# EVIDENCE AND CLINICAL DECISIONS: Asking the Right Questions to Obtain Clinically Useful Answers

**William R. Proffit, DDS, PhD**
Kenan Distinguished Professor, Dept. of Orthodontics, Univ. of North Carolina School of Dentistry

## Abstract

Orthodontists need to know the effectiveness, efficiency and predictability of treatment approaches and methods, which can be learned only by carefully studying and evaluating treatment outcomes. The best data for outcomes come from randomized clinical trials (RCTs), but retrospective data can provide satisfactory evidence if the subjects were a well-defined patient group, all the patients were accounted for, and the percentages of patients with various possible outcomes are presented along with measures of the central tendency and variation. Meta-analysis of multiple RCTs done in a similar way and systematic reviews of the literature can strengthen clinically-useful evidence, but reviews that are too broadly based are more likely to blur than clarify the information clinicians need. Reviews that are tightly focused on seeking the answer to specific clinical questions and evaluating the quality of the evidence available to answer the question are much more likely to provide clinically useful data.

An orthodontist, like all health care providers, wants to know three things about the treatment he or she is providing: its

1. effectiveness (how well it works, i.e., how effective it is in dealing with the patient's problems, taking into account possible negative side effects),

2. efficiency (how cost-effective it is, with cost in its broader sense to include time and effort for the provider and impact on the patient), and

3. predictability (the amount of variation in patient response).

These things can be learned in only one way, by carefully evaluating treatment outcomes. The hierarchy of quality in the evidence for clinical outcomes in orthodontics is shown in Figure 1.[1] This is similar to but differs in two ways from the classic diagram promulgated by the Cochrane Collaboration as the ideal hierarchy for biomedical studies.

First, not all orthodontic conditions can be evaluated with randomized clinical trials (RCT's), so retrospective studies of treatment outcomes are, and will continue to be, important sources of the evidence that underlies our clinical treatment. Although prospective studies without random assignment of patients are sometimes viewed as better than retrospective studies, a well-conducted retrospective study can provide equally reliable information. Some important questions are impossible to answer with RCT's because of

Dr. William R. Proffit, Dept. of Orthodontics, UNC School of Dentistry, Chapel Hill, NC 27599-7450, phone 919-966-4428, FAX 919-843-8864, William_Proffit@dentistry.unc.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ethical considerations, with the limits of orthodontic camouflage versus orthognathic surgery being the best example. It is unacceptable to randomly assign patients to surgery. Many other questions that theoretically could be evaluated in RCT's cannot be answered that way because of the slow pace of orthodontic treatment and the long follow-up required to be sure of the stability of treatment outcomes. It simply costs too much and takes too long, especially when adequate evidence can be obtained from well-conducted retrospective studies.

Second, although it is correct that meta-analysis, combining the data from well-done clinical trials, can strengthen (or weaken) the conclusion from those studies, it is critically important that the data were obtained in a comparable way. This becomes a bigger problem when a "systematic review" of the literature looks beyond clinical trials and incorporates a large number of retrospective reports on a broad topic in an attempt to define guidelines for a broad range of clinical problems. Reviews of that type are more likely to blur than clarify the evidence that clinicians need.

The purpose of this paper is to further illustrate the problems we currently have in evaluating clinical outcome data, and discuss ways to strengthen the evidence from retrospective studies and reduce confusion from overly-extensive reviews of the literature.

## What Does It Take to Obtain Good Retrospective Data?

There are three key criteria for good retrospective (or prospective) data:

1. The subjects were a well-defined patient group, who were selected by pre-treatment characteristics and received specific treatment—not, for example, all the Class II patients treated during a defined time period with a variety of methods. This is a place where the inadequacies of the Angle classification particularly have an effect. There are, of course, multiple types of Class I, II and III malocclusion, and clear conclusions require distinction by facial as well as dental types and consideration of all three planes of space.

   There are far too many bad examples in the orthodontic literature of misleading conclusions due to poorly-selected or biased samples and/or an attempt to answer too broad a question. A good example is a recent study to answer the appropriately focused question, "Are skeletal changes and an improved growth pattern obtained in growing patients treated with a combination of bionator and high pull headgear?". This combination has been considered the most effective form of treatment for long face problems, despite weak evidence from case reports and small samples. An evaluation of records of 24 consecutively-treated patients with this combination were compared to records of untreated patients selected for similar age, gender, vertical skeletal relationships and time intervals between records. The conclusion was surprisingly strong but justified: no long-term skeletal changes were obtained and "Our findings suggest that treatment with bionator and high pull headgear is not recommended for growing patients with hyperdivergent facial patterns when the goal is to decrease the vertical dimension of the face."[2]

2. All the patients, not just the ones judged to represent a successful outcome, are accounted for in the report. Even if care is taken to avoid bias in selecting patients for study, it is almost impossible to find all the post-treatment and/or follow-up records (radiographs, photographs, dental casts) for all the patients eligible for inclusion because of their pretreatment characteristics. When that is the case, an important question is whether the missing patients are systematically different from the ones remaining in the sample. That must be considered and examined to the extent that the data set allows. It is easy to check the age and gender distribution in

the initial and follow-up sample to see if it changed, and often there are other known characteristics of the missing patients that can be compared in the same way.

3. The statistical design and methodology is appropriate. Sample size and the distribution of treatment changes are critical variables. The larger the sample size, the more precise the statistical evaluation can be. The sample size needed to detect a difference can (and should) be calculated in advance. A general guideline for orthodontic outcome studies is that the sample size should be at least 20-25 to have a reasonable chance of detecting statistical significance for clinically relevant changes. 30-40 usually is large enough. Rarely does the sample need to be more than 50, because small changes detected in a large sample, even if statistically significant, are likely to be insignificant clinically. It is important to keep in mind that in groups of treated patients, a few individuals usually have most of the variation—so statistics based on the normal distribution can be misleading and should not be used without verifying the distribution within the sample. Non-parametric statistics often are required.

It also is important to look at the way the data are presented. This should not be just a table of mean ± std. deviation. A box plot that shows the central tendency, estimate of deviation and range within the sample gives a far better perception of the variability within the sample (Figure 2) and guards against the frequent assumption that a change like the mean of the sample represents what should be expected from future patients. In the presence of variable responses, what the clinician really needs to know is the chance of a significant or highly significant improvement for a new patient of the same type who will receive the same treatment. For clinicians, it can be very helpful to be able to say, to yourself and the patient or parent, that this approach has been shown to have a [definite percentage] chance of success (Figure 3).

## Considerations in Evaluating Clinical Trial Outcomes

The same guidelines as those for retrospective studies, of course, also apply to RCTs, but are built into the fabric of the RCT research design. The great advantage of a randomized clinical trial is that confounding factors are equalized during the randomization process, and that it allows the evaluation of influences on pre- and post-treatment differences that had not even been identified before the study began. As long as something is normally distributed within both groups, its effect can be calculated. That does not mean that all clinical trial results truly represent reality. A <5% probability that the result is erroneous means that about one in 20 such studies will be wrong, and multiple comparisons increase the chance of error.

The best outcome data orthodontists have ever had, and perhaps the best data we are likely to obtain, come from the three major trials of early (preadolescent) versus later (adolescent) treatment of excessive overjet.[3,4,5] The research plans for the three trials were not identical, but they are close enough to allow comparison of the findings. All three trials reported the same thing: a small but statistically significant difference in mean a-p growth between children who had a phase of preadolescent treatment and controls who did not, which disappeared during adolescent treatment for all these patients. About 75% of the patients improved during early treatment, the other 25% did not, and cooperation did not account for all the difference.

The RCT results have been widely challenged by clinicians who offer some version of "If you had done it properly [my way], it would have worked" and "You're denying treatment to children who really need it". The current Cochrane Review of orthodontic treatment is

based on eight RCTs, including the three major ones discussed above, and concludes "The evidence suggests that providing early orthodontic treatment for children with prominent upper front teeth is no more effective than providing one course of orthodontic treatment when the child is in early adolescence."[6] The chance that multiple similar trials all got it wrong is not zero, but it is vanishingly small. It is now nearly ten years after the early Class II outcomes were published, and their gradually increasing acceptance leads one to hope for broad acceptance within the next decade.

That does not mean that we now know everything about the timing of Class II treatment. The importance of evaluating all the salient characteristics of a patient's problems has already been noted, and we do not have extensive clinical trial data for treatment of combined Class II and vertical problems. The recent retrospective data for long-face Class II treatment already has been noted.[2] Some short face Class II patients have an impinging overbite and trauma to the palatal tissues, which may be a valid reason for early treatment (I think it is)—but good data to document this simply do not exist.

What should a clinician do for a child with this problem? The answer has to be "Use your best clinical judgment", based on whatever evidence is available coupled with your own clinical experience and that of your teachers. In a broad overview of clinical practice, many treatment decisions must be made in the absence of good evidence. Perhaps a reasonable guideline is that substituting a new and complex treatment method for an older and simpler one, in the absence of compelling evidence, rarely is wise. The important consideration, of course, is the cost-benefit ratio—and sometimes an unproven new method is chosen in dogged pursuit of a level of perfection that may be more of a benefit for the orthodontist than the patient.

Another way to look at what can be learned from the clinical trials is what you can find out by asking specific clinically-focused questions. For example, should an orthodontist expect a shorter treatment time and better result for a mild Class II, and the reverse for a severe Class II? If that is correct, it probably would affect the choice of treatment method and estimates of treatment time. The UNC clinical trial data indicate, surprisingly, that the severity of the initial jaw discrepancy is not a predictor of either post-treatment occlusion or treatment time (Figure 4). Although that has been published,[3] it wasn't highlighted in a paper that largely focused on two phase vs. one phase treatment—and many clinicians still think that initial severity is an important predictor of time and outcome. The existing clinical trial data can and should be used to search for answers to other clinically-focused questions.

## Meta-analyses and Systematic Reviews: Current Problems

Meta-analysis was introduced as a way of combining data from multiple similar randomized clinical trials to obtain a more powerful analysis. As a way to strengthen the conclusions, this is appropriate and valuable—if the clinical trials really were focused on answering the same question or questions, as the orthodontic Class II trials were. Systematic analysis of the literature (systematic review) is an extension of the meta-analysis idea beyond clinical trials, and sometimes a systematic review is called meta-analysis. Whatever it is called, broadening the review to include retrospective reports requires great care in selection of the studies so that they really are similar—and considering retrospective reports as equal to clinical trials rarely is warranted.

At present, some meta-analyses of orthodontic clinical trials, and many if not most systematic reviews of the orthodontic literature, conclude only that the data are weak and further study is needed. Particularly in systematic reviews, a major part of the weakness often is a lack of specifically focused clinical questions and evaluation of the quality of the

currently available data to answer those questions. Often there is more extensive information about the review process than about what the findings might mean in clinical practice.

A recent systematic review of the effects of orthognathic surgery on the oropharyngeal airway nicely illustrates these problems.[7] Its primary goal was to evaluate the extent to which orthognathic surgery (primarily mandibular setback) could predispose patients to sleep apnea or assist in treating it (maxillary, mandibular and/or chin advancement). This was to be judged by studies of the treatment effect on airway dimensions. An extensive (and extensively described) search of publication data bases yielded 59 full articles that were retrieved for review; of these, 22 were included in the study because they met quality criteria based on the way the study was carried out and reported. So far, so good—but only 4 of the 22 studies had 3D data, and the airway is an irregularly shaped three-dimensional space that is almost impossible to evaluate accurately from the 2D cephalometric radiographs used in the other 18 studies. Further, the extent to which the shape of the airway in upright and alert individuals reflects its dimensions during sleep in a prone position is not known, but it would be remarkable if there were no differences. The conclusion: "Three-dimensional studies are important in the near future, … as evidence is lacking on the volume changes after orthognathic surgery." That does not provide any clinically useful information, and leaves one to wonder about the indirect approach to sleep apnea in the first place. To evaluate the effect of orthognathic surgery on sleep apnea, it would make more sense to observe patients in a sleep lab before and after surgery.

The review of the literature that Joondeph did in preparation for his Angle Lecture at the 2012 AAO meeting offers an interesting contrast to the "big picture" systematic review.[8] He focused on the transverse changes associated with various types of orthodontic and surgical treatment, identified five alternative ways to correct a transverse discrepancy, and asked specific questions related to the type of treatment. The full report has not yet been published, but a few examples of specific clinically important questions, and equally specific answers, illustrate the difference with this approach.

Clinically-focused question 1: "Does it matter if your maxillary expansion appliance is attached to bone screws or banded teeth"? The answer from a recent clinical trial: "Both expanders showed similar results, and dental expansion was greater than skeletal expansion in both groups".[9] Question 2: "What is the relationship between transverse expansion in the maxilla and arch perimeter increase?" The answer, from a retrospective study of 21 consecutive palatal expansion cases: "The average perimeter gain was 0.7 times the amount of transverse gain across the maxillary molars, but varied between 0.5 and 0.8".[10] Question 3: "Does that ratio apply to mandibular transverse expansion (where there is no skeletal component, just tooth movement)"? The answer, from a CT modeling study and therefore derived quite indirectly: "Lateral expansion of the mandibular molars results in perimeter gain of 0.3 times the amount of transverse increase".[11] Question 4: "Does transverse expansion of the maxilla improve the amount of maxillary protraction with early facemask therapy?" The answer, from both a clinical trial[12] and a retrospective study[13]: "Changes were the same when using facemask therapy with or without expansion." A reasonable conclusion would be that the answers from the two RCTs and the answer from the consecutively treated patients are credible, and that the answer from only a simulation study (the arch perimeter gain from mandibular molar expansion) is rather dubious without clinical confirmation.

Note the difference from a more typical systematic review: the focus is on the specific aspects of treatment and the quality of the evidence that is available to answer specific questions, not on a broad evaluation of the quality of evidence more generally. Should clinicians care exactly how many papers exist with transverse expansion as their topic, and

how many would be considered worth reviewing on the basis of screening criteria? Not if all that detail gets in the way of answering questions that would directly influence clinical practice. Nevertheless, it must be kept in mind that expert opinion has been used in a report of this type to select the best evidence from the entire body of literature that deals with the topic.

## Conclusions

It is unfortunate that even evidence from multiple RCTs, as for two-phase versus one-phase Class II treatment, still takes so long to really influence clinical practice. RCT data now exist for the effect of maxillary expansion on protraction—simultaneous expansion does not increase protraction[12]—but this has yet to decrease enthusiasm for combining the two procedures, whether or not expansion is needed. Perhaps more clinically-focused presentations of the available evidence and consideration of its quality in direct relationship to answering clinical questions can facilitate the transition to evidence-based practice.

There is a long history in orthodontics of new ideas and methods that were adopted enthusiastically and applied indiscriminately at first, well in advance of adequate evaluation. As a result, within a decade the idea or method was largely discredited and discarded—so now it is rarely used even in the situations where it was eventually shown to be a significant advance. Two examples are sectioning gingival fibers to improve stability after correcting rotations and distraction osteogenesis as a way to correct jaw discrepancies. Both are effective, efficient and predictable in the right circumstances but now are rarely used even when they are indicated.

At present, enthusiastic and extensive reviews of the literature that really do not help clinicians move toward evidence-based practice are common. This is not because the method is bad, but because it does not provide clinically useful answers if the questions are not clinically focused and if the scope of the review goes outside comparable studies. As we work toward evidence-based treatment, the goal should be to ask the right specific questions and obtain answers that do have clinical utility. Otherwise, the evidence-based approach risks being discredited and discarded as so many other new ideas have been.

## Acknowledgments

## References

1. Proffit, WR.; Fields, HW.; Sarver, DM. Contemporary Orthodontics. 5. St Louis: Elsevier; 2013.

2. Freeman CS, McNamara JA, Baccetti L, et al. Treatment effects of the bionator and high-pull facebow combinations followed by fixed appliances in patients with increased vertical dimension. Am J Orthod Dentofac Orthop. 2007; 131:184–95.

3. Tulloch JFC, Proffit WR, Phillips C. Permanent dentition outcomes in a two-phase randomized clinical trial of early Class II treatment. Am J Orthod Dentofac Orthop. 2004; 125:657–67.

4. King GJ, McCorray SP, Wheeler TT, et al. Comparison of peer assessment ratings (PAR) from 1-phase and 2-phase treatment protocols for Class II malocclusion. Am J Orthod Dentofac Orthop. 2003; 123:489–96.

5. O'Brien K, Wright J, Conboy F, et al. Early treatment for Class II division 1 malocclusion with the twin block appliance. Am J Orthod Dentofac Orthop. 2009; 135:573–79.

6. Harrison JE, O'Brien KD, Worthington HV. Orthodontic treatment for prominent front teeth in children. Cochrane Database of Systematic Reviews. 2007; (3) Art No. CD003452.

7. Mattos CT, Vilani GNL, Sant'Anna EF, et al. Effects of orthognathic surgery on oropharyngeal airway: a meta-analysis. Int J Oral Maxillofac Surg. 2011; 40:1347–56. [PubMed: 21782388]

8. Joondeph D. Traverse the transverse In preparation, 2013 publication anticipated.

9. Lagravere MO, Carey JP, Giseon H, et al. Transverse, vertical and anteroposterior changes from bone-anchored maxillary expansion vs traditional rapid maxillary expansion: a randomized clinical trial. Am J Orthod Dentofac Orthop. 2010; 137:304–05.

10. Adkins MD, Nanda RS, Currier GF. Arch perimeter changes on rapid palatal expansion. Am J Orthod Dentofac Orthop. 1990; 97:194–199.

11. Motoyoshi M, Hirabayashi M, Shimazaki T, Namura S. An experimental study on mandibular expansion: increases in arch width and perimenter. Eur J Orthod. 2002; 24:125–130. [PubMed: 12001548]

12. Vaughn GA, Mason B, Moon HB, Turley PK. The effects of maxillary protraction therapy with or without rapid palatal expansion: a randomized clinical trial. Am J Orthod Dentofac Orthop. 2005; 128:299–309.

13. Tortop T, Keykubat A, Yuksei S. Facemask therapy with and without expansion. Am J Orthod Dentofac Orthop. 2007; 132:467–74.

**Figure 1.**
The hierarchy of quality for evidence in decisions for orthodontic patients. This diagram differs from the more frequently-used hierarchy of the Cochrane Collaboration in two ways: (1) it shows a greater recognition of the importance of good retrospective (or non-random prospective) data because randomized clinical trials frequently are impossible or impractical, given the long duration of treatment and longer follow-up required to evaluate the outcomes, and (2) questions the importance of systematic reviews of the literature that, in contrast to well-conducted clinical trials, often are poorly focused and potentially misleading.
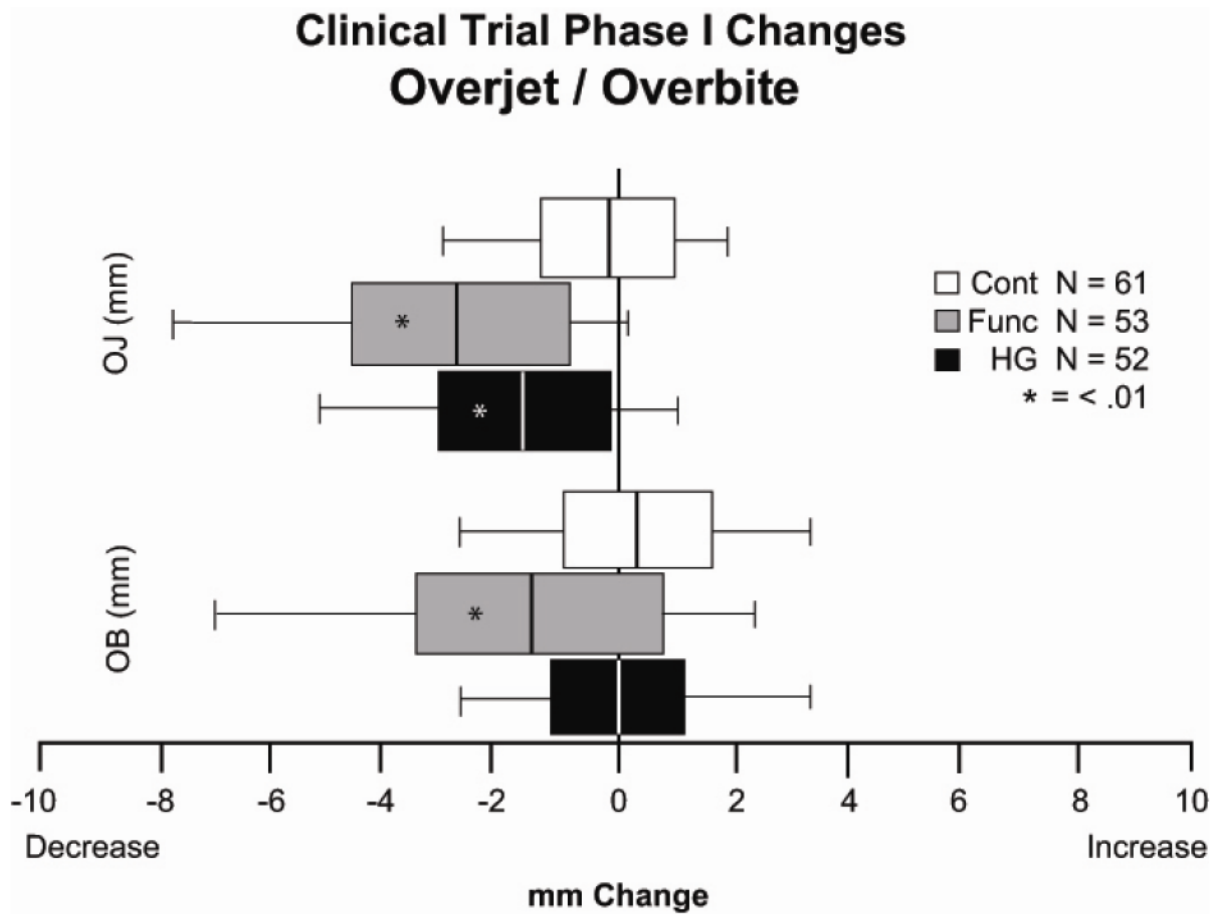
**Figure 2.**
Phase 1 changes in the UNC Class II RCT for overjet and overbite. A, Tabulated data with mean changes, standard deviation and range. B, The same data shown as box plots. Although the numbers for the variability within the sample are there in the table, the variability is shown much more clearly in the plots. Reports often are written with the implication that the mean outcome is what should be expected for future patients—which is incorrect and misleading, as the box plot shows.
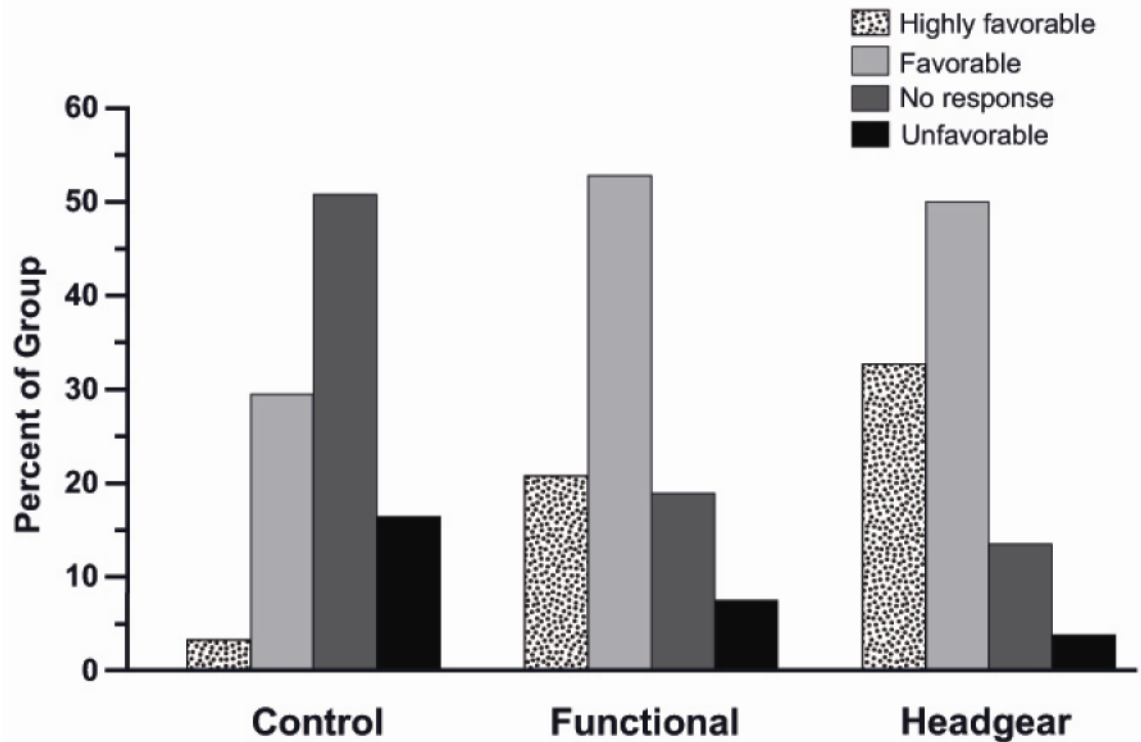
## ANB Change, Phase 1

**Figure 3.**
UNC Class II RCT data for the percentages of children with types of outcomes. It is true that phase 1 treatment produced a significantly different outcome from the controls—but equally true and more important that the chance of a favorable outcome was about 75% for both functional appliance and headgear treatment, and the chance of a highly favorable outcome (the result usually presented in case reports) was much lower.
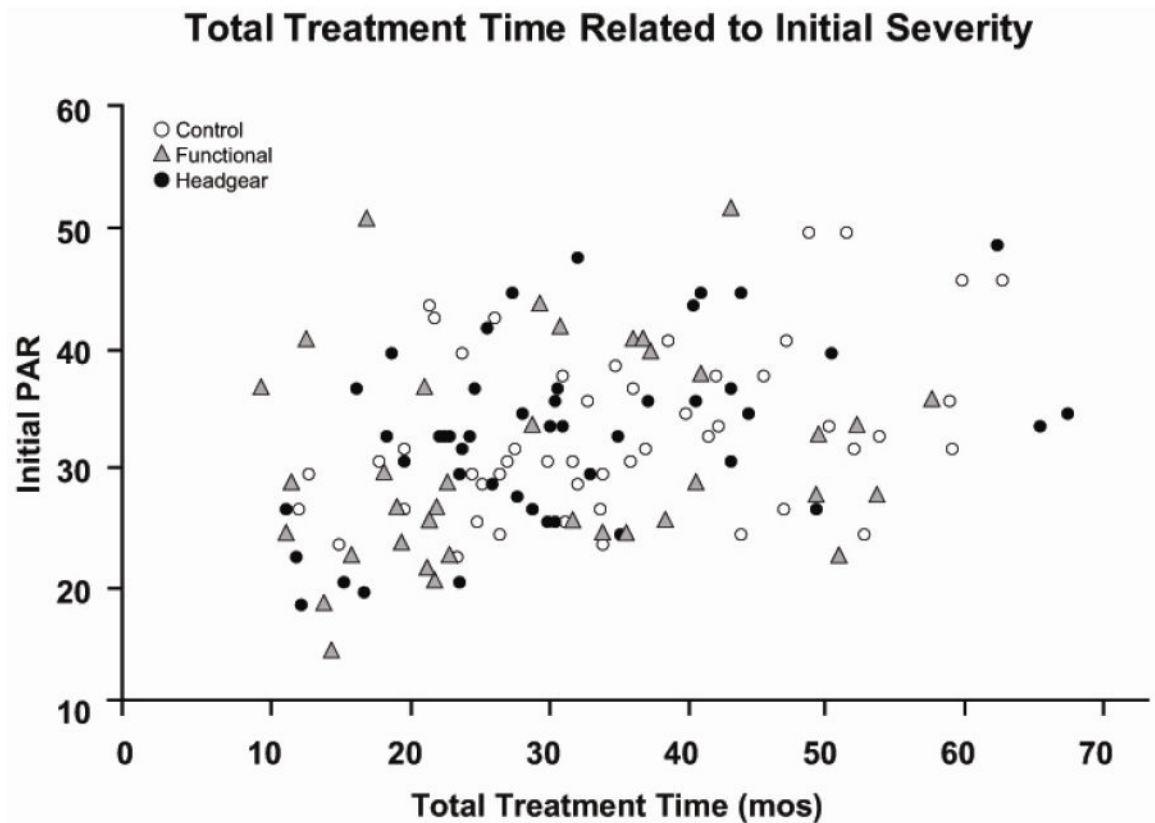
A



Final Outcome (PAR) Related to Initial Severity
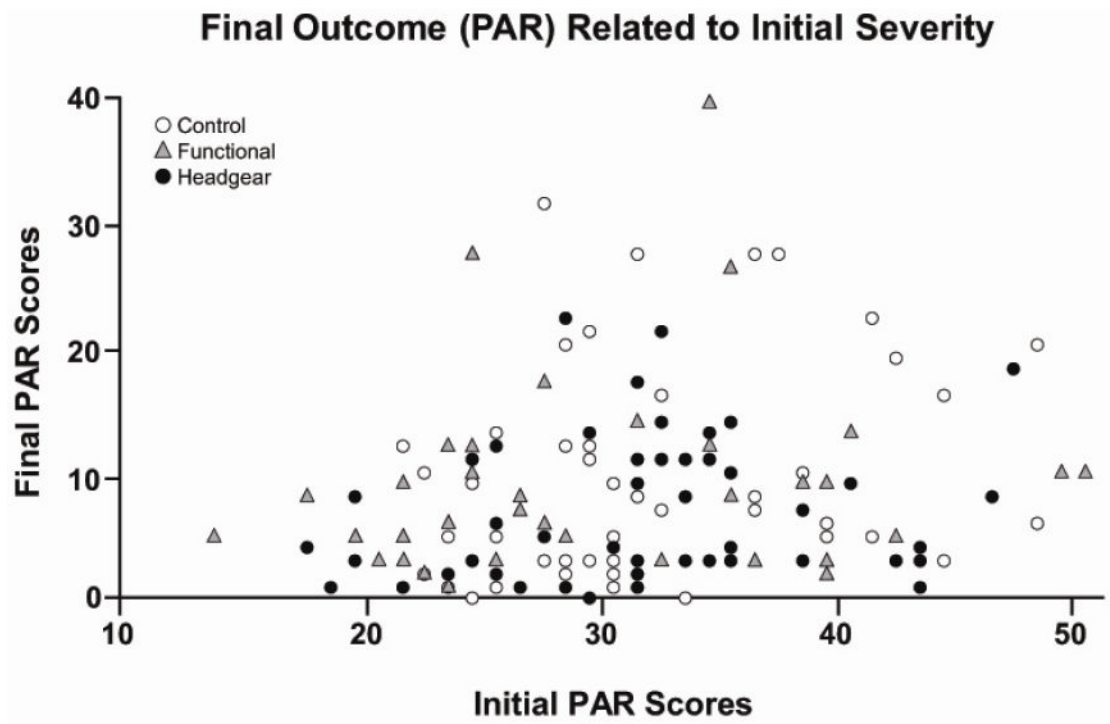
B



**Figure 4.**
The relationship between A, the initial PAR score (before any treatment) and the final PAR
score at the end of phase 2 treatment for all patients and B, total treatment time and initial
PAR score. The data (from the UNC clinical trial) show what many clinicians have observed
about treatment for Class II problems: what looks like a difficult case may respond well to
treatment, and what looks like an easy case may not respond well at all. Now we have
evidence to support that view.