Practical Aspects of Randomization and Blinding in Randomized Clinical Trials

Article *in* Arthroscopy The Journal of Arthroscopic and Related Surgery · December 2003

DOI: 10.1016/j.arthro.2003.09.023 · Source: PubMed

5 authors, including:

Stephen Bridgman
University of East Anglia
65 PUBLICATIONS **2,614** CITATIONS
SEE PROFILE

Lars Engebretsen
University of Oslo
649 PUBLICATIONS **40,239** CITATIONS
SEE PROFILE

Katie Dainty
St. Michael's Hospital and the University of Toronto
92 PUBLICATIONS **2,573** CITATIONS
SEE PROFILE

Nicola Maffulli
Università degli Studi di Salerno
1,826 PUBLICATIONS **57,577** CITATIONS
SEE PROFILE

ISAKOS Scientific Committee Report

Practical Aspects of Randomization and Blinding in Randomized Clinical Trials

Stephen Bridgman, M.D., F.R.C.S., Katie Dainty, M.Sc., Alexandra Kirkley,† M.D., F.R.C.S.C., and Nicola Maffulli, M.D., M.S., Ph.D., F.R.C.S.(Orth)

Abstract: Randomization and blinding are important tools in determining the effectiveness of a new intervention and ensuring the validity of a clinical trial. However, randomness and haphazardness are not equivalent. Randomization cannot overcome poor experimental design or technique. Several types of randomization including historical controls and pseudorandomization are discussed, as well as methods of treatment allocation, stratification, and minimization techniques. The importance of decreasing bias and the advantages and disadvantages of blinding in randomized clinical trials are also covered. **Key Words:** Research methodology—Randomization—Treatment allocation—Bias—Blinding.

Randomized controlled trials are widely recognized as the standard for unbiased assessment of the effects of different medical interventions. The first randomized clinical trial was the British Medical Research Council evaluation of streptomycin in the treatment of tuberculosis in 1947, but the concept dates back as far as 1935 to R. A. Fisher's work in agriculture.

RANDOMIZATION

Randomization is the process by which units, which in health care are the patients, in a clinical trial are "randomly" assigned to receive one of the treatments under evaluation. Randomization is a key tool in an effectively run trial and is mainly used to reduce the bias in assigning patients to study treatment groups.

The purpose of randomization is to avoid bias in the judgement or systematic arrangement of treatment, and to provide a solid basis for statistical analysis such as significance tests. When designing a clinical trial, the effect of treatment being assessed is usually only moderate. Therefore, to be able to detect any treatment effect properly, external variables or "error" must be minimized.

The two main types of error are:

1. Random error caused by sampling: This type of error is unavoidable, and refers to the error resulting from studying a finite sample size to obtain an approximation of the true effect size on the population.
2. Systematic errors or biases: This refers to the influence of an individual's preferences on the allocation of a given patient to a treatment group.

By virtue of these two types of error, clinical trials for the most part need to have large sample sizes and

From the Orthopaedic Surgical Trial Unit (S. B.), Department of Trauma and Orthopaedic Surgery (N. M.), Keele University School of Medicine, Stoke-on-Trent, Staffordshire, England; and Fowler Kennedy Sport Medicine Clinic, University of Western Ontario (A. K., K. D.), London, Ontario, Canada.

† Deceased.

Address correspondence and reprint requests to Nicola Maffulli, M.D., M.S., Ph.D., F.R.C.S.(Orth), Department of Trauma and Orthopaedic Surgery, Keele University School of Medicine, North Staffordshire Hospital, Thornburrow Drive, Hartshill, Stoke-on-Trent, Staffordshire ST4 7QB, England. E-mail: n.maffulli@keele.ac.uk

0749-8063/03/1909-3890$30.00/0

doi:10.1016/S0749-8063(03)00838-7

must be randomized. Randomization ensures that imbalances, if any, arise purely from chance and can be taken into account using the appropriate statistics.

The objective of this article is to consider how to derive and implement a suitable randomization scheme as well as how to incorporate and maintain blinding in orthopaedic clinical trials. As has been realized, the data from nonrandomized, biased trials can be impossible to interpret and, therefore, the results can be inapplicable.

Uncontrolled Trials

In an attempt to give patients the best available care, most clinicians are always willing to try new treatments when they become available. The keen and adventurous physician may try the new treatment on a few patients and develop a theory on its effectiveness through this type of uncontrolled trial. Not only is the treatment studied without any direct comparison to a similar group of patients or more standard therapy, but the new treatment is often given to less seriously ill patients or falsely reported when failures occur. The conclusions drawn from the results of uncontrolled trials can be very misleading, as they may provide a very distorted view of therapy without sound scientific basis.

Historical Controls

Another method of conducting a comparison of a new treatment to a standard therapy has been to use an historical control group. In other words, investigators compare patients treated with the new therapy with a group of previous patients treated with the standard treatment. The major flaw with this type of comparison is that it is impossible to ensure that the comparison is fair. Historical controls are less likely to have clearly defined criteria for inclusion, the quality of the data collected is often inferior and haphazard, criteria responses are not well controlled, and it is very difficult to ensure that all aspects of patient care, other than the treatment under study, remain constant.

In the same way, literature controls, where the control group is made up of patients treated elsewhere and previously reported in the medical literature, offer equally poor comparison of treatments. These provide nothing but innumerable opportunities for the groups to differ significantly in patient selection, care, and data collection methods.

Pseudorandomization

Clinicians have attempted to use a variety of questionable methods of "pseudorandomization." This is more scientifically referred to as systematic assignment. Some common methods include patients allocated to treatments alternately, by date of birth, or by the day of the week of presentation to the clinic. Obviously, the main problem with these types of pseudorandomization is that they lead to bias on the basis of foreknowledge of treatment. The investigator can easily figure out in advance which treatment a patient would receive depending on their information. This leads to a multitude of problems. For example, if the physician already has strong views on a particular treatment, they can delay the referral, or change the order of randomization to suit their own needs. Unless there is no foreknowledge of treatment, the results of a trial will always be suspect. For example, in a trial of nonoperative treatment and surgery in patients with an Achilles tendon rupture, patients were allocated depending on the day they presented. Referring physicians were aware of this, and referred patients accordingly, depending on the treatment they wished for their patients. Clinical trials that use randomization rules that allow any possibility of foreknowledge of treatment cannot truly be termed randomized clinical trials.

Methods of Random Treatment Allocation

There are many acceptable methods for randomly allocating patients to treatment groups. The most direct method is known as simple randomization. This procedure uses a truly random code to assign all patients to receive one of two (or more) treatments. This can either be from a table of random numbers, computer-generated list, or simply by tossing a coin. This method requires large sample sizes (typically greater than 1,500) to allow for the assumption of similarity in important prognostic characteristics and numbers of patients in each arm of the study.

To counteract this sample size problem, many people prefer to use what is known as fixed or random block randomization. This allows blocks of treatments to be permuted into a random order; for example, with a block length of 4 and 2 treatments, the following orderings are possible AABB, ABAB, ABBA, BABA, BAAB, and BBAA. Putting blocks like this together to form a randomization list maintains balance in the treatment groups for every certain number of patients. The main advantage of using block randomization is that, if the trial does not enroll the full number of

patients expected, there will still be an equal or approximately equal number of patients in each treatment group for a balanced block.

With a fixed block length, such as four, the lack of foreknowledge may be compromised. For example, if a physician knows the block length and has already randomized 7 patients, then he or she will be able to calculate the outcome for the eighth patient. To decrease the chance of this foreknowledge bias occurring, variation in block length can be used.

Although a block permutation scheme solves the problem of an overall imbalance in treatment allocations, it does not ensure that patients in the two treatment groups have the same sort of presenting factors that influence prognosis or treatment responsiveness. For example, so far we have not described a way of ensuring that the two groups have, for instance, similar age profiles. An extension of the permuted block design known as stratification will allow for this. Stratification is a technique that divides the patients further into strata based on specific characteristics and allows a separate randomization list to be produced for each stratum. This will overcome group inequality and ensure that patients will not only be balanced overall but also will be balanced within their stratum. In the case of multicenter trials or trials with multiple physicians/surgeons, the effect of the institution/surgeon entering the patient should be considered for stratification. Different institutions can show very different patient response rates and populations for reasons of patient selection and data collection methods.[1] In orthopaedic surgery, stratification for the surgeon should also be considered due to differences in surgical technique and patient care between clinicians.

The stratified approach can be unworkable, however, if there are too many strata being considered. Patients should be stratified based on the smallest number of characteristics possible. A computer should be used for analysis to avoid small block sizes in the allocation process. Stratification should be performed at the beginning of the trial because, when performed retrospectively, it can involve subgroup analyses, which can be controversial.

There are other methods of treatment allocation that are termed "adaptive." In adaptive treatment allocation, future treatment allocations depend on what has already happened. The most general form of this is minimization, where the choice of treatment is the one that minimizes any imbalances between different prognostic variables. For example, the first 20 patients entered in a clinical trial are randomly assigned to treatment groups A or B by simple or block randomization methods. At that time, the pertinent characteristics of all patients in each group are assessed. Each subsequent patient who is entered in the trial is then assigned to a group so that the total difference in the groups is minimized. Sometimes a random element is involved, as in a biased coin method where the probability becomes 2 to 3 in favor of the treatment that is so far least allocated. As long as the treatment cannot be predicted with any degree of certainty, the random element is not required.[2] By and large, minimization is of greatest value in smaller trials, i.e., trials with less than 100 patients where several patient variables are known to be of prognostic importance. It can be useful in larger trials, but it can become complicated and entail large administrative efforts.

Unequal Randomization

In a clinical trial with two treatment arms, it is statistically most efficient to have roughly equal numbers of patients in each group. However, unequal randomization, up to a 3:2 ratio, is a realistic alternative given the need to gain further experience with newer procedures and treatments. A 2:1 or 3:2 ratio can be used without significant loss of statistical power. In fields such as orthopaedic surgery, there is a case for greater use of this type of randomization in trials involving new surgical techniques versus a given standard technique. The trial is usually motivated by some enthusiasm for the new therapy or technique and, therefore, it may be worth putting more than half of the patients in the new treatment group to gain knowledge and experience. There is a reasonable case for more widespread use of unequal randomization, provided that the inequality is not so great as to seriously impair the statistical efficiency of the treatment comparison.[1]

Cluster Randomization

Although interventions are normally at the level of the individual patient, they may be applied at the level of organizations or geographical areas. Areas and organizations represent discrete clusters of individuals. Cluster studies may be appropriate where it is not possible to randomize individuals. They do, however, create extra complexities that need to be carefully considered in the design of studies. In particular, evaluation may be at the level of the individual whereas intervention is at the cluster level, only a few clusters may be available for study, and individual responses may be correlated within clusters. Randomization is the preferred method of allocating clusters to an in-

tervention group. Individuals within a cluster may differ systematically from individuals in other clusters. As a general rule, at least 10 clusters per group will be required if individual level analyses are planned. In orthopaedic sports medicine, an example of a cluster trial is a trial of neuromuscular training to prevent knee ligament injuries.

Methods of Revealing the Treatment Allocation

The most important factor in the method chosen for randomization is its unpredictability. Simple randomization is inherently unpredictable, but has other problems. Block randomization is the most useful system, as long as the block sizes are not identical. If an adaptive system is used, then the lack of predictability will come from the study design and choice of minimization variables.

Once the appropriate scheme has been chosen and the randomization lists have been made, there are a number of methods by which to reveal the randomization. The traditional approach has been "envelope randomization," where numbered envelopes, each containing a treatment allocation, are opened sequentially. The problem with this method is that it is subject to various forms of abuse. An example of a problem would be a trial of multidirectional instability of the shoulder in which patients were randomized to "shrink" or open capsular shift using envelope randomization. Surgeon X thinks young girls do not like the scar of open surgery, and so when he opens an envelope and sees that it says "open capsular shift" he puts it back and chooses another envelope. There are sufficient examples of problems with sealed envelopes, and it is no longer a generally recommended method of randomization for multicenter trials.

A preferred method is to use a central randomization service, or a list held by someone outside the trial. A remote computer or a dedicated telephone randomization service, managed by a trials unit rather than the clinical team responsible for the study, are two good options. After giving the necessary baseline details, the patient is randomized according to the schedule. This method can protect against bias, but has the problem of having to rely on external staff and the problem of different time zones in the case of a multicenter trial.

An alternative to telephone randomization, and probably the method of the future, is Web-based randomization. This has the double advantage of being external to the trial and of not requiring staffing. However, the disadvantages come in equipping the trial center with the appropriate computer technology and the possibility of technical difficulties with computer software.

The timing of the actual revelation should be predetermined and specifically written in the study protocol. The time at which the allocation is revealed, for the most part, will depend on the nature of the study and of the intervention being assessed. For example, in surgical studies it must be decided whether it is most efficient to randomize the patient before surgery or during surgery. In some cases, certain information may have to be confirmed prior to the patient being deemed eligible for randomization, i.e., making sure the patient has good enough bone quality for an uncemented versus cemented implant in a total shoulder arthroplasty trial comparing the two methods.

A new method being used in some centers is prerandomization. In this case, patients' treatment allocation is predetermined before they are even approached about the study. For example, a patient with an Achilles tendon rupture is deemed eligible for a clinical trial comparing surgery versus standard treatment. The trial is using the prerandomization method, and so before the physician approaches the patient about the study it is decided that they will be randomized to the surgical group. The patient is then approached and simply asked whether or not he or she accept surgical treatment of the Achilles tendon rupture. This method of randomization still remains extremely controversial, and a legitimate discussion about its advantages and disadvantages is beyond the scope of this paper.

Regardless of the time at which the envelope is opened, or the telephone call is made, most clinical trials operate under the intention to treat analysis principle. This principle states that all data from all patients (i.e., all who were randomized) is to be included in the analysis, even if they did not receive the treatment allocated to them. Excluding patients from analysis because of poor compliance, loss to follow-up, dropping out, complications, or even crossing over to the alternate treatment group, may seem reasonable at the time, but can lead to a skewed statistical analysis and thus to a misleading interpretation of the results. Sackett et al.[3] described a case in which 16 patients in a study were "not available for follow-up" because of stroke or death. These patients were excluded from the initial analysis. All but one were originally allocated to a surgery group in which 10 patients suffered strokes and 5 others died during or shortly after surgery. When all 16 patients were reen-

tered into the analysis, the data naturally appeared totally different.

BLINDING

Blinding is used in combination with randomization to control for expectation effects in clinical trials. It is a normal feature of most controlled clinical trials intended to be reported in a peer-reviewed journal.[4] Blinding or masking is intended to limit the occurrence of conscious and unconscious bias in the conduct and interpretation of outcomes in a clinical trial. Such bias would arise from the influence of a knowledge of the treatment allocation of a subject and can have an effect on recruitment, allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of outcome measures, the handling of withdrawals, and the exclusion of data from analysis.[5] The essential aim is to prevent identification of the treatments until all such opportunities for bias have passed.

Randomized clinical trials can be either single-, double-, or triple-blind. This means that neither the subjects themselves (single-blind), nor the treating physician/evaluator (double-blind), knows which intervention the subject has been assigned to. A triple-blinded trial is one in which neither the subject nor any of the investigators or sponsors who are involved in the treatment in clinical evaluation are aware of the treatment received. This includes anyone determining subject eligibility, evaluating endpoints, or assessing compliance with the protocol.[5]

If blinding of the treatment assignment is feasible, it is generally advisable for it to be designed so that as few people as possible know the treatment assignment. However, in many cases it is most important and only feasible to ensure blinded evaluation, even if the patient and treating physician are aware of the treatment allocation.[2] This is particularly true of surgical trials.

By using randomization and blinding, not only are study biases reduced but study validity is also improved. It is paramount to balance the need for scientific objectivity with ethical practice in terms of informed consent. Investigators must ensure that subjects are provided with adequate information about the purpose and methods of randomization and blinding and must assess their understanding of these methods before obtaining consent.[6]

Investigators must also consider in advance the conditions under which the patient and/or the investigator blind may be broken, for example, to treat an adverse event or plan for future treatment of the patient. The principal investigator needs to specify a priori in the protocol under which circumstances the code will be broken, where the code is located, who will break it, and how this information will be handled. The subject must also be clear on whom to notify in the event of an emergency in order to maintain the blind as long as possible.

Blinded protocols should always clearly specify who is to be blinded, why, how, and to what. The effectiveness of blinding should be assessed for the patient in a single-blind study and for both the patient and the investigator in double-blind studies. Patients can be asked what drug or surgical treatment they believe they are or were receiving and the basis for their beliefs.[7] The effectiveness of the blinding can then be assessed by estimating whether they are correct more often than would be expected by chance. Because of the strong nature of the placebo effect in drug studies, the great majority of patients in placebo-controlled trials believe they are receiving the actual drug. Likewise in surgical trials, the patients will usually become aware of their particular procedure through incisions or rehabilitation procedures. Although there may be nothing that can be done about patients identifying their treatment, an assessment of the effectiveness can be useful in interpreting study results.

One particular advantage of double-blind studies is that they allow for more objective evaluation of side effects, both by the patient and the physician.[2] This allows investigators to correct for any over-reporting of side effects on active therapy and get a more unbiased estimate of adverse reactions attributable to the treatment itself.

A concern in blinded studies is how a subject may assess an intervention that proves beneficial to him or her after the completion of the trial.[8] It may be argued that those who volunteer for clinical trials deserve assurance that they will receive the intervention proven to be superior in the trial, once their study participation is complete. A convenient solution can be to allow patient cross-over, i.e., from a nonsurgical to a surgical intervention after a certain period of time.

The individual circumstances of each clinical trial make it impossible to give any general rule on blindness that could apply to all trials. However, aspects such as ethics, feasibility, the importance of decreasing bias, and a compromise of variations of these aspects to ensure a sound protocol, should be considered.

Considerable time and effort are required in blinded

trials to ensure that they work properly. Once undertaken, the protocol must be implicitly adhered to in order to maintain the integrity of the results expected of a blinded trial.

Examples Where Blinding Is Effective

In intervention studies in which the outcome designed to measure the effect of the intervention is affected by knowledge of treatment allocation, it is important to blind the subject undergoing the intervention, the treating physician, the evaluator, or preferably all three. These are described as single or double/triple blind designs. In trials regarding medications this can easily be achieved using placebo tablets similar to the "real" tablet in design, taste, and smell.

In PROBE (prospective randomized open blinded endpoint) designs for injury-prevention studies or clinical studies comparing treatment methods, new injuries are classified "blindly" by independent researchers so as to obtain nonbiased, standardized classifications of injury data (e.g., type, degree of seriousness, structures involved).

In studies in which medication is injected into joints, both the person undergoing the treatment and the treating physician may be blinded by having a third person prepare the medication and the placebo substance and the syringes and placing them in similar boxes. In this way, neither the patient nor the treating physician will know whether the patient received real or placebo medication. If this is combined with using an independent person for the final evaluation, this becomes a triple-blinded study.

Examples Where Blinding Is Difficult to Obtain

It is difficult to obtain blinding in drug therapy where the specific taste of one drug is impossible to conceal without changing the formula of the drug. A "double-dummy" technique can then be used whereby the taste is concealed in an additional tablet or liquid similar to the placebo group.

It is difficult to blind clinical studies comparing surgical procedures involving surgical approaches, i.e., open versus arthroscopic procedures. In an open-label trial, both the surgeon and the patient will know which treatment he or she received unless a placebo incision is placed in the skin making a single-blind protocol possible. In those cases, it is particularly important that the investigator's knowledge of the next treatment should not influence the decision to enter the subject into the study. This decision should precede knowledge of the randomized treatment. For these trials, one must use a centralized randomization method to administer the assignment of a randomized treatment. This type of trial requires an independent evaluator to do the subsequent outcome measurement of the patient. Ideally, this evaluator should have no ties to the hospital or clinic where the surgical treatment is taking place or vested interests in the outcome of the trial. In some circumstances, sham surgical procedures have been used successfully to ensure that high-cost, high-risk surgical procedures are being evaluated with maximum objectivity.

CONCLUSIONS

The key points to remember about randomization are:

1. Patients should not be randomized if one treatment option is unfeasible or contraindicated.
2. There should be no foreknowledge by the randomizing clinician about treatment allocation.
3. Randomization must be performed to ensure that blinding, where feasible, is reserved throughout the necessary time period.
4. Once patients are in the trial, they must remain in for the analysis even if they do not receive their allocated treatment. This method of analysis is known as "intention to treat," and is important to reduce bias.

In most clinical trials, blinding is used in combination with randomization to reduce bias in the assessment of outcome.[9] Despite the rarity of deceit in clinical research, examples of incorrect results owing to bias in trials without blinding[10] and with single-blind studies[11] reinforce the value of this methodology. We must realize that, in certain circumstances, patient blinding is just not possible, particularly when the intervention is surgical. Under these circumstances, alternative approaches must be considered to ensure the most reliable methodology.

REFERENCES

1. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. for the CONSORT Group. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001;134:663-694.
2. Pocock SJ. *Clinical trials: A practical approach.* Chichester: Wiley, 1984.
3. Sackett DL, Haynes RB, Tugwell P. *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Boston: Little, Brown, 1985.

4. Holme I. Kvalitetskriterier i intervensjonsstudier. *Norsk Idrettsmedisin* 2002;2:20-24.
5. Stat Med 1999;18:1905-1942.
6. Grady C. Ethical principles in clinical research. In: Gallin JL, ed. *Principles and practice of clinical research.* London: Academic, 2002;15-26.
7. Manolio TA. Design and conduct of observational studies and clinical trials. In: Gallin JL, ed. *Principles and practice of clinical research.* London: Academic, 2002;187-206.
8. Harth SC, Thong Y. Aftercare for participants in clinical research. Ethical consideration in an asthma drug trial. *J Med Ethics* 1995;21:225-228.
9. Califf RM. Large clinical trials: Clinical research institutes. In: Gallin JL, ed. *Principles and practice of clinical research.* London: Academic, 2002;234.
10. Karlkowski TR. Ascorbic acid for the common cold: A prophylactic and therapeutic trial. *JAMA* 1975;231:1038-1042.
11. Henkin RI. A double blind study of the affects of zinc-sulfate on taste and smell dysfunction. *Am J Med Sci* 1976;272: 285-289.

RECOMMENDED READING

1. Altman DG. *Practical statistics for medical research.* CRC Chapman and Hall, 1991.
2. DeMets DL, Friedman LM, Furberg CD. *Fundamentals of clinical trials.* New York: Springer Verlag, 1998.
3. Stevens A, ed. *The advanced handbook of methods in evidence based healthcare.* London: Sage.
4. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol* 1999;52:19-26.
5. Jadad AR. *Randomised controlled trials.* London: BMJ Books, 1998.