Early class III protraction facemask treatment reduces the need for orthognathic surgery: a multi-centre, two-arm parallel randomized, controlled trial

Nicky Mandall, Richard Cousley, Andrew DiBiase, Fiona Dyer, Simon Littlewood, Rye Mattick, Spencer J. Nute, Barbara Doherty, Nadia Stivaros, Ross McDowall, Inderjit Shargill & Helen V. Worthington

Early class III protraction facemask treatment reduces the need for orthognathic surgery: a multi-centre, two-arm parallel randomized, controlled trial

Nicky Mandall[1]*, Richard Cousley[2], Andrew DiBiase[3], Fiona Dyer[4], Simon Littlewood[5], Rye Mattick[6], Spencer J. Nute[7], Barbara Doherty[1], Nadia Stivaros[8], Ross McDowall[9], Inderjit Shargill[10] and Helen V. Worthington[10]

[1]Tameside Hospital NHS Foundation Trust, Lancashire, UK
[2]Peterborough City Hospital, Peterborough, Cambridgeshire, UK
[3]Kent and Canterbury Hospital, Canterbury, Kent, UK
[4]Sheffield Dental Hospital, Sheffield, UK
[5]St Luke's Hospital, Bradford, UK
[6]Newcastle Dental Hospital, Newcastle upon Tyne, UK
[7]Southend Hospital, Westcliff-on-Sea, Essex, UK
[8]Oldham Orthodontic Practice, Oldham, UK
[9]Queen Alexandra Hospital, Portsmouth, UK
[10]School of Dentistry, University of Manchester, Manchester, UK

Objective: To evaluate whether patients who had received early class III protraction facemask treatment were less likely to need orthognathic surgery compared with untreated controls. This paper is a 6-year follow-up of a previous clinical trial. **Design**: Multi-centre 2-arm parallel randomized controlled trial. **Setting**: Eight United Kingdom hospital orthodontic departments. **Participants**: Seventy three 7- to 9–year-old children. **Method**: Patients were randomly allocated, stratified for gender, into an early class III protraction facemask group (PFG) ($n = 35$) and a control/no treatment group (CG) ($n = 38$). The primary outcome, need for orthognathic surgery was assessed by panel consensus. Secondary outcomes were changed in skeletal pattern, overjet, Peer Assessment Rating (PAR), self-esteem and the oral aesthetic impact of malocclusion. The data were compared between baseline (DC1) and 6-year follow-up (DC4). A per-protocol analysis was carried out with $n = 32$ in the CG and $n = 33$ in the PFG. **Results**: Thirty six percent of the PFG needed orthognathic surgery, compared with 66% of the CG ($P = 0.027$). The odds of needing surgery was 3.5 times more likely when protraction facemask treatment was not used (odds ratio = 3.34 95% CI 1.21–9.24). The PFG exhibited a clockwise rotation and the CG an anti-clockwise rotation in the maxilla (regression coefficient 8.24 (SE 0.75); 95% CI 6.73–9.75; $P < 0.001$) and the mandible (regression coefficient 6.72 (SE 0.73); 95% CI 5.27–8.18; $P < 0.001$). Sixty eight per cent of the PFG maintained a positive overjet at 6-year follow-up. There were no statistically significant differences between the PFG and CG for skeletal/occlusal improvement, self-esteem or oral aesthetic impact. **Conclusions**: Early class III protraction facemask treatment reduces the need for orthognathic surgery. However, this effect cannot be explained by the maintenance of skeletal cephalometric change.

Key words: Class III, interceptive treatment, orthognathic surgery, protraction facemask

*Address for correspondence: Nicky Mandall, Orthodontic Department, Tameside Hospital NHS Foundation Trust, Fountain Street, Ashton-under-Lyne, Lancashire OL6 9RW, UK.
Email: Nicky.Mandall@tgh.nhs.uk

Introduction

Orthopaedic treatment for class III skeletal problems is aimed at reducing or re-directing mandibular growth and/or enhancing maxillary growth, and a number of treatments methods have been described. Functional appliances have been considered but retrospective studies suggest they have no skeletal effect (Robertson, 1983; Loh *et al.*, 1985; Kerr *et al.*, 1988, 1989). If the mandible is protrusive, chin cup treatment is an option but there is no reported orthopaedic effect and reduction in chin prominence is primarily because of a downward and backward mandibular rotation (Graber, 1977; Ritucci *et al.*, 1986; Suguwara *et al.*, 1990). Alternatively, using combination chin cup and maxillary protraction headgear has been shown to achieve two degrees of maxillary protrusion, one degree of mandibular retrusion and an overall jaw relation improvement of three degrees (Ishii *et al.*, 1987; Takada *et al.*, 1993).

Early class III protraction facemask treatment

Previous retrospective studies have suggested that protraction headgear has a short-term orthopaedic effect, with an increase in SNA of up to two degrees. ANB also improved in some studies to around three degrees, often secondary to a downwards and backwards mandibular rotation. On average, there was a 6 mm improvement in overjet (Ngan *et al.*, 1992, 1996, 1997; Chong *et al.*, 1996; Gallagher *et al.*, 1998; Pangrazio-Kulbersh *et al.*, 1998; Kajiyama *et al.*, 2000; Cozza *et al.*, 2004). The weight of evidence, although there is some equivocal data (Merwin *et al.*, 1997; Yuksel, 2001), is towards treating children under 10 years old, since immature circummaxillary sutures may favour treatment success (Baccetti *et al.*, 1998, 2000; Kapust *et al.*, 1998; Franchi *et al.*, 2004; Kajiyama *et al.*, 2004). However, the majority of the early class III protraction facemask evidence is retrospective in nature with the inherent problems of selection, information and confounding bias. Retrospective data collection may, therefore, overestimate the clinical treatment effect and should be cautiously interpreted.

A number of systematic reviews and meta-analyses on the existing retrospective and some prospective studies have been carried out (Kim *et al.*, 1999; Jäger *et al.*, 2001; De Toffol *et al.*, 2008; Foersch *et al.*, 2015; Zhang *et al.*, 2015). They suggest a clinically significant mean forward movement at SNA of 1.4 degrees, at SNB, of minus1.3 degrees and an improvement in ANB of 2.6 degrees. However, studies have mainly reported early outcomes and lack a concurrent no treatment/control group.

This is probably because, if an appliance is in routine clinical use, it is not ethically possible to withdraw such treatment as part of a research study. In 2003, protraction facemask treatment was hardly ever used in the United Kingdom (UK) possibly because of a lack of prospective long-term evidence. This gave a unique opportunity to randomize to a protraction facemask group (PFG) or a no treatment/control group (CG) (Mandall *et al.*, 2010, 2012).

The need for orthognathic surgery following early class III protraction facemask treatment

Previous retrospective work from the United States, Hong Kong and Italy (Hagg *et al.*, 2003; Baccetti *et al.*, 2004; Wells *et al.*, 2006; Masucci, 2011) suggests that between two-thirds to three-quarters of patients treated with a protraction facemask maintain a positive overjet or do *not* require orthognathic surgery. In addition, there was an expected re-establishment of class III growth and statistically significant skeletal or occlusal differences between treated patients and historical/or growth study controls were not always seen.

This randomized clinical trial was started in 2003 and long-term follow-up has enabled an assessment of need for orthognathic surgery. Therefore, the primary aim of this report is to prospectively evaluate whether early class III protraction facemask treatment reduces the need for orthognathic surgery. Secondary aims were to evaluate long-term skeletal and occlusal changes, self-esteem and the aesthetic impact of malocclusion.

The null hypothesis stated that early class III protraction facemask does not reduce the need for orthognathic surgery and that there were no differences between treated patients (PFG) and untreated/controls (CG) in terms of skeletal and occlusal changes, self-esteem and oral aesthetic impact of malocclusion.

Subjects and methods

Trial design

The study was a multi-centre two-arm parallel, randomized controlled trial with a 1 : 1 allocation ratio for two groups. No changes were made to the trial after commencement. The original protocol for this trial was published previously (Mandall *et al.*, 2010).

Participants

Eligibility criteria included patients aged between 7 and 9 years old, a class III skeletal problem when assessed

clinically in the retruded contact position, with a clinically retrusive midface and three to four incisors in crossbite in the intercuspal position. A lateral cephalogram was not considered to be ethically justified to screen patients prior to the study, therefore inclusion was based on the above clinical criteria. Patients were excluded if they were of non-white Caucasian origin, had a cleft lip and palate and/or craniofacial syndrome, a maxillo-mandibular planes angle (greater than 35 degrees, measured cephalometrically after registration) or lower face height greater than 70 mm (measured clinically from soft tissue columella to soft tissue menton) (Farkas, 1994) or previous history of temporomandibular joint (TMJ) dysfunction.

Study setting

Patients were recruited through primary school screening and general dental practitioner referrals to eight UK hospital orthodontic departments. The consultant orthodontist at each hospital was designated the principal investigator and was responsible for patient recruitment and consent. Multi-centre and local ethical and Research and Development (R + D) approval was obtained (MREC reference: 03/8/2).

Patients were registered in the trial between February 2003 and July 2005.

Interventions

The clinical intervention for patients in the PFG consisted of rapid maxillary expansion (RME) (DB orthodontics.co.uk) where a bonded maxillary expansion device with 3 mm thickness of acrylic placed over a metal framework (Baccetti *et al.*, 1998). This was modified, if needed, by extending coverage over the upper incisor edges. Bilateral vestibular hooks were placed in the acrylic adjacent to the upper deciduous first molars. The appliance was cemented with glass ionomer cement, but if it prematurely de-bonded, light cure composite was used, following etching of the buccal and palatal cusps of the upper first permanent molars. For patients with posterior cross bites, the expansion screw was activated once per day until the lingual cusps of the upper posterior teeth approximated the buccal cusps of the lower posterior teeth. If no transverse change was required, the maxillary splint was still activated once a day for 7–10 days to disrupt the circum-maxillary sutures.

Elastics connected downwards at 30 degrees from the vestibular hooks on the RME to the crossbar of a vertically adjustable protraction facemask (TP Orthodontics Europe). Patients were asked to wear the protraction facemask for 14 hours per day during the evening and

night. A co-operation calendar was used in an attempt to increase treatment compliance. Extra oral elastics of increasing force were used until a force of 400 g was delivered per side: 3/8″ 8 oz elastics (for 1–2 weeks) then 1/2″ 14 oz elastics followed by : 5/16″ 14 oz elastics (Baccetti *et al.*, 1998). The elastics were crossed over to prevent interference with the lip commissure. The clinical end point was established as the end of active protraction facemask treatment. This was defined as achievement of either a class I incisal relationship or a positive overjet with no anterior crossbite, and a correction of the class III skeletal pattern to a clinically apparent class I skeletal relationship. Once the active treatment had finished, none of the patients in the PFG received any form of retention. A functional appliance is sometimes used to try to maintain the protraction facemask correction, but because our study was specifically looking at the effect of the protraction facemask, the additional use of a functional appliance would have been a confounder. It was not considered possible to withhold upper arch alignment for patient in the PFG as this would be a treatment normally offered at age 12–13 years. Like the CG, some patients in the PFG received an upper fixed appliance after DC3 data collection. None of the patients receiving upper arch alignment had upper arch expansion. Any interceptive extractions or dentoalveolar surgery required to disimpact teeth were also recorded.

Patients in the CG did not receive protraction facemask treatment. However, it was not considered ethical to withhold upper arch alignment from patients in this trial, at around 12–13 years of age. Therefore, some of this group received upper arch fixed appliance treatment, as clinically indicated, after the DC3 data collection time point. None of the patients receiving upper arch alignment had upper arch expansion. Any interceptive extractions or dentoalveolar surgery required to disimpact teeth were also recorded. The treatment required was decided by the treating consultant who was also the principal investigator at each site.

Outcome measures

The primary outcome measure of interest in this 6-year follow-up report was need for orthognathic surgery. Evaluation of clinical need for orthognathic surgery is multifactorial (Kerr *et al.*, 1992; Stellzig-Eisenhauer, 2002) and, as yet, there is no objective measure of need. Therefore, a panel consensus method was used where the seven orthodontic consultant clinicians met and, as a group, came to a decision for each patient regarding clinical need for orthognathic surgery as 'yes' or 'no'. The records were labelled with a blinding ID number so the

panel did not know if the patients were in the PFG or CG. Records consisted of extra oral and intra oral photographs, clinical overjet measurement and cephalometric variables. The following discriminatory criteria were used: Skeletal class III: SNA, SNB, ANB. If surgery was indicated for another reason such as vertical maxillary excess, high or low maxillary mandibular planes angle or skeletal asymmetry this was recorded; Dentoalveolar: Incisor inclinations (compensation), overbite, anterior crossbite; Soft tissues: Facial profile, perialar recession, upper labial flatness, nasolabial angle, facial convexity, chin prominence, chin/neck distance.

Further primary outcomes were skeletal change and occlusal improvement. For skeletal outcomes, lateral cephalograms were traced by an experienced clinician (NS). Maxillary rotational changes and occlusal plane rotation were calculated by the author (IS). Maxillary superimposition on the anterior zygomatic process was carried out using Bjork's structural method (Björk *et al.*, 1977a, 1977b). For occlusal improvement, PAR scores (Richmond *et al.*, 1992) were measured by a calibrated examiner (R McD) (mean difference −0.07; 95% confidence intervals for the difference −1.60 to +1.47). Overjet measurements were recorded from study models, with a steel millimetric ruler, by an experienced examiner (NM).

Psychosocial outcomes were recorded using questionnaires. The Piers Harris questionnaire was used to evaluate self-esteem and has been previously validated (Piers Harris 2, 2002). Oral Aesthetic Subjective Impact Score (OASIS) (Mandall *et al.*, 2000) measures impact of concern about appearance of teeth, including nice comments, unpleasant comments, teasing, avoidance of smiling, covering the mouth because of the teeth and self-perceived aesthetic component of the Index of Orthodontic Treatment Need (Brook *et al.*, 1989).

Data for all outcome measures were collected at: DC1 baseline at trial registration; DC2 at 15 months; DC3 at 3 years; DC4 at 6 years follow-up. This paper reports the 6-year follow-up data. There was no change to the trial outcomes after the trial commenced.

Sample size calculation

The primary outcome of interest, at 6-years follow-up, was whether or not patients needed orthognathic surgery. Previous literature suggested around one-third of patients treated with a protraction facemask would need surgery. However, reliable population based estimates for untreated class III were not available (Proffit *et al.*, 2003). Therefore, at the time, there was no evidence to help us carry out a sample size calculation based on a

difference between the PFG and CG for the orthognathic surgery outcome. We therefore continued with long-term follow-up of the original trial sample.

Peer Assessment Rating data (PAR) (Richmond *et al.*, 1992) as Ngan *et al.* (2000) had shown protraction facemask treatment resulted in a 30% PAR improvement. We estimated a 25% mean PAR improvement as previous retrospective data may have over-estimated treatment success. The control group value was set at 0% PAR improvement as a clinical estimation.

A sample size of 23 in each group (PFG and CG) has a 90% power to detect a difference in means of 0.25 (difference between a PFG mean PAR reduction of 25% and a CG mean PAR reduction of 0%) assuming a common standard deviation of 0.25 using a two group test with a 0.05 two sided significance level. Thus a total sample size of 46 patients was needed for the trial. In order to allow for long-term attrition, 73 patients were recruited.

Random allocation sequence

Written consent was obtained from the patient and parent. The patient was then randomly allocated to the PFG or CG. The randomization list was generated by author (NM) using random number tables. Randomization was restricted in blocks of 10, with stratification according to gender. Stratification ensured separate randomization lists for girls and boys, since their varied growth timing may have confounded the class III skeletal measurements. The randomization sequence was concealed centrally in sealed opaque envelopes labelled with an unique identification number. The Consultant Orthodontist at the participating site telephoned a research assistant at the co-ordinating centre to enrol a patient. The research assistant then revealed the treatment allocation by opening the next consecutive opaque envelope, after the patient was registered in the trial.

Blinding

This trial was single blind, as the researchers making up the panel for the orthognathic surgery decision, those measuring the radiographs and study models and the statistician were blind to the treatment/control allocation until the data were analysed. The records were labelled with a blinding ID number so the researchers did not know if the patients were in the PFG or CG. Ideally, the clinician collecting the records at the 6-year follow-up would also have been blinded. This was not possible in this trial as the original operator would be likely to remember who had received protraction facemask treatment.

Statistical analysis

Overjet and lateral cephalogram measurements were carried out twice and a mean value calculated to reduce random error. SPSS version 22 was used to undertake the analysis. The data were checked for normality and appropriate descriptive statistics were then generated. The changes occurring between DC1 and DC4 were calculated. Chi square statistic was used to compare the proportion of children in the PFG and CG who received upper arch fixed appliance treatment aged 12 years. It was also used to test for an association between receipt of orthodontic treatment and need for surgery and an association between gender and need for surgery. Odds ratios were then calculated for the effect of PFG or CG on the primary outcome, needing surgery or not. Multiple linear regression models were fitted to the dependent variables (DC4) with DC1 data and group as covariates. All analyses were conducted at the 0.05 level of significance, and there was no allowance made for multiple testing. This is considered in the interpretation of the results. Intra-examiner reliability was assessed by re-measuring 20 cephalograms, overjets and PAR scores, 1 week apart. The panel consensus for the orthognathic surgery outcome was re-measured on 20 patients at the end of the assessment day.

Reliability of cephalometric and study model/occlusal measures were assessed using limits of agreement and intra-class correlation coefficients. Panel reliability for the decision regarding need for orthognathic surgery was assessed using weighted kappa. We collected as much data as possible on patients who dropped out of the study to reduce possible assessment bias. If a subject failed to co-operate during protraction facemask treatment or treatment was stopped, the data were still collected. The analysis was undertaken as a per-protocol analysis and no attempt was made to input missing DC4 data.

Results

Participant flow

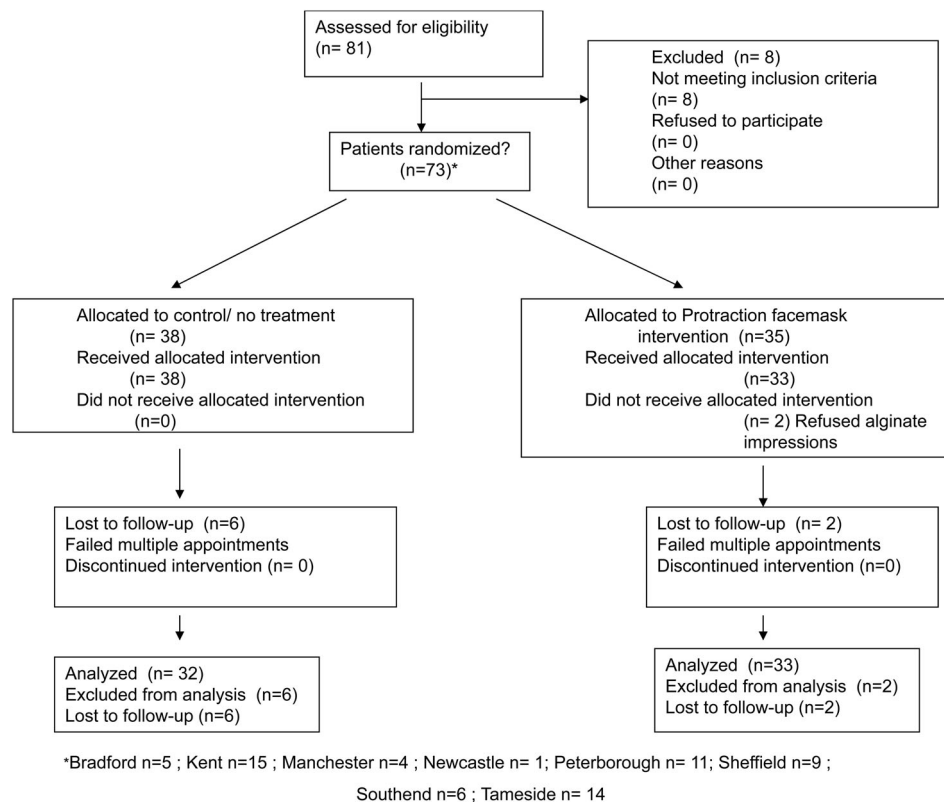The trial profile (Figure 1) shows the number of children randomly allocated to the control or protraction

*Bradford n=5 ; Kent n=15 ; Manchester n=4 ; Newcastle n= 1; Peterborough n= 11; Sheffield n=9 ;
Southend n=6 ; Tameside n= 14

Figure 1   Trial profile

Table 1 Baseline characteristics of patients in the CG and PFG.

	CG Mean (SD)	PFG Mean (SD)
Age (years)	9.0 (0.8)	8.7 (0.9)
Gender *n* (%)		
Male	16 (47.1)	18 (52.9)
Female	22 (56.4)	17 (43.6)
SNA	78.6 (2.5)	78.5 (3.1)
SNB	81.1 (2.8)	80.5 (3.2)
ANB	−2.5 (2.0)	−2.0 (1.9)
Sn/MxP	8.0 (3.0)	8.4 (3.6)
MMangle	26.0 (5.0)	26.2 (4.3)
%LFH	54.3 (2.6)	54.8 (1.8)
UI/MxP	110.5 (9.6)	109.1 (5.5)
LI/MdP	87.6 (7.2)	87.0 (6.5)
Inter-incisal angle	136.4 (11.6)	138.5 (9.4)
Overjet (mm)	−2.3 (1.1)	−2.2 (1.3)
Weighted PAR	32.5 (10.2)	33.6 (8.6)
Forward mandibular displacement on closure %	52.6	52.9
Piers Harris self-esteem total mean score (SD)	49.5 (8.1)	51.4 (5.8)
OASIS mean score (SD)	21.2 (7.5)	20.6 (6.7)

facemask group and the number/reasons for dropouts at 6-year follow-up.

Baseline data

Complete records were available for 65 patients out of 73 representing an 89% attendance at 6-year follow-up. As expected, the randomization process ensured pre-treatment equivalence and there were no apparent baseline differences between the PFG and the CG at baseline for age, gender, cephalometric values and questionnaire data (Table 1). The mean age of patients at DC4 was 15.0 years (SD 10.3 months) in the PFG and 15.3 years (SD 10.1 months) in the CG. In the PFG, there were 17 boys (52%) and 16 girls (48%) and in the CG there were 15 boys (47%) and 17 girls (53%). On average treatment time for the PFG had been 8.6 months (SD 3.5 months). All data were collected for the PFG and CG at 15 months There was high intra-examiner reliability for cephalometric and study model/occlusal measurements (ICC range 0.72–1.00), although the limits of agreement were wide (Table 2). Root mean square (rms) values suggested that random error was within acceptable limits (rms: PAR = 1.35;

Table 2 Reliability [Limits of agreement and Intra-class Correlation Coefficients (ICC)].

	Mean difference	Limits of agreement	ICC
SNA	0.3	1.1, −0.5	0.99
SNB	0.2	1.0, −0.6	0.99
ANB	0.1	1.1, −0.9	0.95
Maxillary plane rotation	0.2	−4.0, 4.4	0.81
Occlusal plane rotation	0.1	−6.9, 7.1	0.72
MM angle	0.0	0.6, −0.6	1.00
Percentage lower face height	0.3	1.9, −1.3	0.95
Upper incisor/ maxillary plane	0.0	1.0, −1.0	1.00
Lower incisor/mandibular plane	−0.1	0.5, −0.7	1.00
Inter-incisal angle	0.2	1.2, −0.8	1.00
Overjet	0.2	1.4, −1.0	0.98
Weighted PAR	0.1	3.9, −3.7	0.98

Table 3 Receipt of orthodontic treatment, or extractions alone, age 12 years.

	Control group (CG)	Protraction facemask group (PFG)
No orthodontic treatment	16	19
Upper fixed appliance non-extraction	5	5
Upper fixed appliance with extractions	7	7
Upper or lower arch extraction only	4	2

overjet = 0.13 mm; cephalometric values range 0.053–0.10 degrees).

Numbers analyzed for each outcome and subgroup analyses

Patient were initially assigned to the CG ($n = 38$); however, six were lost to follow-up so the final number for analysis in this group was $n = 32$. In the PFG, 35 patients were initially assigned but two were lost to follow-up, therefore $n = 33$ were analysed. The $n = 2$ patients in the PFG who had refused treatment because they were unable to have alginate impressions were kept in the analysis. The primary analysis was therefore carried out as a per-protocol analysis.

Reduced need for orthognathic surgery in the PFG

We found most cases were very easy to assess for need for orthognathic surgery with only 7 out of 63 generating significant discussion. Of these seven cases, the panel consensus was that four needed surgery and three did not. Thus, there was a fairly equal distribution of 'surgery' or 'no surgery' decisions in the borderline cases. Importantly, early protraction facemask treatment was successful in reducing the need for orthognathic surgery. In the CG, 21 (66%) were considered to be in need of

orthognathic surgery compared with only 12 (36%) thought to need surgery in the PFG ($P = 0.026$). (Percentage difference between groups = 30%: 95% CI 6% to 53%). This can be expressed as an unadjusted relative effect odds ratio of 3.34 (95% CI 1.21–9.24). This means that the odds of needing surgery is 3.5 times more likely when a protraction face mask is not used. The chi square statistic showed no association between gender and need for orthognathic surgery: 55% of females and 46% of males needed orthognathic surgery ($P = 0.62$).

Comparison of the number of patients treated with fixed appliances between the PFG and CG in Table 3 showed no statistically significant difference between groups (Chi square P value = 0.75). There was no association between: the receipt of orthodontic treatment and the need for surgery ($P = 1.0$); between extraction or non-extraction upper fixed appliance treatment and the need for surgery ($P = 0.32$).

No differences in antero-posterior cephalometric outcomes

Previous statistically significant results for earlier data collection time points are shown in Table 4 (DC2: 15 months and DC3: 3 years follow-up). The data can then be compared with the DC4 data 6 years follow-up presented in this paper.

Table 4 Statistically significant effects of PFG compared with the CG at DC2 and those maintained at DC3 (in bold).

	PFG DC2 (degrees)	PFG DC3 (degrees)
SNA	**1.1**	0.7
SNB	**−1.5**	−0.7
ANB	**2.6**	1.4
Maxilla rotation	**4.4** down and backwards	**4.1** down and backwards
Functional occlusal plane		
Rotation	**4.5** up and forwards	**2.8** up and forwards
MM angle	**1.6**	0.4
L/MdP	**−3.7**	−0.8
Overjet (mm)	**4.1**	**2.5**
% weighted PAR (difference between PFG improvement and CG worsening)	**40.8**	**29.4**

Table 5 Cephalometric changes DC1 to DC4.

	DC4 → DC1	DC4 → DC1		
	Mean change (SD) CG	Mean change (SD) PFG	Regression coefficient (SE) 95% CI	*P* value
SNA	0.9 (2.9)	0.6 (4.3)	0.34 (0.90) −1.45 to 2.14	0.70
SNB	1.6 (2.3)	0.6 (4.3)	1.01 (0.87) −0.73 to 2.74	0.25
ANB	−0.7 (2.6)	0.0 (3.3)	−0.87 (0.72) −2.30 to 0.56	0.23
Sn/MxP	0.1 (2.5)	0.2 (4.4)	−0.30 (0.84) −1.98 to 1.39	0.73
Maxillary rotational change	5.5 (3.5)*	2.7 (2.2)**	8.24 (0.75) 6.73 to 9.75	**<0.001**
Mandibular rotational change	5.4 (3.0)*	1.4 (2.7)**	6.72 (0.73) 5.27 to 8.18	**<0.001**
MMangle	−1.0 (3.6)	−0.7 (4.4)	−0.29 (0.97) −2.23 to 1.65	0.77
%LFH	1.2 (1.8)	−0.1 (2.7)	1.07 (0.50) 0.06 to 2.07	**0.04**
UI/MxP	6.8 (10.2)	7.1 (7.4)	0.79 (1.63) −2.46 to 4.04	0.63
LI/MdP	−2.7 (5.2)	−0.4 (5.3)	−2.14 (1.23) −4.59 to 0.31	0.09
Inter-incisal angle	−3.1 (11.4)	−6.0 (8.0)	2.01 (2.17) −2.33 to 6.34	0.36
Overjet (mm)	1.7 (3.0)	3.0 (3.0)	−1.16 (0.76) −2.69 to 0.37	0.14
Weighted PAR	10.2% improved	27.7% improved	4.36 (3.76) −3.17 to 11.88	0.22
Piers Harris	2.1 (8.1)	1.8 (7.8)	−1.26 (1.78) −4.84 to 2.32	0.48
OASIS	−1.3 (10.3)	−3.5 (10.7)	2.13 (2.64) −3.16 to 7.41	0.42

*Rotation upwards and forwards/anti-clockwise.
**Rotation downwards and backwards/ clockwise.
Bold *P* value denotes statistically significant differences between PG and CG.
CG = Control Group.
PFG = Protraction Facemask Group.

All analyses at DC4 were defined a priori. There were no statistically significant differences for antero-posterior skeletal change between the PFG and CG from baseline to DC4 (6 years follow-up) (Table 5). On average, over time, SNA had moved forwards less than 1° in both the PFG and CG (*P* = 0.70). SNB moved forwards 0.6° in the PFG and 1.6° in the CG (*P* = 0.25). ANB became more class III in the control group by 0.7° but there was no worsening or improvement of the class III skeletal pattern in the PFG (*P* = 0.23). There was a tendency, although not statistically or clinically significant, for the lower incisors to retrocline in the CG (2.7°) compared with the PFG (0.4°).

In the PFG, treatment eliminated a forward mandibular displacement if it was present at baseline (Table 6). This may suggest that these patients had an enhanced improvement in B point because of postural changes. Therefore, within the PFG, we compared the SNB improvement in patients with and without forward mandibular displacement at baseline. Improvement at SNB was no better for patients whose mandibular displacement had been eliminated by protraction facemask treatment compared with treated patients who still had a forward displacement at DC4. (*P* = 0.09).

In contrast to the lack of long-term changes at A and B point, statistically significant changes were observed particularly for maxillary rotation, but also for mandibular rotation. The maxilla rotated clockwise by 2.7 degrees in

the PFG and rotated anti-clockwise by 5.5 degrees in the CG (*P* < 0.001). The mandible rotated 1.4 degrees clockwise in the PFG and 5.4 degrees anti-clockwise in the CG (*P* < 0.001). These rotational changes were not reflected in any clinically significant increase in lower face height or MM angle in the PFG compared with the CG.

Study models and occlusal outcomes

The mean overjet at DC4 in the PFG was +0.8 mm and in the CG was −0.6 mm. The difference between groups was not statistically significant and suggests a marginal treatment effect (*P* = 0.14). Of potential importance is that on average the PFG were maintaining a positive overjet at 15 years old (but only just). When the data were dichotomized (*a priori* analysis), 21 (68%) of the PFG had a positive overjet at DC4 compared with 15 (48%) of the CG (*P* = 0.20). At DC4, PAR improvement was 27.7% in the PFG compared with a 10.2% improvement in the CG. This did not reach statistical significance (*P* = 0.22).

No differences in Piers Harris (self-esteem) and OASIS psychosocial outcomes

The Piers Harris score (Table 5) were compared between PFG and CG from DC1 to DC4. There were tiny changes in self-esteem over time and no statistically

significant increase in self-esteem as a result of protraction facemask treatment. For the OASIS scores, both groups tended to have a reduced impact of their malocclusion over time with the PFG being less concerned about their aesthetic appearance (−3.5 points) and the CG less concerned by −1.3 points ($P = 0.43$) (Table 5).

Study drop-outs

As with any long-term follow-up, some patients dropped out of this study ($n = 8$), resulting in a risk of attrition bias. For example, the drop-outs may be the patients whose treatment was less successful, so biasing the data towards an enhanced treatment effect. Therefore, the baseline characteristics of the patients remaining in the study and the dropouts were compared. The start age, gender, SNA, SNB, ANB and overjet of the dropouts were no different to those patients still in the study, which suggests that there is no attrition bias (P value 0.47–0.92).

Harms

No serious harms were observed. No adverse events or side effects were reported.

Discussion

Reduced need for surgery in the PFG

Our main finding was that patients in the PFG were statistically significantly less likely to need orthognathic surgery compared with the CG. In the PFG, two-thirds of patients did *not* need surgery and 68% maintained a positive overjet at age 15 years. These data are similar to previous literature with reports ranging from 66% to 75% of patients having a positive overjet in the mid-teens and no clinical need for surgery (Hagg *et al.*, 2003; Baccetti *et al.*, 2004; Wells *et al.*, 2006; Masucci *et al.*, 2011).

The data from this trial also showed that one-third of patients in the CG did *not* need surgery. This information is now useful for discussions with class III patients of all ages, although it cannot be used to predict whether individuals will need orthognathic care.

Antero-posterior cephalometric measures of SNA, SNB and ANB did not show long-term differences between the PFG and the CG, and this lack of long-term treatment effect is similar to other studies (Westwood *et al.*, 2003; Masucci *et al.*, 2011). This is probably not surprising, as most clinicians would agree class III skeletal growth will re-establish itself in the long-term.

However, it then becomes more difficult to explain why the need for orthognathic surgery was so much lower in the PFG than the CG.

The explanation may result from the accumulation of multiple skeletal and occlusal protraction treatment effects, on their own insignificant, but together they shift the clinical decision away from surgery. For example on average, in the PFG group compared with the CG there was a marginal advantage of nearly 1 degree ANB and maintenance of a positive overjet of around 1 mm.

It could be suggested that maxillary and mandibular growth rotations appear to be more important than previously thought, and may contribute to explaining the reduced need for surgery in the PFG. It could be hypothesized that the clockwise rotation of the maxilla and mandible in the PFG compared with the anti-clockwise direction in the CG gives a marginally more favourable facial profile. However, it is not possible to otherwise attempt to explain the reduced need for surgery in the PFG because, on average, there was no long-term convincing skeletal effect.

Weighted PAR cannot be used to try to explain the differences in surgical need between the PFG and the CG. PAR scores for PFG and CG were not statistically significantly different and both groups showed improvement. The improvement is mainly attributed to the heavily weighted overjet improvement from −2.2 mm to 0.8 mm in PFG with 68% of patients achieving a positive overjet at DC4. Improvement in the CG also occurred which was surprising and 48% of this group had a positive overjet at 15 years of age which cannot be attributed to upper arch alignment carried out during the observation period. It might also be suggested, although a weak hypothesis, that as there was a tendency for the lower incisors in the CG to be slightly more retroclined, marginally more dentoalveolar compensation may be occurring with biological attempts to maintain a positive overjet despite the underlying class III skeletal pattern.

The treatment outcome in the PFG was also very variable indicated by the wide standard deviations. This suggested wide biological variability in treatment response that is also found with many other orthodontic interventions. The inherent difficulty of ascertaining predictors of treatment success still remains a challenge for the orthodontic research community

Psychological outcomes

Very small trends towards increased self-esteem and reduced impact of malocclusion were seen in both the PFG and the CG. However, an association between improved self-esteem or reduced impact of malocclusion

Table 6 Forward mandibular displacement on closure at DC1 and DC4.

	DC1 (*n*)		DC4 (*n*)	
	No	Yes	No	Yes
CG	14	18	21	11
PFG	15	18	30	3

following early protraction facemask treatment is not supported. This is perhaps not surprising as the skeletal and dental outcomes were similar in PFG and CG and the trend towards differences between groups was not great enough to influence the psychological outcomes. It is noteworthy that the standard deviations for the OASIS questionnaire were very wide suggesting large variation in the impact of malocclusion. Future studies might wish to investigate the reasons why some children become more and some less concerned about their malocclusion over time.

Study design and limitations of the trial

Gender was recognized as a potential confounding factor at the beginning of the study and was addressed by stratified randomization to minimize the differences in treatment effect that may occur because of pubertal growth spurt timing. The effect of gender was not looked at in the planned statistical analyses, apart from whether the children needed surgery or not. It is acknowledged that further class III growth is likely to occur in both groups after the age of 15 years when our data were collected. However, we did not feel it was ethical to further delay data collection as some patients in the study were over 16 years old and wished to start pre-surgical orthodontics. Therefore, it is possible that one-third of patients needing orthognathic surgery in the PFG may be an under-estimate, particularly for the boys. This potential under-estimate will be similar in the PFG and the CG because of randomization.

Some participants were offered upper fixed appliances, as indicated, between the DC3 (12 years old) and DC4 (15 years old) data collections. As this was a real world study, it would have been unethical to withhold upper arch alignment, fixed appliances as clinically indicated, or extractions (in either arch) to disimpact teeth. The data showed that similar numbers of children received such treatment in the PFG and the CG showing equivalence between groups (Table 3). We have also shown that the use of upper fixed appliances either with or without extractions was not the confounder that we thought it

would be, because the need for orthognathic surgery was not influenced by this variable.

Clear cut objective criteria on clinical need for orthognathic surgery cannot be decided upon because of its multifactorial nature. Additionally every patient is unique and there is an aesthetic and subjective element to the decision, particularly in borderline cases. We therefore decided that the most scientifically robust method would be to record this outcome as 'yes' or 'no' based on a panel consensus using multiple clinical and radiological factors. It is acknowledged that with more ideal circumstances, the panel of orthodontists would be composed of clinicians not involved in the study. It is also noteworthy that there was a discordance between the long-term skeletal changes and the clinical judgement of need for orthognathic surgery which supports more emphasis on clinical assessment rather than relying too heavily on cephalometric measures.

The presence or absence of an anterior mandibular displacement on closure at the start of protraction facemask treatment will always be an important factor to consider. As Gravely (1984) suggested that no forward mandibular displacement persisted cephalometrically between retruded and intercuspal position, we continued to take only one lateral cephalogram in the intercuspal position throughout the study. Our analysis did not show any additional SNB improvement in patients whose forward mandibular displacement had been eliminated by protraction facemask treatment compared with those without pre-treatment displacement ($P = 0.08$). Table 6 shows the number of patients with forward mandibular displacement at the start and at 15 years in each group. Protraction facemask treatment was highly effective in eliminating a forward mandibular displacement on closure and it could be hypothesized that this is another factor that might contribute to a reduced need for orthognathic surgery in the PFG.

All patients in the protraction facemask group had RME prior to placing protraction forces. Patients with posterior cross bites were expanded to allow for some overcorrection as previously described. Patients without posterior crossbite turned the expansion screw once a

day, for 10 days. At the time of starting this trial, a meta analysis by Jäger *et al.* (2001) suggested that SNA and ANB changes were significantly greater when RME was used to release the circum-maxillary sutures. Therefore, we decided to use RME on every treated patient, which also standardized the clinical intervention and was at the time evidence based. Since then, Vaughan *et al.* (2005) and Tortop *et al.* (2007) have shown that the success of protraction headgear is not influenced by the use of RME. More recently, there is equivocal evidence for the effect of alternating rapid maxillary expansion and constriction with protraction facemask treatment on the forward movement of A point (Masucci *et al.*, 2014; Liu *et al.*, 2015).

An *a priori* sample size calculation was not possible for the primary outcome (need for orthognathic surgery) because of the lack of epidemiological data of need for orthognathic surgery in the untreated general population and this is a recognized limitation of the trial. Also to consider is the possibility that multiple statistical testing would result in a type 1 error (showing there is a difference between groups when there is not). This was considered unlikely because most of the comparisons between the groups were not statistically significant.

It is recognized that PAR is a measure of occlusal outcome only and other indices such as the Index of Complexity, Outcome and Need (ICON) (Daniels and Richmond, 2000) encompass multiple treatment factors. PAR was used in this trial because previous data were used as the basis for the sample size calculation and this would also enable comparison with previous research. Need for orthognathic surgery was evaluated using a panel consensus or consultant orthodontists. It is recognized that this method has limitations, as there is subjectivity compared with the use of a validated index. For future assessment of need for orthognathic surgery, the Index of Orthognathic Functional Treatment Need (IOFTN) (Ireland *et al.*, 2014) would be considered to be more robust.

The results of this randomized clinical trial may be generalized because the data were collected in the 'real world' clinical environment across multiple sites. The protocol described may delivered in a primary or secondary care treatment setting. The patients who participated in the study were considered to show a treatment response that could be replicated in the wider population under 10 years of age.

Given that the long-term follow-up data showed no skeletal effect in the PFG, it is suggested for the future, that more skeletal methods of class III protraction, such as bone anchored maxillary protraction, may be more effective in the long-term. A multi-centre randomized clinical trial is currently being carried out to evaluate this.

Conclusions

- Early class III protraction facemask treatment reduces the need for orthognathic surgery from two-thirds (CG) to one-third (PFG);
- Early protraction facemask treatment-related improvements in SNA, SNB and ANB were not maintained at 6 years follow-up. However, there was a statistically significant long-term clockwise rotation in the maxilla and mandible in the PFG;
- Sixty eight per cent of patient in the PFG presented with a positive overjet at age 15 years and
- Early protraction facemask treatment does not seem to confer a clinically significant psychosocial benefit.

Acknowledgements

References

Baccetti T, Franchi L, McNamara JA. Treatment and posttreatment craniofacial changes after rapid maxillary expansion and facemask therapy. *Am J Orthod Dentofacial Orthop* 2000; **118**: 404–413.

Baccetti T, Franchi L, McNamara Jr JA. Cephalometric variables predicting the long-term success or failure of combined rapid maxillary expansion and facial mask therapy. *Am J Orthod Dentofacial Orthop* 2004; **126**: 16–22.

Baccetti T, McGill JS, Franchi L, McNamara JA, Tollaro I. Skeletal effects of early treatment of Class III malocclusion with maxillary expansion and face-mask therapy. *Am J Orthod Dentofacial Orthop* 1998; **113**: 333–343.

Björk A, Skieller V. Growth of the maxilla in three dimensions as revealed radiographically by the implant method. *Br J Orthod* 1977a; **4**: 53–64.

Björk A, Skieller V. Cephalometric growth analysis of the maxilla. *Trans Eur Ortho Soc* 1977b; **7**: 209–233.

Brook PH, Shaw WC. The development of an index of orthodontic treatment priority. *Eur J Orthod* 1989; **11**: 309–320.

Chong Y-H, Ive JC, Artun J. Changes following the use of protraction headgear for early correction of Class III malocclusion. *Angle Orthod* 1996; **66**: 351–362.

Cozza P, Marino A, Mucedero M. An orthopaedic approach to the treatment of Class III malocclusions in the early mixed dentition. *Eur J Orthod* 2004; **26**: 191–199.

Daniels C, Richmond S. The development of the Index of Complexity, Outcome and Need (ICON). *J Orthod* 2000; **27**: 149–162.

De Toffol L, Pavoni C, Baccetti T, Franchi L, Cozza P. Orthopedic treatment outcomes in class III malocclusion. *Angle Orthod* 2008; **78**: 561–573.

Farkas LG. Anthropometry of the Head and Face. 2nd edn. New York: Raven Press. 1994.

Foersch M, Jacobs C, Wriedt S, Hechtner M, Wehrbein H. Effectiveness of maxillary protraction using facemask with or without maxillary expansion: a systematic review and meta-analysis. *Clinical Oral Investigations* 2015; **19**: 1181–1192.

Franchi L, Baccetti T, McNamara JA. Postpubertal assessment of treatment timing for maxillary expansion and protraction therapy followed by fixed appliances. *Am J Orthod Dentofacial Orthop* 2004; **126**: 555–568.

Gallagher RW, Miranda F, Buschang PH. Maxillary protraction: treatment and posttreatment effects. *Am J Orthod Dentofacial Orthop* 1998; **113**: 612–619.

Graber LW. Chin cup therapy for mandibular prognathism. *Am J Orthod* 1977; **72**: 23–41.

Gravely JF. A study of the mandibular closure path in Angle class III relationship. *Br J Orthod* 1984; **11**: 85–91.

Hagg U, Tse A, Bendeus M, Rabie ABM. Long-term follow-up of early treatment with reverse headgear. *Eur J Orthod* 2003; **25**: 95–102.

Ireland AJ, Cunningham SJ, Petrie A, Cobourne MT, Acharya P, Sandy JR, Hunt NP. An index of Orthognathic Functional Treatment Need (IOFTN). *J Orthod* 2014; **41**: 77–83.

Ishii H, Morita S, Takeuchi Y, Nakamura S. Treatment effect of combined maxillary protraction and chincap appliance in severe skeletal Class III cases. *Am J Orthod Dentofacial Orthop* 1987; **92**: 304–312.

Jäger A, Braumann B, Kim C, Wahner S. Skeletal and dental effects of maxillary protraction in patients with Angle Class III malocclusion a meta-analysis. *J Orofac Orthop* 2001; **62**: 275–284.

Kajiyama K, Murakami T, Suzuki A. Evaluation of the modified maxillary protractor applied to Class III malocclusion with retruded maxilla in early mixed dentition. *Am J Orthod Dentofacial Orthop* 2000; **118**: 549–559.

Kajiyama K, Murakami T, Suzuki A. Comparison of orthodontic and orthopedic effects of a modified maxillary protractor between deciduous and early mixed dentitions. *Am J Orthod Dentofacial Orthop* 2004; **126**: 23–32.

Kapust A, Sinclair PM, Turley PK. Cephalometric effects of face mask/expansion therapy in Class III children: a comparison of three age groups. *Am J Orthod Dentofacial Orthop* 1998; **113**: 204–212.

Kerr WJS, Miller S, Dawber JE. Class III malocclusion: Surgery or orthodontics? *Br J Orthod* 1992; **19**: 21–24.

Kerr WJS, TenHave TR. A comparison of three appliance systems in the treatment of Class III malocclusion. *Euro J Orthod* 1988; **10**: 203–214.

Kerr WJS, TenHave TR, McNamara JA. A comparison of skeletal and dental changes produced by Function Regulators (FR-2 and FR-3). *Eur J Orthod* 1989; **11**: 235–242.

Kim JH, Viana MAG, Graber T, Omerza F, Begole E. The effectiveness of protraction face mask therapy: a meta-analysis. *Am J Orthod Dentofacial Orthop* 1999; **115**: 675–685.

Liu W, Zhou Y, Wang X, Liu D, Zhou S. Effect of maxillary protraction with alternating rapid palatal expansion and constriction vs expansion alone in maxillary retrusive patients: a single-center, randomized controlled trial. *Am J Orthod Dentofacial Orthop* 2015; **148**: 641–651.

Loh MK, Kerr WJS. The function regulator III: effects and indications for use. *Br J Orthod* 1985; **12**: 153–157.

Mandall NA, Cousley R, DiBiase A, Dyer F, Littlewood S, Mattick R, Nute S, Doherty B, Stivaros N, McDowall R, Shargill I, Worthington H. Is early class III protraction facemask treatment effective? A multicentre, randomized, controlled trial: 15-month follow-up. *J Orthod* 2010; **37**: 149–161.

Mandall NA, Cousley R, DiBiase A, Dyer F, Littlewood S, Mattick R, Nute S, Doherty B, Stivaros N, McDowall R, Shargill I, Ahmad A, Walsh T, Worthington H. Is early class III protraction facemask treatment effective? A multicentre, randomized, controlled trial: 3-year follow-up. *J Orthod* 2012; **39**: 176–185.

Mandall NA, McCord JF, Blinkhorn AS, Worthington HV, O'Brien KD. Perceived aesthetic impact of malocclusion and oral self-perceptions in 14–15-year-old Asian and Caucasian children in Greater Manchester. *Eur J Orthod* 2000; **22**: 175–183.

Masucci C, Franchi L, Defraia E, Mecedero M, Cozza P, Baccetti T. Stability of rapid maxillary expansion and facemask therapy: a long-term controlled study. *Am J Orthod Dentofacial Orthop* 2011; **140**: 493–500.

Masucci C, Franchi L, Giuntini V, Defraia E. Short-term effects of a modified Alt-RAMEC protocol for early treatment of Class III malocclusion: a controlled study. *Orthodontics and Craniofacial research* 2014; **17**: 259–269.

Merwin D, Ngan P, Hagg U, Yiu C, Wei SHY. Timing for effective application of anteriorly directed orthopedic force to the maxilla. *Am J Orthod Dentofacial Orthop* 1997; **112**: 292–299.

Ngan P, Hagg U, Yiu C, Merwin D, Wei SHY. Treatment response to maxillary expansion and protraction. *Eur J Orthod* 1996; **18**: 151–168.

Ngan P, Hagg U, Yiu C, Wei SHY. Treatment response and long-term dentofacial adaptations to maxillary expansion and protraction. *Sem Orthod* 1997; **3**: 255–264.

Ngan P, Wei SHY, Hagg U, Yiu CKY, Merwin D, Stickel B. Effect of protraction headgear on Class III malocclusion. *Quintessence Int* 1992; **23**: 197–207.

Ngan P, Yiu C. Evaluation of treatment and posttreatment changes of protraction facemask treatment using the PAR index. *Am J Orthod Dentofacial Orthop* 2000; **118**: 414–420.

Pangrazio-Kulbersh V, Berger J, Kersten G. Effects of protraction mechanics on the midface. *Am J Orthod Dentofacial Orthop* 1998; **114**: 484–491.

Piers Harris 2. Piers Harris Children's Self concept Scale. 2002: 2nd Edition Western Psychological Services, 12031 Wilshire Boulevard, Los Angeles, CA 90025–1251.

Proffit WR, White RP, Sarver DM. Contemporary treatment of dentofacial deformity. Mosby 2003, Chapter 16, p508 Table 16.

Richmond S, Shaw WC, O'Brien KD, Buchanan IB, Jones R, Stephens CD, Roberts CT, Andrews M. The development of the PAR Index (Peer Assessment Rating): reliability and validity. *Eur J Orthod* 1992; **14**: 125–139.

Ritucci R, Nanda R. The effect of chin cup therapy on the growth and development of the cranial base and midface. *Am J Orthod Dentofacial Orthop* 1986; **90**: 475–483.

Robertson NRE. An examination of treatment changes in children treated with the function regulator of Fränkel. *Am J Orthod* 1983; **83**: 299–310.

Stellzig-Eisenhauer A, Lux C, Schuster G. Treatment decision in adult patients with Class III malocclusion: Orthodontic therapy or orthognathic surgery. *Am J Ortho Dentofacial Orthop* 2002; **122**: 27–37.

Sugawara J, Asano T, Endo N, Mitani H. Long-term effects of chincap therapy on skeletal profile in mandibular prognathism. *Am J Orthod Dentofacial Orthop* 1990; **98**: 127–133.

Takada K, Petdachai S, Sakuda M. Changes in dentofacial morphology in skeletal Class III children treated by a modified maxillary protraction headgear and a chin cup: a longitudinal cephalometric appraisal. *Eur J Orthod* 1993; **15**: 211–221.

Tortop T, Keykubat A, Yuksel S. Facemask therapy with and without expansion. *Am J Orthod Dentofacial Orthop* 2007; **132**: 467–474.

Vaughan GA, Mason B, Moon HB, Turley PK. The effects of maxillary protraction therapy with or without rapid palatal expansion: a prospective randomised clinical trial. *Am J Orthod Dentofacial Orthop* 2005; **128**: 299–309.

Wells AP, Sarver DM, Proffit WR. Long-term efficacy of reverse pull headgear therapy. *Angle Orthod* 2006; **76**: 915–922.

Westwood PV, McNamara JA, Baccetti T, Franchi L, Sarver DM. Long-term effects of Class III treatment with rapid maxillary expansion and facemask therapy followed by fixed appliances. *Am J Orthod Dentofacial Orthop* 2003; **123**: 306–320.

Yuksel S, Ucem TT, Keykubat A. Early and late facemask therapy. *Eur J Orthod* 2001; **23**: 559–568.

Zhang W, Hong-Chen Q, Yu M, Zhang Y. The effects of maxillary protraction with or without rapid maxillary expansion and age factors in treating class III malocclusion: a meta analysis. *PLoS One*; 2015 June.