Original Article

Randomized controlled trial: the gold standard or an unobtainable fallacy?

Lars Bondemark* and Sabine Ruf**

*Department of Orthodontics, Malmö University, Malmö, Sweden and **Justus-Liebig-University, Giessen, Germany

Correspondence to: Sabine Ruf, Zentrum für ZMK, University of Giessen, Schlangenzahl 14, 35392 Giessen, Germany. E-mail: sabine.ruf@dentist.med.uni-giessen.de

Summary

Background: This article is the result of a debate at the European Journal of Orthodontics Open Session in 2013 in Reykjavik, Iceland.

Objective: The aim of this article is to highlight some of the strengths and weaknesses of clinical orthodontic research, with particular emphasis on randomized controlled trials (RCT). The ultimate aim of improving clinical orthodontic research in general.

Design: This article is organized into two sections with arguments for and against RCTs. The backgrounds to evidence-based evaluation and the level or quality of evidence in trials are discussed. The article emphasizes what makes high quality clinical research, and gives practical advice including examples of tips and potential pitfalls for those undertaking clinical research.

Results and Conclusion: The overriding message is constructive and it is hoped that the article serves as an aid in evaluating, designing, conducting, and reporting clinical research.

Part I: RCTs are for orthodontics!

Lars Bondemark

Evidence-based evaluation

Scientific assessment in health care aims to identify interventions that offer the greatest benefits for patients while utilizing resources in the most efficient way. Scientific assessment is needed in health care both for established methods and for new medical innovations. Implementing evidence-based health care means that decisions are supported by the best available scientific evidence from rigorous trials as a complement to other knowledge and to input from patients and caregivers (1). Important health care decisions that concern a patient’s health should always proceed from the best available scientific evidence (2).

The evaluation begins with a clinically relevant question, followed by an efficient literature search and finally an evaluation of the evidence, applying strict rules for reliability and validity. This refers to a conscious and systematic effort to design clinical treatment in accordance with the best possible scientific evidence (3, 4).

Level of evidence

In an evidence-based approach to evaluation of effectiveness, the randomized controlled trial (RCT) is the acknowledged standard and is considered to generate the highest level of evidence, followed by controlled trials (5). Trials without controls, case series, case reports, and finally expert opinions, generate low or insignificant evidence (Figure 1).

Randomized controlled trial

The first published RCT appeared in the 1948 article entitled ‘Streptomycin treatment of pulmonary tuberculosis’ (6). Then, by the late 20th century, RCTs were recognized as the standard method for ‘rational therapeutics’ in medicine.

During the last two decades there is about 2.4 million clinical trials considering medicine (excluding dentistry) while for dentistry excluding orthodontics there exists almost 200,000 clinical trials. For orthodontics the corresponding amount of trials is 30,000. In spite of RCTs represents strong evidence and considered the ‘gold standard’ of clinical trials, there are only 5 per cent RCTs in medicine and the proportion is 4.5 per cent in dentistry as well as in orthodontics. However, during the last 20 years an important increase of RCTs is shown in orthodontics (Figure 2).

The RCT requires careful and rigorous planning and coordination and it must be remembered that all scientific methods have their weaknesses. To avoid easy mistakes of the research methodology, not only regarding RCTs, the following approach can be useful: consider how many subjects that will be needed, that is estimate the material...
size by a power analysis; use a correct and proper randomization; and use the intention-to-treat (ITT) which means that all subjects, successful or not, are included in the final analysis. Consequently, if any subjects dropout from the trial, or not respond to the treatment, these shall still be recorded and counted as unsuccessful.

Randomization
The randomized allocation of subjects ensures that both known and unknown determinants are evenly distributed among the different study groups. This minimizes bias in assessment of differences in effects between two or more treatment alternatives. Thus, the idea with randomization is that the treatment will be the only thing that will constitute a significant difference between the patient groups.

There is evidence in the literature that randomization and RCT-quality often are suboptimal, that is often clinical trials described as RCTs are not really RCTs (7). The main problem is that the randomization is not assessed correctly. If the random distribution had been carried out with flip of a coin or throwing a dice, these procedures carries the risk of tampering, and one might be tempted to remake the procedure so it fits. Furthermore, allocation based on deterministic measure like date of birth, day of week or file number are not considered true randomization. The best way to randomize is to allow an independent person or centre to assess the randomization and assignment protocol, and thus, not involve any persons associated with the trial in the randomization procedure.

Consort
To improve the reporting of RCTs, a group of scientists and editors has published the Consolidated Standards of Reporting Trials (CONSORT 2010). The CONSORT statement enables the readers to understand a trial’s design, conduct, analysis and interpretation, and to assess the validity of its results. It emphasizes that this can only be achieved through complete transparency from authors. Consequently, CONSORT 2010 for example helps the authors to improve the reporting of a two-parallel design RCT by using a checklist and flow diagram. The most up-to-date revision of the CONSORT statement can be freely viewed and downloaded from the website: http://www.consort-statement.org/consort-statement/.

Important notes
Although, considerable weight is placed on the evidence from RCTs, this research method is not appropriate to answer every question. Valuable information can also be obtained from other levels of evidence and each has its role to play in providing evidence about the treatment we provide for our patients. However, if a clinical research question regarding evaluation of different treatment methods is aimed to be properly answered it is recommended primarily to perform an RCT to minimize bias. Consequently, using RCTs will diminish clinicians or patients preference for certain treatment, and most important by the random allocation process, confounding factors, that is factors over which we have no control during the trial, will affect the various constituent groups equally.

While some claim that only evidence from RCTs should be considered, others maintain that the study design should be determined by the research question to be addressed. RCTs may be expensive and time-consuming and can sometimes be inappropriate for ethical reasons, especially when the control subjects remain untreated. Therefore, it is important to acknowledge that well-designed prospective and retrospective studies may also provide valuable evidence (8). However, because the limitations inherent in these study designs, the results must be carefully analyzed and interpreted with caution (8). In addition, health register data can be used for studies but also health registers have weaknesses. Establishment of health registers requires the approval of both authorities and patients. The data entry of diagnoses and treatment outcomes must be consistent and accurate, and hereby time consuming for the clinician and involve further costs. There is also a risk of neglect to report treatment failures similar to what have been experienced in retrospective studies where often only the successful cases are recorded and analyzed.

The study design of choice is often a subject of lively and detailed discussions. Not seldom, the RCT methodology has been condemned as study design and very often these opinions of RCTs are expressed by prominent scientists or eminent professors who support their arguments on their own so-called experience. Such a form of ‘eminence-based evidence’ is not productive and has very little to do with evidence-based medicine (EBM) or evidence-based dentistry (EBD). Moreover, there are some misconceptions about evidence-based evaluations such as all evidence that has not been scientifically evaluated should be dismissed. It is important to bear in mind that ‘lack of evidence does not necessarily imply lack of effect’. Instead, there is need for further relevant evaluations. Clearly, new well-designed RCTs and non-randomized studies can achieve important support to reliable evidence in orthodontics.

Recommendations
If a clinical research question regarding evaluation of different treatment methods is aimed to be properly answered it is recommended primarily to perform an RCT to minimize bias. Consequently, using RCTs will diminish clinicians or patients preference for certain treatment, and most important by the random allocation process, confounding factors, that is factors over which we have no control
during the trial, will affect the various constituent groups equally. When comparing different interventions the following approaches can be strongly recommended:

- create a relevant question—use PICO which means: P-population, I-intervention, C-control, and O-outcome. Below an example of a research question that follows PICO: ‘Is Quad-helix treatment (intervention) more cost-effective (outcome) than expansion plate treatment (control) in 8–10 year-old patients with unilateral posterior crossbite (population)?’
- use RCT design with proper randomization
- plan and coordinate the trial carefully—have patience since it takes time to run an RCT
- have sufficient amount of subjects, that is make a sample size estimation—large materials are preferable to small, thus, with small materials there is greater risk of chance due to unknown or confounding factors that may interfere the outcomes
- use valid and reliable methods
- follow the CONSORT statement
- use blinding if possible
- use the ITT approach to evaluate the results

**Part II: RCTs are not for orthodontics!**

Sabine Ruf

When I was asked by the EJO Editor David Rice to take part in the Open session debate and was assigned the role to argue against RCTs, it was clear that I had the role of the Devil’s advocate. Taking the position against the accepted norms and arguing in favour of those types of research condemned by EBM and EBD is a challenging job which cannot be accomplished alone. Therefore, in the following, I have ‘engaged’ researchers from inside and outside the field of orthodontics in the argumentative discussion process.

**Randomized controlled trials**

When thinking about the basic goal of research within any given field, there is no doubt that it is to seek for the truth, for whatever question is at hand. This naturally also applies to orthodontic research. But how do we find the truth? According to the hierarchy of evidence for the evaluation of health care outcomes ([9](#))—the best way for seeking the truth are RCTs. They are considered as the gold standard because they deliver the highest level of evidence, due to their potential to limit all sorts of bias. According to the ‘Classical EBM ideology’ the role of RCTs is beyond any question, which is why Sackett et al., ([10](#)) recommended: ‘if you find a study was not randomized, we’d suggest that you stop reading it and go on to the next article’.

Sackett’s statement was based on the assumption, that a RCT does not only limit bias, but is entirely bias-free. However, time has evolved and scientists in all medical and dental disciplines have realized that this is not the case. Therefore, according to the currently accepted ‘New EBM ideology, RCTs may minimize, but do not eliminate bias’ ([11](#)). In addition Kapchuk ([12](#)) in his article ‘The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf’ pointed out, that ‘RCTs can introduce their own deviations from truth’. Despite this fact, there are still ‘way too many EBM advocates who are way too ready to blindly wield EBM like a mighty sword without understanding its limitations’ ([13](#)). And these advocates from my point of view are still too frequent in the field of orthodontics. In contrast, I think, that RCTs are not useful for orthodontic research, at least not for the crucial clinical questions of the specialty.

**RCTs in orthodontics**

Looking back, for example on what RCTs have told us about the treatment of Class II malocclusions, Meikle ([14](#)) concluded: ‘if one asks whether RCTs have achieved their intended objective, or provided knowledge not previously available from retrospective studies or animal experimentation, the answer would have to be no’. This statement becomes even more important if one considers that Class II malocclusions are the topic of the specialty with the largest amount of RCTs available.

In search of an answer, why RCTs are not ‘working’ for crucial clinical orthodontic questions, we have to recall the primary goal of RCTs, which is to test whether an intervention works by comparing it to a control condition. Such a control condition could be no treatment, a placebo treatment or an alternative treatment. And here lies the problem, because orthodontics is a device-driven specialty! ‘Orthodontic treatments are not a series of pills that can be administered at random and evaluated blindly’ ([15](#)). The opposite is the case, because in orthodontics even ‘invisible’ appliances are visible to the patients and even if hardly visible, in any case they are perceivable for the patient. This in turn excludes the possibility of orthodontic placebo treatments at least for the majority of the questions.

Of course there is always the alternative to design a RCT with an untreated control group. However, even if we would put aside ethical concerns, I can only think of only a few classical clinical orthodontic research questions that would warrant such a RCT study design. The latter is even more true if we stay realistic and do not ignore the decades of available clinical experience in orthodontics. This includes to unconditionally asking for untreated control groups in ridiculous conditions in which no control group is required to derive a clinical relevant and valid answer. Some facts are simply not in need of a Level I evidence-based proof! This applies for example to the fact that malocclusions, at least severe ones do not self-correct. In this context, I can recommend reading the satirical call for an RCT on parachute use to prevent death and major trauma related to gravitation challenge ([16](#)). Such an RCT is urgently needed as ‘parachute effectiveness has never been proven by RCT!’ ([16](#)).

**Bias immanent to RCTs**

In summary, for the purpose of orthodontic research on the RCT level, we are basically only left with the comparison of different treatment modalities. Setting up such an orthodontic RCT is no doubt possible, but does it always make sense? The answer to this question will depend on the kind of truth we are actually seeking, because ‘a trial is not a neutral condition! Thus, the effects under clinical conditions are different from those under experimental and double blind conditions’ ([17](#)). Why? Because, of the effect of ‘active research participation’ ([18](#)), which indicates that, ‘trial recruitment and retention depend on a set of convictions forged largely as a result of contextual factors peripheral to the intervention, including the friendliness and helpfulness of research centre staff’. Furthermore, ‘participants in clinical trials are actively involved in shaping the research process, rather than passive recipients of treatment. Thus, the outcomes of trials, notably those involving contact interventions’, which are daily business in orthodontics, ‘should be regarded not as a matter of fact, but as products of complex environmental, social, interpretive and biological processes’.

Already an informed consent mandatory for every patient taking part in a RCT may affect the treatment response to an extent that might be clinically most relevant. In 1994, Bergmann et al. ([19](#)) shortly before informed consent became mandatory in France, sat up a RCT on the effect of informed consent on analgesic activity of pain killers. They included 49 consecutively hospitalized patients with...
mild to moderate cancer pain in their study and randomly assigned them to be or not to be informed, that they were part of a RCT. In a second step, all patients in both groups were randomly assigned to the pain killer (naproxen) or placebo groups. The results (Figure 3) with respect to the analgesic activity in the ‘normal’ uninformed group were as expected—the pain killer reduced the pain while the placebo did not. The informed consent changed these effects dramatically—with informed consent both the placebo and the pain killer reduced the cancer pain significantly more and for a longer period than under the ‘normal’ clinical conditions. Such effects are not a problem unique to Medicine, because each and every trial in orthodontics also requires an informed consent and in addition, as indicated above, can neither be blinded nor double blinded. In turn both the patients and the operators will adapt their behaviour and attitudes, which will influence the results to a clinically significant extent also in for orthodontics. A very good example for this interrelation can be seen in the RCT by Sandler et al. (20). Their material comprised of 51 patients randomly assigned to either headgear wear or mid-palatal implant insertion to reinforce anchorage during fixed appliance treatment. Before treatment both groups had similar PAR-scores. The results revealed, comparable amounts of PAR-score reduction in both groups thus, both the headgear and the palatal implant were equally effective. There was also no difference in treatment length between the groups—thus also the efficiency of the two treatment modalities was equal. But is this really true? It is some sort of truth, but not clinical truth, it is a truth modified by the trial condition itself because, as Sandler et al. (20) stated—the headgear patients surprised the clinicians with the speed and efficiency of the method…cooperation was beyond the level that would normally be expected…we believe that we witnessed the Hawthorne effect!’

The Hawthorne effect refers to a phenomenon whereby individuals improve or modify their behaviour in response to their awareness of being observed (21). We must be aware of the fact, that RCTs in orthodontics are extremely susceptible to Hawthorne effects, because every orthodontic treatment success depends to some extent on patient cooperation. So given a trial changes cooperation, all orthodontic treatment modalities requiring cooperation cannot be tested reliably using a RCT design, at least not if we are seeking clinically relevant truth. And that is maybe ‘why RCTs in orthodontics have not achieved their intended objective’—as Meikle (14) stated already in 2005.

So we have to accept, that for many orthodontic research questions it will either not be possible at all or not be sensible to conduct a RCTs, because of the difficulties associated with undertaking them (22):

- high costs,
- ethical problems,
- informed consent, parents’ consent,
- bias problems,
- clinicians/patients preference for certain treatment,
- recruiting sufficient patients,
- overestimation of eligible/willing to participate,
- etc…

Alternatives to RCTs

The above mentioned arguments naturally raise the question where to go from here, or in other words, what kind of research to perform in the future. ‘The popular belief, that only RCTs produce trustworthy results and that observational studies are misleading does a disservice to patient care, clinical investigation, and the education of health care professionals’ (23).

So walking down the hierarchy of evidence ladder in our mind (Figure 1), the next type of studies to reflect about would be cohort studies or in other words non-randomized prospective studies. However, the only main difference between cohort studies and RCTs is the type of allocation, which is not random. This in turn means, that from the point of view of orthodontics the majority of the aforementioned disadvantages of RCTs apply to cohort studies as well. So this brings us down to the level of case–control studies and retrospective studies. I personally fully agree with Johnston (15) that we can ‘only move forward in orthodontics by looking back’. However, not in the way we have done in times past but by improving the quality of retrospective clinical studies. ‘Today’s most pressing clinical question is the recognition and, to the extent possible, the elimination of various biases that beset the average retrospective investigation’ (15).

If we compare RCTs, cross-sectional studies, cohort studies and case–control studies it become clear, that the three types of observational studies have the potential for a higher external validity than RCTs or in other words a higher degree to which the derived data can be extrapolated to the general population, which is of course desirable. On the other hand, they have the disadvantage of a lower internal validity, because it is much more difficult to eliminate and/or control different sorts of bias and confounding variables (24). So how can we improve retrospective studies? Instead of investing millions of euros, dollars, or pounds in RCTs only, we could alternatively or additionally invest in a sort of International Orthodontic Registry in which the registration of orthodontic cases with clearly defined malocclusion characteristics would be compulsory. Nevertheless, also such registry studies are of course not bias-free but they are a primary source for population-based case–control studies with several advantages, the most important ones being (25):

- the possibility of large sample sizes,
- data collected in a systematic manner,
- studies can be undertaken much faster,
- the costs are relatively low,
- studies of rare effects and rare diseases are possible,
- no informed consent is required and
- the analysis of long-term effects is much easier.

Figure 3. Changes (mm) in VAS pain levels after naproxen or placebo intake in patients with and without informed consent [adapted from Bergmann et al. (19)].
So, concluding it can be said, that ‘experiment, observation, and mathematics, individually and collectively, have a crucial role in providing the evidential basis for modern therapeutics. Arguments about the relative importance of each are an unnecessary distraction. Hierarchies of evidence should be replaced by accepting—indeed embracing—a diversity of approaches’ (26).

References