The Online
ANGLE ORTHODONTIST
An International Journal of Orthodontics and Dentofacial Orthopedics
78th Anniversary 1930–2008

TABLE OF CONTENTS

The Angle Orthodontist: **Vol. 72, No. 6, pp. 571–575.**

National Institute of Dental and Craniofacial Research Efficacy Trials of Bionator Class II Treatment: A Review

Thomas Jacobs, DDS, MS;[a] Pornrachanee Sawaengkit, DDS, MS[b]

ABSTRACT

Reviewing publications from two North American functional appliance efficacy trials on Class II/1 malocclusions, we conclude that preadolescents show first molar correction in 13% of untreated cases, 38% of bionator-treated cases, and 50% of combined headgear/biteplane–treated cases. Also, bionator and headgear reduce ANB differences at nearly equivalent rates (degrees/year). Thirty-two percent of untreated patients show favorable reductions (>0.5 degrees/y), compared with 80% of patients treated by either bionators or headgear. Successful outcomes, however, are highly dependent on the severity of the malocclusion. Using either headgear or bionators, clinicians should expect successful Class II molar correction in over 57% of preadolescents presenting with half cusp molar discrepancies. When left untreated, 30% of such patients are reported to attain a Class 1 molar relation. But only 5% of cases presenting with greater than half cusp discrepancies are found to resolve themselves. The bionator appliance, though, corrects 30% of such cases, and combined headgear/biteplane treatment corrects 40%.

KEY WORDS: Functional appliance, Review, Bionator, Class II malocclusion, Efficacy.

Accepted: May 2002. Submitted: February 2002

INTRODUCTION Return to TOC

In 1988, responding to growing criticism that no consensus on the efficacy of different orthodontic treatments existed, the National Institute of Dental and Craniofacial Research (NIDCR) awarded 5-year grants to two North American orthodontic centers to study the comparative effectiveness of Balters' bionator appliance and headgear in the early treatment of Class II, division 1 malocclusions. The participating clinics and principal investigators were the University of Florida (UFLORIDA) (G. J. King) and the University of North Carolina Chapel Hill (UNC) (W. R. Proffit).

A measure of the efficacy of a procedure answers the question, Does it work? This is usually determined by randomized controlled trials, using carefully chosen patient populations treated under ideal conditions. In contrast, effectiveness is a measure of how well an efficacious procedure works in the field. The pool of patients seen by clinicians is in general far more diverse than that in controlled study groups, showing a larger range in condition severity, patient compliance, and other possibly confounding factors. Effectiveness then is generally estimated by retrospective analysis of a number of individual efficacy studies.

A major difficulty in efficacy trial design is the identification of useful outcome measures. Antezak-Bouckoms[1] proposes that both short-term and long-term measures be used. Short-term outcomes are immediate treatment results—for instance, specific occlusal or skeletal relationships (or both). In contrast, long-term outcomes might include why patients seek and clinicians recommend particular treatments, incidence of temporomandibular joint dysfunction, facial esthetics, improved masticatory function, degree of root resorption, or long-term treatment stability. The careful reader of any clinical study needs to understand what the outcome measure reported is, its normal range, and whether it is appropriate for the patients under his or her care.

This article begins by discussing how the NIDCR efficacy trials differed from previous clinical studies. We then detail the methods, materials, and outcome measures used. Finally, results from each of the two studies are summarized. The investigators of these clinical trials have themselves collaborated on a review article.[2] But that work reports on only the craniofacial growth effects of headgear and bionator—not their efficacy. The information presented here then is a summary of the preliminary and final reports published thus far by each of the study groups on this particular topic.

NIDCR-FUNDED EFFICACY TRIALS OF FUNCTIONAL APPLIANCES Return to TOC

The NIDCR efficacy trials began with one group publishing a retrospective, literature study that identified problems in the reasoning and procedures used in previous clinical trials of functional appliances and recommended less biased alternatives.[3,4] The question to be answered was whether the experimental methodology used in these earlier studies biased the reported results. In other words, Should the reader question the accuracy of the published conclusions? To determine this, the authors reviewed all clinical reports of Class II treatment with functional appliances published between 1980 and 1987 in four peer-reviewed journals—the American Journal of Orthodontics, Angle Orthodontist, British Journal of Orthodontics, and the European Journal of Orthodontics. Particularly, to be included in the review, clinical studies should have involved more than 10 patients. Each article was evaluated by several reviewers by looking for possible errors and biases in the methodologies used. A summary of their findings is given below, with each section beginning with the categorical question raised by the reviewers.

How were patients assigned to treatment groups? Prior studies were found to assign patients to groups by (1) clinician's beliefs and preferences, (2) random allocation, and (3) nonrandom methods such as choosing consecutively enrolled patients. Most variables affecting orthodontic treatment outcome are unknown, and random assignment of subjects to study groups reduces their effect statistically. Although time consuming and labor intensive, this research design minimizes both known and unknown biases.

Moreover, prospective, randomized controlled trials are routinely required in medicine to determine the efficacy of competing therapies.

How was the control or comparison population selected? The comparison group was found to be (1) a historic population-based data set such as the Michigan or Burlington growth studies, (2) an untreated population either concurrent or previously observed, (3) patients treated with an alternative orthodontic appliance, or (4) absent—no control group.

The authors recommend the use of a concurrent, untreated population for comparison. It was argued that contemporary controls need to be used because of secular trends. The term “secular trends” describes the changes in growth and development of successive generations of people living in the same region. Some observed populations show increased average heights and weights with each generation. Also, the onset of puberty can occur at progressively younger ages. But secular trends in several European populations have stabilized, and in fact, some non-European populations have even shown reductions in body size and delayed sexual maturation across generations. Thus, the concern appears valid for treatment studies on localized preadolescent and adolescent populations.

Was the clinical trial prospective or retrospective? Prospective studies enroll the patients at the start of the study, and follow them up to determine the treatment outcome. In contrast, retrospective studies analyze the outcomes of patients already treated. A common criticism of retrospective studies is that they are biased when only those patients completing treatment favorably are included. But this concern can be resolved by randomly selecting a study group from among all the patients treated—including favorable and unfavorable outcomes.

Were patient gender, severity of initial condition, and maturational development status considered in the selection process? It is important that these confounding variables be noted and controlled when grouping patients into study groups. Certainly, gender and initial condition are definable. There are however many measures of skeletal and dental maturity—often tedious and having reproducibility issues. At present, there is no agreement on a measure of maturational age and its applicability to orthodontic treatment.

Were patients treated by only one clinician or by several? Differences in clinician ability, patient management skills, and experience with the study appliances are critical variables to consider when designing orthodontic outcome trials—a proficiency bias. To address this, it was proposed that patients be randomly rotated among the clinical staff at each appointment.

Alternatively, for moderate sample sizes, one experienced practitioner might treat all the patients.

How many patients were studied? Was the statistical test used appropriate for the sample size and number of variables? Pilot studies, using recorded nominal values of the outcome metrics data, should be used to determine required sample sizes when planning a study. Also, the calculated minimal detectable difference for all outcome variables should be larger than the instrumentation error.

This literature review concluded that small sample sizes, weak research designs, and ambiguous and incomplete information made it difficult to agree with the conclusions of most of the articles studied. It was recommended that prospective, randomized clinical trials with parallel untreated control groups be followed in future efficacy studies of orthodontic treatments.

But simply assigning patients randomly to treatment groups is only one aspect of a well-designed clinical study. Sackett[5] proposes four other characteristics. These are (1) are all pretreatment patients recognizably the same at the start? (2) are all clinically relevant outcomes reported? (3) are the patients reported equivalent to those in one's own clinic—age, gender, skeletal age, dental age, medical history, and severity of malocclusion? and (4) are all patients entering the study accounted for at the end?

UFLORIDA Clinical Studies[6–8]

A prospective, randomized controlled trial was completed to evaluate the efficacy of both bionator and combined headgear/biteplane treatment in correcting Class II malocclusions.

From a total of 6428 clinical screenings, 1207 children were found eligible for participation. The selection criteria used were (1) chronologic age of 9–10 years, (2) first permanent molars fully erupted, (3) emergence of not more than three permanent cuspids or premolars, (4) a positive overbite and overjet, and (5) an average bilateral Class II molar relationship of at least one-half cusp. Molar classification was scored in 1/4 cusp increments. In asymmetric cases where one molar relation was less than a half-cusp and the other side larger, the average of the cusp differences for two sides was used to accept or reject subjects.

The study protocol was explained to parents of the 1207 candidates. Patients would be randomly assigned to one of three groups (1) untreated or observation only, (2) headgear with anterior biteplane appliance, or (3) bionator functional appliance. Depending on mandibular plane angle (SN-GoGn), the headgear appliance used could be either cervical pull (SN-GoGn $\leq$ 40 degrees) or high pull (SN-GoGn > 40 degrees). In addition, a maxillary biteplane with labial bow and molar circumferential wires would be worn. Its purpose was to disclude the posterior teeth. Patients were required to wear the headgear for at least 14 hours each day and the biteplane continuously—except while cleaning, eating, or playing contact sports. Bionators would be constructed in the same way as Balters' device was, protruding the mandible to a Class I relation with minimal vertical opening. Mandibular occlusal acrylic was preferentially removed during treatment to permit lower molar eruption. It was required that bionator patients follow the same routine as the headgear patients, except that the appliance was to be worn for 22 hours each day rather than for 14 hours.

Informed consent for the trial was signed by 360 parents (29%), and initial records were taken. A prestudy review dismissed 29 patients from the study because of screening errors, and an additional six patients withdrew before treatment started. The average age of the population was 10 years.

The remaining 325 children were stratified into three selection groups and randomly assigned to one of the three treatment paths. One selection group was the severity of the Class II molar relationship. This variable ranged from mild (bilateral 1/2 cusp), moderate (one side > 1/2 cusp), to severe (bilateral full cusp).

A second grouping was the mandibular plane angle (SN-GoGn). This variable ranged from low (<30 degrees), normal (30–40 degrees), to high (>40 degrees). The third criterion was whether prestudy orthodontic treatment was required to correct posterior crossbite, incisor eruption, and oral habits or to alter maxillary incisor angulation.

Investigators would later include gender, ethnicity, and age as prerandomization selection criteria. Forty-nine patients would leave the study after initial records leaving 276 starting treatments. Of these, 248 successfully completed the study—control (n = 79), headgear (n = 90), and bionator (n = 79). The initial mean Class II molar discrepancy for each group was three-fourths cusp bilaterally.

Active treatment ended when either the Class II molar relation was corrected or at most 3 years had elapsed. This was followed by a 12-month retention period or follow-up period (or both). During the first 6 months, half of the treated patients were randomly instructed to continue wearing their appliances overnight (about 10 h/d). The remaining half discontinued appliance wear completely. Appliance wear ended for all patients at the end of this period; final records were taken at least 6 months later.

When using bilateral Class II molar correction as an outcome measure, the untreated observation group showed an 8% success rate, the bionator group 44%, and the headgear group 62% immediately after treatment.

After the 12-month follow-up period, 13% of untreated patients had self-corrected, and 38% of bionator-treated and 50% of headgear-treated patients had maintained their Class 1 molar relations.

Limits to each appliance's corrective range were also observed by the investigators. Bionator patients requiring half cusp correction showed a 57% success rate but those requiring greater correction showed only a 30% success rate. By comparison, combined headgear/biteplane treatment successfully corrected 70% of patients requiring half cusp correction and 40% of those presenting with discrepancies greater than a half cusp. Finally, untreated patients with half cusp molar relations showed correction 30% of the time, and

only 5% of patients with greater discrepancies showed complete correction.

UNC Clinical Studies[9–11]

A prospective, randomized controlled trial was conducted to evaluate the effectiveness of headgear and bionator activator appliances to correct Class II malocclusions. The headgear face-bow was attached to upper first molar tubes and anchored with combined headcap and neck straps. The functional appliance was a modified Balters' bionator. On average, this appliance protruded the mandible 4–6 mm and increased the vertical dimension minimally.

The outcome measure used in the trial was the annualized ANB difference—the rate of change in ANB (degrees/year). This metric was defined because the bionator and headgear appliances were to be used for only 15 months. After that, a phase of fixed wire appliance treatment would begin.

Using published nominal values, the authors decided that 1.15 degrees was the minimally detectable ANB difference to expect. Assuming this mean difference, a significance level of 0.01, and a power of 0.90, it was calculated that at least 40 patients would be required in each of the three study groups.

To collect this required sample set, the UNC clinicians screened 2164 patients. Of these, 207 met the trial's admission criteria. These selection criteria were (1) overjet > 7 mm, (2) all incisors erupted, (3) no second molars erupted, and (4) a prepubertal skeletal age. The maturational age of patients was determined by Bowden's analysis of hand-wrist radiographs. This technique maps observed radiographic ossification patterns onto a normalized skeletal maturity growth curve. On this standard curve, the start, peak, and end of the adolescent growth spurt are marked. Only patients having maturational ages at least 1 year before the peak height velocity on the Bowden curve were considered candidates. Also, a panoramic oral radiograph was taken to confirm that no permanent teeth were absent. Of the initial 207 patient candidates, 192 remained after applying these exclusionary criteria.

The average age of this group was 9.8 years.

From this population, 175 agreed to enter the study. This sample set was separated into male and female groups. Next, the two groups were blocked into sets of six, and patients were randomly assigned from each set of six to one of the three study groups—control, bionator, or straight-pull headgear.

The average age and gender were statistically the same in all groups. Also, all groups had equivalent skeletal and dental features. This was determined by statistical tests for mean differences across eight cephalometric variables. These particular measures were (1) overjet (mm), (2) overbite (mm), (3) ANB, (4) Witts (mm), (5) Sn-GoGn, (6) anterior/posterior facial height (mm), (7) maxillary length (mm), and (8) mandibular length (mm).

A total of 166 patients completed the trial—control (n = 61), headgear (n = 52), and bionator (n = 53). Nine patients were dismissed from the study for the following reasons: (1) initial screening errors, (2) never returning after initial records, (3) moving, or (4) not cooperating with clinicians. The study period ended for all patients after 15 months of therapy.

The annualized ANB change outcome was categorized into four levels (1) a highly favorable change (>1.5 degrees/y decrease), (2) a favorable change (1.5–0.5 degrees/y decrease), (3) no change (±0.5 degrees/y), and (4) an unfavorable change (>0.5 degrees/y increase).

The UNC study concludes that when left untreated, Class II patients will show favorable ANB changes in 30% of the cases and highly favorable changes in only 2%. Patients treated with straight-pull headgear can expect favorable ANB changes in 50% of cases and highly favorable changes in 30%. Finally, bionator-treated patients will display favorable ANB changes in 55% of cases and highly favorable changes in 20%.

DISCUSSION AND CONCLUSIONS Return to TOC

A prospective, randomized controlled study at UFLORIDA on 248 children showed that bionator and combined headgear/biteplane treatment corrected Class II molar relations in 38% and 50% of patients, respectively. By comparison, untreated patients showed molar correction in only 13% of patients.

The UNC prospective, randomized controlled study on 166 children concluded that when left untreated, Class II patients will show favorable ANB changes in 32% of cases. A favorable ANB change was defined as a reduction in ANB of greater than 0.5 degrees/y. Patients treated with either bionators or straight-pull headgear can expect similar favorable ANB changes in 75% and 80% of cases, respectively.

The UFLORIDA group presents a well-designed and illuminating randomized controlled trial. The UNC study group uses an outcome measure (annualized ANB difference) that is difficult to use in clinical work. On first reading, the UNC study might suggest that bionators

and headgear are equally effective in Class II treatment. But using Class II molar correction as an outcome measure, the UFLORIDA study shows this to be false. Bionator treatment appears to be about 75% as effective as combined headgear/biteplane treatment in correcting Class II molar relationships.

A multivariant analysis of headgear- and bionator-treated patients in the UFLORIDA study showed no effect of gender, initial mandibular plane angle, chronologic age, or skeletal age on success rate.[8] But patient compliance correlated positively with favorable outcome. Success rate also decreased with increasing malocclusion severity.

For several reasons, the usefulness of these two randomized controlled trials in forming clinical guidelines is severely limited. Although the NIDCR should be praised for initiating the studies, it should also be faulted for not demanding collaboration between groups as a condition for funding. This has resulted in the collection of data by two groups using methods too dissimilar to allow meaningful data synthesis in regard to common measures of clinical efficacy.

Because sample sizes cannot be combined, the conclusions here are based on at most 90 treated patients for each appliance, which is a rather small number.

We believe that the preferred research path to follow is to agree on common methods beforehand—not one but many—and gather a larger sample size. This could be accomplished by either a collaborative effort between academic centers or, most economically, a consortium of funded private practices.

REFERENCES Return to TOC

1. Antezak-Bouckoms A. The anatomy of clinical research. *Clin Orthod Res.* 1998; 1:75–79. [PubMed Citation]

2. Ghafari J, King GJ, Tulloch JFC. Early treatment of Class II, division 1 malocclusion—comparison of alterative treatment modalities. *Clin Orthod Res.* 1998; 1:107–117. [PubMed Citation]

3. Tulloch JFC, Medland W, Tuncay OC. Methods used to evaluate growth modification in Class II malocclusion. *Am J Orthod Dentofacial Orthop.* 1990; 98:340–347. [PubMed Citation]

4. Tulloch JFC. Bias and variability in clinical research. *Clin Orthod Res.* 1998; 1:94–96. [PubMed Citation]

5. Sackett DL. On identifying the best therapy. In: Trotman CA, McNamara JA, eds. *Orthodontic Treatment: Outcome and Effectiveness.* Craniofacial Growth Series Vol. 30. Ann Arbor, Mich: Center for Human Growth & Development, The University of Michigan; 1994:7–19.

6. Keeling SD, King GJ, Wheeler TT, McGorray S. Timing of Class II treatment: rationale, methods, and early results of an ongoing randomized clinical trial. In: Trotman CA, McNamara JA, eds. *Orthodontic Treatment: Outcome and Effectiveness.* Craniofacial Growth Series Vol. 30. Ann Arbor, Mich: Center for Human Growth & Development, The University of Michigan; 1994:81–112.

7. Keeling SD, Wheeler TT, King GJ, Garvan CW, Cohen DA, Cabass S, McGorray SP, Taylor MG. Anteroposterior skeletal and dental changes after early Class II treatment with bionators and headgear. *Am J Orthod Dentofacial Orthop.* 1998; 113:40–50. [PubMed Citation]

8. Wheeler TT, McGorray SP, Dolce C, Taylor MG, King GJ. Effectiveness of early treatment of Class II malocclusion. *Am J Orthod Dentofacial Orthop.* 2002; 121:9–17. [PubMed Citation]

9. Tulloch JFC, Phillips C, Proffit WR. Early vs. late treatment of Class II malocclusion: preliminary results from the UNC clinical trial. In: Trotman CA, McNamara JA, eds. *Orthodontic Treatment: Outcome and Effectiveness.* Craniofacial Growth Series Vol. 30. Ann Arbor, Mich: Center for Human Growth & Development, The University of Michigan; 1994:113–138.

10. Tulloch JFC, Philips C, Koch G, Proffit WR. The effect of early intervention on skeletal pattern in Class II malocclusion: a randomized clinical trial. *Am J Orthod Dentofacial Orthop.* 1997; 111:391–400. [PubMed Citation]

11. Tulloch JFC, Profitt WR, Phillips C. Influences on the outcome of early treatment for Class II malocclusion. *Am J Orthod Dentofacial Orthop.* 1997; 111:533–542. [PubMed Citation]

[a]Cicero, Ill

[b]Faculty of Dentistry, Mahidol University, Bangkok, Thailand

Corresponding author: Thomas Jacobs, DDS, MS, 1908 South 50th Avenue, Cicero, IL 60804 (E-mail: tjacobs1@mindspring.com)