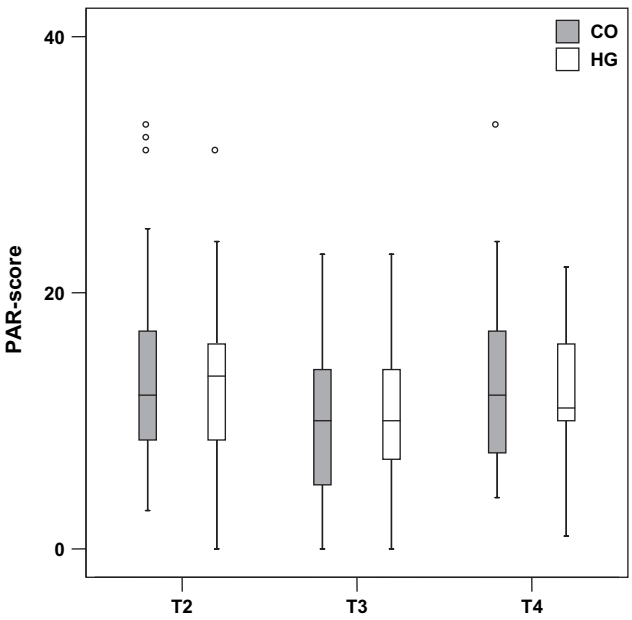

Figure 1 Box plot showing the Peer Assessment Rating (PAR) scores in the headgear (HG) and control (CO) groups.

No differences were found between genders, but the increase in PAR from T3 to T4 was significant for males [mean increase 4.0 (SD 4.67; $P = 0.002$)].

No significant differences were found between the HG and control groups in LII at any time period (Table 2, Figure 4a).

By separating the groups according to the extraction of permanent teeth, a significant difference in incisor irregularity was found at T0. The patients who had permanent extractions in the second phase of treatment had a greater incisor irregularity at T0 (6.84 mm; SD 3.78), when compared with the group without extractions (3.79 mm; SD 2.27; $P = 0.0001$). At T3, there was an opposite trend; incisor irregularity was greater in the group without (3.36 mm; SD 1.64) than in the group with (2.31 mm; SD 1.47; $P = 0.045$) extractions. The tendency was similar at T4 but not significant (Table 2, Figure 4b).

After separating the groups according to gender, a significant increase in LII in males from T3 to T4 was observed [mean difference 0.64 (SD 1.11; $P = 0.03$)].

A larger intercanine distance was found in the HG group (27.93 mm; SD 1.97), when compared with the control group at T1 (25.76 mm; SD 1.71; $P = 0.005$). At T2, the corresponding values for the HG and controls were 26.63 mm (SD 2.42) and 26.30 mm (SD 1.92; $P = 0.038$), respectively. Males had a larger intercanine distance at T0 (26.46 mm; SD 2.03) when compared with females (25.01 mm; SD 2.10; $P = 0.001$).

No significant differences were found in the AC/IOTN between the HG and control groups at T4 ($P = 0.702$). After pooling the HG and control patients, the AC/IOTN scores were compared with the PAR Index at T4. The finding was that patients with higher PAR scores also had significantly poorer aesthetic scores ($P = 0.023$, Figure 5). No associations were found between AC/IOTN scores and LII.

Discussion

The purpose of this study was to assess long-term stability by evaluating occlusion, lower anterior segment, and intercanine distance changes in a group treated early with cervical HG and in a control group. A further aim was to evaluate the aesthetic outcome of the dental arches. The most important consideration was to determine if treatment started in the early mixed dentition provides more stable results compared with single-phase treatment delayed until adolescence.

After the two-phase treatment, the average PAR score was approximately at the same level for patients who underwent early treatment, as for those who did not. These findings are in agreement with the research of King *et al.* (2003) and Tulloch *et al.* (2004). The PAR score in the

Table 1 Weighted Peer Assessment Rating (PAR) Index after follow-up periods of 4 (T2), 8 (T3), and 13 (T4) years and PAR Index change between T3 and T4 in the headgear (HG), control (CO), extraction (EX), no-extraction (NO-EX), headgear no-extraction (HG NO-EX), headgear extraction (HG EX), control no-extraction (CO NO-EX), and control extraction (CO EX) groups.

	T2		T3		T4		T3–T4 difference		
	X	SD	X	SD	X	SD	X	SD	*P**
HG	13.28	6.85	9.96	5.75	11.08	4.80	2.70	4.57	0.095
CO	14.33	7.91	9.7	6.12	13.04	7.29	3.75	4.94	0.003
P	0.597		0.877		0.408		0.579		
NO-EX	13.04	6.98	8.83	5.29	10.38	5.08	3.05	5.33	0.016
EX	16.77	8.51	12.85	6.87	18.11	7.13	4.22	3.19	0.004
P	0.112		0.032		0.001		0.546		
HG NO-EX	12.74	6.88	9.71	5.92	9.90	4.31	2.25	5.01	0.244
HG EX	19.50	0.71	12.50	3.54	17.00	1.41	4.50	2.12	0.205
P	0.186		0.525		0.050		0.565		
CO NO-EX	13.36	7.20	7.84	4.44	10.69	5.63	3.54	5.67	0.044
CO EX	16.27	9.22	12.91	7.44	18.43	8.18	4.14	3.58	0.023
P	0.327		0.026		0.015		0.802		

*P**, the value for the difference in the PAR scores between T3 and T4.

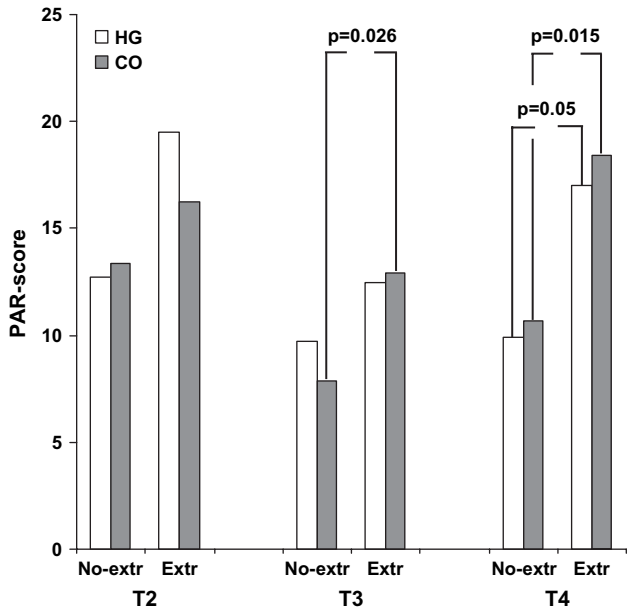

Figure 2 Bar chart showing Peer Assessment Rating (PAR) scores at follow-up of 4 (T2), 8 (T3), and 13 (T4) years in the headgear (HG) and control (CO), extraction (Extr) and no-extraction (No-extr) subgroups.

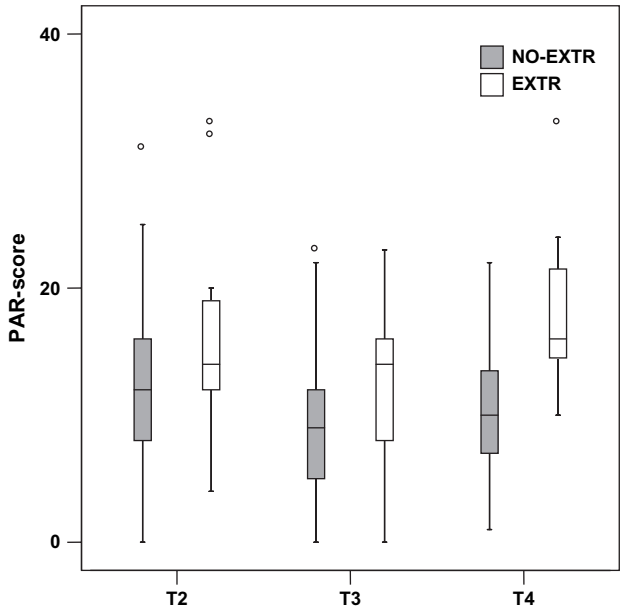

Figure 3 Box plot showing the Peer Assessment Rating (PAR) scores in the extraction (EXTR) and no-extraction (NO-EXTR) subgroups (the headgear and control groups combined) at follow-ups of 4 (T2), 8 (T3), and 13 (T4) years.

present study was reduced after the second phase of treatment for both groups, but increased at the 13 year follow-up. However, worsening of the PAR scores was significant for the control group when the follow-ups at T3 and T4 were compared. The reduction in PAR scores at T3 was most probably due to the fixed appliance treatment in those patients. Fixed appliance treatment effectiveness *per se* was not analyzed, as a second phase of treatment was undertaken only in 27 per cent of the HG and in 57 per cent of the control group. This was due to the fact that in some patients the alignment of the teeth and improved occlusion after the first phase treatment was sufficient, and no further treatment was considered necessary (Pirttiniemi *et al.*, 2005).

A significant difference was found in the HG group between subjects with and without extractions at T4 and also analogous differences in PAR scores in the control group. The PAR score values were higher in patients treated with extractions when compared with those treated without, but an increase in PAR scores occurred in both groups.

Table 2 Little's Irregularity Index (mm) measured in the headgear, control, extraction, and no-extraction groups before (T0) and after follow-up periods of 2 (T1), 4 (T2), 8 (T3), and 13 (T4) years.

	T0		T1		T2		T3		T4	
	X	SD	X	SD	X	SD	X	SD	X	SD
Headgear	3.97	2.16	2.78	1.91	3.22	1.56	3.39	1.52	3.39	2.11
Control	5.06	3.54	2.45	1.87	3.25	1.82	2.87	1.75	3.13	2.57
P	0.143		0.485		0.950		0.263		0.759	
No-extraction	3.79	2.28	2.36	1.37	3.12	1.47	3.36	1.64	3.41	2.40
Extraction	6.84	3.78	3.42	2.93	3.67	2.37	2.31	1.47	2.60	2.38
P	0.0001		0.055		0.311		0.045		0.406	

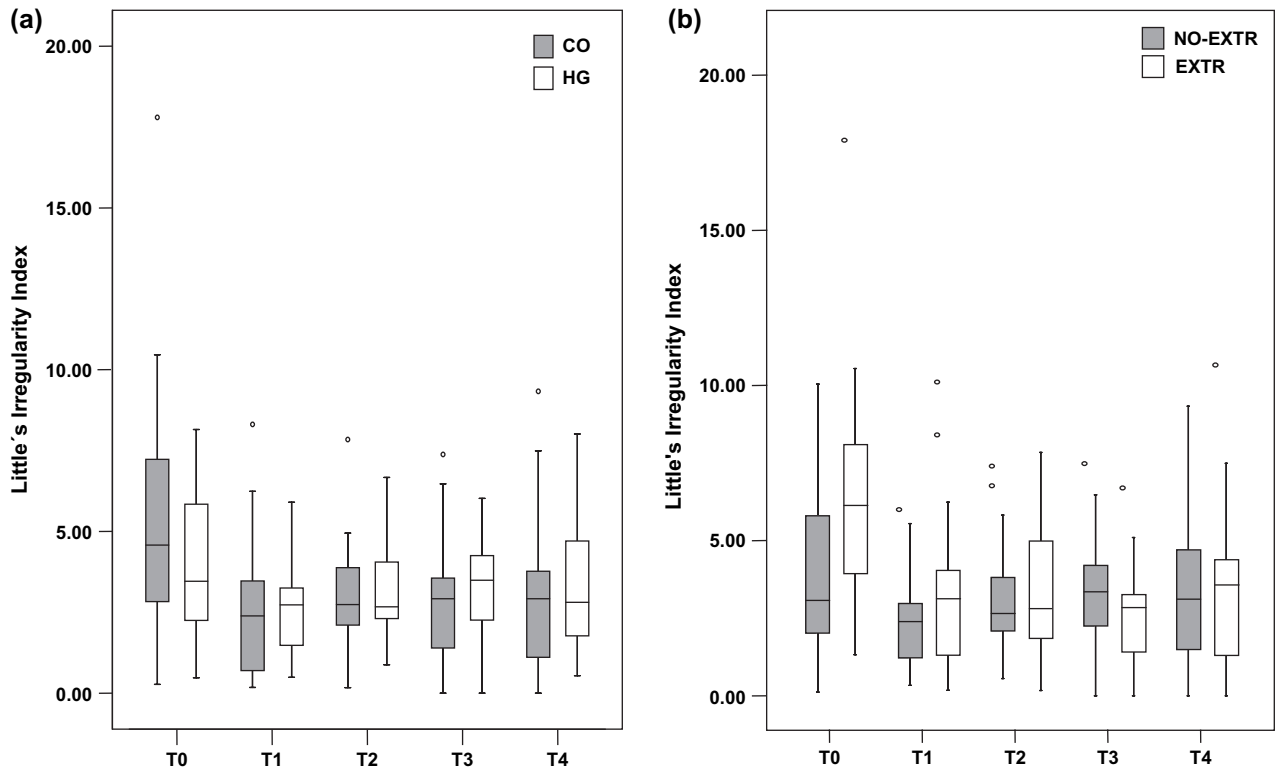

Figure 4 Box plot showing Little's Irregularity Index (a) in the headgear (HG) and control (CO) groups and (b) in the extraction (EXTR) and no-extraction (NO-EXTR) groups (the HG and CO groups combined) before treatment (T0) and after follow-up periods of 2 (T1), 4 (T2), 8 (T3), and 13 (T4) years.

When crowding was analyzed according to the extraction of permanent teeth, a significantly higher LII value was seen at T0. It could be speculated that lower incisor irregularity at an early age might be a diagnostic tool for planning extractions during later orthodontic treatment. After completion of the second phase of treatment, LII was poorer in patients without extractions. The difference in alignment of the lower anterior segment may be due to the fact that all extraction cases had a second phase of treatment. Little *et al.* (1988) found that the most relapse in incisor alignment occurred during the first 10 years of retention and subsequently continued but to a lesser degree. Paquette *et al.* (1992) observed that half of their non-extraction and three-quarters of their extraction patients had a relapse in incisor irregularity. Rossouw *et al.* (1999) and Erdinc *et al.* (2006) found no significant differences in treatment outcomes when groups treated with or without extractions were compared. Kahl-Nieke *et al.* (1995) noted greater post-retention mandibular crowding in subjects treated without extractions compared with those treated with

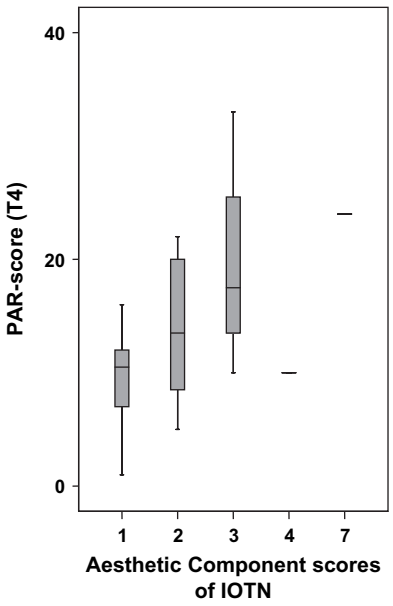

Figure 5 Box plot showing the association between the Peer Assessment Rating (PAR) scores and Aesthetic Component of the Index of Orthodontic Treatment Need (IOTN) at the 13 year follow-up (T4) (headgear and control groups combined).

extractions, after a mean follow-up time of 16 years. The US PAR weighting does not take into account the lower anterior segment, and therefore it was interesting that incisor irregularity was relatively small, while PAR scores were higher, in the patients with extractions. This suggests that other occlusal factors tend to relapse more than lower incisor stability in patients with a history of extractions, when compared with those without extractions.

The only significant difference between the HG and control groups was a larger lower intercanine distance in the HG group at the end of the first phase, with the change becoming insignificant at later time points. The increase in intercanine distance was probably due to early expansion as a result of the HG treatment. Males had a significantly larger intercanine dimension than females at T0 but at the later time points the difference was not significant.

A significant increase in PAR score and also an increase in LII were found at T3-T4 in males in the present study. Ormiston *et al.* (2005), in their investigation on treatment stability, found that in males a sustained period of growth was closely related to increased instability. The influence of dentofacial growth and rotation of the jaws on occlusion was evaluated by Björk and Skieller (1972), who speculated on the possible role of mandibular growth on the changes in lower anterior dental position.

The general conclusion, based on the existing literature (Little, 1990; Sadowsky *et al.*, 1994; Kahl-Nieke *et al.*, 1995; Rossouw *et al.*, 1999; Ormiston *et al.*, 2005), is that the degree of post-retention anterior crowding as well as the relapse in occlusion is unpredictable and variable.

There were no differences between the studied groups in dental aesthetics at T4. Thus, it can be concluded that neither treatment modality is superior. An interesting finding was observed when comparing the AC/IOTN and the US-weighted PAR score, which does not score the lower anterior segment. Poorer PAR scores were associated with a higher AC/IOTN. With LII, however, this type of association was not found. Thus, it can be concluded that in this study the examiners relied more on the occlusion and upper anterior segment alignment than on lower incisor alignment, when dental aesthetics were evaluated.

Although some studies have attempted to find pre-treatment factors associated with post-retention crowding, most have failed to show any reliable predictors of stability. PAR scores and LII in treated patients tend to deteriorate over time. Many factors play important roles in stability, such as treatment and retention methods, compliance, and growth.

Conclusions

No significant differences were found for occlusal outcome between patients who had undergone early HG treatment and a corresponding control group, when evaluated at a 13 year follow-up. Lower PAR scores were observed in the group treated without extraction of permanent teeth but relapse in occlusion occurred in both groups, treated with or without extractions. The greatest irregularity in lower incisor alignment at the beginning of treatment occurred in the group later treated with extractions. Neither treatment method was superior when long-term aesthetics was examined. The results of the present study indicate that the occlusion and upper anterior segment alignment are considered more important for aesthetics than lower incisor alignment.

Address for correspondence

Viktorija Krušinskienė
Clinic of Orthodontics
Faculty of Odontology
Kaunas University of Medicine
J. Lukšos-Daumanto 6
LT 50106 Kaunas
Lithuania
E-mail: viktesv@gmail.com

Funding

Research grant from the European Orthodontic Society.

Acknowledgement

We wish to thank Dr Ahti Niinimaa for assistance with the statistical analysis.

References

Björk A, Skieller V 1972 Facial development and tooth eruption. An implant study at the age of puberty. American Journal of Orthodontics 62: 339–383

DeGuzman L, Bahiraei D, Vig K W L, Vig P S, Weyant R J, O'Brien K 1995 The validation of the Peer Assessment Rating index for malocclusion severity and treatment difficulty. American Journal of Orthodontics and Dentofacial Orthopedics 107: 172–176

Erdinc A E, Nanda R S, Işiksal E 2006 Relapse of anterior crowding in patients treated with extraction and nonextraction of premolars. American Journal of Orthodontics and Dentofacial Orthopedics 129: 775–784

Evans R, Shaw W C 1987 Preliminary evaluation of an illustrated scale for rating dental attractiveness. European Journal of Orthodontics 9: 314–318

Ghafari J, Shofer F S, Jacobsson-Hunt U, Markowitz D L, Laster L L 1998 Headgear versus functional regulator in the early treatment of Class II, division 1 malocclusion: a randomized clinical trial. American Journal of Orthodontics and Dentofacial Orthopedics 113: 51–61

Haruki T, Little R M 1998 Early versus late treatment of crowded first premolar extraction cases: postretention evaluation of stability and relapse. The Angle Orthodontist 68: 61–68

Kahl-Nieke B, Fischbach H, Schwarze C W 1995 Post-retention crowding and incisor irregularity: a long-term follow-up evaluation of stability and relapse. British Journal of Orthodontics 22: 249–257

Kiekens R M A, Maltha J C, van't Hof M A, Kuijpers-Jagtman A M 2006 Objective measures as indicators for facial esthetics in white adolescents. The Angle Orthodontist 76: 551–556

King G J, McGorray S P, Wheeler T T, Dolce C, Taylor M 2003 Comparison of peer assessment ratings (PAR) from 1-phase and 2-phase treatment protocols for Class II malocclusions. American Journal of Orthodontics and Dentofacial Orthopedics 123: 489–496

Little R M 1975 The irregularity index: a quantitative score of mandibular anterior alignment. American Journal of Orthodontics 68: 554–563

Little R M 1990 Stability and relapse of dental arch alignment. British Journal of Orthodontics 17: 235–241

Little R M, Riedel R A, Årtun J 1988 An evaluation of changes in anterior mandibular alignment from 10 to 20 years postretention. American Journal of Orthodontics and Dentofacial Orthopedics 93: 423–428

Mäntysaari R, Kantomaa T, Pirttiniemi P, Pykäläinen A 2004 The effects of early headgear treatment on dental arches and craniofacial morphology: a report of a 2 year randomized study. European Journal of Orthodontics 26: 59–64

O'Brien K 2003 Effectiveness of early orthodontic treatment with the Herbst or Twin-block appliances: a randomized, controlled trial. American Journal of Orthodontics and Dentofacial Orthopedics 124: 128–137

Ormiston J P, Huang G J, Little R M, Decker J D, Seuk G D 2005 Retrospective analysis of long-term stable and unstable orthodontic treatment outcomes. American Journal of Orthodontics and Dentofacial Orthopedics 128: 568–574

Paquette D E, Beattie J R, Johnston Jr L E 1992 A long-term comparison of nonextraction and premolar extraction edgewise therapy in 'borderline' Class II patients. American Journal of Orthodontics and Dentofacial Orthopedics 102: 1–14

Pirttiniemi P 2005 The effects of early headgear treatment on dental arches and craniofacial morphology: an 8 year report of a randomized study. European Journal of Orthodontics 27: 429–436

Richmond S, Roberts C T, Andrews M 1992 The PAR Index (Peer Assessment Rating): methods to determine outcome of orthodontic treatment in terms of improvements and standards. European Journal of Orthodontics 14: 180–187

Rossouw P E, Preston C B, Lombard C 1999 A longitudinal evaluation of extraction versus nonextraction treatment with special reference to the posttreatment irregularity of the lower incisors. Seminars in Orthodontics 5: 160–170

Sadowsky C, Schneider B J, BeGole E A, Tahir E 1994 Long-term stability after orthodontic treatment: nonextraction with prolonged retention. American Journal of Orthodontics and Dentofacial Orthopedics 106: 243–249

Sinclair P M, Little R M 1983 Maturation of untreated normal occlusions. American Journal of Orthodontics 83: 114–123

Tulloch J F C, Proffit W R, Phillips C 2004 Outcomes in a 2-phase randomized clinical trial of early Class II treatment. American Journal of Orthodontics and Dentofacial Orthopedics 125: 657–667