

# RANDOMIZED CONTROLLED CLINICAL TRIALS IN ORTHOPEDICS: DIFFICULTIES AND LIMITATIONS

Eduardo Angeli Malavolta<sup>1</sup>, Marco Kawamura Demange<sup>2</sup>, Riccardo Gomes Gobbi<sup>3</sup>, Marta Imamura<sup>4</sup>, Felipe Fregni<sup>5</sup>

## ABSTRACT

Randomized controlled clinical trials (RCTs) are considered to be the gold standard for evidence-based medicine nowadays, and are important for directing medical practice through consistent scientific observations. Steps such as patient selection, randomization and blinding are fundamental for conducting a RCT, but some addi-

tional difficulties are presented in trials that involve surgical procedures, as is common in orthopedics. The aim of this article was to highlight and discuss some difficulties and possible limitations on RCTs within the field of surgery.

**Keywords** - Evidence-Based Medicine; Orthopedics; Randomized Controlled Trials as Topic; Comparative Study

## INTRODUCTION

Randomized controlled clinical trials (RCTs) are considered to be the gold standard for evidence-based medicine today<sup>(1-3)</sup>. They have the aim of creating a basis for supporting the use of one therapeutic method or another by means of analyzing different interventions, or through comparing interventions with placebo. Evidence-based medicine thus has the objective of directing medical conduct through consistent scientific observations.

The level of evidence in clinical studies can be classified in the following manner, ordered from the lowest to the highest level: case reports, case-control studies, prospective observational studies and RCTs. The main problems in observational studies are selection bias and lack of a control group. In this respect, the main advantage of RCTs is precisely the randomization, which diminishes the chances of “confounder” effects and selection bias.

For clinical trials to have the potential to suggest to physicians that a certain therapeutic method is the one most indicated in a given situation, they need to present internal and external validity. In fact, studies have to answer two essential questions posed by clinicians: “are the results valid?” (which can be assessed through the internal validity of the study); and “are the results applicable to my patient?” (which can be assessed through the external validity of the study). If a study presents internal validity, this means that the study design was appropriate for answering the question posed in relation to the sample studied. In turn, external validity implies the potential for the results from that trial to be extrapolated to other populations, i.e. such that the results are valid not only for the participants in that study, but also for others<sup>(4)</sup>. Possessing adequate internal validity is a condition sine qua non for possessing external validity: in other words, methodological problems that affect the study

1 – Attending Physician in the Shoulder and Elbow Group, Institute of Orthopedics and Traumatology, School of Medicine, University of São Paulo (USP), São Paulo, Brazil.

2 – MSc in Medicine. Attending Physician in the Knee Group, Institute of Orthopedics and Traumatology, School of Medicine, University of São Paulo (USP), São Paulo, Brazil.

3 – Attending Physician in the Knee Group, Institute of Orthopedics and Traumatology, School of Medicine, University of São Paulo (USP), São Paulo, Brazil.

4 – PhD in Medicine. Collaborating Professor in the School of Medicine, University of São Paulo (USP); Physician in the Division of Physical Medicine and Rehabilitation, Institute of Orthopedics and Traumatology, School of Medicine, University of São Paulo (USP), São Paulo, Brazil.

5 – PhD in Medicine. Head of the Neuromodulation Laboratory and Clinical Research Teaching Center, Harvard Medical School, Boston, MA

Work performed at the Medical Investigation Laboratory for the Musculoskeletal System (LIM 41), Department of Orthopedics and Traumatology, School of Medicine, University of São Paulo (USP).

Correspondence: Rua Ovídio Pires de Campos, 333, Cerqueira Cesar – 05403-010 – São Paulo, SP – E-mail: eduardomalavolta@hotmail.com

Work received for publication: December 20, 2010; accepted for publication: March 21, 2011.

The authors declare that there was no conflict of interest in conducting this work

This article is available online in Portuguese and English at the websites: [www.rbo.org.br](http://www.rbo.org.br) and [www.scielo.br/rbort](http://www.scielo.br/rbort)

itself make it impossible for its results to be extrapolated, since these results may not be valid even for the sample studied<sup>(5)</sup>.

Designing a perfect study with minimal bias may not be feasible, because this may make it unviable to carry out the study (for example, through greatly increasing the cost or the number of individuals that would have to be included). For this reason, for most studies, a balance has to be sought between feasibility and methodological rigor, with a design in which great care is taken to avoid errors that might invalidate the study internally. Steps such as patient selection, randomization and blinding are fundamental and give rise to certain additional difficulties in trials that involve surgical procedures, as is common in orthopedics. The aim of this paper was to highlight and discuss some difficulties and possible limitations of RCTs within the field of surgery.

## DIFFICULTIES AND LIMITATIONS

### Design and initial question

The design of any study should be chosen based on a specific question that corresponds to the problem to be investigated. A very broad initial question makes it difficult to calculate the sample size and choose the outcome variables. For example: what is the result from uncemented total knee arthroplasty? If the question is more specific, this helps to delineate the study design: “Does the functional result over the first year after total knee arthroplasty differ between uncemented and cemented procedures?” In the initial question, the population of interest, primary outcome, length of follow-up until the primary outcome and the comparison group should be defined.

Some study designs present limitations in surgical trials. For example, crossover designs cannot be implemented. In studies within orthopedics, the intervention is often permanent and there is no “washout” as occurs with use of medications. In addition, patients cannot be used as their own controls for two successive interventions in the case of single organs. In the case of double organs (surgery on both knees, for example), patients at the time of the second intervention are not the same as at the time of the first intervention, because they enjoy the benefits from the previous intervention. However, one possible design

consists of comparing effects between double organs if the level of disease is similar, which rarely occurs.

Furthermore, it needs to be emphasized that the use of placebo groups in surgical trials frequently presents significant ethical limitations, and this may even be impossible to implement<sup>(6)</sup>.

### Sample calculation

Another important limitation that surgical trials frequently present consists of the small numbers of cases studied and its consequent implication for the statistical power of the study. The power of a study depends on the sample size, the homogeneity of the sample, the homogeneity of the results (standard deviation of the results) and the differences between the means of the results in each group. In surgical studies, it is often difficult to reach the minimum number of patients (sample size) needed to attain a statistical power of 80%, the percentage that is considered to be the minimum acceptable in most clinical trials. This occurs for a variety of reasons: the cost of the study (usually higher when surgical procedures are involved) may make it unviable to reach the ideal number; the low frequency of some pathological conditions and consequently of surgical procedures on these conditions; and difficulties in calculating what the ideal sample size should be, especially when new surgical procedures are involved, because of the imprecision in estimating the magnitude of the treatment effect; among others. As an example, Freedman et al<sup>(7)</sup> showed that only 9% of orthopedic trials included an a priori calculation of the sample size.

In this light, proposals to conduct multicenter trials are often made, with the aim of expanding the numbers of individuals. In such situations, bias relating to standardizing the intervention may arise. It is difficult to affirm that surgical interventions made by different centers are the same, given that there is variability between surgeons and many other variables are included. For example, there are the types of anesthesia, differences in anesthesia equipment, differences in surgical equipment, the methods of sterilizing surgical materials, the nursing team’s expertise in postoperative care, the organization of the team at the surgical center (thus influencing the duration of surgical procedures), differences in the availability of technology (such as the quality of technological equipment

for radiology and microscopy) and differences in the bacterial flora between hospitals, among various other factors (both detectable and undetectable) that influence the final result. It may even be argued that these biases would influence both arms of the study, and would be diluted with increasing numbers of individuals. This may be true in some cases, but it may generate an inverse effect in others, when certain factors cause bias in only one of the arms (use of radiology, when necessary in only one of the surgical techniques studied, for example). Moreover, in cases in which increases in the numbers of individuals may dilute all this variance, an enormous sample is needed, thus implying a study of dimensions that are often impracticable from a financial point of view. On the other hand, RCTs with large samples are generally sponsored by the industry and are conducted by researchers with links to these companies, which creates conflicts of interest and considerably increases the risk of bias<sup>(8, 9)</sup>.

On the other hand, increasing the heterogeneity of a study will increase its external validity. In fact, one of the factors that explain the extremely favorable results seen in some small studies within orthopedics may be the presence of excellent conditions at the place where the study was conducted that would be difficult to reproduce in other locations.

One other important point is that even if studies present low power, they may be valuable when aggregated in meta-analyses. Likewise, such studies may supply preliminary data for calculating sample sizes and assessing the feasibility of conducting similar studies in the future.

It needs to be emphasized that calculating the sample size should always be done before starting the study: it should never be done subsequently or with the aim of validating the result.

### **Type I error**

Type I error (or alpha error) consists of concluding that a difference exists between the study groups when in reality it does not exist (also called a false positive). It is usually considered acceptable for the chance that this error could occur to be up to 5%. One common cause of this error comes through carrying out many statistical tests on different hypotheses, until finding a positive result. The best way to avoid this situation is

to define in advance what statistical test will be used at the end of the study, along with the main question of the study, thereby minimizing the number of tests for the primary outcome. Another situation that may favor this error is to use multiple outcomes with many variables, which increases the final number of tests. However, for secondary outcomes, the number of tests does not influence the type I error of the study.

In surgical RCTs, it is common to use several outcome measurements, given that the results from a surgical procedure can be analyzed in different manners: pain scales, functional scales, quality-of-life scales, satisfaction scales and complementary examination scales. Once again, the study design should be directed towards answering just one specific question. The other outcomes should be analyzed as secondary outcomes.

### **Type II error**

Type II error consists of concluding that there is no difference between the study groups when in reality a difference exists (false negative). By convention, the limit for occurrences of this error is generally set at a chance of 20% (in other words, a study power of 80%). The causes of type II error include insufficient numbers of patients in the study, generally because of lack of sample size calculation or error in this calculation, or difficulty in obtaining sufficient numbers of subjects, as discussed above. In a review of the literature on orthopedic trauma, Lochner et al<sup>(10)</sup> showed that the incidence of this problem was 90.52% among a total of 117 randomized studies evaluated. In small studies, in which the researchers conclude that there is a high possibility of type II error, one solution is to increase the homogeneity of the study through diminishing the variance and increasing the power of the study. However, this measure reduces the external validity of the study.

### **Intention to treat**

In clinical trials, it is suggested that the analysis on the data obtained should be done based on the group for which the patient was selected. This principle is called intention-to-treat analysis. This method for assessing the results makes it possible to protect the randomization, which is a fundamental point in relation to RCTs.

In surgical trials, following this principle may ge-

nerate strange or even incongruent responses, depending on how this method is applied. For example, a patient drawn for non-surgical treatment who, for any reason, then undergoes surgical treatment should, according to the intention-to-treat principle, still be analyzed as a non-surgical case. If this patient happens to present infection at the surgical site, the results from such a study would show “occurrence of infection of the surgical site” as a “complication from non-surgical treatment”. Hence, the intention-to-treat principle should be applied using data prior to performing the rescue surgery, or else the results will be biased.

Although drop-outs are a frequent problem in clinical research, this problem is different in orthopedics, since the intervention is generally applied fully to the patients (thus differing from studies on drugs, in which the patients can drop out at any time, from the beginning to the end of the study). In this respect, studies within orthopedics present an additional advantage. Investigators in this field should therefore undertake frequent measurements if there is any drop-out or crossing of treatments.

In a recent study, Herman et al<sup>(11)</sup> observed that only 16.4% of RCTs within orthopedics between 2005 and 2008 had adequately applied this principle. Most of these studies had excluded the patients lost during the follow-up, from the final statistical analysis. This was notably done in trials that involved surgical procedures. Omission of these data may lead to bias, because it affects the integrity of the randomization.

### External validity

When extrapolation of the results from a clinical trial is desired, the premise taken is that the intervention undertaken or that is intended will be the same at all locations or for all physicians in different locations. This is because both the formula of the drug and its pharmacokinetics and pharmacodynamics are the same, regardless of who is prescribing it. The problem is that the external validity of RCTs is often low. This occurs in surgical trials for three basic reasons: the surgeon, the environment or the patient.

Interventions made by one surgeon are not necessarily identical to those made by another. That is, no matter how reproducible a technique is, it is not identical. Moreover, new surgical techniques

generally depend on a learning curve, and this may vary for each technique and for each surgeon. In this way, even for the same surgeon, operations performed at different times may differ significantly. Thus, results from a surgical technique obtained by one group of authors may be different from those obtained by others, without implying methodological failings or random effects but, rather, as a result of the fact that the interventions are not identical. This unmeasurable phenomenon may imply that there is a limitation to the external validity of surgical studies, especially for results that are produced by one or just a few surgeons.

The environment of a teaching hospital, where many studies are conducted, also may not be representative of the general population. Furthermore, patients participating in studies tend to receive differentiated attention in relation to ordinary patients. Finally, the patients who agree to enter such studies and those who fit within the often restrictive inclusion criteria may not be representative of the general population.

Nonetheless, the particular features of surgical trials as described above do not justify avoidable methodological failures. As reported by Ahmad et al<sup>(12)</sup>, in a study on osteoarthritis of the hip and knee, researchers failed to describe details about their study. While the surgical procedure used was described in all studies, the pre and postoperative care and the anesthesia used were described in only 7%, 50% and 13%, respectively. Lack of information of this nature seriously compromises the generalization of the data and, consequently, the external validity.

### Recruitment

The acceptance rate for participation in a surgical RCT is generally less than 50%, and the main reasons for not participating that patients give are their preference for one of the arms of the treatment, their discontent with the randomization and the possibility of higher expenses<sup>(13)</sup>. New and experimental surgical procedures cause patients to be anxious and apprehensive to a greater extent than if they were faced with a new medication. This may be because although most people have already taken some form of medication during their lives, most of them have never been through surgery. It may also be because of real worry about sequelae caused by unsuccessful

surgery. Thus, recruitment of subjects for surgical trials is more difficult, which increases the study period required and often forces the sample size downwards. Although there is no ideal way to deal with this difficulty, conducting the study with linkage to large renowned institutions may stimulate acceptance. Small variations in established techniques or use of new surgical procedures to treat pathological conditions for which no effective treatment exists may also stimulate people to enter the study. No information on risks or existing evidence should ever be omitted from the free and informed consent statement, which it is mandatory for all patients to sign before entering the study, as required by the institutional ethics committees.

Another important problem regarding recruitment is selection bias. Patients who agree to participate in the study may have characteristics differing from those who are receiving treatment at a clinical center. Within this context, it is important that the group that did not agree to participate in the study should be compared with the group that did agree to this. If there are differences in any factors, these should be included in a covariance analysis.

### **Randomization**

The random nature of patients requires that both surgical techniques should be performed with observance of randomization. Any limitation on carrying out one of the techniques (only on certain days of the week, only by a certain surgeon, etc) implies a risk of harming this principle. For example, in cases of emergency or trauma, in which the procedure is complex and is done by a particular surgeon, the randomization would be subject to the availability of the physician and thus allocation according to convenience would be required, thereby impairing the randomization.

When the study involves comparison between different implants, further difficulties arise. Ideally, the draw should be made at the time of the surgery. However, there are operational and financial disadvantages in this, resulting from transportation and sterilization of the materials, especially in the case of implants supplied as a batch, when they do not form part of the hospital's arsenal (both of the implants under examination should be available and sterilized within the operating room). On the other hand, if the draw is done earlier, there is a considerable risk of

loss of blinding, given that a series of individuals (in the materials center and nursing team) will know which group the subject is in. Furthermore, if the surgeon's blinding is removed, he could introduce bias, even if unconsciously.

Another difficulty in the randomization occurs when inclusion of a given patient can only be defined after intraoperative evaluation. For example, in a study on meniscal sutures, the patient can only be included in the study if the suture is possible. Thus, the team needs to be prepared to perform randomization during the operation or to use a preoperative randomization method that does not create imbalance between the groups if a patient is excluded during the operation.

Among the RCTs within orthopedics that were published between 1988 and 2000, only 41% presented randomization that was appropriate, according to Bhandari et al<sup>(14)</sup>.

### **Protection of randomization (concealment)**

The randomization methods that are most used (sealed envelopes and lists generated using computer software) are susceptible to bias. Regarding sealed envelopes, they should be sealed in such a way that opening them without tearing them is impossible, and should be made of opaque material such that seeing through the envelope by holding it up to the light is impossible<sup>(1)</sup>. Lists generated using software need to be kept protected by the principal researcher throughout the study period, under the care of someone who is not participating in the trial. The ideal methods for protecting the randomization list would be those provided by outsourced companies that are available 24 hours a day, on the internet or by telephone. Use of dates of birth or hospital registration numbers are not considered to be valid forms of randomization, not are lists that are open for everyone to read.

As stated above, to protect the randomization in trials that compare two surgical procedures, it should preferably be done inside the operating theater, after anesthesia, thus diminishing the chance of unmasking the preoperative data-gathering and the patients themselves.

### **Blinding**

Blinding or masking is an important part of conducting RCTs, in order to minimize bias. Double-blind studies (both the patient who receives the intervention

and the physician or the researcher conducting the study are blinded) are the type of RCT most used. Sometimes a triple-blind design is suggested (with additional blinding for the person who analyzes the data and the person who writes the text)<sup>(15)</sup>. These terms may cause confusion for readers, and the recommendation that is currently most accepted is that descriptive reports on who was blinded in the study should be provided<sup>(16)</sup>.

Blinding in non-pharmacological trials has already been shown to be more laborious than in pharmacological trials, in a comparative study by Boutron et al<sup>(17)</sup>. The difficulties relating to blinding in RCTs that involve interventions relating to surgical procedures are described below.

#### a) Blinding of the surgeon

Two situations may be presented: studies comparing two surgical interventions and studies comparing a surgical intervention with non-surgical treatment.

For the surgeon who performs the operation, if the operative techniques differ (because of different access routes or implants), blinding is impossible<sup>(18)</sup>. It could be argued that if the postoperative data gathering was blinded, a source of bias would be minimized, but if the surgeon believes more in one of the techniques, he might make greater efforts towards this one, thus inducing randomization errors. If the operative technique is identical, and the intervention consists of introducing an additional factor (injection of growth factor, postoperative medication, use of a new suture, etc), blinding may be possible, even for the surgeon, provided that the intervention allows blinding with a placebo or has the same physical characteristics as the control.

If the study involves a non-surgical group, obviously it is not possible for the surgeon to be blinded.

It needs to be emphasized that in studies in which the surgeon is not blinded, some data gathered during the surgical procedure (for example, the volume of bleeding) are subject to data-gathering bias.

#### b) Blinding of the patient

This form of blinding is very difficult to achieve in comparisons between surgical and non-surgical treatments, for obvious reasons.

When the aim of the study involves comparison between two different surgical techniques, there is a

new difficulty when the access routes are different. During the follow-up, if the patient has access to the radiographic examinations (which usually occurs), differences between the implants used may also be noticed and, in this way, the reliability of the data to be analyzed may be altered.

Studies with practically “perfect” blinding of patients do exist, like the study described by Moseley et al<sup>(19)</sup>. In this clinical trial on the efficacy of knee arthroscopy for treating osteoarthritis of the knee, three groups were evaluated: arthroscopic debridement, simple lavage of the joint and placebo surgery in which only the skin incisions were made (sham surgery). The placebo group watched a video simulation of the surgery during their procedure. However, the ethical implications of studies like this are evident, and it is rare to obtain approval for such studies from research ethics committees and from patients<sup>(20, 21)</sup>. It needs to be borne in mind that if such a study has any blinding problem at other stages (such as in relation to allocation or data gathering), placebo surgery will become completely anti-ethical. Moreover, it should be remembered that surgery does not consist of just an incision and skin suturing. If the surgery generates other signs (such as joint effusion or hematoma, etc) during the postoperative period, the blinding may be compromised and placebo surgery (just an incision, for example) may not be justified.

If there is difficulty with blinding, the research group should consider alternative solutions, such as the use of objective outcomes (like laboratory measurements) or short-term outcomes in which it is more feasible to have the patient blinded for a shorter period, for example during the hospital stay.

#### c) Blinding of the assessor

The independent assessor, who is generally a physician or physiotherapist who is not participating directly in the study, is an important player in RCTs within orthopedics. Functional scales, which are used in virtually all studies of this type, are applied by this assessor. The assessor may present loss of blinding, especially if the procedures under evaluation are carried out using different access routes, which gives rise to surgical scars that lead to identification of the group. Masking the scar in all the evaluation, through

the use of appropriate clothing, is one way to ensure this principle, but in daily practice this may not be achieved, especially during physiotherapy sessions.

When different surgical techniques require different rehabilitation protocols, the blinding may again be lost if the physiotherapist who is in attendance at the regular sessions is the same one who applied the functional scales.

In a systematic review, Poolman et al<sup>(22)</sup> observed that only 50% of the RCTs within orthopedics implemented this form of blinding. In addition, they showed that the effect of the treatment was significantly greater in the studies without blinding thereby making it clear that assessment bias was present.

#### d) Blinding of the data assessor

This blinding is the simplest type to obtain. It is enough for the results spreadsheet not to contain a description of the group to which the patient belongs, but only numbers, and it can be achieved in virtually all RCTs. However, no article published in an orthopedics journal cited this method between the years 1988 and 2000, according to Bhandari et al<sup>(14)</sup>.

### Adherence

Differing from the various clinical conditions that require continual regular checks over long periods, and possibly throughout the patient's life, such as in cases of hypertension or diabetes, orthopedic interventions are often "curative". Acute conditions (fractures) and even chronic conditions (osteoarthritis undergoing arthroplasty, for example) present significant improvement of symptoms over the short and medium terms and, in the absence of complications, it may be difficult to keep up the follow-up for a long period, thus giving rise to loss of subjects.

Another difficulty relating to patient adherence to orthopedic surgical protocols is the need for rehabilitation. Taking a pill at home demands less from patients than does leaving home and traveling to the physiotherapy location, when the patient is in pain after undergoing surgery. Since lack of adequate rehabilitation or dropping out from rehabilitation is often an exclusion criterion of the study, this may compromise the final result from the surgery. Therefore, this

problem should be borne in mind when developing the protocol. Facilitating and simplifying the postoperative procedures as much as possible stimulates better patient adherence.

### FINAL REMARKS

RCTs have excellent application for assessing new medications. Today, for a new medication to be approved by the United States Food and Drug Administration (FDA)\*, it has to go through assessments at several levels (studies on animals, safety analysis and clinical studies). Before regularization of the product, so-called phase III studies involve thousands of patients in a randomized manner. Even after the drug has been made commercially available, monitoring of even bigger patient samples continues. In the case of surgical interventions, there is no standardization of specific rules for their approval, and RCTs are not required for a given surgical technique to be incorporated into clinical practice. It is usually enough to have case series showing good clinical results, for the technique to be put into use.

RCTs present various advantages, especially with regard to diminished bias in data gathering and analysis, thus justifying their great prestige within medical research. However, the limitations and difficulties in applying their principles to studies within the field of surgery are far from few. These particular features give rise to disadvantages when seeking to publish in general medical journals, where the competition for editorial space is fierce.

One important point is that many of the difficulties in conducting RCTs can be resolved through alternative solutions, like some that have been discussed in this paper. In other cases, researchers should be aware of the limitations and assess whether the data gathered will be valid. It is worth emphasizing that open observational studies are also important, especially at the initial stages of testing a new intervention. However, in such cases, well-designed data analysis techniques are fundamentally important, with construction of multivariate regression models and controls for potential confounding factors.

Concluding, orthopedists should prioritize conduc-

\* U.S. Food and Drug Administration (FDA) website; Investigational New Drug (IND) Application. Available at: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/InvestigationalNewDrugINDApplication/default.htm>

ting RCTs whenever feasible. Although knowledge derived from other study designs that are considered to present a lower level of evidence (case-control, cohort, case series, descriptions of techniques and

specialists' opinions) is also important, the bias present in these studies and the validity of their results should be interpreted critically and, if possible, should be analyzed together with the results from RCTs.

## REFERENCES

- Zlowodzki M, Jonsson A, Bhandari M. Common pitfalls in the conduct of clinical research. *Med Princ Pract.* 2006;15(1):1-8.
- Chaudhry H, Mundi R, Singh I, Einhorn TA, Bhandari M. How good is the orthopaedic literature. *Indian J Orthop.* 2008;42(2):144-9.
- Soucacos PN, Johnson EO, Babis G. Randomised controlled trials in orthopaedic surgery and traumatology: overview of parameters and pitfalls. *Injury.* 2008;39(6):636-42.
- Portney LG, Watkins MP. Validity in experimental design. In: Portney LG, Watkins MP, editores. *Foundations of clinical research - applications to practice.* 3rd ed. New Jersey: Pearson Prentice Hall; 2009. p. 161-91.
- Paradis C. Bias in surgical research. *Ann Surg.* 2008;248(2):180-8.
- Black N. Evidence-based surgery: a passing fad? *World J Surg.* 1999;23(8):789-93.
- Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. *J Bone Joint Surg Br.* 2001;83(3):397-402.
- Bhandari M, Jönsson A, Bühren V. Conducting industry-partnered trials in orthopaedic surgery. *Injury.* 2006;37(4):361-6.
- Lynch JR, Cunningham MRA, Warme WJ, Schaad DC, Wolf FM, Leopold SS. Commercially funded and United States-based research is more likely to be published; Good-quality studies with negative outcomes are not. *J Bone Joint Surg.* 2007;89(5):1010-8.
- Lochner HV, Bhandari M, Tornetta P. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg.* 2001;83(11):1650-5.
- Herman A, Botser IB, Tenenbaum S, Chechick A. Intention-to-Treat Analysis and Accounting for Missing Data in Orthopedic Randomized Clinical Trials. *J Bone Joint Surg Am.* 2009;91(9):2137-43.
- Ahmad N, Boutron I, Moher D, Pitrou I, Roy C, Ravaud P. Neglected external validity in reports of randomized trials: the example of hip and knee osteoarthritis. *Arthritis Rheum.* 2009;61(3):361-9.
- Abraham NS, Young JM, Solomon MJ. A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. *Surgery.* 2006;139(4):469-83.
- Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in *The Journal of Bone and Joint Surgery* from 1988 through 2000. *J Bone Joint Surg Am.* 2002;84(3):388-96.
- Gotzsche PC. Blinding during data analysis and writing of manuscripts. *Control Clin Trials.* 1996;17(4):285-90.
- Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet.* 2002;359(9307):696-700.
- Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol.* 2004;57(6):543-50.
- Simunovic N, Devereaux PJ, Bhandari M. Design considerations for randomised trials in orthopaedic fracture surgery. *Injury.* 2008;39(6):696-704.
- Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med.* 2002;347(2):81-8.
- Heckerling PS. Placebo surgery research: a blinding imperative. *J Clin Epidemiol.* 2006;59(9):876-80.
- Hornig S, Miller FG. Is placebo surgery unethical? *N Engl J Med.* 2002;347(2):137-9.
- Poolman RW, Struijs PAA, Krips R, Sierevelt IN, Marti RK, Farrokhyar F, et al. Reporting of outcomes in orthopaedic randomized trials: Does blinding of outcome assessors matter? *J Bone Joint Surg Am.* 2007;89(3):550-8.