The role of clinical research in orthodontics

Sheldon Baumrind DDS, MS

This paper* argues the need for rigorous clinical research in orthodontics and outlines some of the conceptual problems encountered in designing and conducting clinical orthodontic studies.

In the present context, the term "clinical research" designates and is limited to investigations of change through time in actual human beings — either treated patients or untreated control subjects. Such investigations are perforce complex and difficult. Why then should they be attempted? Would it not be more efficient for investigators to limit their studies to more easily controlled, simplified systems, such as rats or cats or dogs or monkeys — or even tissue cultures?

One cogent answer to these questions is contained in a statement on the adequacy of experimental models which is credited to Einstein: "Things should be kept as simple as possible," he said,"—but not simpler." If the experimental models one studies understate or misrepresent the complexity of the systems in which one is really interested, the answers one arrives at are likely to be simplistic rather than simple.

In any research study, clinical or basic, the problem of deciding whether the mean difference between two or more samples is real is confounded by variability — by what in clinical terms we call "individual differences". If there were no variation within the populations from which the samples are drawn, it would only be necessary to measure one member of each sample in order to decide which sample has the larger mean — because in that event the value for the single individual would be equal to the sample mean. Alternatively, when populations are characterized by substantial individual variation, the danger exists that the means of the groups of individuals which an investigator happens to sample may by chance differ markedly from the means of the populations from which they are drawn.

Two strategies may be used to facilitate the identification of mean differences in the presence of high variability. One is to increase sample size, (although as Ackerman[1] pointed out, this course of action may in some instances be effectively impossible). The other strategy is to simplify the study by structuring it in such a way as to artificially constrain and reduce variability. An investigator may, for example, study tooth movement in same-sex litter mate Sprague-Dawley rats, or study responses to tensile forces in tissue cultures of cloned fibroblasts. One can even disregard the biological component of the treatment equation entirely as is done, for example, when one studies the mechanical properties of different kinds of archwire in vitro on a biomaterials laboratory bench. Such investigations may have real value, but the use of simplified models always raises a question as to whether the results can properly be generalized to real life clinical situations.

*Adapted from an invited lecture presented to the Harvard Society for the Advancement of Orthodontics, June 8,1991.

By structuring the conditions of any experiment in either basic or clinical science, one can indeed reduce variability and thus make it easier to detect small differences between group means. But the more we constrain the experimental conditions, the more we reduce the domain to which the outcomes of studies can appropriately be generalized. If, for example, we are able to establish by a carefully conducted structured experiment that a certain mechanism of gene transcription occurs under certain closely specified conditions in a strain of E. coli, can we reasonably infer that the same mechanism operates under other conditions in mouse or man? Or, if experiments establish that a tissue culture composed of cells of fibroblastic origin responds in some certain way to stretching or to a change in oxygen concentration, can orthodontists reasonably infer that the same mechanisms operate when a closing loop arch is activated? Clearly not! Indeed it may be reasonable to say that the findings of structured experiments can never appropriately be generalized to real world experience without first conducting additional real world (i.e., clinical) investigations. To do otherwise would be to engage in blind leaps of faith concerning the representativeness of our experimental designs and models. And the more experimental conditions are controlled to reduce variability, the greater becomes the difficulty in applying experimental findings directly to real world, clinical situations.

An additional problem is that cogent arguments can be made concerning the ethics of conducting structured clinical experiments in the kinds of long-term therapeutic situations which interest orthodontists.[2,3,4] One telling argument is that since therapists have an absolute and transcendent obligation as professionals to deliver for each patient the treatment which they believe best for that patient, no subject can ethically be randomized to one of two possible treatments unless there is true uncertainty as to which of the two treatments is in the patient's best interest.[5] For the same reason, any experimental design that asks a clinician to treat a patient against the clinician's own professional biases is inappropriate at best. And even if ethical reservations could be overcome, it would clearly be of only minimal scientific value to accumulate data on how patients fare under treatments not considered optimal at the time they are delivered.

Another concern is that therapists have a categorical obligation to collect, interpret and apply a full range of clinically relevant information in planning and delivering treatment for each individual patient. For this reason, structured designs which attempt to blind the therapist with respect to the knowledge of any consequential aspect of a patient's history, diagnosis or treatment plan are unethical (in addition to being unrep-resentative of the usual conditions of clinical practice).

Alternative models

Fortunately an alternative approach is available that avoids many of the problems of structured experimental designs. It has been called the "naturalistic" or "observational" model. When this approach is used, the investigator (always driven by a previously-determined set of well-specified questions), first defines appropriate samples of subjects and measurements. The investigator then observes and measures the subjects through time, while perturbing the processes of growth and treatment as little as possible. Records and data are accumulated and hypotheses are tested using the logical methods of inferential statistics.

This naturalistic approach, by relinquishing the idea of complete control over the objects of experimentation, accepts increased variability within the experimental groups and makes it more difficult to discern small but true differences between the means of experimental groups. Yet by the same token, those findings of naturalistic research which are in fact perceivable through the haze of real world variability are more directly applicable to the treatment of real world patients than are findings from structured models. Moreover, in addition to supplying information about mean effects, naturalistic studies can yield relatively unbiased estimates of individual variability associated with specific clinical interventions. *Such estimates of the variability associated with specific types of clinical intervention are in themselves extremely valuable to clinicians. Indeed it can tenably be argued that at the present level of our knowledge, they frequently constitute the most important findings from clinical studies.*

The thrust of this essay is not to argue that structured experiments are inappropriate or without value. Quite the contrary. Basic investigations into biological mechanisms, animal simulation studies of craniofacial development and response to therapeutic intervention, and biomechanical studies on the properties of dental materials are all valuable in their own rights. In addition, all can yield important heuristic insights and serve to constrain and define the most important paths for subsequent clinical trials. But the author does maintain that direct observational studies of treatment planning, progress and outcome are the necessary and ultimate tests of the utility of biological and biomechanical theories as applied to clinical practice.

Clinical studies

Clinical studies may be conducted on either a retrospective or a prospective basis.

In general, the greatest advantages of retrospective studies are that they can be conducted relatively rapidly (since one can use the records of patients whose treatment is already complete), that the investigator is protected against the circumstance of "subject drop-out" during the course of treatment, and that they are relatively economical. The major disadvantages of retrospective studies are that there is no truly satisfactory way to control for selection bias, that the available records are rarely complete, and that the observations one can make are constrained by the limitations of the treatment records of clinicians whose primary concerns were focused on issues other than those of greatest interest to the research investigation.

The advantages and disadvantages of prospective studies are in essence the inverse of those of retrospective studies. Provided that ethically and logistically satisfactory plans for random assignment to treatment can be developed, prospective trials afford an opportunity to control for selection bias and to define and control the records acquisition process. The main disadvantages of prospective trials are that they are expensive and that a great deal of time must inevitably elapse between project initiation and the point at which data on most of the main outcome variables become available for analysis.

Perhaps the most important and least appreciated advantage of prospective trials over retrospective studies turns on the issue of selection bias. Selection bias, the more general class of which the susceptibility bias[6] recently discussed by Johnston and co-workers[7] is an important component, occurs when subjects are assigned to experimental groups on the basis of properties that could influence experimental outcome but those properties are not taken into account in the experimental design. It is a problem in all non-randomized designs but occurs in particularly blatant form in the kinds of studies which constitute the body of retrospective clinical research in orthodontics.

The problem manifests itself in the following manner: One of the most important motivations in clinical orthodontic research is the desire to identify and partial out the effects of treatment. Indeed almost all reports of orthodontic clinical studies are couched in terms of differences in treatment effect between different types of therapeutic intervention. For example, we wish to discover the difference in effect on mandibular plane of high pull head gear and cervical traction. Or, say, the difference in stability of lower anterior alignment of extraction and non-extraction treatment. But when retrospective studies are used to study these or similar questions, they do not generally take into consideration the reasons why different subjects were assigned to different treatments by the treating clinicians.

If, for example, the same clinician treated different patients using different methods (as is true in most published orthodontic treatment comparisons), we are entitled to ask why some subjects were assigned to one treatment and some to another. One would certainly hope and expect that clinicians do not make their assignments to treatment arbitrarily. But if they do not, then how can the part of the outcome that was treatment-method specific be differentiated from the part which derived from the very same pre-existing differences that caused some subjects to be assigned to one treatment and some to another?

Alternatively, if a retrospective study compares subjects who were treated by one clinician using one technique with subjects who were treated by another clinician using another technique, how can differences in outcome resulting from differences between techniques be distinguished from differences in outcome associated with differences in clinical skill? One must conclude from serious consideration of these problems that retro-spective studies usually do not measure and report treatment effects but rather changes which occur during treatment and which are more properly called treatment-associated effects. Information about such changes is valuable to orthodontists and well conducted retrospective studies certainly do have an important role, but if we are to quantify treatment effects per se, some way must be found to randomize patients before treatment begins.

The conduct of clinical trials *always* raises ethical questions. Indeed it has been argued that there may be a fundamental contradiction between the roles of the research scientist and the clinician because the scientist is primarily focused on ascertaining the efficacy of the treatment, while clinician is primarily concerned with delivering the best possible care to each individual patient. Although the present author believes that this apparent contradiction can be overcome, it seems clear that any protocol that attempts to lock the patient and clinician into a pre-set treatment regimen by restricting the freedom of the therapist or the patient to make in-course corrections places the interests of the patient in jeopardy.

The patient's interests are also placed at risk by any design in which the therapist and his or her professional associates are blinded with respect to

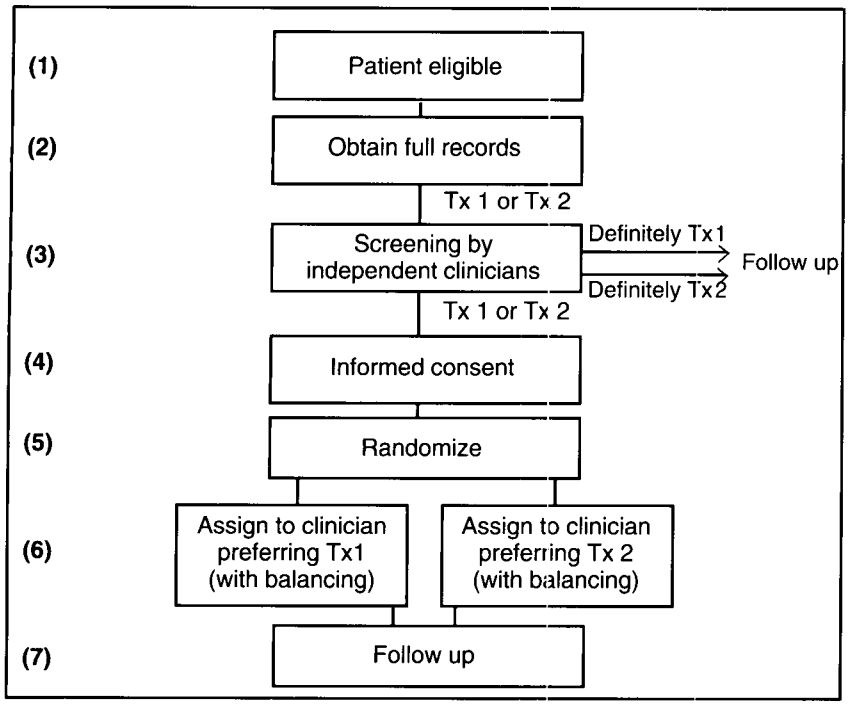

Figure 1
Design of a randomized clinical trial with clinician-preferred treatment

any information about the patient or the therapeutic process which could affect the treatment outcome. Parenthetically, this risk to the patient from the blinding of the therapist to any relevant facts about the patient's condition is likely to be much greater where treatment continues through time, as in chronic illness or orthodontics, than when the experiment involves only brief interventions, such as the short-term administration of a drug or other therapeutic modality. Finally, the patient has the right to receive treatment from a therapist who believes he or she is delivering the best treatment available and the therapist should never be asked to administer treatment which is contrary to his or her own professional convictions.

It seems clear that the requirements of ethical conduct of clinical trials in orthodontics can be much more easily met in naturalistic/observational designs than in structured ones. But if one is to attempt to identify true treatment effects, the problem of randomizing ethically must be solved. The recently reported design of the writer's prospective study which we describe as a randomized clinical trial with clinician-preferred treatment[8] represents one solution to this problem. In general, the new design is applicable in situations in which two treatment alternatives are being compared (where treatments 1 and 2 may for example be extraction vs. non-extraction, or early vs. late, or extra-oral traction vs. functional appliance).

The first step in the screening process (see Figure 1) is to assess the eligibility of the patient to be included in the trial. The criteria for selection may include such factors as age, sex, type of malocclusion stage of eruption etc. In step 2, full records are obtained for each patient who meets the basic demographic criteria for inclusion in the study and a complete set of cephalogrammetric and study cast measurements is generated. The records are then assessed independently by several competent clinicians (three to five), each of whom has expressed willingness to treat the patient (step 3). Each member of the panel evaluates the patient independently and indicates his or her treatment preference. If the independent assessments of the clinicians are in agreement, the patient is treated by the agreed upon method without randomization. If, and only if, disagreement exists among the panel members concerning the type of treatment preferred, informed consent is obtained (step 4) and the patient is randomized to one of the two treatments (step 5). After randomization, the patient is treated by one of the clinicians who during the evaluation step opted in favor of the treatment to which the patient was later randomized (step 6).

The new design appears to have several advantages which are frequently not found in other clinical randomization schemes. First, an objective clinician-administered test assures that each randomized subject is truly a borderline case. In consequence, within the limits of our current knowledge and skills, no subject is ever randomized to a clearly inappropriate treatment. Second, all available information is placed at the disposal of the clinician without blinding. Third, no clinician is ever asked to treat against his or her clinical preferences. Fourth, because each clinician treats according to his or her own preferences, the patient receives the full advantage of the placebo effect associated with the clinician's conviction about the appropriateness of the treatment method.

The new design has been used in a prospective clinical trial now in progress at the graduate orthodontic clinics at UCSF and the University of the Pacific. Progress has been generally satisfactory although not without some problems, chief of which is the fact that the method is quite labor intensive. As a corollary to its use, a consequential volume of data on the decision-making processes of clinicians has been accumulated and made available for study.

A final question

We now consider briefly a final question. Under the best of conditions, what information can we expect to obtain from clinical trials? The naive

answer is that one would wish to find out which treatment is best for the correction of the malocclusion under study. But the best method of treatment for the correction of one aspect of a malocclusion may not be best for the correction of other aspects and it is not always clear what relative weights should be assigned to the several variables we use for assessing the goodness of treatment outcome. In the treatment of Class II malocclusions, for example, is controlled reduction of overjet more or less important than long-term stability of mandibular incisor and canine alignment? Is achieving a stable molar relationship in neutroclusion more or less important than minimizing upper anterior root resorption? Is achieving improved masticatory function more or less important than achieving optimal facial esthetics?

Clearly, what one concludes is "best" in complicated multidimensional studies depends to a significant degree on the choice and weighting of outcome variables.

The task of clinical research is to provide rigorous and objective information on the changes which are observed in specific measurable parameters in response to therapeutic interventions of different sorts. Deciding how the available objective information should be applied to the treatment of any patient *always* involves an evaluation of the appropriateness of the information to the special requirements of the case at hand. Such an evaluation is properly made by a skilled clinician and an informed patient acting in concert. Clinical decisions, by their very nature, vary with the particulars of each case and always have strong subjective components.

This state of affairs is not limited to evaluations for orthodontic treatment. It characterizes all fields of medical decision-making and is, in fact, the essence of "doctoring". Consider for example a generic real world problem. Is the use of a low risk treatment which produces 70% improvement in all cases better than or worse than a higher risk treatment which produces almost complete improvement in 95% of cases but also yields increased disability in the remaining 5%? In answering questions of this sort, it is both conceivable and appropriate that the same conscientious clinician could made different decisions for different patients.

The fact that clinical decision-making always incorporates strong subjective components is recognized intuitively if not explicitly by most experienced therapists. Some serious and dedicated clinicians even believe that the problems of treatment are so complex and intractable that clinical studies of the types advocated above will never be of palpable assistance in the actual treatment of patients. They point out that there are many more different variables involved in treating an orthodontic patient than there are in launching a rocket to the moon. Some assert that each patient is different from all others — a law unto himself or herself, and that therefore orthodontics will always remain an art and not a science. The present writer recognizes the thrust of these assertions but contends that they represent the creation of a false and unnecessary dichotomy between art and science. In actuality, all the modern arts, from film to architecture to brain surgery draw on, incorporate and benefit enormously from relevant elements of science. As orthodontic clinicians, we must do the same — but we must keep our goals and expectations appropriate.

We conclude, therefore, that except in special and very limited circumstances, clinical studies in orthodontics cannot and should not be expected to reveal categorically which of two or more treatments is better in a global sense. They can and should be expected to supply valid and reliable information about the mean effects of different treatments. But more important, they should supply information about the usual individual variability of human growth, development and response to therapeutic intervention. Clinical research should also yield refined, reliable, and accurate estimates of the physiologic, economic, and time costs of various treatment strategies when those strategies are carried out by clinicians acting at particular levels of skill and competence. Information of this kind is vitally needed by clinicians, patients, and the public in order that appropriate decisions can be made for planning individual treatment and for formulating public health policy.

Acknowledgments

The investigations by the author which are discussed in this paper are supported by NIH-NIDR Grant #DE08713).

Author Address

Dr. Sheldon Baumrind
University of California San Francisco
Dept. of Growth and Development
Box 0640
School of Dentistry
San Francisco, CA 94143

S. Baumrind is Professor of Growth and Development/Radiology and Orthopedic Surgery, University of California San Francisco.

References

1. Ackerman JL. Orthodontics: Art, Science or Transcience? Angle Orthod 1974; 44:243-250.
2. Hellman S, Hellman DS. Of mice and not men. Problems of the randomized clinical trial. N Engl J Med 1991;324:1585-1589.
3. Schafer A. The ethics of the randomized clinical trial. N Engl J Med 1982;307:719-24.
4 Burchell HB. Vicissitudes in clinical trial research. Subjects, Participants, Patients. Controlled Clinical Trials, 1992;13:185-189.
5. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med 1987;317,3:141-145.
6 Horowitz R, McFarlane M. Brennan T, Feinstein A. The Role of Susceptibility Bias in Epidemiologic Research. Arch Intern Med 1985;145:909-912.
7. Paquette DE, Beattie JR, Johnston LE. A long-term comparison of nonextraction and premolar extraction edgwise therapy in "borderline" Class II patients. Am J Orthod Dentofac Orthop 1992;102:1-14.
8. Korn EL, Baumrind S. Randomized Clinical Trials with Clinician-Preferred Treatment. Lancet 1991;337:149-152.