SPECIAL ARTICLE

Randomized clinical trials in orthodontics: Reality, dream, or nightmare?

Giliana Zuccati,[a] Carlo Clauser,[a] and Roberto Giorgetti[b]
Florence and Siena, Italy

Randomized controlled trials (RCTs) are considered the highest level in the hierarchy of evidence for treatment effectiveness. However, RCTs have also been criticized for various shortcomings. The purpose of this article was to review the most common criticisms against RCTs and answer them based on the principles of scientific inquiry, so that orthodontists can build their evidence-based practice on the best scientific research. In the era of evidence-based medicine, designing RCTs is the challenge for researchers in orthodontics. (Am J Orthod Dentofacial Orthop 2009;136:634-7)

Randomized controlled trials (RCTs) are considered the highest level in the hierarchy of evidence for treatment effectiveness; observational studies, because of their potential for bias, provide a lower level of evidence. In RCTs, the subjects are randomly allocated to 2 or more groups: the experimental group is treated, and the control group receives no treatment (negative control), conventional treatment (active control), or a placebo.[1,2]

Geoffrey Marshall directed a multicenter RCT in 1948 to evaluate the benefits of streptomycin with a "rigorously planned investigation with concurrent controls."[3] The empirical use of streptomycin and exaggerated claims for gold treatment had persisted over 15 years. This double-blind randomized trial showed that streptomycin was effective and initiated a new era in the fight against pulmonary tuberculosis and in clinical research.

Nevertheless, criticisms have been raised against RCTs. For example, RCTs are expensive and time-consuming, and they have not achieved their objectives or provided knowledge not already available from retrospective studies or animal experimentation. "What is particularly interesting is that knowledge based on years of clinical experience has been disregarded and then announced as if it was something completely new."[4]

The purpose of this article was to review the most common criticisms of RCTs and attempt to answer them based on principles of scientific inquiry, so that orthodontists can build their evidence-based practice on the best scientific research.

ORTHODONTIC RCTs ARE UNETHICAL

In the early decades of the 19th century, bloodletting was the usual treatment for pneumonia. Broussais used up to 100,000 bloodsuckers in a year.[5] However, Louis[6] had some doubt and began to study the efficacy of the therapy by delaying the bloodletting in a control group. His data were criticized and considered rather weak, the numeric analysis gave only probabilities without certainties, and his samples were small. Was it ethical, his critics asked, to delay treatment for patients suffering from pneumonia?

In orthodontics, randomization or treatment delay has been considered unethical because the usual treatment is assumed to be effective; however, many entrenched beliefs have been undermined by sound clinical research.[7-9] Obviously, patients have the right not to participate in an RCT, and special informed consent is mandatory. But when there is doubt as to which therapy is better, RCTs are the most ethical tool. This principle is universally recognized in oncology and cardiology, when the main outcome is the patient's survival, and the RCT is the standard experimental design.

IT IS DIFFICULT TO AMASS A LARGE SAMPLE AND CONTROLS WITH SIMILAR MALOCCLUSION

The choice of sample size depends on the difference to be observed, the variability in the population, and the power of the test, rather than on the experimental design, whether a case-control study or an RCT. Patient recruitment is a serious problem if the condition of interest is rare or the studied variable has large variations. Retrospective studies might give useful information

[a] Private practice, Florence, Italy.
[b] Professor and chair, Department of Orthodontics, School of Dental Medicine, University of Siena, Siena, Italy.
The authors report no commercial, proprietary, or financial interest in the products or companies described in this article.
Reprint requests to: Giliana Zuccati Clauser, Via Masaccio 173, Florence, Italy; e-mail, gilzuccati@libero.it.
Submitted, April 2009; revised and accepted, May 2009.
0889-5406/$36.00

doi:10.1016/j.ajodo.2009.06.001

about variability that can be used to determine the correct sample size of a planned RCT.

An adequate size is needed to allow randomization to equal out potentially confounding variables. The sample should be large enough to detect a clinically significant difference between the test and control groups and to prevent type II error.[10] Significant results have been obtained by an RCT comparing 2 kinds of palatal expanders despite small sample size in a limited age range.[11]

GROWTH STUDY DATA CAN PROVIDE AN ADEQUATE CONTROL GROUP

According to this argument, a control group derived retrospectively from growth study data is indistinguishable from that derived prospectively from untreated Class II Division 1 patients. Normative data are reported for boys and girls at each year of age, and patients can be matched for age, sex, and treatment time.[12] The results do not appear to differ, no matter which method is used, and statistical testing for differences between 2 types of control groups was not significant, as was true for the Twin-block groups.[13]

This approach is not valid and introduces bias. Imagine 2 control groups, A and B, with the same mean: there cannot be a statistical difference between A and B. If the mean in the test group is different, it might well be statistically different from group A but not from group B (which share the same mean) if the standard deviation of group A is less than that of group B. Even a large overlap does not guarantee the same statistical result. But the most serious bias is that a statistical comparison entails the assumption that the 2 samples were extracted from the same population by a random procedure.[1,2] If the assumption is not met, the statistical decision is biased. Historic data differ from simultaneous sampling for many variables that cannot be controlled and might affect the result. Statistics might discriminate between chance and treatment effects if the assumptions are met (as they are in well-designed RCTs): otherwise, bias, instead of treatment, might be responsible for the difference observed. This makes the results difficult to analyze and the conclusions uncertain when the allocation of subjects to the treatment and control groups is not rigorously randomized.

RETROSPECTIVE STUDIES GIVE THE SAME RESULTS AS RCTs WITHOUT RANDOMIZING

This is the most insidious prejudice. Retrospective studies on the effectiveness of streptomycin during 15 years did not provide conclusive evidence. Only RCTs allowed for widespread use of this effective new drug, thus saving thousands of lives.[3] It is not surprising that retrospective studies and RCTs lead to the same conclusions in most cases, even if a trend of retrospective studies to overestimate the treatment effect has been assessed in medicine. However, the results of observational studies need to be confirmed.[14] There is a fundamental difference between opinion and evidence, or between belief and knowledge: this upgrade is the value added by RCTs to clinical research. Moreover, sometimes the results are not equivalent. For example, Tulloch's RCTs reached different conclusions from previous studies with weaker designs and were confirmed by other groups of researchers.[7-9] Therefore, Tulloch's results should be considered conclusive.

Unfortunately, sometimes retrospective studies, although well conducted, are misleading; unavoidable sources of bias complicate the interpretation and undermine the reliability of any experimental design except appropriately blind RCTs as far as the treatment effect is concerned.

VALID EVIDENCE REQUIRES RETHINKING SOME CURRENTLY POPULAR RESEARCH DESIGNS

The research of Tulloch et al[7] established that early treatment for most children with Class II malocclusion is no more effective and considerably less efficient than later 1-stage treatment during adolescence; this goes against deeply rooted current opinions. Some objected that Moyers et al[15] detailed 6 Class II types with different treatment plans for most of them, and that a correct research design would have assigned patients to the bionator or the bite plate and headgear on the basis of individual diagnosis.

The study of Tulloch et al[7] showed that the usual treatment approach does not enhance growth significantly; this does not exclude the different effects that could be observed in selected subgroups. Such doubts do not invalidate the study of Tulloch et al but might justify further research with new and more refined RCTs to answer more subtle questions.

ONE RCT IS NOT SUFFICIENT TO PROVIDE CONCLUSIVE EVIDENCE

An RCT might have been incorrectly designed and therefore not conclusive. Retrospective studies are never conclusive, no matter how well designed and performed, because they are intrinsically vulnerable to bias.[1,2] Observational prospective studies can be conclusive to establish the efficacy of a treatment under special circumstances, as in the case of osseointegrated implants for edentulous patients.[16] Even a well-designed RCT might not be considered conclusive if the results are different from all previously accumulated

knowledge. In these cases, a rule of thumb that makes the reader more comfortable with a clinical decision is to trust results based on more conservative statistical criteria.

A P value <0.01 warrants the reliability of a countercurrent result even if one bears in mind the tenet of Bayes' theorem: according to this, $P<0.05$ is not sufficient to corroborate a conclusion associated with a low a priori probability.[17]

RCTs ARE SUSCEPTIBLE TO BIASES OF COMPLIANCE AND LONG-TERM ATTRITION

Not all subjects comply with the regimen to which they are assigned, and, for studies that require long follow-up periods, there is a natural tendency for a high dropout rate. Moreover, because it takes such a long time to complete prospective trials, the procedure that was investigated might not even be used, or the question might be obsolete.[4]

These problems affect every type of study design. On the other hand, many new therapies have been abandoned because they proved to be ineffective or even harmful after a short period of enthusiasm. If long-term objectives are to be achieved, long studies must be carried out. There is no plausible alternative.

The Consolidated Standards of Reporting Trials (CONSORT) guidelines are an effective tool to prevent many types of flaws in RCTs. These guidelines can even be used to evaluate other types of studies and to identify and discuss possible sources of bias.[18]

RCTs CANNOT INTERCEPT RARE OR UNEXPECTED COMPLICATIONS

The RCT is the gold standard to assess the efficacy of a therapy, but infrequent complications are better studied by surveys. Case reports and case series might be adequate to warn against risks. Thilander et al[19] studied the complications of implant placement in a small group of adolescents; that study has been useful to prevent widespread use of implants in growing adolescents.

THE RESULTS OF RCTs CANNOT BE APPLIED TO EVERYDAY CLINICAL PRACTICE BECAUSE THE VALUE OF THE EVIDENCE FOR CLINICIANS FROM RCTs IS QUESTIONABLE

Any well-designed RCT answers 1 specific question on 1 variable, usually measuring the outcome of a treatment. The decision of how to treat a patient is often based on evaluation of many variables, and the treatment objectives might not coincide with the same criteria used in an RCT. Evidence-based dentistry is a comprehensive approach to oral health: the operator's skills and the patient's needs and preferences are parts of it, but even 1 RCT can provide reliable information that is helpful in making a sound clinical decision.

There is a need for RCTs with valid methodologies to assess the efficacy of orthodontic therapies: adequate sample size based on power calculations, adequate sequence of randomization with allocation concealment, blind outcome assessment, completeness of follow-up, and accounting for dropouts. The quality of RCTs can improve if the CONSORT guidelines are followed.[18]

The methodologic issue is of paramount importance, but the value of a perfectly designed RCT might be completely annulled if the treatment is not performed according to the state of the art. Presenters and authors have been invited to clearly provide the level of scientific evidence that underpins their beliefs.[20] If it is not on the level of a systematic review or an RCT, this belief should not be supported, but this might be too extreme. If retrospective studies are well designed and their results are consistent, they are helpful in the clinical decision-making process when well-designed RCTs are unavailable. For instance, Marshall et al[21] sought to specifically answer whether maxillary expansion is stable in the long term. The best evidence we are basing our clinical practice on comes from 1 retrospective controlled trial and 1 prospective controlled trial of adolescent subjects; additional controlled trials are needed to add to our knowledge of long-term expansion stability. When reliable data from RCTs are not available, clinical decisions are based on the interpretation of observational studies, even if selection bias and comparison with biased control groups undermine their reliability. Moreover, observational studies assist in identifying hypotheses that can be tested by RCTs and provide important information to estimate adequate sample sizes.

CONCLUSIONS

In an era focused on evidence-based medicine, studies with an RCT design are the challenge for researchers in orthodontics. RCTs are desirable because they produce the highest level of scientific evidence for evaluation of treatment effectiveness. They are the coming reality and can no longer be viewed as dreams or nightmares.

REFERENCES

1. Brunette DM. Critical thinking, understanding and evaluating dental research. Chicago: Quintessence; 1996. p. 150.
2. Pini Prato GP, Esposito M. Come cercare, leggere e interpretare l'articolo scientifico. Società Italiana di Parodontologia. Milan: Masson; 2005. p. 58-9.

3. Streptomycin treatment of pulmonary tubercolosis: a Medical Research Council investigation. BMJ 1948;2:769-82.
4. Meikle MC. Guest editorial: what do prospective randomized clinical trials tell us about the treatment of Class II malocclusions? A personal viewpoint. Eur J Orthod 2005;27:105-14.
5. Cohen IB. The triumph of numbers: how counting shapes modern life. New York: W.W. Norton; 2005. p. 109.
6. Louis PCA. Recherche sur les effets de la saignée dans plusieurs maladies inflammatoires. Archives Générales de Medicine 1828; 18:321-36.
7. Tulloch JFC, Proffit WR, Phillips C. Outcomes in a 2-phase randomized clinical trial of early Class II treatment. Am J Orthod Dentofacial Orthop 2004;125:657-67.
8. Dolce C, McGorray SP, Brazeau L, King GJ, Wheeler TT. Timing of Class II treatment: skeletal changes comparing 1-phase and 2-phase treatment. Am J Orthod Dentofacial Orthop 2007;132:481-9.
9. O'Brien K. Is early treatment for Class II malocclusion effective? Results from a randomized controlled trial. Am J Orthod Dentofacial Orthop 2006;129(Suppl):S64-5.
10. Rinchuse DJ, Sweitzer EM, Rinchuse DJ, Rinchuse DL. Understanding science and evidence-based decision making in orthodontics. Am J Orthod Dentofacial Orthop 2005;127:618-24.
11. Oliveira NL, Da Silveira AC, Kusnoto B, Viana G. Threee-dimensional assessment of morphologic changes of the maxilla: a comparison of 2 kinds of palatal expanders. Am J Orthod Dentofacial Orthop 2004;126:354-62.
12. Hayes JL. Design flaws in some randomized controlled trials. Am J Orthod Dentofacial Orthop 2008;134:466-7.
13. Trenouth MJ. Design flaws in some randomized controlled trials. Am J Orthod Dentofacial Orthop 2009;135:141.
14. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Eng J Med 2000;342:1887-92.
15. Moyers RE, Riolo ML, Guire KE, Wainright RL, Bookstein FL. Differential diagnosis of Class II malocclusions. Am J Orthod 1980;78:477-94.
16. Adell R, Lekholm U, Rockler B, Brånemark PI. A 15-year study of osseointegrated implants in the treatment of the edentulous jaw. Int J Oral Surg 1981;10:387-416.
17. Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. Ann Intern Med 1999;130:995-1004.
18. Turpin DL. CONSORT and QUOROM guidelines for reporting randomized clinical trials and systematic reviews. Am J Orthod Dentofacial Orthop 2005;128:681-6.
19. Thilander B, Odman J, Grondahl K, Friber B. Osseointegrated implants in adolescents: an alternative in replacing missing teeth? Eur J Orthod 1994;16:84-95.
20. Huang G. Fasten your seat belts for the bumpy ride to evidence-based practice. Am J Orthod Dentofacial Orthop 2005;127:4-5.
21. Marshall SD, English JD, Huang GJ, Messersmith ML, Riolo ML, Shroff B, et al. Long-term stability of maxillary expansion. Am J Orthod Dentofacial Orthop 2008;133:780-1.