

L.E. Johnston Jr

## Let them eat cake: the struggle between form and substance in orthodontic clinical investigation

**Authors' affiliations:**

Lysle E. Johnston Jr, Department of Orthodontics and Pediatric Dentistry, The University of Michigan

**Correspondence to:**

Lysle E. Johnston Jr, DDS, MS, PhD  
Department of Orthodontics and Pediatric Dentistry  
The University of Michigan  
Ann Arbor, MI 48109-1078  
USA  
Tel: +1 734 747 3814  
Fax: +1 734 763 8100  
E-mail:lejrr@umich.edu

**Dates:**

Accepted 16 July 1998

**To cite this article:**

Clin. Orth. Res. 1, 1998; 88–93  
Johnston Jr LE:

Let them eat cake: the struggle between form and substance in orthodontic clinical investigation

Copyright © Munksgaard 1998

ISSN 1397-5927

**Abstract:** Events of the past decade or so argue that there is little support in the literature for much of the specialty's treatment portfolio. The resulting call for 'evidence-based' treatment (not to mention the obvious intellectual bankruptcy of much of the clinical literature) has prompted many to argue that the randomized clinical trial (RCT) represents the future of orthodontic clinical investigation. The RCT, after all, is medicine's gold standard; what more is there to say? A popular, but ultimately divisive, corollary of this mimicry is the smug tendency to discount all other sources of data. In the face of a need for information, this attitude is also a wasteful conceit: in the end, the RCT can be applied only to a very narrow spectrum of orthodontic questions. Randomization implies equal susceptibility. Any prospective participant would have to be informed of this equality as part of the informed consent process. Unfortunately, it would be nearly impossible to enroll fully-informed subjects into any study whose alternatives are of markedly different morbidity: extraction versus non-extraction or orthodontics versus surgery. Thus, when measured against the most vexing clinical questions, the orthodontic RCT is almost by definition an amusing diversion – expensive, but relatively trivial in scope. Like it or not, it seems reasonable to conclude that most of the specialty's comparative clinical data will have to be generated by way of non-randomized designs in which care is taken to minimize the various known sources of bias. There probably is no other way.

**Key words:** clinical trial; prospective; retrospective

*In the activities of clinical therapy ... the formulation of hypotheses and counter-hypotheses is a trivial creative challenge. It takes absolutely no insight to ask the obvious question, 'I wonder whether bypass surgery is good for people with coronary artery disease?' It also takes no insight to establish a counter-hypothesis .... The truly creative problems of clinical management are in methodology, not in hypothesis formulation. Where the devil can we get appropriate groups of people and data to compare the therapies in an unbiased, objective way so that the questions will get effective, useful answers? (1)*

The present communication is an after-the-fact addition to Prospective Clinical Research in Orthodontics, a meeting designed to showcase yet another round of progress reports from 'the first controlled studies' in orthodontics. Given that the quasi-experimental Saint Louis University/University of Michigan Class II treatment comparisons are productive, albeit philosophically discrepant charter members of this pioneering cohort, their lowly estate is significant and worthy of extended comment.

Like children playing at being adults, many in orthodontics feel that the road to scientific maturity lies in imitating medicine, no matter what the cost and regardless of outcome. There is a tendency, therefore, to see the means of investigation as an end in itself and, by extension, to discount all other sources of data. For example, it is a common strategy to show slides depicting some sort of *scala naturae* in which clinical research designs ascend from the slime of retrospection to the perfection of the randomized clinical trial (RCT). At a time when the specialty is under attack from a variety of barbarian hordes and its main defenses are timely, adequately designed clinical studies, the blanket assertion that truly meaningful data can come only from an RCT is self-serving and inappropriate; it is the equivalent of saying, 'let them eat cake'. A few dyspeptic comments are in order.

History teaches that orthodontists tend to be professionally successful, regardless of their educational background or choice of appliances; few look to 'science' as a source of practical clinical guidance. Perhaps as a result, disagreement was, until recently, largely recreational and for the most part limited to superficial tactical details, such as type of appliance (edgewise, bioprogressive, headgear, bionator, etc.), slot size and angulation, inter-bracket distance, and the like. It was harmless sport because, at the same time, there was general agreement concerning the basic strategies of treatment. Extraction,

for example, was seen as an answer to crowding and protrusion; surgery, as an answer to severe skeletal discrepancies. Times have changed. For a variety of reasons – specialist/generalist turf wars, fear of litigation, busyness problems – entire treatment strategies have come under attack, both from within and without the specialty. These attacks have taken their toll.

Some of our critics, for example, argue that bicuspid should never be extracted or that surgery is tantamount to malpractice. The bitterness of the disagreement and the fervor of the participants is a threat to the public health; it threatens the well being of the specialty; it distorts clinical judgement. A major problem with the various revisionist arguments is that they call for alternatives that have never been shown to be as *orthodontically* effective as the treatments they would replace. For example, one would have to achieve 12 mm of *stable* expansion to equal the 5 mm of space per quadrant created routinely by bicuspid extraction (the width of a bicuspid minus the anchorage lost during space closure). Wishful thinking aside, the literature argues that contemporary 'arch-development' methods fall far short of this requirement, as do fanciful functional-appliance substitutes for surgery ('bite forward ... forever'). Accordingly, it is important that we examine carefully the arguments against extraction and surgery, because, if they are true, it would mean that we have no safe, effective treatment for a large segment of the orthodontic population. Unfortunately, in contrast to superficial technical details, many of our basic orthodontic strategies are surprisingly difficult to study in real time.

Given a desire to compare, say, arch leveling with 0.018-inch and 0.022-inch slots or early treatment with headgear and bionator, a careful 'retrospective' comparison probably will provide data that are 'good enough', both because it is easy to obtain samples that were similar before treatment and because the contrasts themselves are only of mild-to-moderate importance. In contrast, if one wants to document the supposed superiority of say, non-extraction edgewise orthodontic treatment (better function, more pleasing profiles, greater stability, etc.), one cannot simply round up a few ex-patients and compare outcomes. Comparisons between groups formed on the basis of a clinician's treatment assignment tend to be compromised by 'susceptibility bias', a systematic segregation driven by the simple fact that many types of malocclusion seem uniquely susceptible to a specific treatment. For example, if extraction and non-extraction patients were systematically different before treatment (crowded

and protrusive versus spaced and 'flat'), differences after treatment would defy interpretation. To effect a meaningful comparison, therefore, outcomes must be compared in patients who initially were similar and thus equally susceptible to both treatments.

Clearly, a well-designed randomized clinical trial would serve to eliminate susceptibility bias, along with a variety of other biases, both known and unknown. A key prerequisite to the conduct of a clinical trial is that there be true uncertainty about the relative efficacy of the various arms of the trial ('equipose'; (2)). Individual orthodontists, however, commonly do not share the uncertainty of the refereed literature. As a result, they might not agree with, and thus would be unable to render, a treatment dictated by random assignment. Baumrind's method of 'clinician-preferred' treatment assignment (3) seems to address these requirements and problems by providing an empirical definition of uncertainty and by ensuring that treatment is prosecuted by a clinician who happens to agree with the randomization. It is my opinion, however, that this ingenious approach does not solve the problem of informed consent.

Because great pains must be taken to determine and document equal susceptibility to all treatments, a participant in a randomized clinical trial is usually subjected to a more thorough analysis than is the average orthodontic patient. From the standpoint of hypothesis testing, therefore, it is probable that the RCT would serve to eliminate bias and to equalize experimental conditions. From the standpoint of a prospective subject, however, randomization can never equalize the anticipated pain, discomfort, and inconvenience of the various arms of the trial. Unfortunately, many of the most pressing questions in orthodontics – those asked in the clinic and in court, rather than those offered up for ritual slaughter at meetings and in monographs – involve treatments that differ greatly in terms of anticipated pain, discomfort, and inconvenience. Thus, if all things (except the treatments to be compared) really are equal *and the patient knows that they are equal*, how can a sample ever be recruited?

What sort of subjects, for example, would agree to surgery if they understood – i.e. were fully-informed – that they could be treated equally well orthodontically? ('Mrs. Jones, your daughter can be treated orthodontically or surgically; however, we flipped a coin and it came up *surgery*. Sign here.') Some no doubt would agree to participate, perhaps in the hope that they will be randomized to the less morbid arm of the trial. If they are lucky, fine; if

not, they can always withdraw. Unfortunately, even if a few surgery patients or a few extraction patients happen to make it this far, there is one last hurdle: would fully-informed subjects permit themselves to be randomized to a surgery or extraction arm if pains were taken to elicit a preference? As noted by Kodish et al. (4), 'the autonomy principle dictates that patients' personal values and motivations be given the highest priority in reaching a treatment decision.' I would argue, therefore, that many (but surely not all) non-trivial orthodontic trials can succeed only to the extent that a prospective subject's right to informed consent is abridged. Many disagree with this judgement, especially those who seek to bring the enlightenment of randomized trials to those toiling in the darkness of retrospective investigation.

Elsewhere in this issue, Baumrind has noted that he is 'not greatly concerned' by my ethical reservations. As I understand it, he feels that empirical uncertainty (i.e. among-clinician disagreement) is not the same thing as equal susceptibility. Even in the face of uncertainty (as defined by evenly divided opinion), one treatment will probably turn out to be better than the other. Based on this fine distinction between uncertainty and equal effect, it is argued that it would be misleading/unnecessary to tell prospective subjects that they probably would be served equally well by either arm of the trial. I doubt that this argument can be sustained.

In tennis, if you cannot see the ball out, you have to call it in. In the clinic, however, if opinion is so evenly divided that one treatment option is about as popular as the other, I would argue that the two must be assumed, at least provisionally, to be equally effective. If it is assumed that the two treatments are equal, then this information must be shared with prospective subjects. Indeed, in Baumrind's approach, among-clinician disagreement is the main eligibility requirement. If eligible subjects also must be equally susceptible, then from the standpoint both of the study and its participants, disagreement also defines equal susceptibility. QED. More to the point, if prospective subjects were informed of the among-clinician disagreement that defined their eligibility and were pressed for a preference, how many would allow themselves to be assigned the more morbid option?

It should come as no great surprise, therefore, that the only trials that have managed to honor the details of their original application are those designed to compare technical details, not basic strategies. Although a mean-spirited observer might question the need for 10 years of support

for each of three studies of 'headgear versus functionals', others apparently feel that the mere existence of these trials signals a scientific coming-of-age for orthodontic clinical research. Unfortunately, in its orthodontic incarnations, the RCT is a somewhat debased gold standard.

In the *Wizard of Oz*, Professor Marvel asked the four awed supplicants to 'pay no attention to that man behind the curtain'. In the world of the orthodontic randomized clinical trial (surely the metaphorical equivalent of the Emerald City), we are asked by other professors to be equally awed and equally selective in our attention. For example, we are asked (or at least expected) to ignore the lack of blinding during and after treatment, not to mention the probable impact of the so-called 'Hawthorne effect'. The goodness of a treatment outcome depends on the orthodontist's efforts and the patient's cooperation, both of which might well be elevated to supra-normal levels by dint of both parties' known participation in a clinical trial. Moreover, because of the time and money invested in each of a relatively small coterie of subjects, orthodontic RCTs sometimes feature vigorous, but scientifically questionable, steps to achieve exalted levels of patient compliance (e.g. periodic telephone calls to restore flagging enthusiasm and cooperation). Indeed, even the act of taking the occasional peek at the preliminary data ('the headgear patients are doing fine, but the functionals are lagging behind ...'), presumably for the purpose of issuing breathless communiqués and news flashes at various meetings throughout the world, is seen by some as a violation of the RCT ground rules. Moreover, as alluded to elsewhere in this volume, there are a number of practical problems that may prove equally damaging to the successful completion of a non-trivial orthodontic RCT.

Orthodontic treatments take years to finish and even longer to evaluate. Thus, a truly useful RCT would have a time frame (15–20 years) approaching that of a major medical trial (and exceeding the residual professional life of most of the current principal investigators). Indeed, an orthodontic trial also would probably outlive its working hypothesis. Most damaging, however, is the certainty of sample attrition and the attendant loss of statistical power. If you start with 300, a one-in-ten response is adequate; if you start with 30, you end up with an exceptionally expensive case report. Ultimately, orthodontic investigators will have to assess

their goals and priorities. The quest for man-powered flight was ennobling, expensive, and the subject of several fascinating Public Broadcasting specials; however, it did little for the person faced with, say, the rigors of a coach/class flight from Detroit to Dallas. Is it to be form or substance? As noted by Feinstein (5):

*A ... misconception is to give randomization credit for certain scientific standards and precautions for which it is really not responsible .... The misconception just cited – which confuses the tactic of randomized assignment and the strategy of a scientific plan – is particularly important, because many of the desirable scientific features associated with randomized clinical trials ... are really attributable to advance scientific planning, not to randomization. These desirable features can therefore be obtained with suitable planning even when randomization is not used.*

In the end, if 'many of the desirable scientific features associated 'with randomized clinical trials' can be achieved without randomization, then there is no need to limit one's perspectives to narrow, trivial problems merely because they lend themselves easily to the RCT genre. The tail need not wag the dog.

Based on the reasoning outlined in this brief communication, the Saint Louis University/University of Michigan response to the NIDR request for applications opted for the substance and promise of a partially retrospective design over the form and pretence of a randomized clinical trial. Our implementation of Feinstein's call for 'suitable planning' took the form of a statistical algorithm designed to identify samples of patients who, at the start of treatment, were similar with respect to the features that normally dictate the choice of treatment strategies (crowding, protrusion, overjet, and the like). Given this approach, we were able to go well beyond the narrow scope of a prospective trial to examine treatment alternatives that could never be randomized to fully-informed subjects (e.g. extraction versus non-extraction; orthodontics versus surgery). The results of these studies will be examined elsewhere in this issue. In the meantime, I would argue that, for orthodontics, 'the first controlled studies' are much like teaching a dog to play the piano – an interesting achievement, but hardly a benchmark against which all other musical activities must be judged.

## Abstrakt

Fälle in den letzten zehn Jahren lassen erkennen, daß es in der Literatur nur wenig Unterstützung für einen Großteil des Behandlungsportfolios in der Kieferorthopädie gibt. Die daraus resultierende Nachfrage nach 'auf Beweisen basierender' Behandlung (abgesehen vom offensichtlichen intellektuellen Bankrott eines Großteils der klinischen Literatur) hat dazu geführt, daß viele behaupten, die klinische Zufallsstudie (Randomized Clinical Trial: RCT) sei die Zukunft der klinischen Versuche in der Kieferorthopädie. Die RCT ist schließlich auch der goldene Standard in der Medizin; was gibt es da noch zu sagen? Eine beliebte, letztendlich aber umstrittene Folge dieser Nachahmung ist die selbstgefällige Neigung dazu, alle anderen Datenquellen zu unterschätzen. Angesichts des Bedarfs an Informationen ist diese Einstellung außerdem eine verschwenderische Einbildung: letztendlich kann die RCT nur bei einem sehr engen Spektrum von kieferorthopädischen Fragen angewendet werden. Eine zufällige Zuweisung setzt die Gleichheit der Voraussetzungen voraus. Jeder potentielle Teilnehmer müßte als Teil des Prinzips der informierten Zustimmung über diese Gleichheit unterrichtet werden. Leider würde es nahezu unmöglich sein, voll informierte Personen in irgendeine Studie aufzunehmen, deren Behandlungsalternativen große Unterschiede in den Schadensgraden beinhalten: Extraktion im Vergleich zur Nichtextraktion oder Kieferorthopädie im Vergleich zur Kieferchirurgie. Verglichen mit den kompliziertesten klinischen Fragen ist somit die kieferorthopädische RCT schon von der Definition her eine amüsante Ablenkung-teuer, aber vom Umfang her relativ belanglos. Ob es uns gefällt oder nicht, es scheint eine vernünftige Schlußfolgerung zu sein, daß ein Großteil der klinischen Vergleichsdaten in der Kieferorthopädie durch Designs der nicht-zufälligen Zuweisung gewonnen werden muß, wobei darauf geachtet werden muß, daß die verschiedenen bekannten Neigungen minimiert werden. Wahrscheinlich gibt es keine andere Möglichkeit.

## Resumen

Eventos en esta última década demuestran que hay poco apoyo en la literatura con respecto a todo lo que incluye la carpeta de tratamientos de la especialidad. La reclamación del tratamiento basado en la evidencia (por mencionar la falta de material intelectual da la literatura clínica) ha impulsado a muchos a argumentar que la investigación clínica aleatorizada (ICA) representa el futuro de la investigación clínica en la ortodoncia. La ICA es, después de todo la norma en investigaciones médicas. Una tendencia popular, pero divisiva, es la consecuencia de estra mimica a descontar todas las otras fuentes de datos. Ante la necesidad urgente de información, esta actitud es una vanidad malgastadora: a un extremo, la ICA solo se puede aplicar a un espectro muy limitado en las preguntas hechas en investigaciones de ortodoncia. Aleatorizado (al azar) implica igualdad de susceptibilidad. Cualquier participante prospecto tiene que ser informado de esta igualdad como parte del proceso de consentimiento informado. Desafortunadamente, es casi imposible obtener participantes totalmente informados en cualquier investigación que conlleve alternativas notablemente diferentes: extracción contra no-extraer o cirugía contra ortodoncia. Así que, cuando se compara con las preguntas clínicas más exigentes, el uso de ortodoncia de la ICA es casi por definición una proposición costosa, pero relativamente trivial en alcance. Nos guste o nó, parece razonable concluir que la mayoría le dos datos clínicos comparativos de la especialidad tendran que ser generados por medio

de diseños no-aleatorizados en el cual se tome precaución para minimizar las varias fuentes desconocidas de prejuicios. Probablemente, no hay ninguna otra manera.

## 抄録

過去10年余りを振り返ってみると、矯正治療として行われていることの多くは、文献で支持されていないという議論がある。その結果、（臨床研究の文献の多くが明らかな知的破産を起こしている事は言うまでもなく）証拠にもとづく治療が要求されるようになり、無作為割付臨床試験（RCT）こそが将来の臨床矯正学の研究を代表するものであるという議論を促すことになった。RCTは実際のところ、医学の黄金律であって、それ以上何かを付け加える必要があるだろうか？この受け売りの、人気のある、しかし最終的には分裂を生じる当然の結末は、RCT以外のすべての源のデータを割り引いて考えるという一人よがりな傾向である。情報のニーズに直面している今、このような態度は無駄の多いうぬぼれでもある。さらに、RCTは矯正学上の諸問題の非常に狭い範囲にのみ適用が可能である。無作為割り付けとは、等しい感受性を意味している。いかなる前向き試験の参加者も、この平等性について、インフォームド・コンセントを得るプロセスにおいて、説明を受けるべきであるが、残念ながら、十分な説明を受けた者に、抜歯か非抜歯か、矯正治療か外科手術かを選択する際、別の治療法との間で罹患率に大きな差が生じるような試験に参加させることはほとんど不可能に近いであろう。したがって、最もやっかいな臨床的疑問に取り組むに際して、矯正学におけるRCTは、ほとんど定義上、高価ではあるが、ほとんど論じるに値しない、気ままな道楽に過ぎない。好むと好まざるとにかかわらず、矯正学の比較臨床データの大半は、様々な既知の源のバイアスが入り込む余地ができるだけ小さくなるように注意を払いながら、無作為割付によらない実験計画を用いて作成しなくてはならないだろう。おそらく、そうする以外に方法はないと思われる。

## References

1. Feinstein AR. The clinician as scientist. In: McNamara Jr JA, Carlson DS, Vig PS, Ribbens KA, editors. *Science and Clinical Judgment in Orthodontics. Monograph 18, Craniofacial Growth Series*. Ann Arbor: Center for Human Growth and Development, The University of Michigan; 1986. pp. 1-14.

2. Freedman B. *Equipoise and the ethics of clinical research*. *N Engl J Med* 1987;**31**:141–5.
3. Korn EL, Baumrind S. *Randomised clinical trials with clinician-preferred treatment*. *Lancet* 1991;**337**:149–52.
4. Kodish E, Lantos JD, Siegler M. *The ethics of randomization*. *CA: A J Clin* 1991;**41**:180–6.
5. Feinstein AR. *An additional basic science for clinical medicine: III. The challenges of comparison and measurement*. *Ann Intern Med* 1983;**99**:705–12.