The effect of early intervention on skeletal pattern in Class II malocclusion: A randomized clinical trial

J. F. Camilla Tulloch, BDS, FDS, D.Orth.,[a] Ceib Phillips, PhD, MPH,[b] Gary Koch, PhD,[c] and William R. Proffit, DDS, PhD[a]
Chapel Hill, N.C.

Early treatment for Class II malocclusion is frequently undertaken with the objective of correcting skeletal disproportion by altering the growth pattern. Because the majority of previous studies of growth modification for Class II malocclusion have been based on retrospective record reviews, the efficacy of such an approach has not been well established. In this controlled clinical trial, patients in the mixed dentition with overjet ≥ 7 mm were randomly assigned to either early treatment with headgear, or modified bionator, or to observation. All patients were observed for 15 months with no other appliances used during this phase of the trial. The three groups, who were equivalent initially, experienced statistically significant differences ($p < 0.01$) in skeletal change. There was considerable variation in the pattern of change within all three groups, with about 80% of the treated children responding favorably. Although patients in both early treatment groups had approximately the same reduction in Class II severity, as reflected by change in the ANB angle, the mechanism of this change was different. The headgear group showed restricted forward movement of the maxilla, and the functional appliance group showed a greater increase in mandibular length. The permanence of these skeletal changes and their impact on the subsequent treatment remains to be evaluated. (Am J Orthod Dentofac Orthop 1997;111:391-400.)

The majority of patients with Class II malocclusions have some sort of skeletal imbalance. Because Class II malocclusion becomes apparent early in the mixed dentition, the possibility of growth modification and the optimal timing for treatment are both questions of considerable clinical interest. Given a young patient with a noticeable overjet, the choices are early treatment to modify jaw growth, later treatment to camouflage the jaw discrepancy through tooth movement, or for the most severe, surgical correction of the skeletal relationship. Although the goals of each approach are the same, namely, to improve facial and dental appearance, maintain or enhance the oral health, and establish a stable and functional occlusion, the treatment approaches are very different. To date there has been little systematic evaluation of the effectiveness, benefits or costs of early as compared with later treatment, and the uncertainty about the benefit of early treatment for Class II malocclusion continues.

During the last decade (1980-1990), more than 130 articles reporting data on groups of 10 or more patients with Class II malocclusions were published in the four most widely circulated English language orthodontic journals.[1] Collectively, these articles provided information on 14 different appliance systems or treatment approaches. Although nearly all studies reported successful correction of the occlusion, a previous review of this literature[2] cautioned against the emerging consensus that early treatment can modify growth. The majority of reports during this period use the retrospective selection of cases, which tends to bias samples in favor of positive treatment findings. In addition, even the best of these studies suffer from methodologic limitations such as sample sizes too small to detect clinically important differences, inappropriate or no control or comparison groups, failure to record or adjust for major confounding variables such as gender, age at start, or duration of treatment, and failure to account for patients who start in treatment but for whom no subsequent data is available. Although more recent studies that use prospective designs have since been published,[3-9] these either report on small samples, do not use concurrent untreated patients as controls, or do not address the question

From the University of North Carolina.
Supported in part by NIH grant DE-08708 from the National Institute of Dental Research and by the Orthodontic Fund, Dental Foundation of North Carolina.
[a]Professor, Department of Orthodontics.
[b]Research professor, Department of Orthodontics.
[c]Professor, Department of Biostatistics.
Reprint requests to: Dr. J. F. Camilla Tulloch, Department of Orthodontics, University of North Carolina School of Dentistry, Chapel Hill, NC 27599-7450.

0889-5406/97/$5.00 + 0 **8/1/73340**

Table I. Criteria used to identify patients eligible for the trial

Inclusion criteria	*Exclusion criteria*
Overjet ≥ 7 mm	Congenital syndromes or defects
All permanent incisors and first molars erupted	Obvious facial asymmetry
All permanent teeth (excluding third molars) developing as seen on panoramic radiograph	Extreme vertical disproportion
1 year prepeak-height velocity as judged from the hand/wrist radiograph	Prior orthodontic treatment including space maintainers or habit appliances

of the effectiveness of early orthodontic treatment or the possibility of growth modification.

This study, which is part of a larger study of the benefit of early treatment for Class II malocclusion, was designed to determine whether clinically relevant and statistically significant changes in expected growth can be produced. The intent of the first phase of the trial was to assess the effect of early treatment in a broad way, addressing the questions: Does early treatment affect the relationship of the maxilla to the mandible or can treatment results be accounted for by dentoalveolar changes alone; and is there a difference in the effect produced by different appliances? In this article the focus is on the methods that allow the statistical evaluation of the skeletal (growth) effects of two simple but contrasting approaches to growth modification, as compared with the changes seen in a group of equivalent patients being observed without treatment.

METHOD

Research Design

A parallel, randomized clinical trial was designed to cover the three phases of orthodontic treatment frequently used to treat Class II malocclusion: early treatment in the mixed dentition (phase 1), comprehensive treatment in the permanent dentition (phase 2), and retention after treatment (phase 3). This report describes only the skeletal and dental changes observed during phase 1. In phase 1 of the trial, each child was randomly assigned by using a stratified block randomization, with gender as the stratification factor, to one of three groups, headgear, functional appliance, or observation only. Randomization was performed within gender in blocks of six patients with Proc Plan in SAS.[10] Fifteen months was set as the time for comparison of treatment versus observation during phase 1, though phase 1 treatment could continue beyond this time point if clinical objectives were not achieved.

Eligibility Criteria

The criteria for inclusion in the trial (Table I) were chosen to ensure that patients would be in the mixed dentition, have the more severe Class II malocclusions, all permanent teeth developing (excluding third molars), and growth potential throughout phase 1. A hand-wrist film was taken to determine the child's skeletal maturity with the method of Bowden,[11] and patients were included only if they were referenced at least 1 year before peak height velocity on the growth curve. The baseline clinical records were examined to exclude patients with clinically obvious facial asymmetry, more than 2 standard deviations from normal vertical proportionality,[12] for whom delaying treatment for 15 months would compromise care, or those who had had prior orthodontic treatment, including space maintainers and habit appliances.

Early Treatment Protocol

A simple standard approach was chosen for each treatment group, so that a broad range of presenting clinical conditions could be accommodated. A combination headgear was used with a supershort outer bow (ending approximately at the mesial of the molar tubes) and adjusted to deliver between 8 and 10 ounces to the headcap and with the neck strap force just sufficient to prevent buccal flaring of the upper molars. The functional appliance was a modified bionator with the bite taken with 4 to 6 mm of protrusion and minimal vertical opening. Reactivation of the functional appliance, when necessary, was by construction of a new appliance. All appliances were delivered within 1 month of the patient's initial records, and no other appliances were used during phase 1. All patients were seen by the same orthodontist at 6- to 8-week intervals for the treatment groups and at 6, 12, and 15 months for the observation only group.

Standardized Records

Records were obtained for each patient by the same research technician, before treatment and at the end of 15 months of treatment or observation. All cephalograms were taken in the natural head position with the patient seated and teeth in occlusion (leaf gauges were not used). Each cephalogram was traced and digitized by the same research technician. Because the molar bands were not removed at the end of phase 1, the technician was not masked as to these patients' treatment group.

Measurements were made with an x-y coordinate system, established with the sella nasion (S-N) line rotated down 6° anteriorly as the horizontal reference, and the vertical reference a perpendicular through sella. The 139-point digitizing model included 89 anatomic landmarks. The method error for the landmarks used ranged from 0.24 (sella) to 1.2 mm (pogonion). The reliability of landmark location and digitizing, as indicated by the intraclass correlation statistic, ranged from 0.89 for overbite to 1.00 for anterior face height. A restricted set of 12 cephalometric measures was used to describe the position and relationship of the maxillary and mandibular skeletal and dental units. We limited the number of variables used because many cephalometric measures provide overlap-

ping information. The use of only a few measures also decreases the likelihood of false positive findings occurring by chance.

Sample Size Considerations

The primary focus of phase 1 of the trial was the impact of treatment on growth. An earlier metaanalysis by Mills[13] showed the averaged annualized change in SNPg for nine growth studies of treated and untreated patients to be 15%. A preliminary sample size of 40 patients per group was calculated as necessary to detect a difference in the mean value between any two of the three groups equal to a doubling of the annualized change in SNPg. Conservative significance and power levels (alpha = 0.01 and power = 0.90) were chosen to reflect the broad scope of the expected data analyses. Additional patients were recruited because of the anticipated losses over the 10 years of the trial.

Patient Recruitment

The first patient was recruited in August 1988 and the last in November 1993. The recruitment of patients is summarized in Table II. Fifteen-month records were obtained on the 166 children completing phase 1 and on 3 of the patients transferred to a parallel ongoing observational study.

Analysis

Analysis of data, based on only the patients in a clinical trial who actually received the treatment rather than those who were assigned to receive the treatment, may give invalid results.[14] Accordingly, two groups of patients were defined for the purpose of analysis, an "intent to treat" (ITT) sample, which comprised all patients (n = 180) who were randomized and for whom baseline records were available, and an "efficacy analyzable" (EA) sample (n = 166) who were the patients completing the phase 1 protocol. Because 11 of the ITT sample only had initial records, their 15-month measures were imputed with regression coefficients from least squares regression analyses that included baseline measures, treatment group, and gender as explanatory factors. All ITT analyses, including baseline equivalency assessments, were performed with the child's initial group assignment. The descriptive and inferential statistics reported in the text and tables are for the EA sample only, because there were no differences in the findings for the ITT and EA groups. The level of significance was set at 0.01 for all analyses.

The equivalence of the groups at baseline was confirmed with Mantel Haensel row mean score statistics for age and Bowden summary score of skeletal maturity, and Mantel Haensel chi-square for gender and presence of bilateral Class II molar relationship. A multivariate analysis of variance on maxillary, mandibular, interjaw, and dental measures was used and confirmed by a one-way analysis of variance on each measure. The three groups were quite similar in all characteristics at baseline (Tables III and IV) with a higher proportion of boys than girls in all three groups, a limited age range, and generally a rather severe malocclusion.

Table II. The flow of patients through recruitment, enrollment and treatment sequences in Phase 1 of the trial

Number of patients	*Activity*
2164 screened	General screening for orthodontic patients, plus special screening in response to announcements about trial
207	Met clinical screening criteria
192	3 with congenitally missing teeth 12 (11 girls) skeletal age too advanced
	⇓
	Informed consent signed, patients randomized to group.
	⇓
180 with initial records	12 did not attend for initial records 4 patient/parent objected to assigned group 4 cited travel difficulties 4 no reason given
175 started trial	5 patients transferred for treatment elsewhere 2 objected to assigned group 3 clinical decision against group assignment
166 completed phase 1	9 did not complete first phase of trial 3 later determined ineligible (2 missing permanent teeth) (1 skeletal asymmetry) 2 protocol violations with other appliances placed 1 moved; 1 refused treatment; 2 did not return

The outcome of interest in this phase of the trial was the difference in the change in skeletal and dental relationship experienced by the two early treatment groups as compared with the observation only group. To adjust for minor differences in the time each patient was observed and to facilitate comparison with other studies, the results are presented as the annualized change in the various cephalometric measures chosen to characterize skeletal size, position, or relationship. A factorial analysis of variance, with gender and treatment group as the explanatory factors, was used to assess the annualized change in each cephalometric measure. For those cephalometric measures with a statistically significant overall model and a significant difference among treatments from the analysis of variance model (Appendix 1), comparisons between pairs of treatment groups were performed with the least squares comparisons in Proc GLM in SAS.[10] Although calculation of missing 15-month data was performed for only 11 of the 180 patients in the ITT group, this imputation does violate the assumption of random sampling. Stratified Kruskal Wallis tests, which are more robust to imputation, were used to confirm the analysis of variance results in the ITT sample.

RESULTS

The results of the first phase of the trial are presented in Table V. The overall analysis of vari-

Table III. Comparison at baseline of the three groups formed by random allocation

	Control (n = 61)				*Functional (n = 53)*				*Headgear (n = 52)*				
	Mean	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*p value*
Age (years)	9.4	1.2			9.4	1.0			9.4	1.0			0.94
Bowden (scale)	17.4	2.1			18.1	1.8			17.8	2.0			0.19
Overjet (mm)	8.4	2.0			8.3	2.2			8.3	2.3			0.93
Bilateral Class II molars			57	93			47	89			47	90	0.65
Male			35	57			30	57			31	60	0.94

Table IV. Descriptive statistics for the pretreatment cephalometric values for the three groups

	Control (n = 61)		*Functional (n = 53)*		*Headgear (n = 52)*	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Maxillary skeletal						
SNA (degrees)	83.3	3.8	82.2	3.8	83.2	3.3
Mx unit length (mm)	88.3	4.3	88.8	4.3	87.9	4.1
A to N perp (mm)	−0.7	3.7	−1.8	3.7	−0.8	3.2
Mandibular skeletal						
SNB (degrees)	77.0	3.4	75.9	3.7	77.1	3.1
Md unit length (mm)	107.2	5.3	107.7	4.8	107.4	5.1
Pg to N perp (mm)	−11.3	6.4	−13.6	6.7	−11.9	5.7
Skeltal relationship						
ANB (degrees)	6.3	2.0	6.3	2.1	6.0	1.8
Unit difference (mm)	19.0	3.3	18.9	3.8	19.5	3.6
A-B difference (mm)	9.3	3.5	9.4	3.4	9.0	3.3
Dental relationship						
Overjet (mm)	8.4	2.0	8.3	2.2	8.3	2.3
Overbite (mm)	5.2	1.9	5.4	2.3	5.2	2.3
Interincisal angle	119.8	6.8	121.3	8.0	121.4	6.9

Mx = Maxillary; *perp* = perpendicular; *Md* = mandibular. *P* values for Wilks' Lambda were 0.43 (maxillary skeletal), 0.24 (mandibular skeletal), 0.84 (skeletal relationship), and 0.92 (dental relationship).

ance model was statistically significant (Appendix 1, $p < 0.01$) for all the outcomes, except for maxillary unit length ($p = 0.48$), pogonion to NP ($p = 0.08$), and incisor angulation ($p = 0.87$). For all measures, neither gender nor the interaction between gender and treatment were significant explanatory effects (Appendix 1). This indicates that the pattern of change was not different for boys and girls within the three groups, and there was no overall difference between boys and girls for any of the cephalometric measures.

The comparison between groups (Table V) indicate that early treatment, at least as carried out in this trial, did, on average, reduce the severity of the Class II skeletal discrepancy, no matter whether a headgear or a functional appliance was used, but the mechanism by which these changes occurred depended on the treatment group. The headgear group had the forward movement of the maxilla restricted or reversed when compared with the control or functional appliance groups, whereas the patients in the functional appliance group showed increased forward positioning of the mandible with increased mandibular unit length as compared with the control group. Although there was no specific effort to reduce the overjet in this phase of the trial, the average overjet did reduce in the two treatment groups, though not in the control group. The mean changes can be visualized in the composite tracings presented in Fig. 1.

The variability of individual responses is often obscured by a focus on mean changes, and it should be noted that despite the statistically significant average differences, there was wide variation within all three groups. The box plots (Fig. 2) of annualized change during phase 1 of the trial give a graphic display of the range of response for selected measures. The data on the control patients provide a vivid representation of the changes that occurred without any treatment during this time. As Fig. 2

Table V. Descriptive statistics for the annualized change in cephalometric measures occurring during phase 1 of the trial for each of the three groups. *p* values associated with the assessment of the average annualized change within each group are given together with the *p* values associated with the comparison of the average annualized changes between groups

	Control (n = 61)			*Functional (n = 53)*			*Headgear (n = 52)*			*p values** contrasts between group*		
	Mean	*SD*	*p value**	*Mean*	*SD*	*p value**	*Mean*	*SD*	*P value**	*C vs F*	*C vs Hg*	*F vs Hg*
Maxillary skeletal												
SNA (degrees)	0.26	1.17	0.07	0.11	1.26	0.52	−0.92	1.11	0.0001	0.46	0.0001	0.0001
Mx unit length (mm)	1.34	1.47	0.0001	1.46	1.55	0.0001	1.03	1.60	0.0001	—	—	—
A to N perp (mm)	0.21	1.17	0.14	0.05	1.25	0.79	−0.25	1.13	0.0001	0.41	0.0001	0.0001
Mandibular skeletal												
SNB (degrees)	0.43	0.90	0.0002	1.07	0.91	0.0001	0.15	0.88	0.25	0.003	0.08	0.0001
MD unit length (mm)	2.36	1.17	0.0001	3.69	1.47	0.0001	2.97	1.82	0.0001	0.0001	0.03	0.02
Pg to N perp	0.81	1.72	0.0003	1.14	1.90	0.0001	0.20	1.76	0.43	—	—	—
Skeletal relation												
ANB (degrees)	−0.17	0.73	0.13	−0.93	0.99	0.0001	−1.07	0.73	0.0001	0.0001	0.0001	0.40
Unit difference (mm)	1.02	1.37	0.0001	2.23	1.60	0.0001	1.94	1.89	0.0001	0.0001	0.003	0.37
A-B difference (mm)	−0.22	0.98	0.09	−1.23	1.23	0.0001	−0.86	0.91	0.0001	0.0001	0.002	0.08
Dental relationship												
Overjet (mm)	−0.09	0.98	0.58	−2.66	1.81	0.0001	−1.50	1.36	0.0001	0.0001	0.0001	0.0002
Overbite (mm)	0.40	1.13	0.03	−1.55	2.16	0.0001	−0.05	1.00	0.86	0.0001	0.11	0.0001
Interincisal angle	0.80	4.31	0.22	0.32	4.52	0.65	0.71	4.31	0.29	—	—	—

**p* values in support of the null that the average annualized change is zero.
***p* values are not reported for contrasts between groups unless the overall model for that measure in the two-way ANOVA with interaction and the explanatory factor of treatment in the main effects only model were significant. *p* values reported are in support of the null that the average annualized change is the same in the two groups being compared.

shows, although there was, on average, no change in overjet for the control group, there was a considerable range in response (50% of the children showing an increase in overjet and 50% a reduction). Nearly all the patients in the two treatment groups showed a reduction in the ANB angle, but this was also true for a large majority of patients in the control group. Although on average the SNA angle decreased for the headgear group, it actually increased for some of these patients. The range of response for the SNA angle was also considerable for both the control and functional appliance groups. In a similar way, although the average change in mandibular length was greatest in the functional appliance group, approximately 20% of this group experienced less growth than the mean growth experienced by the control group.

DISCUSSION

Currently, there is little consensus among clinicians about the merits of alternative treatments for patients with Class II malocclusion. It is interesting to reflect that during the past 30 years, the most frequent approach to growth modification in the United States has been extraoral force (headgear) to restrict or redirect the growth of the upper jaw. At the same time, the European approach has more generally favored the forward positioning of the lower jaw (functional appliance therapy) to stimulate mandibular growth. The important questions about the two approaches are whether either really works; whether the two approaches produce differential effects; even if early intervention does change growth, will this change be sustained; and, finally, will early treatment make a difference to the patients' subsequent management or to their long-term outcomes.

The randomized clinical trial (RCT), first used in healthcare to evaluate the treatment of tuberculosis,[15] has become accepted as the standard for comparing alternative treatment approaches. The strength of the RCT derives from the method used to allocate patients to alternative groups. Randomization offers the best chance, though still no guarantee, of achieving equivalence across comparison groups of all factors, both known and unknown, that might confound or modify the effect of treatment. Although theoretically possible, alternative approaches such as matching patients on factors thought to be influential, for example, gender, age, or severity of the presenting condition, are cumbersome, require large numbers of patients to achieve the desired matches, and ignore the important consideration of pretreatment equivalence across factors that are not even suspected of being important.[16,17] The balancing effect produced in a randomized trial allows the dif-

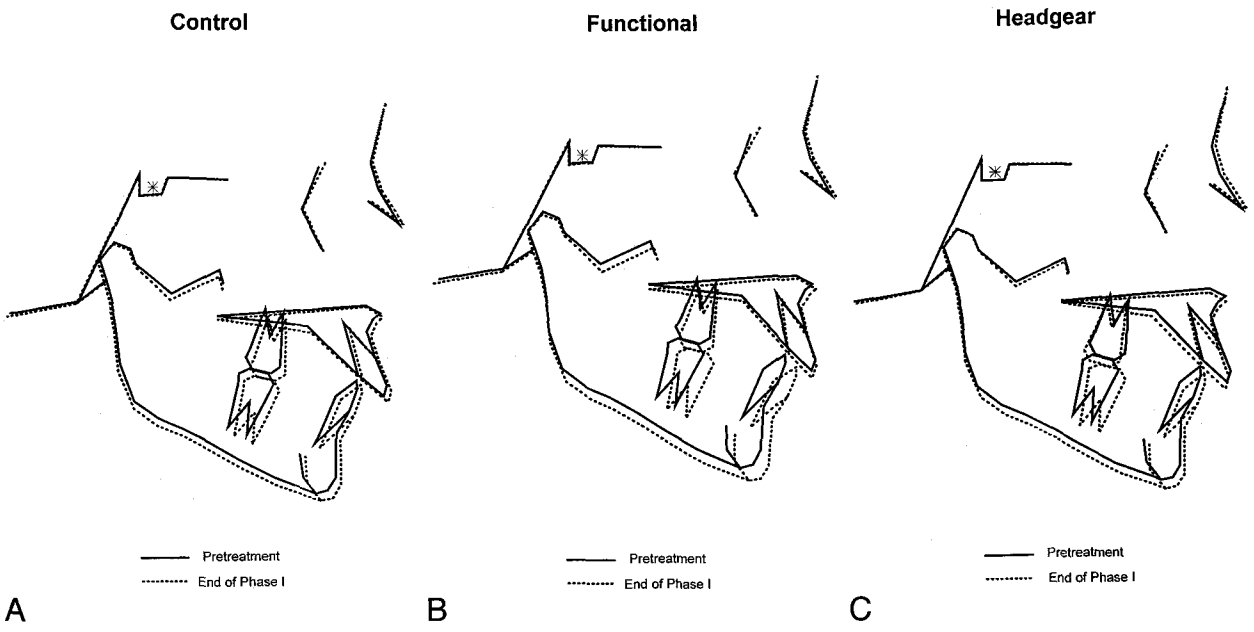

Fig. 1. Composite tracings of mean changes occurring in three groups during phase 1 of trial **(A)** control patients who had observation only during this period, **(B)** functional appliance (modified bionator) group, **(C)** headgear group (face-bow to first molar bands with headcap and neck strap). No other treatment was undertaken during this phase of trial.

ferences in outcome to be more safely attributed to differences in treatment rather than differences in the characteristics of the patients.

Although the RCT is now widely accepted in healthcare evaluations, there are some difficulties in applying this method to an orthodontic problem. First and foremost is the ethical necessity that the alternative treatments not only be equally appropriate (given current knowledge), but also that patients and parents be fully informed of the risks of the alternatives. Our focus in the University of North Carolina clinical trial has been on the genuine dilemma of whether growth modification actually does occur during early treatment, and if it does, whether this makes a difference to the ultimate outcomes that patients seek from treatment. Such a trial is ethical because we really do not know the answer,[2,18,19] and important because of the increased emphasis on early treatment in current clinical practice. There have been many instances in medical practice where treatments have been widely endorsed on the basis of weak research design, only to be found to be either ineffective or even harmful when evaluated in carefully controlled randomized trials.[20,21] The maxim that enthusiastic treatment reports tend to have no controls whereas well-controlled treatment reports tend to have no enthusiasm[21,22] surely holds as true for orthodontic care as it does for general medical practice.

An additional difficulty arises in orthodontic trials due to the relatively long time period before the outcomes of care can be assessed. Because the alternative timing of treatment (early versus late) pertains to different age groups, patients need to be entered into a trial at the start of the early treatment period (8 to 9 years) and, for a full evaluation, observed to an age corresponding to 1 or 2 years after the completion of comprehensive treatment (16 to 18 years) depending on the complexity of the case. Simply selecting and comparing groups of patients who start treatment at different ages ignores the issue of pretreatment equivalency. If patients who attend for and receive treatment early are different from those who seek or receive treatment later, it becomes difficult to interpret whether the differences in outcome are due to differences in the treatments or differences in the patients.

One important question when interpreting any clinical study is the extent to which the study population reflects the broader population who seek treatment for similar problems. The area from which these children were recruited is one of the

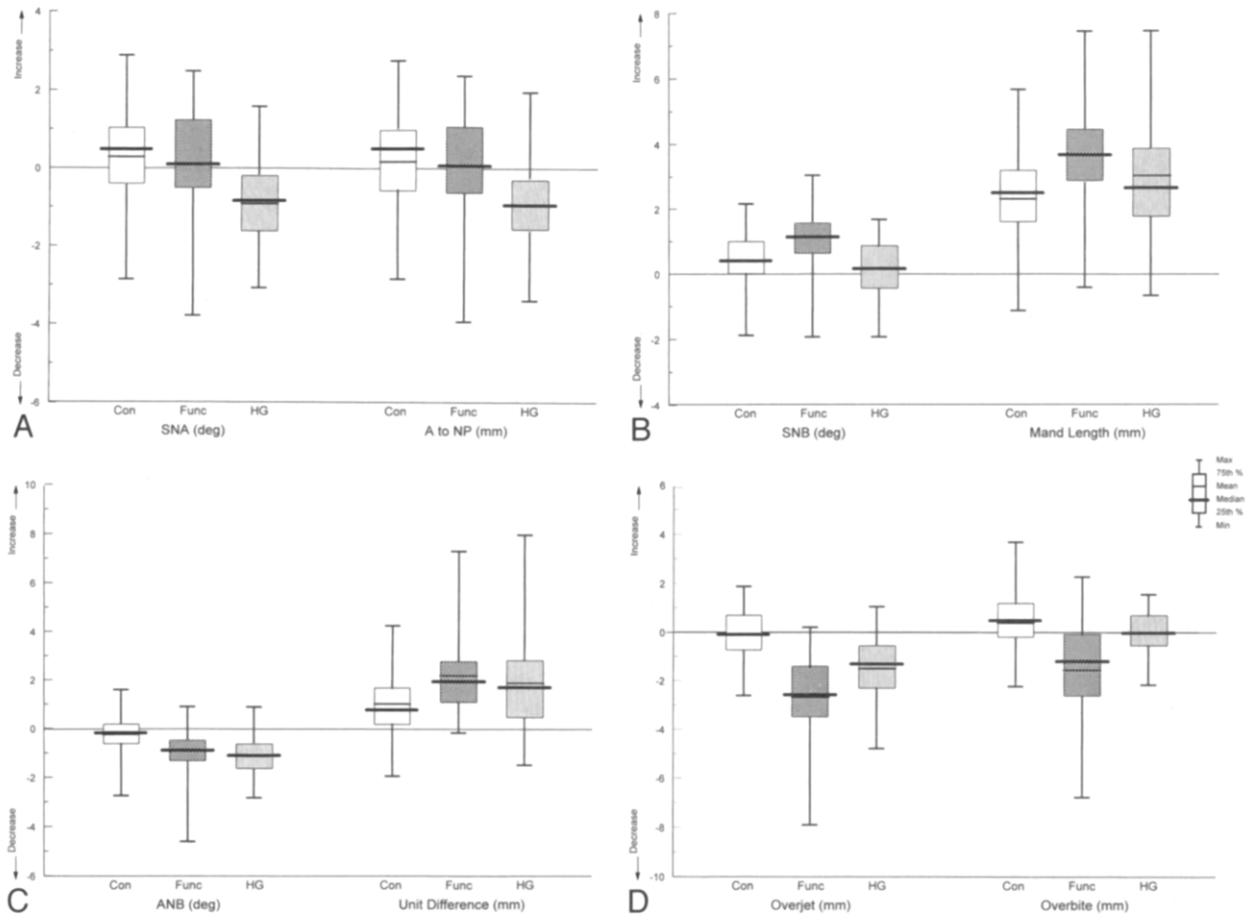

Fig. 2. Comparison of selected skeletal and dental changes occurring in three groups during phase 1 of trial. Box plots show distribution (median, 25th and 75th percentiles and range) of values within each group. This display is used instead of mean and standard deviations that suggest symmetry to data that may or may not reflect true distribution. However, these displays indicate there is little skewness in any of these measures.

"Standard Metropolitan Areas" delineated by the U.S. Bureau of the Census, and data that allows it to be compared with other such areas in the United States are available. However, the children enrolled in the trial include both those who were seeking orthodontic care in a university orthodontic department, and those responding to announcements about the trial. As such, they may reflect some differences from children seeking care in private offices. In addition, the findings from this study are based on data from white children only. Although it would have been desirable to have included other ethnic groups, the well-documented morphologic differences between groups would have greatly increased the variability and hence the number of patients needed for study. Sufficient patients were not available to permit the formation of independent subgroups. The response of children in different ethnic subgroups may prove to be an important area for future research.

The results of the trial show growth modification, in terms of modest changes in the skeletal jaw relationship of preadolescent children occurring during a period of early orthodontic treatment. Though small, the changes are in the direction that conventional clinical wisdom has suggested. This finding supports the rationale for differential diagnosis of Class II problems, and suggests that different appliances may be more or less suitable for different patients. However, it would be highly inappropriate to infer from this study that all children treated with headgear or functional appliances will show a predictable favorable growth response. There is great variability seen in the response of all

three groups. Some of the children in the control group showed favorable growth and their skeletal Class II relationship improved to a greater extent than the mean changes seen in the treatment groups. Some children in the headgear and functional appliance groups did not improve, and a few actually got worse.

A second point to stress is that the differences between the groups are small, particularly in comparison to the variability seen in untreated patients. This highlights the importance of using a prospective study design with an untreated group to serve as a control. Retrospectively collected samples are likely to include many patients who grew favorably, regardless of the appliance used. Patients who fail to show the desired response are seldom available for evaluation, because they are usually transferred to another appliance or another treatment approach, or treatment is discontinued, leaving only those patients who demonstrated the favorable changes available for study. The predominant use of this type of patient group in orthodontic publications[23,24] has tended to exaggerate the magnitude of the treatment/growth effect reported. Taking this in conjunction with the high probability of publication bias, that is, the tendency of authors to submit and editors to accept studies with positive rather than null or negative findings, together with the repeated publication of data from the same samples,[25] may have inappropriately raised our expectations of the possibilities for growth modification.

The phase 1 treatment outcomes from the University of North Carolina trial do not yet answer the most important questions: Does it really make a difference in the long-term whether treatment is started sooner or later, providing effective treatment occurs at some point? It is possible that the small skeletal changes produced by early treatment will be sustained over time and the subsequent management of and treatment outcomes for these patients will be different from those children for whom treatment was delayed. It is also possible that the effects of early treatment will gradually diminish so that the early and late treatment groups are indistinguishable by the age of 16 or so.[26] The impact of early treatment on subsequent growth, patient management, future treatment, and the long-term outcomes of care is the focus of the second phase of this trial. The question of the optimum timing for the start of treatment for children with Class II malocclusion is a complex one and, like many complex questions, requires the careful collection and analysis of data before the risks and benefits of treatment can be evaluated in terms of health gains versus resources used.

CONCLUSIONS

1. The randomized clinical trial is an efficient way to study the impact of early orthodontic treatment.
2. Children with Class II malocclusion experience considerable variation in growth during the preadolescent period, both with and without treatment.
3. Early treatment with either headgear or functional appliance therapy can both reduce the severity of a Class II skeletal pattern. With either approach, there is about a 75% chance of improvement in the jaw relationship.
4. On average, headgear produces greater change in the maxilla, whereas functional appliance therapy produces greater mandibular change, but there is considerable variation in the effect with both appliance systems.
5. Whether these early changes will be sustained and will make a difference in the patients' subsequent management and treatment outcomes remains to be evaluated.

We thank Lyna Rogers and Deborah Price for technical assistance throughout this project.

REFERENCES

1. Simon LS. A quantitative analysis of the measurements used to define and describe Class II malocclusion and the effects of treatment on growth. [Master's thesis.] Chapel Hill: Department of Orthodontics, School of Dentistry, University of North Carolina, 1993.
2. Tulloch JFC, Medland W, Tuncay OC. Methods used to evaluate growth modification in Class II malocclusion. Am J Orthod Dentofac Orthop 1990;98:340-7.
3. Schiavoni R, Grenga V, Macri V. Treatment of Class II high angle malocclusions with the Herbst appliance: a cephalometric investigation. Am J Orthod Dentofac Orthop 1992;102:393-409.
4. Firouz M, Zernik J, Nanda R. Dental and orthopedic effects of high-pull headgear in treatment of Class II, Division 1 malocclusion. Am J Orthod Dentofac Orthop 1992;102:197-205.
5. Nelson C, Harkness M, Herbisson P. Mandibular changes during functional appliance treatment. Am J Orthod Dentofac Orthop 1993;104:153-61.
6. Ghafari J, Jacobsson-Hunt U, Shofer FS, Laster LL. Changes in arch width in early treatment of Class II, Division 1 malocclusions. Am J Orthod Dentofac Orthop 1994;106:496-502.
7. Sidhu MS, Kharbanda OP, Sidhu SS. Cephalometric analysis of changes produced by a modified Herbst appliance in the treatment of Class II division 1 malocclusion. Br J Orthod 1995;22:1-12.
8. Keeling SD, Garvan CW, King GJ, Wheeler TT, McGorry S. Temporomandibular disorders after early Class II treatment with bionators and headgears: results from a randomized controlled trial. Semin Orthod 1995;1:149-64.
9. Ghafari JG, Shofer FS, Laster LL, Markowitz DL, Silverton S, Katz SH. Monitoring growth during orthodontic treatment. Semin Orthod 1995;1:165-75.
10. SAS Institute Inc. SAS/STAT User's guide. Ver. 6, 4th ed., Vol. 2. Cary (NC): SAS Institute Inc., 1989.
11. Bowden BD. Epiphyseal changes in the hand/wrist area as indicators of adolescent stage. Aust Orthod J 1976;4:87-104.
12. Fields HW, Proffit WR, Nixon WL, Phillips C, Stanek E. Facial pattern differences in long face children and adults. Am J Orthod 1984;85:217-23.
13. Mills JRE. Clinical control of craniofacial growth: a skeptic's viewpoint. In: McNamara JA, Ribbens KA, Howe RP, editors. Clinical alteration of the growing face. Monograph 14, Craniofacial Growth Series. Ann Arbor: Center for Human growth and Development, University of Michigan, 1983.
14. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received: Is it really an option? Stat Med 1991;10:1595-605.
15. Daniels M, Hill AB. Chemotherapy of pulmonary tuberculosis in young adults: an analysis of the combined results of three Medical Research Council Trials. Br Med J 1952;1:1162-8.

16. Ingelfinger JA, Mosteller F, Thibbodeau LA, Ware JH. Reading a report of a clinical trial. Chap. 10. Biostatistics in medicine. New York: Macmillan Publishing Co., 1983.
17. Sackett DL. Readers guides for therapy: Was the assignment of patients to treatment randomized? ACP J Club 1991;A-12 (editorial).
18. King GJ, Keeling SD, Hocevar RA, Wheeler TA. The timing of treatment for Class II malocclusion in children: a literature review. Angle Orthod 1990;60:87-97.
19. Chate RA. The burden of proof: a critical review of orthodontic claims made by some general practitioners. Am J Orthod Dentofac Orthop 1994;106:96-105.
20. Mosteller F, Frazier HS. Evaluating medical technologies. In: Frazier HS. Mosteller F, editors. Medicine worth paying for: assessing medical innovations. Cambridge (MA): Harvard University Press, 1995.
21. Sackett DL, Cook DH. Can we learn anything from small trials. In: Warren KS. Mosteller F, editors. Doing more good than harm: the evaluation of health care interventions. Ann N Y Acad Sci 1993;703:25-32.
22. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparison of therapy I: medical. Stat Med 1989;8:441-5.
23. Tulloch JFC, Antczak-Bouckoms AA, Tuncay OC. A review of clinical research in orthodontics. Am J Orthod Dentofac Orthod 1989;95:499-504.
24. Johnston LE, Paquette DE, Beattie JR, Cassidy DW, McCray JF, Killiany DM. The reduction in susceptibility bias in retrospective comparisons of alternative treatment strategies. In: Vig KD, Vig PS, editors. Clinical research as the basis for clinical practice. Monograph 25, Craniofacial Growth Series. Ann Arbor: Center for Human Growth and Development, University of Michigan, 1983.
25. Dickersin K, Yuan-I M. Publication bias: the problem that won't go away. In: Warren KS, Mosteller F, editors. Doing more good than harm: the evaluation of health care interventions. Ann N Y Acad Sci 1993;703:135-48.
26. Livieratos FA, Johnston LE. A comparison of one-stage and two-stage nonextraction alternatives in matched Class II samples. Am J Orthod Dentofac Orthop 1995;108:118-31.

APPENDIX

Appendix table

Measures		*DF*	*Sum of squares*	*Mean square*	*Type I* *Source*	*DF*	*Sum of squares*	*Mean squares*
Skeletal								
SNA (deg)	Model	5	48.53	9.71	Gender	1	0.14	0.14
	Error	160	227.24	1.42	Treatment	2	45.79	22.89
					Treatment * Gender	2	2.73	1.36
A to N perp (mm)	Model	5	46.65	9.33	Gender	1	0.03	0.03
	Error	160	226.80	1.42	Treatment	2	44.04	22.02
					Treatment * Gender	2	2.57	1.29
Mx unit length (mm)	Model	5	6.70	1.34	Gender	1	0.09	0.09
	Error	160	387.15	2.42	Treatment	2	5.14	2.57
					Treatment * Gender	2	1.48	0.74
SNB (deg)	Model	5	24.77	4.95	Gender	1	0.07	0.07
	Error	160	129.69	0.81	Treatment	2	23.44	11.72
					Treatment * Gender	2	1.25	0.62
Pg to N perp (mm)	Model	5	31.78	6.36	Gender	1	0.03	0.03
	Error	160	516.53	3.23	Treatment	2	24.31	12.16
					Treatment * Gender	2	7.43	3.72
Md unit length (mm)	Model	5	61.26	12.25	Gender	1	0.08	0.08
	Error	160	364.56	2.28	Treatment	2	52.32	26.16
					Treatment * Gender	2	8.87	4.43
ANB (deg)	Model	5	29.42	5.88	Gender	1	0.08	0.08
	Error	160	108.11	0.68	Treatment	2	27.50	13.75
					Treatment * Gender	2	1.85	0.92
Unit difference (mm)	Model	5	56.26	11.25	Gender	1	0.00	0.00
	Error	160	417.86	2.61	Treatment	2	46.11	23.06
					Treatment * Gender	2	10.15	5.07
A-B difference (mm)	Model	5	30.76	6.15	Gender	1	0.00	0.00
	Error	160	184.84	1.16	Treatment	2	29.17	14.59
					Treatment * Gender	2	1.58	0.79
Dental								
Overbite (mm)	Model	5	117.43	23.49	Gender	1	0.67	0.67
	Error	160	367.94	2.30	Treatment	2	115.77	57.89
					Treatment * Gender	2	0.98	0.49
Overjet (mm)	Model	5	195.73	39.15	Gender	1	1.97	1.97
	Error	160	316.57	1.98	Treatment	2	190.07	95.03
					Treatment * Gender	2	3.70	1.85
Interincisal angle	Model	5	70.3	14.06	Gender	1	21.96	21.96
	Error	160	316.57	1.98	Treatment	2	190.07	95.03
					Treatment * Gender	2	3.70	1.85

The degrees of freedom, Type I sum of squares, and mean square terms for the two-way analysis of variance with interaction for each of the annualized change measures are provided. The type I SS are given so that F statistics for the hypotheses of interest from both the full model with interaction term and the reduced main effects model can be calculated.

$$F(\text{Gender*Tx}) = \frac{\text{MeanSquare(Gender*Tx)}}{\text{MeanSquare(Error)}}$$

$$F(\text{Gender}) = \frac{\text{SS(Gender)/df(Gender)}}{[\text{SS(Error)} + \text{SS(Gender*Tx)}]/[\text{df(Error)} + \text{df(Gender*Tx)}]}$$

$$F(\text{Treatment/Gender}) = \frac{\text{SS(Tx)/df(Tx)}}{[\text{SS(Error)} + \text{SS(Gender*Tx)}]/[\text{df(Error)} + \text{df(Gender*Tx)}]}$$