REVIEW ARTICLE

Randomized Controlled Trials

Part 17 of a Series on Evaluation of Scientific Publications

Maria Kabisch, Christian Ruckes, Monika Seibert-Grafe, Maria Blettner

SUMMARY

Background: In clinical research, randomized controlled trials (RCTs) are the best way to study the safety and efficacy of new treatments. RCTs are used to answer patient-related questions and are required by governmental regulatory bodies as the basis for approval decisions.

Methods: To help readers understand and evaluate RCTs, we discuss the methods and qualitative requirements of RCTs with reference to the literature and an illustrative case study. The discussion here corresponds to expositions of the subject that can be found in many textbooks but also reflects the authors' personal experience in planning, conducting and analyzing RCTs.

Results: The quality of an RCT depends on an appropriate study question and study design, the prevention of systematic errors, and the use of proper analytical techniques. All of these aspects must be attended to in the planning, conductance, analysis, and reporting of RCTs. RCTs must also meet ethical and legal requirements.

Conclusion: RCTs cannot yield reliable data unless they are planned, conducted, analyzed, and reported in ways that are methodologically sound and appropriate to the question being asked. The quality of any RCT must be critically evaluated before its relevance to patient care can be considered.

▶ Cite this as:
Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M: Randomized controlled trials: part 17 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2011; 108(39): 663–8. DOI: 10.3238/arztebl.2011.0663

Interdisziplinäres Zentrum Klinische Studien (IZKS), Universitätsmedizin der Johannes-Gutenberg-Universität Mainz: Dipl.-Biomath. Kabisch, Dipl.-Math. Ruckes, Dr. med. Seibert-Grafe

Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsmedizin der Johannes-Gutenberg-Universität Mainz: Prof. Dr. rer. nat. Blettner

Clinical research lays the groundwork for progress in medicine and is an indispensable prerequisite for evidence-based medicine. Randomized controlled clinical trials (RCTs) are the gold standard for ascertaining the efficacy and safety of a treatment. RCTs can demonstrate the superiority of a new treatment over an existing standard treatment or a placebo. In clinical research RCTs are used to answer patient-related questions, and in the development of new drugs they form the basis for regulatory authorities' decisions on approval. Alongside meta-analyses, high-quality RCTs with a low risk of systematic error (bias) provide the highest level of evidence (1, 2).

The aim of this article is to provide an introduction into the methods and quality requirements of RCTs in order to help the reader understand and evaluate publications that present the results of such studies. Since RCTs are by definition interventional, often investigating drugs or medical devices, ethical and legal aspects will also be discussed.

The discussion here corresponds to expositions of the subject in numerous textbooks (3–5) but also reflects the authors' own experience of planning, conducting and analyzing RCTs. To aid understanding, some methodological issues are illustrated by reference to a published trial, the ALIFE study (Anticoagulants for LIving FEtuses). The fundamental principles of methodology and statistical analysis for all studies, including RCTs, have been expounded in earlier articles in this journal's series on evaluation of scientific publications (6–11).

The results of the ALIFE study were published in the New England Journal of Medicine in April 2010 (12) and presented in the "Studies in Focus" series of the German-language edition of *Deutsches Ärzteblatt* in July 2010 (13). In this study, women who had had two or more miscarriages were assigned randomly to one of three treatment groups: aspirin plus heparin, aspirin alone, or placebo. The primary objective of the study was to investigate the efficacy of the different treatments as shown by the rate of live births.

Objectives

The basis of every RCT is the study protocol that describes the medical/scientific background, the risk:benefit assessment, the study design, the study methods, and the overall planning, conduct and

analysis (14). The primary study question, i.e., the primary objective, results from the medical/scientific rationale for the study.

To answer the primary study question, a primary endpoint is required. This is a parameter measured or observed that is recorded at a defined time and can be assumed to reflect the effect of a treatment. The endpoint may be clinical, e.g., the live birth rate in the ALIFE study.

In a confirmatory study hypotheses are formulated *a priori* according to the primary study question. If the primary objective of the trial is to demonstrate the superiority of a new treatment over an existing treatment or placebo, then the initial assumption (null hypothesis) is that the two treatments do not differ in efficacy. Based on statistical analysis the null hypothesis can be retained or must be rejected in favor of the alternative hypothesis. The alternative hypothesis is assumed when a statistically significant difference is ascertained between the two treatments. (A detailed description of methods for statistical evaluation is given in an earlier article in this series [15].)

The primary study question is accompanied by one or more ancillary study questions, i.e., secondary objectives. The secondary endpoints investigate other effects of the treatment, e.g., the occurrence of adverse events or the influence on biomarkers. In the ALIFE study, the secondary endpoints included the rate of miscarriage, the premature birth rate, and the rate of maternal thrombopenia.

From the statistical viewpoint it is vital to distinguish between the primary and secondary study questions, because the number of study subjects depends solely on the primary endpoint (16). Study planning includes calculation of the number of subjects necessary for detection by statistical analysis of a minimally relevant difference in efficacy, from the clinical viewpoint, between the treatments. The number of patients is therefore crucial for the statistical power of a study. (Sample size calculation is described in detail in a previous article in this series [17].)

In the ALIFE study a difference of 15% in live birth rate was assumed between the combination of aspirin plus heparin and aspirin alone or placebo. In order to demonstrate the postulated positive effect of the combination therapy, 364 women were enrolled in the trial.

Study design

In trials with randomized and controlled design (e.g., a two-armed study with parallel groups), the effects of the study treatment (intervention) are compared with those of a control treatment and the patients are randomly assigned to the two groups. The patients in the control group receive either another treatment or a placebo. The ALIFE trial is a three-armed parallel group study to establish whether the combination treatment or the monotherapy improve the live birth rate compared with placebo. The use of placebos in clinical trials is ethically justified provided that no standard treatment is available. If comparison with placebo is indispensable for methodological reasons, it can be justified as long as patients will not be harmed (18). That is the case, for example, if the study is of only short duration or if the severity of disease permits postponement or interruption of treatment.

As in any study of human subjects, the study population of an RCT must be clearly defined. Precise inclusion and exclusion criteria are elaborated to ensure that only eligible patients are recruited. The study participants must be homogeneous with regard to demographic characteristics, disease state, and possibly even comorbidity and comedication.

To ensure "fair" comparison between the treatments, the different study groups must be truly comparable. This can be achieved by standardization of, for example, the time(s) of intake of the study medication and the methods used to measure clinical parameters, but most important for comparability is randomization of the participants.

Randomization

In RCTs the patients are randomly assigned to the different study groups. This is intended to ensure that all potential confounding factors are divided equally among the groups that will later be compared (structural equivalence). These factors are characteristics that may affect the patients' response to treatment, e.g., weight, age, and sex. Only if the groups are structurally equivalent can any differences in the results be attributed to a treatment effect rather than the influence of confounders. If the confounders are known, structural equivalence of the patient groups can be attained by stratified randomization *(Box)*.

In the ALIFE study the patients were assigned to the three treatment groups with a randomization ratio of 1:1:1. They were randomized taking account of the prognostic factors of age (<36 years or ≥36 years) and number of miscarriages (2 or ≥3), and because the study was multicentric they were stratified by study center. If patients were allocated to treatment groups by conscious or unconscious selection for prognosis-related characteristics, rather than randomly, this could lead to biased treatment comparison and distorted results (selection bias).

The assignment to study groups must not be in any way predictable. Predictability of group allocation is avoided by ensuring the study staff are unaware to which treatment the next patient will be allotted. Alternating assignment to the different treatments is not truly random.

Blinding

Bias is avoided not only by randomization but also by blinding. A study may be double blind, single blind, or open.

In a double-blind study neither patient nor study physician knows to which treatment the patient has been assigned. Double-blind studies are advantageous if knowledge of the treatment might influence the course and therefore the results of the study. Thus it is

particularly important that the study physician is blinded to treatment if the endpoints are subjective. Blinding of patients to their treatment is important, for example, if their attitude could potentially affect their reliability in taking the test medication (compliance) or even their response to treatment.

If only one party, either patient or study physician, is blinded to the treatment, the study is called single blind; a study with no blinding is described as open. The highest possible degree of blinding should be chosen to minimize bias.

Analysis population

The data subjected to statistical analysis in RCTs are those gathered from patient populations defined in the study protocol. The primary population for analysis is the so-called intention-to-treat (ITT) population, comprising all randomized patients. In analysis according to the ITT principle, patients are allocated to the group to which they were randomized, thus retaining the advantages of randomization such as structural equivalence. Because the ITT population includes all patients who were randomized, the data for analysis include some patients whose treatment was interrupted, prematurely discontinued, or did not take place at all. The analysis strategy for ITT data is therefore conservative, i.e., the treatment effect tends to be underestimated (19), regardless of whether the primary endpoint represents an improvement or a deterioration. Many studies define a modified ITT (mITT) population, which may for example comprise the patients who received at least a defined amount of study treatment.

An alternative strategy is to restrict analysis to the data from the per-protocol (PP) population. Patients in whom study conduct deviated from the protocol are excluded from analysis. These so-called protocol violations include, for example, failure concerning the application of inclusion or exclusion criteria and incorrect administration of the study treatment. In analysis according to the PP principle, patients are allocated to the treatment groups depending on the treatment they actually received. Because the PP population includes only those patients who completed the study according to the protocol, the results may be distorted in favor of the investigational intervention (19).

To assess the robustness of the study findings, PP evaluation is carried out as a sensitivity analysis if the ITT population is the patient population for the primary efficacy analysis (16). If the results of PP and ITT evaluation of the primary endpoint are very similar, they can be regarded as reliable. Should this not be the case, the possible reasons for the discrepancy between the results of the ITT and PP analyses must be discussed in the results section of the publication.

The data of the ALIFE study, particularly the primary endpoint, were statistically evaluated on the basis of the ITT population. The rates of live births in the three treatment groups did not differ significantly *(Table 1)*. Analysis according to the PP principle confirmed this finding. Neither aspirin and heparin combined nor aspirin alone were demonstrated to have a greater effect than placebo on the live birth rate.

Quality standards and legal requirements in Germany

Clinical trials have to be performed according to national and international regulations. The Declaration of Helsinki, first formulated by the World Medical Association in 1964 and revised several times in the intervening years (20), lays down fundamental ethical principles for research on human beings. Trials investigating drugs and medical devices have to comply with the relevant German laws for drugs—the German Medicines Act (AMG; for German text see Bundesgesetzblatt I p. 2262)—and the GCP regulation (GCP-Verordnung [21]), and for devices the Medical Devices Act (MPG; for German text see Bundesgesetzblatt I p. 983), revised in March 2010. The GCP regulation, which came into force in 2004, made adherence to good clinical practice (GCP) a legal requirement in Germany (21). GCP Guideline ICH-E6 of 1997 forms the basis for European Directives 2001/20/EG and 2005/28/EG, on which in turn the GCP regulation is based (14). The aim of GCP is to protect study participants and ensure high quality of study data.

In 2004 the International Committee of Medical Journal Editors made registration of a clinical trial in a public registry a precondition for its publication (22). The professional code of conduct for physicians in Germany demands that every study in human subjects be submitted to the responsible ethics committee for approval. Drug trials and most studies of medical

BOX

Stratified randomization

If the stratification factors sex (male, female) and age (<18 years, ≥18 years) are to be considered and 150 patients are to be randomized in a ratio of 1:1 into the active treatment and placebo groups (2×75 patients), then randomization has to be performed for each separate subgroup (stratum). Two stratification factors, each with two values, yield four strata (male and <18 years, male and ≥18 years, female and <18 years, female and ≥18 years).

	Active treatment	Placebo
Male and <18 years	10	10
Male and ≥18 years	16	17
Female and <18 years	24	23
Female and ≥18 years	25	25
Total	75	75

TABLE1

Results of the ALIFE study (adapted from [12])

	Aspirin plus Heparin	Aspirin alone	Placebo	p-value
Intention-to-treat population n	123	120	121	
Live births n (%)	67 (54.5)	61 (50.8)	69 (57.0)	0.63
Relative risk (95% CI)	0.96 (0,76–1,19)	0.89 (0.71–1.13)	1.00	
Absolute difference in live birth rates (95% CI) %	-2.6 (-15.0–9.9)	-6.2 (-18.8–6.4)		

Relative risk and absolute difference were calculated for the comparisons between aspirin plus heparin and placebo and between aspirin alone and placebo. The p-value applies to all treatment group comparisons.
95% CI, 95% confidence interval

TABLE 2

Minimal requirements for a publication reporting a randomized controlled trial (adapted from [23])

Study design	Description of study design (e.g., parallel group comparison)
Study population	Specification of inclusion and exclusion criteria for patients
Treatments	Detailed account of treatments and their application in each intervention group and control group
Objectives	Precise formulation of primary and secondary objectives/study questions
Endpoints	Clear definition of primary and secondary endpoints
Sample size	Description of how the required number of study participants was determined
Randomization	Description of type of randomization of patients to treatment groups (e.g., stratified randomization)
Blinding	Specification of degree of blinding (e.g., double blind)
Analysis population	Number of patients analyzed in each treatment group and definition of population for analysis (e.g., ITT)
Results	Presentation of the results for all primary and secondary endpoints for each treatment group
Adverse events	Details of all major adverse events for each treatment group
Interpretation	Interpretation of the results, taking into account the study question, possible causes of bias, the current state of knowledge, and other researchers' publications on the same topic
Generalizability	Discussion of the applicability of the study results to general patient care

devices require not only approval from the local ethics committee but also from regulatory bodies at the federal level (Federal Institute for Drugs and Medical Devices [BfArM] or Federal Institute for Vaccines and Biomedicines, Paul-Ehrlich-Institut [PEI]). The applications have to be accompanied by the study protocol, the information to be supplied to the patients, the consent form for participation, and confirmation that adequate insurance has been arranged.

Trials of drugs and medical devices also have to be registered with state authorities. There are legally defined obligations to report suspected unexpected serious adverse reactions or early termination of a study, and the final study report must also be submitted. The Federal Data Protection Act (BDSG; for German text see Bundesgesetzblatt I p. 2814) and the AMG obligate researchers to pseudonymize all person-related data that are gathered, documented, stored, and analyzed in the course of a clinical trial. In other words, information revealing the identity of a patient (name or initials) must be replaced by a code. Only patients who have agreed in advance to the recording, storage, processing and dissemination of their data may participate in a clinical study.

Discussion

Any publication of an RCT must lucidly describe the planning, conduct, and analysis of the study. The CONSORT statement provides a minimum set of recommendations for reporting RCTs (23). The most important aspects that have to be described in the publication are listed in *Table 2*. The progress of patients through an RCT and the numbers of patients whose data were analyzed can be depicted in a flow diagram *(Figure)*.

The study results and their interpretation must be discussed in detail in the study report and any subsequent publication, and the limitations of the methods used should be described, all with reference to the study design, the recent literature, and the current state of knowledge. Critical discussion plays a decisive part in clinical evaluation of the results. In the publication of the ALIFE study, the findings were compared with those of other RCTs investigating the effects of heparin on reduction of miscarriages and inconsistencies were discussed. Ultimately, the available study data did not justify the recommendation of anticoagulants for women with recurring miscarriages.

Although RCTs are the gold standard with regard to level of evidence, their generalizability, i.e., the extent to which their results can be extrapolated to the wider patient population (external validity) is often questioned, because standardized and controlled study conditions do not adequately reflect clinical reality. Moreover, the patients selected for a study are not necessarily representative, in that those seen in routine daily practice will often have numerous comorbidities and comedications. After marketing approval of a new treatment, phase-IV studies are carried out to establish its efficacy and safety in a larger and more heterogeneous population; as a rule these studies are RCTs.

Epidemiological studies, e.g., cohort studies, are particularly suitable for detection of infrequent adverse effects.

Conclusion

RCTs are the best type of study for determining whether there is a causal relationship between intervention and effect (24). Recent discussions in the scientific community and the new Law on the Reorganization of the Pharmaceutical Market (AMNOG; for German text see Bundesgesetzblatt I p. 2262), which regulates the use of drugs and medical devices, clearly show that RCTs are still the standard for demonstrating efficacy and safety so that a new treatment can be approved for use in patients. However, it seems clear that post-marketing studies comparing new and established treatments are still required.

The IZKS Mainz is supported by the grant "Clinical Trial Centers [Klinische Studienzentren], no. FK 01KN1103, IZKS Mainz" from the Federal Ministry of Education and Research.

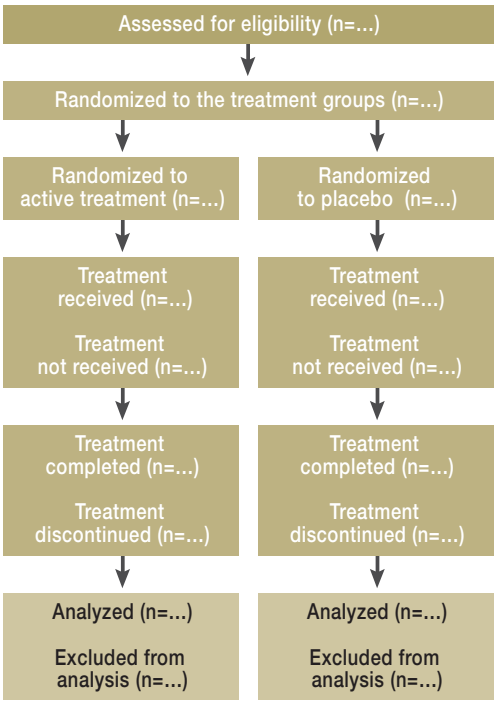
Acknowledgment
The authors are grateful to Daniel Wachtlin of the Interdisciplinary Center for Clinical Trials (IZKS) Mainz for helpful discussions.

Conflict of interest statement
The authors declare that no conflict of interest exists.

Manuscript received on 23 February 2011, revised version accepted on 28 June 2011.

Translated from the original German by David Roseveare.

FIGURE



Patient flow in a randomized controlled trial (adapted from [23])

REFERENCES

1. Harbour R, Miller J: A new system for grading recommendations in evidence based guidelines. BMJ 2001; 323: 334–6.
2. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M (2001): Oxford Centre for Evidence-based Medicine Levels of Evidence. www.cebm.net/levels_of_evidence.asp
3. Meinert CL: Clinical Trials: Design, conduct, and analysis. Oxford University Press: New York 1986.
4. Shein-Chung C, Jen-Pei L: Design and analysis of clinical trials: concepts and methodologies. John Wiley & Sons: New Jersey 2004.
5. Schumacher M, Schulgen G: Methodik klinischer Studien: Methodische Grundlagen der Planung, Durchführung und Auswertung. Berlin, Heidelberg: Springer-Verlag 2008.
6. Röhrig B, du Prel, Blettner M: Study Design in medical research: Part 2 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(11): 184–9
7. Röhrig B, du Prel du J, Wachtlin D, Blettner M: Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(15): 262–8.
8. Spriestersbach A, Röhrig B, du Prel J, Gerhold-Ay A, Blettner M: Descriptive statistics: the specification of statistical measures and their presentation in tables and graphs: part 7 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(36): 578–83.
9. Sauerbrei W, Blettner M: Interpreting results in 2×2 tables: part 9 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106(48): 795–800.
10. Victor A, Elsäßer A, Hommel G, Blettner M: Judging a plethora of p-values: how to contend with the problem of multiple testing: part 10 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107(4): 50–6.
11. Zwiener, I, Blettner M, Hommel G: Survival analysis: part 15 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2011; 108(10): 163–9.
12. Kaandorp SP, et al.: Aspirin plus heparin or aspirin alone in women with recurrent miscarriage. NEJM 2010; 362: 1586–96.
13. Heinzl S: Können ASS und Heparin wiederholte Fehlgeburten verhindern? Dtsch Arztebl 2010; 107(27): A-1355.
14. ICH E6: Guideline for good clinical practice. London UK: International Conference on Harmonization 1996; CPMP/ICH/135/95.
15. du Prel J, Röhrig B, Hommel G, Blettner M: Choosing statistical tests: part 12 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107(19): 343–8.
16. ICH E9: Statistical Principles for Clinical Trials. London UK: International Conference on Harmonization 1998; CPMP/ICH/363/96.
17. Röhrig B, du Prel JB, Wachtlin D, Kwiecien R, Blettner M: Sample size calculation in clinical trials: part 13 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107(31–32): 552–6.
18. ICH E10: Choice of Control Group and Related Issues in Clinical Trials. London UK: International Conference on Harmonization 2000; CPMP/ICH/364/96.
19. Ellenberg JH: Intention-to Treat Analysis. In: Redmond C, Colton T (eds): Biostatistics in Clinical Trials. New Jersey: John Wiley & Sons: 2001.
20. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. 2008. www.wma.net/en/30publications/10policies/b3/17c.pdf.
21. Verordnung über die Anwendung der Guten Klinischen Praxis bei der Durchführung von klinischen Prüfungen mit Arzneimitteln zur Anwendung am Menschen (GCP-Verordnung – GCP-V). GCP-Verordnung vom 9. August 2004 (BGBl. I S. 2081), die zuletzt durch Artikel 4 der Verordnung vom 3. November 2006 (BGBl. I S. 2523) geändert worden ist.
22. Clinical trial registration: a statement from the International Committee of Medical Journal Editors (ICMJE). Lancet 2004: 364: 911–2.

23. Schulz KF, Altman DG, Moher D: CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.

24. Windeler J: Bedeutung randomisierter klinischer Studien mit relevanten Endpunkten für die Nutzenbewertung. In: Diskussionsforum zur Nutzenbewertung im Gesundheitswesen: Begriffsdefinitionen und Einführung. 2007. www.gesundheitsforschung-bmbf.de/_media/DLR_Nutzenbewert_07–11–22_Druckversion.pdf.

Corresponding author
Prof. Dr. rer. nat. Maria Blettner
Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI)
Universitätsmedizin der Johannes-Gutenberg-Universität Mainz
Obere Zahlbacher Str. 69
55131 Mainz, Germany
blettner-sekretariat@imbei.uni-mainz.de

KEY MESSAGES

- In clinical research, randomized controlled trials are the gold standard for demonstrating the efficacy and safety of a new treatment.
- Randomized controlled trials cannot yield robust data unless they are planned, conducted, and analyzed in ways that are methodologically sound and appropriate to the question being asked.
- Methods to avoid bias, such as randomization and blinding, can help to prevent distortion of the study results.
- The robustness of the results is tested by statistical analysis of the data from patient populations defined *a priori*.
- The quality of a randomized controlled trial depends crucially not only on adherence to methodological standards but also on strict compliance with the protocol regarding the clinical conduct of the study.

RESEARCH PRIMER

An Introduction to the Fundamentals of Randomized Controlled Trials in Pharmacy Research

Sherilyn Houle

INTRODUCTION

The randomized controlled trial (RCT) is regarded as one of the most valued research methodologies for examining the efficacy or effectiveness of interventions.[1] Randomized trials are most often associated with studies of drug effectiveness; however, they have also been successfully applied to research questions related to provision of care by pharmacists.[2-5]

This article is not intended to be an exhaustive guide on performing an RCT, but rather an introduction to the major concepts and approaches involved in designing and conducting an RCT. Readers requiring additional information are referred to more comprehensive publications[6-9] and are encouraged to consult with researchers experienced in this area before undertaking a study. Throughout this paper, individuals enrolled in a study are referred to as "patients", the individuals delivering the study (e.g., pharmacists in practice research) are referred to as the "investigators", and those responsible for study design and analysis are referred to as the "study team".

DESIGN OF A RANDOMIZED CONTROLLED TRIAL

An RCT is a prospective study following patients forward in time. After agreeing to participate, patients are randomly allocated to one or more interventions or a control group and are followed until a finite date or the occurrence of one or more outcomes of interest. The basic RCT design—the parallel group design—is illustrated in Figure 1. Readers should be aware that RCTs may also utilize a cross-over design, where patients alternate between groups, or a factorial design, where different levels of intervention or control are applied to different groups.

As with all study designs, RCTs have both strengths and limitations (Table 1).

RESEARCH QUESTIONS BEST ADDRESSED BY AN RCT

RCTs are appropriate to address questions related to efficacy (performance under ideal and controlled circumstances) or

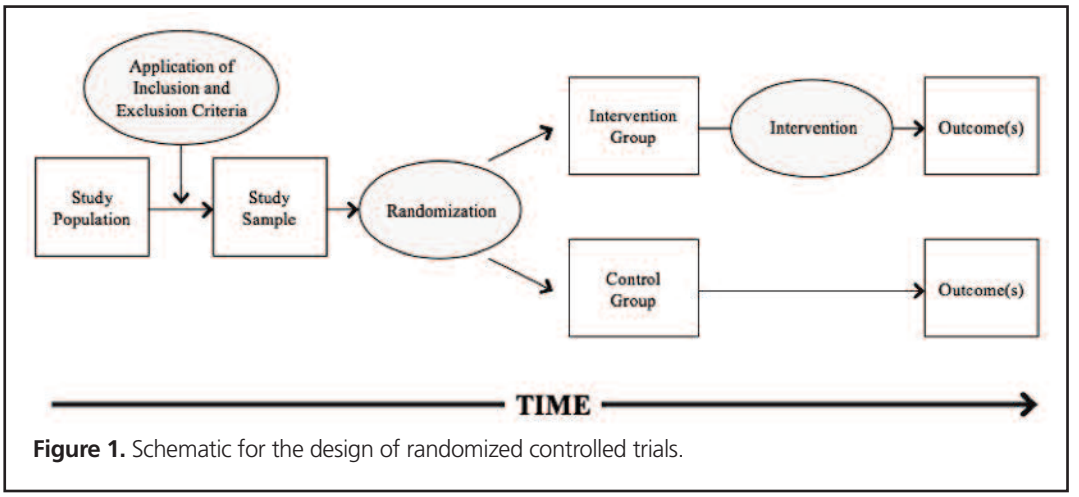


Figure 1. Schematic for the design of randomized controlled trials.

Table 1. Strengths and Limitations of Randomized Controlled Trials

Strengths	Limitations
Ability to evaluate causal relationships	Higher cost than observational studies
High internal validity (the extent to which differences between intervention and control groups can be attributed to the intervention), due to minimized bias within the study	Limited external validity and generalizability, due to strict inclusion and exclusion criteria and application of interventions by protocol
Investigator control over patient exposure	Ethical considerations related to assigning patients to particular care approaches
Prospective data collection, which allows for standardization of exposure and outcome collection	Generally shorter-duration follow-up than observational studies
Attempted balance, through randomization, between known and unknown confounding factors between groups	Inefficiency of detection of rare or delayed outcomes, due to smaller sample size and shorter-duration follow-up than observational studies

effectiveness (performance under "real-world" conditions).[10] Their suitability for these purposes is due to prospective data collection, investigator-controlled application of the intervention, use of a concomitant control group, and randomization to balance, on average, known and unknown confounding factors at the beginning of the study. The following are 2 sample research questions:

- Efficacy of drug therapy: What is the efficacy of atorvastatin 20 mg daily versus rosuvastatin 10 mg daily in the 5-year secondary prevention of major cardiovascular events in patients with type 2 diabetes mellitus?
- Effectiveness of pharmacist intervention: What is the effectiveness of a pharmacist medication review program versus usual care for inpatients aged 65 and older who experienced a fall while in hospital in terms of the occurrence of falls during the first 6 weeks after discharge?

STEPS IN DESIGNING AND CONDUCTING AN RCT

Gathering the Research Team

As with other clinical research designs, it would be highly unusual for an RCT to be conducted without the establishment of a team encompassing a breadth of expertise. Team members often include those with clinical expertise in the area, researchers with experience in RCTs to ensure methodological rigour, statisticians, and practitioners who will serve as investigators to ensure that any potential barriers to effective participation are proactively identified and addressed. Multi-investigator or multicentre studies may also benefit from a project coordinator to provide investigator training and ensure standardized application of the study protocol among sites.

Determining the Research Question

The crafting of a specific research question that adheres to the acronym PICOT (Patients, Intervention, Control, Outcome, Timing) is a crucial step, as it will guide the design of the study and will affect the generalizability and clinical relevance of the findings. Note that both of the example research questions above include all elements of PICOT. Readers requiring additional guidance in crafting a research question are referred to a previous article in this series.[11]

Defining Inclusion and Exclusion Criteria

As with cohort and case–control studies,[12] restrictions are commonly applied to the population eligible for recruitment into an RCT. Inclusion and exclusion criteria should be carefully determined and should strike a balance between generalizability and minimization of bias. Criteria that are too narrow may create challenges in identifying a suitable number of patients and may make the results less generalizable, whereas criteria that are too broad may create challenges in detecting the true effect of the intervention for a given population. For example, some patient populations may exhibit greater non-adherence or may respond differently to a particular drug because of physiologic differences secondary to age, race, or other factors. Readers are referred to other sources for guidance on crafting inclusion and exclusion criteria.[8,9,13]

Randomization

Randomization is the allocation of patients to study groups by chance. The intended function of randomization is to balance known and unknown confounding factors between intervention and control groups, thus minimizing their impact on the relationship between the intervention and the outcomes observed. Although the majority of RCTs apply patient-level randomization (whereby each patient is individually randomized to a group), this approach poses challenges for studying certain interventions. Referring to the sample research question on an inpatient medication review program, let's assume that the same pharmacist is providing care to patients in both the intervention and the control groups. We might expect that the care provided

to patients in the control group may, intentionally or unintentionally, be influenced by any training that the pharmacist has received in fall-risk reduction or a greater awareness of fall-risk management as a consequence of participating in the program. In scenarios where the control group may be inadvertently exposed to parts of the intervention, a situation referred to as "contamination", cluster randomization may be preferred. Cluster randomization applies randomization at the level of the investigator or site. In the case of the sample research question, cluster randomization could be applied at the level of the pharmacist, ward, or hospital, if multiple investigators or sites are involved. All patients of that pharmacist, ward, or hospital would therefore be part of the same treatment group. The benefit that this approach confers in terms of minimizing the risk of contamination is balanced by the drawbacks of requiring a larger number of patients if the study is to achieve statistically significant differences in outcomes between groups and by the challenges of engaging investigators who may be less inclined to participate if they are randomized to provide care only to the control group.

Simple randomization can be performed through the use of sealed envelopes, each containing a group allocation (with a new envelope being opened as each patient is enrolled) or through the use of randomization software. However, sealed envelopes are rarely used, because of the potential for investigator manipulation. Many RCTs employ more complex techniques, such as blocking or stratified randomization. In brief, block randomization ensures that a relatively equal number of patients are randomized to each group as the study progresses, whereas stratified randomization is used to ensure balancing of key baseline characteristics between groups rather than relying on chance. For example, if falls are known to be more common among seniors who are taking drugs with anticholinergic effects,[14] the randomization scheme can be designed to ensure that an equal proportion of patients on any of a prespecified list of drugs are assigned to each group. Serial randomization (randomization into one group followed by randomization into another subgroup) or a factorial design (randomization to combinations of more than one type or level of intervention or control) can also be employed. Regardless of approach, efforts must be made to track randomization and to ensure that the randomization scheme is applied only once per patient. Specifics on the design of the randomization strategy should be concealed from investigators to reduce the risk of selection bias, which is the selective enrolment of patients when they are perceived to be more likely to be enrolled into one group versus the other.

Determining and Delivering the Intervention

In RCTs, it is important to ensure that the intervention is specifically defined and consistently delivered. The protocol should clearly define the timing of intervention and monitoring visits, the drug dosages to be used with protocols for dose modification (primarily for drug effectiveness or safety trials), the measurement tools to be used, and any protocols or guidelines to inform decisions involving professional judgment. Studies involving multiple investigators should include a formal training component to ensure similar knowledge of the condition under study, the intervention protocol, and study processes. Individuals with clinical expertise in the area should be involved at this stage to ensure that the intervention is clinically sound and safe, as it is unethical to expose patients to unnecessary risk or to an intervention that is unlikely to have a sufficient chance of benefit.[15]

Selecting the Control

The control determines the *additional* exposure offered by the intervention. As such, the control must be selected with care. Researchers must first consider whether to use a placebo control or an active control. A placebo control is just that—no active therapeutic effect is anticipated—whereas an active control involves some therapeutic effect, but one that is believed to be at a lower magnitude than the intervention or the current gold standard. In practice research it is unethical for health care professionals to refuse to provide any care to patients in a control group; therefore, the control is generally an active control, in the form of usual care. In usual care, the patient's care team functions as it generally would in the absence of a trial. The only difference is that the intervention is not applied in addition to usual care. As mentioned by Tsuyuki,[13] it is important that the study team determine what usual care looks like in each of the investigators' practices, as pre-existing differences may affect the results observed.

Determining and Measuring Outcomes

Although trials generally identify a primary outcome, from which the sample size needed to detect that outcome at a statistically significant magnitude is calculated, secondary outcomes can be considered if they are likely to also be influenced by application of the intervention. Outcomes can take many forms—clinical (individually or as a composite of multiple related outcomes), economic, process evaluation, patient knowledge or satisfaction, among others—depending on the intervention under study. All relevant outcomes should be identified a priori (i.e., before the study begins) to ensure that data-collection tools capture all required information and to provide statistical rigour. Outcome measurement should be as standardized as possible, through the drafting of specific definitions of what constitutes an outcome (e.g., in the PROVE-IT trial,[16] myocardial infarction was defined as "the presence of symptoms suggestive of ischemia or infarction, with either electrocardiographic evidence [new Q waves in two or more leads] or cardiac-marker evidence of infarction, according to the American College of Cardiology definition"), the consistent use of electronic and validated diagnostic devices, and the use of piloted and validated surveys or tests. As studies with a prospective design, RCTs offer the benefit of allowing data to be collected at

the level of detail required for analysis, unlike retrospective studies or studies of administrative data, where researchers are limited to information that has already been collected. Although detail in patients' baseline characteristics, exposures, and outcomes is beneficial, researchers must ensure that data-collection tools are not overly cumbersome, which might discourage participation and complete outcome collection by investigators.

Blinding Participants and Investigators

To minimize opportunities for bias, some form of blinding is often employed. Recall that there can be up to 3 groups of individuals involved in research: the patients, the investigators, and the study team. Single-blind studies are those in which one of these groups (often the patients) is blinded to patient-group allocations, while double-blind and triple-blind designs incorporate blinding of the investigators and/or the study team analyzing the results. The highest level of blinding possible is preferred, but not all interventions can be adequately blinded. For example, blinding in the drug therapy effectiveness study mentioned above could be achieved by formulating a placebo tablet of the same appearance as the atorvastatin tablet dispensed by an individual not involved in data collection; however, blinding as to care interventions is impossible, since patients and investigators must be aware of the group allocation in order to apply or receive the intervention.

Ethical Considerations

Readers are referred to a previous article in this series for details on ethical requirements in research.[15] In addition to these, RCTs require specific additional considerations that are highlighted here. Information provided to patients as part of the informed consent process must clearly describe what randomization is (e.g., like a coin flip) and must ensure that patients are aware of what allocation to the intervention or control group(s) involves. Additionally, in RCTs of care provided by health professionals, ethics boards may require that the intervention, if found to be beneficial, be offered to patients in the control group at the end of the study, so as to avoid imparting a health advantage to those randomized to the intervention group.

Collecting the Data

As discussed earlier, the prospective design of RCTs allows for the collection of data that are believed to be of value in answering the research question. As randomization is not guaranteed to equally balance all patient characteristics across groups, even if stratification is employed, data on demographic characteristics, medical history, medication use, and lifestyle considerations should be collected to evaluate the degree of balance of these factors across groups. Whenever possible, data should be collected as continuous rather than categorical or binary values (e.g., for smoking status, documenting the average number of cigarettes smoked per day is more valuable than documenting whether or not patients smoke [yes/no] or determining whether they smoke 0–10, 11–20, or ≥ 21 cigarettes per day). Again, a balance must be established between comprehensiveness and practicality, with collection of data for only those characteristics or outcomes believed by the study team to potentially influence the relationship between intervention and outcome. Ideally, outcome data would be collected by an individual blinded to the patient's group allocation, to reduce the risk of subjective interpretation by the investigator or biased reporting by the patient, especially if the patient is providing feedback on the quality of care provided or satisfaction with care.

Determining Sample Size and Analyzing the Data

A thorough discussion of sample size calculations and data analysis methods is beyond the scope of this article and will be addressed in a future paper in this series. Of most importance to note here is that the data analysis strategy and statistical plan should be clearly established before the study is initiated. Study teams are strongly encouraged to consult with a statistician and/or a researcher experienced in RCT data analysis when designing the study and evaluating the findings. Again, it is important to ensure that all relevant outcomes to be evaluated are specified before implementing the study and also to ensure that the most valued outcome is prespecified as the primary outcome in order to appropriately inform the sample size required to detect a reasonable difference between groups.

The study team must also consider whether outcomes will be evaluated on an intention-to-treat or per-protocol basis. In intention-to-treat analyses, each patient's outcomes are analyzed as part of the group to which they were randomized, whether or not they completed the study or received an intervention outside of the study protocol. Per-protocol analysis only counts those patients who completed the study as specified in the protocol. The intention-to-treat approach is generally preferred, as it preserves the benefits of randomization, given that patients choosing to withdraw from the study or receive care outside of the protocol may differ from those who complete the study as specified. Readers requiring additional information on statistical tests and outcome analysis are referred to research and statistics textbooks in this area.[8,9,17]

Sample sizes are estimated on the basis of the expected effect size of the intervention versus the control on the primary outcome. If the expected effect size cannot be estimated from previous studies in the area, it may be necessary to conduct a pilot of the RCT to estimate the effect size. Readers are referred to other articles for specific information on the design and utility of pilot studies.[18,19]

Disseminating and Reporting Results

RCTs, by the nature of their study design, are less prone to bias than observational studies, and therefore the findings from

RCTs strongly support causal inferences. Because of this ability, it is expected that their findings will be reported and disseminated regardless of the outcome observed. To facilitate the consistent and transparent reporting of RCTs, the study should be added to a research registry before patient enrolment begins (e.g., www.clinicaltrials.gov), with study methods and results presented in the manner recommended by the CONSORT statement (available at www.consort-statement.org). Dissemination of study results may take the form of poster or oral presentations at academic or professional conferences, publication in a peer-reviewed journal, discussion via professional networks, or communication with decision-makers. The dissemination strategy should also be preplanned wherever possible, with flexibility to adapt the strategy according to stakeholders' needs. This process of knowledge translation, to be addressed in a future paper of this series, is a key step toward ensuring the uptake and integration of new knowledge into practice.

CONCLUSION

The RCT is a valuable tool in various aspects of research in health care, from drug safety and effectiveness to studies of health professional interventions. To sufficiently isolate the impact of the intervention on the outcome, RCTs must be thoughtfully designed and conducted and must involve team members with expertise across all relevant clinical and methodological aspects.

References

1. Guyatt G, Haynes B, Jaeschke R, Meade MO, Wilson M, Montori V, et al. The philosophy of evidence-based medicine. In: Guyatt G, Rennie D, Meade MO, Cook DJ, editors. *Users' guides to the medical literature: essentials of evidence-based clinical practice.* 2nd ed. New York (NY): McGraw-Hill; 2008. p. 5-16.
2. Santschi V, Chiolero A, Colosimo AL, Platt RW, Taffe P, Burnier M, et al. Improving blood pressure control through pharmacist interventions: a meta-analysis of randomized controlled trials. *J Am Heart Assoc*. 2014;3:e000718.
3. Santschi V, Chiolero A, Paradis G, Colosimo AL, Burnand B. Pharmacist interventions to improve cardiovascular disease risk factors in diabetes: a systematic review and meta-analysis of randomized controlled trials. *Diabetes Care.* 2012;35(12):2706-17.
4. Nkansah N, Mostovetsky O, Yu C, Chheng T, Beney J, Bond CM, et al. Effect of outpatient pharmacists' non-dispensing roles on patient outcomes and prescribing patterns. *Cochrane Database Syst Rev.* 2010;(7):CD000336.
5. Sinclair HK, Bond CM, Stead LF. Community pharmacy personnel interventions for smoking cessation. *Cochrane Database Syst Rev*. 2004; (1):CD003698.
6. Navaneethan SD, Palmer SC, Smith A, Johnson DW, Strippoli GFM. How to design a randomized controlled trial. *Nephrology* (*Carlton*). 2010; 15(8):732-9.
7. Smith A, Palmer S, Johnson DW, Navaneethan S, Valentini M, Strippoli GFM. How to conduct a randomized trial. *Nephrology* (*Carlton*). 2010; 15(8):740-6.
8. Friedman LM, Furberg CD, DeMets DL. *Fundamentals of clinical trials.* 4th ed. New York (NY): Springer; 2010.
9. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research*. 3rd ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2007.
10. Singal AG, Higgins PDR, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol.* 2014;5:e45.
11. Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. *Can J Hosp Pharm.* 2014;67(1):31-4.
12. Gamble JM. An introduction to the fundamentals of cohort and case–control studies. *Can J Hosp Pharm*. 2014;67(5):366-72.
13. Tsuyuki RT. Designing pharmacy practice research trials. *Can J Hosp Pharm*. 2014;67(3):226-9.
14. Woolcott JC, Richardson KJ, Wiens MO, Patel B, Marin J, Khan KM, et al. Meta-analysis of the impact of 9 medication classes on falls in elderly persons. *Arch Intern Med*. 2009;169(21):1952-60.
15. Loewen P. Ethical issues in pharmacy practice research: an introductory guide. *Can J Hosp Pharm*. 2014:67(2):133-7.
16. Cannon CP, Braunwald E, McCabe CH, Rader DJ, Rouleau JL, Belder R, et al. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N Engl J Med*. 2004;350(15):1495-504.
17. Pagano M, Gauvreau K. *Principles of biostatistics*. 2nd ed. Pacific Grove (CA): Duxbury/Thomson Learning; 2000.
18. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract*. 2004;10(2):307-12.
19. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol.* 2010;10(1):1.

Sherilyn Houle, BSP, PhD, is Assistant Professor with the School of Pharmacy, University of Waterloo, Waterloo, Ontario.

Competing interests: None declared.

Address correspondence to:
Dr Sherilyn Houle
School of Pharmacy
University of Waterloo
200 University Avenue West
Waterloo ON N2L 3G1

e-mail: sherilyn.houle@uwaterloo.ca

This article is the eighth in the *CJHP* Research Primer Series, an initiative of the *CJHP* Editorial Board and the CSHP Research Committee. The planned 2-year series is intended to appeal to relatively inexperienced researchers, with the goal of building research capacity among practising pharmacists. The articles, presenting simple but rigorous guidance to encourage and support novice researchers, are being solicited from authors with appropriate expertise.

Previous articles in this series:

Bond CM. The research jigsaw: how to get started. *Can J Hosp Pharm*. 2014;67(1):28-30.

Tully MP. Research: articulating questions, generating hypotheses, and choosing study designs. *Can J Hosp Pharm*. 2014;67(1):31-4.

Loewen P. Ethical issues in pharmacy practice research: an introductory guide. *Can J Hosp Pharm.* 2014;67(2):133-7.

Tsuyuki RT. Designing pharmacy practice research trials. *Can J Hosp Pharm*. 2014;67(3):226-9.

Bresee LC. An introduction to developing surveys for pharmacy practice research. *Can J Hosp Pharm*. 2014;67(4):286-91.

Gamble JM. An introduction to the fundamentals of cohort and case–control studies. *Can J Hosp Pharm*. 2014;67(5):366-72.

Austin Z, Sutton J. Qualitative research: getting started. *Can J Hosp Pharm*. 2014;67(6):436-40.

RESEARCH SERIES

Designing a research project: randomised controlled trials and their principles

J M Kendall

Emerg Med J 2003;**20**:164–168

The sixth paper in this series discusses the design and principles of randomised controlled trials.

The randomised control trial (RCT) is a trial in which subjects are randomly assigned to one of two groups: one (the experimental group) receiving the intervention that is being tested, and the other (the comparison group or control) receiving an alternative (conventional) treatment (fig 1). The two groups are then followed up to see if there are any differences between them in outcome. The results and subsequent analysis of the trial are used to assess the effectiveness of the intervention, which is the extent to which a treatment, procedure, or service does patients more good than harm. RCTs are the most stringent way of determining whether a cause-effect relation exists between the intervention and the outcome.[1]

This paper discusses various key features of RCT design, with particular emphasis on the validity of findings. There are many potential errors associated with health services research, but the main ones to be considered are bias, confounding, and chance.[2]

Bias is the deviation of results from the truth, due to systematic error in the research methodology. Bias occurs in two main forms: (a) *selection bias*, which occurs when the two groups being studied differ systematically in some way, and (b) *observer/information bias*, which occurs when there are systematic differences in the way information is being collected for the groups being studied.

A *confounding* factor is some aspect of a subject that is associated both with the outcome of interest and with the intervention of interest. For example, if older people are less likely to receive a new treatment, and are also more likely for unrelated reasons to experience the outcome of interest, (for example, admission to hospital), then any observed relation between the intervention and the likelihood of experiencing the outcome would be confounded by age.

Chance is a random error appearing to cause an association between an intervention and an outcome. The most important design strategy to minimise random error is to have a large sample size.

These errors have an important impact on the interpretation and generalisability of the results of a research project. The beauty of a well planned RCT is that these errors can all be effectively reduced or designed out (see box 1). The appropriate design strategies will be discussed below.

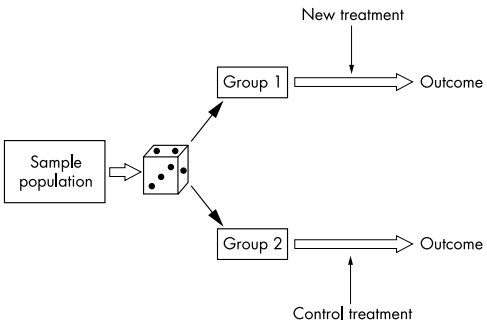


Figure 1 The randomised control trial.

GETTING STARTED: DEVELOPING A PROTOCOL FROM THE INITIAL HYPOTHESIS

Analytical studies need a hypothesis that specifies an anticipated association between predictor and outcome variables (or no association, as in a *null hypothesis*), so that statistical tests of significance can be performed.[3] Good hypotheses are specific and formulated in advance of commencement (a priori) of the study. Having chosen a subject to research and specifically a hypothesis to be tested,

Box 1 Features of a well designed RCT

- The sample to be studied will be appropriate to the hypothesis being tested so that any results are appropriately generalisable. The study will recruit sufficient patients to allow it to have a high probability of detecting a clinicaly important difference between treatments if a difference truly exists.
- There will be effective (concealed) randomisation of the subjects to the intervention/control groups (to eliminate selection bias and minimise confounding variables).
- Both groups will be treated identically in all respects except for the intervention being tested and to this end patients and investigators will ideally be blinded to which group an individual is assigned.
- The investigator assessing outcome will be blinded to treatment allocation.
- Patients are analysed within the group to which they were allocated, irrespective of whether they experienced the intended intervention (intention to treat analysis).
- Analysis focuses on testing the research question that initialy led to the trial (that is, according to the a priori hypothesis being tested), rather than "trawling" to find a significant difference.

Correspondence to:
Dr J M Kendall, North Bristol NHS Trust, Frenchay Hospital, Frenchay Park road, Bristol BS16 1LE, UK;
frenchayed@cableinet.co.uk

preparation should be thorough and is best documented in the form of a protocol that will outline the proposed methodology. This will start with a statement of the hypothesis to be tested, for example: "...that drug A is more efficacious in reducing the diastolic blood pressure than drug B in patients with moderate essential hypertension." An appropriate rationale for the study will follow with a relevant literature review, which is focused on any existing evidence relating to the condition or interventions to be studied.

The subject to be addressed should be of clinical, social, or economic significance to afford relevance to the study, and the hypothesis to be evaluated must contain outcomes that can be accurately measured. The subsequent study design (population sampling, randomisation, applying the intervention, outcome measures, analysis, etc) will need to be defined to permit a true evaluation of the hypothesis being tested. In practice, this will be the best compromise between what is ideal and what is practical.

Writing a thorough and comprehensive protocol in the planning stage of the research project is essential. Peer review of a written protocol allows others to criticise the methodology constructively at a stage when appropriate modification is possible. Seeking advice from experienced researchers, particularly involving a local research and development support unit, or some other similar advisory centre, can be very beneficial. It is far better to identify and correct errors in the protocol at the design phase than to try to adjust for them in the analysis phase. Manuscripts rarely get rejected for publication because of inappropriate analysis, which is remediable, but rather because of design flaws.

There are several steps in performing an RCT, all of which need to be considered while developing a protocol. The first is to choose an appropriate (representative) sample of the population from which to recruit. Having measured relevant baseline variables, the next task is to randomise subjects into one of two (or more) groups, and subsequently to perform the intervention as appropriate to the assignment of the subject. The pre-defined outcome measures will then be recorded and the findings compared between the two groups, with appropriate quality control measures in place to assure quality data collection. Each of these steps, which can be tested in a pilot study, has implications for the design of the trial if the findings are to be valid. They will now be considered in turn.

CHOOSING THE RIGHT POPULATION

This part of the design is crucial because poor sampling will undermine the generalisability of the study or, even worse, reduce the validity if sampling bias is introduced.[4] The task begins with deciding what kind of subjects to study and how to go about recruiting them. The *target* population is that population to which it is intended to apply the results. It is important to set inclusion and exclusion criteria defining target populations that are appropriate to the research hypothesis. These criteria are also typically set to make the researchers' task realistic, for within the target population there must be an accessible/appropriate sample to recruit.

The sampling strategy used will determine whether the sample actually studied is representative of the target population. For the findings of the study to be generalisable to the population as a whole, the sample must be representative of the population from which it is drawn. The best design is *consecutive* sampling from the accessible population (taking every patient who meets the selection criteria over the specified time period). This may produce an excessively large sample from which, if necessary, a subsample can be randomly drawn. If the inclusion criteria are broad, it will be easy to recruit study subjects and the findings will be generalisable to a comparatively large population. Exclusion criteria need to be defined and will include such subjects who have conditions which may contraindicate the intervention to be tested, subjects who will have difficulty complying with the required regimens, those who cannot provide informed consent, etc.

Summary: population sampling

- The study sample must be representative of the target population for the findings of the study to be generalisable.
- Inclusion and exclusion criteria will determine who will be studied from within the accessible population.
- The most appropriate sampling strategy is normally consecutive sampling, although stratified sampling may legitimately be required.
- A sample size calculation and pilot study will permit appropriate planning in terms of time and money for the recruitment phase of the main study.
- Follow CONSORT guidelines on population sampling.[6]

In designing the inclusion criteria, the investigator should consider the outcome to be measured; if this is comparatively rare in the population as a whole, then it would be appropriate to recruit at random or consecutively from populations at high risk of the condition in question (*stratified* sampling). The subsamples in a stratified sample will draw disproportionately from groups that are less common in the population as a whole, but of particular relevance to the investigator.

Other forms of sampling where subjects are recruited who are easily accessible or appropriate, (*convenience* or *judgmental* sampling) will have advantages in terms of cost, time, and logistics, but may produce a sample that is not representative of the target population and it is likely to be dificult to define exactly who has and has not been included.

Having determined an appropriate sample to recruit, it is necessary to estimate the size of the sample required to allow the study to detect a clinically important difference between the groups being compared. This is performed by means of a *sample size calculation*.[5] As clinicians, we must be able to specify what we would consider to be a clinically significant difference in outcome. Given this information, or an estimate of the effect size based on previous experience (from the literature or from a pilot study), and the design of the study, a statistical adviser will be able to perform an appropriate sample size calculation. This will determine the required sample size to detect the pre-determined clinically significant difference to a certain degree of power. As previously mentioned, early involvement of an experienced researcher or research support unit in the design stage is essential in any RCT.

After deciding on the population to be studied and the sample size required, it will now be possible to plan the appropriate amount of time (and money) required to collect the data necessary. A limited pilot of the methods is essential to gauge recruitment rate and address in advance any practical issues that may arise once data collection in the definitive study is underway. Pilot studies will guide decisions about designing approaches to recruitment and outcome measurement. A limited pilot study will give the investigator an idea of what the true recruitment rate will be (not just the number of subjects available, but also their willingness to participate). It may be even more helpful in identifying any methodological issues related to applying the intervention or measuring outcome variables (see below), which can be appropriately addressed.

RANDOMISATION: THE CORNERSTONE OF THE RCT

Various baseline characteristics of the subjects recruited should be measured at the stage of initial recruitment into the trial. These will include basic demographic observations, such as name, age, sex, hospital identification, etc, but more importantly should include any important prognostic factors. It will be important at the analysis stage to show that these potential

confounding variables are equally distributed between the two groups; indeed, it is usual practice when reporting an RCT to demonstrate the integrity of the randomisation process by showing that there is no significant difference between baseline variables (following CONSORT guidelines).[6]

The random assignment of subjects to one or another of two groups (differing only by the intervention to be studied) is the basis for measuring the marginal difference between these groups in the relevant outcome. Randomisation should equally distribute any confounding variables between the two groups, although it is important to be aware that differences in confounding variables may arise through chance.

Randomisation is one of the cornerstones of the RCT[7] and a true random allocation procedure should be used. It is also essential that treatment allocations are *concealed* from the investigator until recruitment is irrevocable, so that bias (intentional or otherwise) cannot be introduced at the stage of assigning subjects to their groups.[8] The production of computer generated sets of random allocations, by a research support unit (who will not be performing data collection) in advance of the start of the study, which are then sealed in consecutively numbered opaque envelopes, is an appropriate method of randomisation. Once the patient has given consent to be included in the trial, he/she is then irreversibly randomised by opening the next sealed envelope containing his/her assignment.

An alternative method, particularly for larger, multicentre trials is to have a remote randomisation facility. The clinician contacts this facility by telephone when he is ready to randomise the next patient; the initials and study number of the patient are read to the person performing the randomisation, who records it and then reads back the randomisation for that subject.

Studies that involve small to moderate sample sizes (for example, less than 50 per group) may benefit from "blocked" and/or "stratified" randomisation techniques. These methods will balance (where chance alone might not) the groups in terms of the number of subjects they contain, and in the distribution of potential confounding variables (assuming, of course, that these variables are known before the onset of the trial). They are the design phase alternative to statistically adjusting for confounding variables in the analysis phase, and are preferred if the investigator intends to carry out subgroup analysis (on the basis of the stratification variable).

Blocked randomisation is a technique used to ensure that the number of subjects assigned to each group is equally distributed. Randomisation is set up in blocks of a pre-determined set size (for example 6, 8, 10, etc). Randomisation for a block size of 10 would proceed normally until five assignments had been made to one group, and then the remaining assignments would be to the other group until the block of 10 was complete. This means that for a sample size of 80 subjects, exactly 40 would be assigned to each group. Block size must be blinded from the investigator performing the study and, if the study is non-blinded, the block sizes should vary randomly (otherwise the last allocation(s) in a block would, in effect, be unconcealed).

Stratified randomisation is a technique for ensuring that an important baseline variable (potential confounding factor) is more evenly distributed between the two groups than chance alone might otherwise assure. In examining the effect of a treatment for cardiac failure, for example, the degree of existing cardiac failure will be a baseline variable predicting outcome, and so it is important that this is the same in the two groups. To achieve this, the sample can be stratified at baseline into patients with mild, moderate, or severe cardiac failure, and then randomisation occurs within each of these "strata". There is a limited number of baseline variables that can be balanced by stratification because the numbers of patients within a stratum are reduced. In the above example, to stratify also for age, previous infarction, and the co-existence of diabetes would be impractical.

Summary: randomisation

- The random assignment of subjects into one of two groups is the basis for establishing a causal interpretation for an intervention.
- Effective randomisation will minimise confounding variables that exist at the time of randomisation.
- Randomisation must be concealed from the investigator.
- Blocked randomisation may be appropriate for smaller trials to ensure equal numbers in each group.
- Stratified randomisation will ensure that a potential baseline confounding variable is equally distributed between the two groups.
- Analysis of results should occur based on the initial randomisation, irrespective of what may subsequently actually have happened to the subject (that is, "intention to treat analysis").

Sample attrition ("drop outs"), once subjects have consented and been randomised, may be an important factor. Patients may refuse to continue with the trial, they may be lost to analysis for whatever reason, and there may be changes in the protocol (or mistakes) subsequent to randomisation, even resulting in the patient receiving the wrong treatment. This is, in fact, not that uncommon: a patient randomised to have a minimally invasive procedure may need to progress to an open operation, for example, or a patient assigned to medical treatment may require surgery at a later stage. In the RCT, the analysis must include an unbiased comparison of the groups produced by the process of randomisation, based on all the people who were randomised; this is known as analysis by *intention to treat*. Intention to treat analysis depends on having outcomes for all subjects, so even if patients "drop out", it is important to try to keep them in the trial if only for outcome measurement. This avoids the introduction of bias as a consequence of potentialy selectively dropping patients from previously randomised/balanced groups.

APPLYING THE INTERVENTION AND MEASURING OUTCOME: THE IMPORTANCE OF BLINDING

After randomisation there will be two (or more) groups, one of which will receive the test intervention and another (or more) which receives a standard intervention or placebo. Ideally, neither the study subjects, nor anybody performing subsequent measurements and data collection, should be aware of the study group assignment. Effective randomisation will eliminate confounding by variables that exist at the time of randomisation. Without effective blinding, if subject assignment is known by the investigator, bias can be introduced because extra attention may be given to the intervention group (intended or otherwise).[8] This would introduce variables into one group not present in the other, which may ultimately be responsible for any differences in outcome observed. Confounding can therefore also occur after randomisation. Double blinding of the investigator and patient (for example, by making the test treatment and standard/placebo treatments appear the same) will eliminate this kind of confounding, as any extra attentions should be equally spread between the two groups (with the exception, as for randomisation, of chance maldistributions).

While the ideal study design will be double blind, this is often difficult to achieve effectively, and is sometimes not possible (for example, surgical interventions). Where blinding is possible, complex (and costly) arrangements need to be made to manufacture placebo that appears similar to the test drug, to design appropriate and foolproof systems for packaging and labelling, and to have a system to permit rapid unblinding in the event of any untoward event causing the patient to become unwell. The hospital pharmacy can be invaluable in

Summary: intervention and outcome

- Blinding at the stage of applying the intervention and measuring the outcome is essential if bias (intentional or otherwise) is to be avoided.
- The subject and the investigator should ideally be blinded to the assignment (double blind), but even where this is not possible, a blinded third party can measure outcome.
- Blinding is achieved by making the intervention and the control appear similar in every respect.
- Blinding can break down for various reasons, but this can be systematically assessed.
- Continuous outcome variables have the advantage over dichotomous outcome variables of increasing the power of a study, permitting a smaller sample size.

organising these issues. Blinding may break down subsequently if the intervention has recognisable side effects. The effectiveness of the blinding can be systematically tested after the study is completed by asking investigators to guess treatment assignments; if a significant proportion are able to correctly guess the assignment, then the potential for this as a source of bias should be considered.

Once the intervention has been applied, the groups will need to be followed up and various outcome measures will be performed to evaluate the effect or otherwise of that intervention. The outcome measures to be assessed should be appropriate to the research question, and must be ones that can be measured accurately and precisely. Continuous outcome variables (quantified on an infinite arithmetic scale, for example, time) have the advantage over dichotomous outcome variables (only two categories, for example, dead or alive) of increasing the power of a study, permitting a smaller sample size. It may be desirable to have several outcome measures evaluating different aspects of the results of the intervention. It is also necessary to design outcome measures that will detect the occurrence of specified adverse effects of the intervention.

It is important to emphasise, as previously mentioned, that the person measuring the outcome variables (as well as the person applying the intervention) should be blinded to the treatment group of the subject to prevent the introduction of bias at this stage, particularly when the outcome variable requires any judgement on the part of the observer. Even if it has not been possible to blind the administration of the intervention, it should be possible to design the study so that outcome measurement is performed by someone who is blinded to the original treatment assignment.

QUALITY CONTROL

A critical aspect of clinical research is quality control. Quality control is often overlooked during data collection, a potentially tedious and repetitive phase of the study, which may lead subsequently to errors because of missing or inaccurate measurements. Essentially, quality control issues occur in clinical procedures, measuring outcomes, and handling data. Quality control begins in the design phase of the study when the protocol is being written and is first evaluated in the pilot study, which will be invaluable in testing the proposed sampling strategy, methods for data collection and subsequent data handling.

Once the methods part of the protocol is finalised, an operations manual can be written that specifically defines how to recruit subjects, perform measurements, etc. This is essential when there is more than one investigator, as it will standardise the actions of all involved. After allowing all those involved to study the operations manual, there will be the opportunity to train (and subsequently certify) investigators to perform various tasks uniformly.

Ideally, any outcome measurement taken on a patient should be precise and reproducible; it should not depend on the observer who took the measurement.[4] It is well known, for example, that some clinicians in their routine medical practice record consistently higher blood pressure values than others. Such interobserver variation in the setting of a clinical trial is clearly unacceptable and steps must be taken to avoid it. It may be possible, if the trial is not too large, for all measurements to be performed by the same observer, in which case the problem is avoided. However, it is often necessary to use multiple observers, especially in multicentre trials. Training sessions should be arranged to ensure that observers (and their equipment) can produce the same measurements in any given subject. Repeat sessions may be necessary if the trial is of long duration. You should try to use as few observers as possible without exhausting the available staff. The trial should be designed so that any interobserver variability cannot bias the results by having each observer evaluate patients in all treatment groups.

Inevitably, there will be a principal investigator; this person will be responsible for assuring the quality of data measurement through motivation, appropriate delegation of responsibility, and supervision. An investigators' meeting before the study starts and regular visits to the team members or centres by the principal investigator during data collection, permit communication, supervision, early detection of problems, feedback and are good for motivation.

Quality control of data management begins before the start of the study and continues during the study. Forms to be used for data collection should be appropriately designed to encourage the collection of good quality data. They should be user friendly, self explanatory, clearly formatted, and collect only data that is needed. They can be tested in the pilot. Data will subsequently need to be transcribed onto a computer database from these forms. The database should also be set up so that it is similar in format to the forms, allowing for easy transcription of information. The database can be pre-prepared to accept only variables within given permissible ranges and that are consistent with previous entries and to alert the user to missing values. Ideally, data should be entered in duplicate, with the database only accepting data that are concordant with the first entry; this, however, is time consuming, and it may be adequate to check randomly selected forms with a printout of the corresponding datasheet to ensure transcription error is minimal, acting appropriately if an unacceptably high number of mistakes are discovered.

Once the main phase of data collection has begun, you should try to make as few changes to the protocol as possible. In an ideal world, the pilot study will have identified any issues that will require a modification of the protocol, but inevitably some problem, minor or major, will arise once the study has begun. It is better to leave any minor alterations that are considered "desirable" but not necessary and resist the inclination to make changes. Sometimes, more substantive issues are highlighted and protocol modification is necessary to strengthen the study. These changes should be documented and disseminated to all the investigators (with appropriate changes made to the operations manual and any re-training performed as necessary). The precise date that the revision is implemented is noted, with a view to separate analysis of data collected before and after the revision, if this is considered necessary by the statistical advisor. Such revisions to the protocol should only be undertaken if, after careful consideration, it is felt that making the alteration will significantly improve the findings, or not changing the protocol will seriously jeopardise the project. These considerations have to be balanced against the statistical difficulty in analysis after protocol revision.

...SOME FINAL THOUGHTS

A well designed, methodologically sound RCT evaluating an intervention provides strong evidence of a cause-effect

Summary: quality control

- An inadequate approach to quality control will lead to potentially significant errors due to missing or inaccurate results.
- An operations manual will allow standardisation of all procedures to be performed.
- To reduce interobserver variability in outcome measurement, training can be provided to standardise procedures in accordance with the operations manual.
- Data collection forms should be user friendly, self explanatory, and clearly formatted, with only truly relevant data being collected.
- Subsequent data transfer onto a computerised database can be safe guarded with various measures to reduce transcription errors.
- Protocol revisions after study has started should be avoided if at all possible, but, if necessary, should be appropriately documented and dated to permit separate analysis.

relation if one exists; it is therefore powerful in changing practice to improve patient outcome, this being the ultimate goal of research on therapeutic effectiveness. Conversely, poorly designed studies are dangerous because of their potential to influence practice based on flawed methodology. As discussed above, the validity and generalisability of the findings are dependent on the study design.

Early involvement of the local research support unit is essential in developing a protocol. Subsequent peer review and ethical committee review will ensure that it is well designed, and a successful pilot will ensure that the research goals are practical and achievable.

Delegate tasks to those who have the expertise; for example, allow the research support unit to perform the randomisation, leave the statistical analysis to a statistician, and let a health economist advise on any cost analysis. Networking with the relevant experts is invaluable in the design phase and will contribute considerably to the final credence of the findings.

Finally, dissemination of the findings through publication is the final peer review process and is vital to help others act on the available evidence. Writing up the RCT at completion, like developing the protocol at inception, should be thorough and detailed[9] (following CONSORT guidelines[6]), with emphasis not just on findings, but also on methodology. Potential limitations or sources of error should be discussed so that the readership can judge for themselves the validity and generalisability of the research.[10]

Further reading

Sackett DL, Haynes RB, Guyatt GH, *et al*. *Clinical epidemiology: a basic science for clinical medicine.* 2nd edn. Toronto: Little, Brown, 1991.

Sackett DL, Richardson WS, Rosenberg W, *et al*. *Evidence-based medicine: how to practice and teach EBM*. Edinburgh: Churchill Livingstone, 1997.

Polgar S. *Introduction to research in health sciences.* 2nd edn. Edinburgh: Churchill Livingstone, 1991.

Bland M. *An introduction to medical statistics.* Oxford: Oxford Medical Publications, 1987.

REFERENCES

1 **Sibbald B**, Roland M. Understanding controlled trials: Why are randomised controlled trials important? *BMJ* 1998;**316**:201.
2 **Pocock SJ**. *Clinical trials: a practical approach*. Chichester: Wiley, 1984.
3 **Hulley SB**, Cunnings SR. *Designing clinical research—an epidemiological approach*. Chicago: Williams and Wilkins, 1988.
4 **Bowling A**. *Researching methods in health: investigating health and health services*. Buckingham: Open University Press, 1997.
5 **Lowe D**. *Planning for medical research: a practical guide to research methods*. Middlesborough: Astraglobe, 1993.
6 **Begg C**, Cho M, Eastwood S, *et al*. Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA* 1996;**276**:637–9.
7 **Altman DG**. Randomisation. *BMJ* 1991;**302**:1481–2.
8 **Schultz KF**, Chalmers I, Hayes RJ, *et al*. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.
9 **The Standards of Reporting Trials Group**. A proposal for structured reporting of randomised controlled trials. *JAMA* 1994;**272**:1926–31.
10 **Altman DG**. Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* 1996;**313**:570–1.

CE

Randomized Controlled Trials
Methodological Concepts and Critique

Susan S. Salmond

Randomized controlled trials, also known as true experiments or intervention studies, are considered to be the gold standard research design for demonstrating a cause-and-effect relationship between an intervention and an outcome. This article will describe key methodological concepts that make the randomized controlled trial this gold standard. Practicing from an evidence-based perspective requires practitioners to be able to critique the strengths and weaknesses of a research study in order to make decisions about adoption of the intervention into one's practice area. Key components to the critique of a randomized controlled trial are defined and the process is illustrated by a critique of Gallo and colleagues article, *A Study of Naloxone Effect on Urinary Retention in the Patient Receiving Morphine PCA*, published in this issue.

Randomized controlled trials or true experiments constitute the most rigorous research design for intervention studies. In an intervention study, the investigator is interested in determining whether a cause–effect relation exists between treatment and outcome(s) (Sibbaid, & Roland, 1998; Stommel & Wills, 2004). To achieve this, there must be comparison groups where individuals in each group receive a different level of exposure to the intervention. The outcomes can then be compared to determine whether the intervention makes a difference.

This article will highlight the key principles underlying intervention studies and then examine components of methodological rigor with an accompanying critique of Gallo, DuRand, and Pshon's (2008) study of the effect of naloxone on urinary retention in the patient receiving morphine via patient-controlled analgesia (PCA) to illustrate the key components. In the study, Gallo et al. were interested in showing a cause–effect relation between administration of low-dose naloxone and complications of urinary retention. The study is a single factor, posttest-only experimental design as illustrated in Figure 1. The target population consisted of orthopaedic surgical patients who were randomly assigned to two groups: the experimental or intervention group and the control group. The intervention group received naloxone in conjunction with PCA morphine and the control group received PCA morphine with no naloxone. It was hypothesized that administration of low-dose intravenous naloxone would decrease the incidence of urinary retention, need for catheterization, and hospital length of stay. The researchers identified appropriate outcome (dependent) variables and indicated objective approaches to measure these outcomes.

Control: The Guiding Principle in Randomized Controlled Trial Design

Randomized controlled trials are considered the "gold standard" for intervention studies, as this design introduces the notion of *control*, thereby minimizing bias. Control is a systematic process of design that isolates the effect of the intervention being tested by ruling out other potential "causes" of the effects under study. Other causes, also called rival causes, are referred to as confounding variables in that they provide an alternative explanation for the effect on the dependent variable.

Planning for the design of a study requires identification of confounding factors that could create rival hypotheses so that these factors can be appropriately controlled for in the design. A thorough literature search on the selected dependent variable assists in identifying potential confounding factors. In the study by Gallo et al., a review of the literature on urinary retention and morphine PCA and low-dose naloxone would be called for. The results of this often appear in the review of the literature section of the manuscript or may not be evident except through examination of the research design.

Review of the empirical literature links postoperative urinary retention to various factors including a history of urinary problems, advanced age (especially over the age of 70), male gender, the total amount of fluid replacement over a 24-hr postoperative period, type of anesthesia (greater incidence with epidural analgesia), pain management medications and dosage (greater

Susan S. Salmond, EdD, RN, Interm Dean, School of Nursing, University of Medicine and Dentistry of New Jersey, Newark, NJ.

The author has no significant interest, financial or otherwise, to any company that might have an interest in the publication of this educational activity.

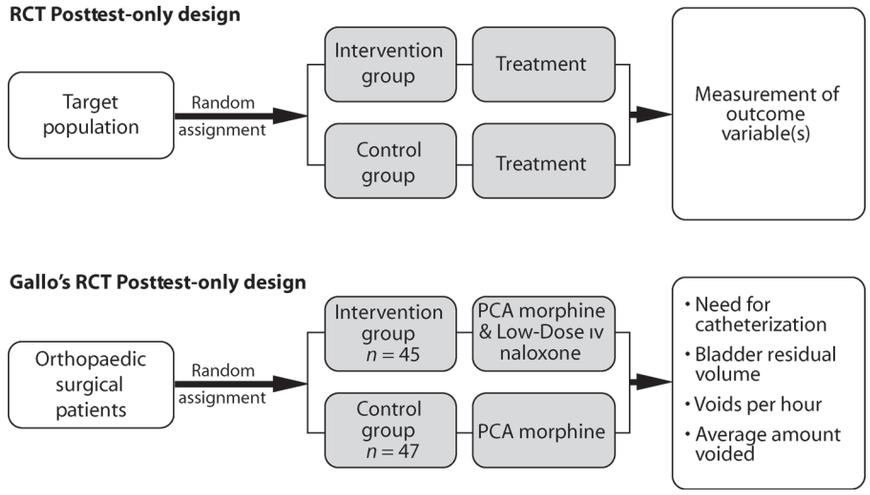


FIGURE 1. Randomized controlled trial (RCT) posttest-only design. *Note.* IV = intravenous; PCA = patient-controlled analgesia.

with narcotic analgesia), and administration with PCA (Fernandes, Carmo, Costa, & Saraiva, 2007; Koch, Grinberg, Gary & Farley, 2000; O'Riordan, Hopkins, Ravenscroft, & Stevens, 2000; Sarasin, Walton, Singh, & Clark, 2006; Wynd, Wallace, & Smith, 1996). Gallo and colleagues accounted for these variables in their research design. Sampling criteria excluded individuals over age 65 and those with a preexisting urinary condition. Random assignment was used which would likely distribute the other variables, and the researchers compared the groups for equivalence.

Randomization: The Process for Achieving Control

A valid comparison of the intervention and control groups requires that the two groups be as alike as possible on the critical demographic and clinical variables that could pose alternative hypotheses for the outcome effect. Moreover, in all intervention studies, there may be confounding effects from factors that the researchers are not even aware of. Randomization is the preferred approach for equalizing conditions and differences of these potentially confounding factors across the different arms of treatment (Melnyk & Fineout-Overholt, 2005). Through randomization, the two groups (experimental and control) are as similar as possible, with the major difference remaining between them being their exposure to different treatments (Stommel & Wills, 2004). Randomization eliminates both conscious bias and unconscious bias associated with the selection of a treatment for a given patient (Stanley, 2007). For example, without randomization, it could be possible to bias the results by assigning patients with anticipated "noncomplicated" surgical procedures to the treatment group. These patients would likely have shorter operative times and require less narcotic analgesia postoperatively; therefore fewer urinary retention complications. In this example, failure to randomize would predispose to imbalances in critical baseline characteristics between the two study groups, thus introducing selection bias and generating rival hypotheses for the occurance of decreased urinary retention.

Kerlinger and Lee (2000) emphasize the critical importance of control in intervention studies to ensure the equivalence of the comparison groups on all factors but the intervention variable itself. They indicate that the best approach to achieving this control is to *randomize whenever possible*. Nonrandomized studies have been reported to overestimate or underestimate treatment effects (Bhandari & Haynes, 2005).

Randomization can be accomplished by selecting participants at random or assigning participants to treatment groups at random. It is not always possible to select participants at random because researchers have access to limited populations and selecting random participants significantly increases the time needed to obtain an adequate sample size. When random selection is not used it is important to randomly assign participants to either the treatment or control group. If treatment groups receive the treatment at different times and with different investigators, times and investigators can be randomly assigned as well.

An additional protection against bias during randomization is *concealment of allocation*. Allocation is concealed when investigators enrolling subjects cannot determine in advance the treatment assignment of the next patient to be enrolled into the study. This prevents possible selection bias whereby the investigator can systematically allocate patients to the intervention or control group on the basis of some subjective perhaps unconscious bias. Use of a random-numbers table successfully conceals allocation and prevents biased assignment or response to assignment awareness.

Blinding

Blinding is another methodological safeguard to ensure that individuals (participants, treatment team members, or data collectors/outcome assessors) are unaware of the treatment (intervention or control) that has been

received/administered/assessed until the study is complete so that differential treatment or a biased assessment of outcome (detection bias) is reduced or eliminated (Licciardone & Russo, 2006). Participants or the clinical team/research team that knows the group assignment may respond differently on the basis of group assignment, thereby biasing the results. Single blinding is when the participant is blinded to which treatment (intervention or control) is being administered. In double blinding, neither the participants nor the data collectors/observers know which subjects are assigned to treatment and control conditions. In triple-blind studies, none of the study participants, the healthcare professionals treating them, or the persons who observe or record the outcomes are aware of the treatment assignment (Stommel & Wills, 2004). Lack of blinding has been associated with increased or decreased magnitudes of observed treatment effects, especially when differences between treatments are small (Poolman et al., 2007). Gallo and colleagues did not use any blinding in their design, thus introducing the possibility of bias, especially at the point of outcomes measurement.

Critique of Randomized Controlled Trials

Although experimental designs/randomized controlled trials are the most rigorous research design, it cannot be assumed to be of good quality. The reader of any intervention study needs to critique the study of interest to make conclusions as to the rigor of the particular study and its applicability for use in the reader's own clinical setting. Table 1 summarizes key components to be evaluated and critiques the study by Gallo et al. on these methodological components.

Clear Research Design

The first group of methodological factors relate to the clarity and appropriateness of the research design. The dependent/outcome variables should be clearly identified with reliable and valid approaches to measuring the variables specified. In critiquing the outcome variable, consider whether all clinically important outcomes were measured. Similarly, the treatment arms (in this case control group and experimental group) must be fully described not only to facilitate reproducibility in future studies but also to allow for comparison across studies. Melnyk and Fineout-Overholt (2005) point out that in some intervention studies, there is a significant difference in the time spent and the degree of patient/researcher interaction between the experimental and control groups, leading to the potential of an effect because of the differential attention factor. In critiquing the treatment arms, evaluate the control group treatment for its similarity with the intervention treatment. Sampling criteria and sample size are important to evaluate in the research design. What are the inclusion and exclusion criteria? Has a power analysis been done in advance to estimate the needed sample size? A small sample may lack the power needed to demonstrate a treatment effect and result in a false negative. This is known as a type II error—the researcher accepts the null hypothesis (that no difference exists) when it should have been rejected (as a real treatment effect exists). Most treatments offer a small additional benefit and therefore require larger samples. Too small a sample could mean that no benefit is detected even though there is actually a small positive effect.

Approaches for Ensuring Validity of the Findings

The key question in internal validity is whether the observed changes can be attributed to the intervention (i.e., the cause) and not to other possible causes ("alternative explanations" for the outcome). Processes used to control for alternative explanations include randomization, concealment of allocation, and blinding.

Ensuring validity requires attention to follow-up assessment, loss of participants during the study, the similarity and differences in the control and intervention group, and whether participants were in fact analyzed within the group to which they were originally assigned. Follow-up assessments need to be conducted long enough to study the effects of the intervention. An educational intervention that tests only after implementation of the education does not have an adequate follow-up to determine the extent of the effect of the educational intervention.

Examine the original sample size and compare it with the sample size used for data analysis to determine the mortality rate. The percentage of subjects who did not complete the study should not exceed 20%. In cases of mortality and noncompliance, a statistical approach to preserving randomizations is to ensure that participants are analyzed in the groups to which they were originally assigned—a process known as intention to treat. By specifying how to handle noncompliant patients and analyzing their data in the groups they were originally allocated to, randomization is preserved.

Group equivalence is important in assuring that the treatment, not difference among the groups, caused the effect. With small sample sizes, the intended effect of randomization may not be achieved. In addition, subject mortality may disrupt group equivalence. Examination of the statistical equivalence or distribution of key factors between groups is done to assure the effectiveness of randomization or the statistical equivalence across groups.

Determing the Meaningfulness of Results

Intervention studies should report both statistical significance and clinical relevance of the findings. This can be accomplished by reporting on the magnitude or size, the precision, and the clinical meaningfulness of the findings.

In an intervention study, the effect size indicates how effective the experimental treatment was. Effect size is calculated by subtracting the mean of the control group from the mean of the experimental group and dividing the resultant by the pooled standard deviation. Small, medium, and large effects are designated as 0.2, 0.5, and 0.8, respectively. As effect sizes are not dependent on sample size, they are a critical indicator of the magnitude of the experimental intervention (Melnyk & Fineout-Overholt, 2005).

TABLE 1. KEY COMPONENTS TO THE CRITIQUE OF RANDOMIZED CONTROLLED TRIALS

Critique components	Review of Gallo et al.'s study
Clear and appropriate research design	
1 Outcome/Dependent Variables	
a. The dependent or outcome variable(s) is clearly defined.	Primary outcome variable Urinary retention Secondary outcome variables Length of stay Pain levels
b. A *standardized* approach for measuring the outcome variable(s) is identified and allows for clear comparison of findings from the groups.	Urinary output # of patient voids Total voiding amount Need for catheterization Catheterization/no catheterization Catheterization time Amount residual urine Bladder distention Subjective feelings of bladder fullness Bladder ultrasound Length of stay Minutes from time of admission to time of discharge Pain/pain control 0–10 Numeric/visual analog scale Total morphine delivered
c. All clinically important outcomes are measured?	Yes. Clinically important outcomes were measured.
2 Experimental Treatment/Intervention	
a. The intervention(s) or treatment(s) that will be manipulated is clearly identified to ensure reproducibility.	Two groups: Control and intervention Control group: PCA morphine Intervention group: PCA morphine with 0.1 mg iv naloxone every 4 hr
b. The control group intervention is similar in time and interaction with the patient to control for effects secondary to time and attention.	The control group intervention was similar in time and interaction with the patient.
3 Sampling	
a. The target population/sample is clearly specified.	Target population: Orthopaedic patients from one acute care facility undergoing surgery and receiving postoperative morphine PCA.
b. Sampling criteria (inclusion and exclusion criteria) clearly depict the sample. Inclusion–exclusion criteria can be used to control for Confounding variables. For example, by excluding those with a history of urinary problems are controlling for a rival hypothesis.	Inclusion criteria/sample: Ninety-seven orthopaedic patients younger than 65 years scheduled for elective orthopaedic shoulder, hip, or knee surgery and agreeing to participate. Exclusion criteria: History of urinary problems (bladder or prostate cancer, renal dialysis), documented naloxone allergy, or opioid dependency.
c. Can adequate sample size is used. This can be done in advance by performing a power analysis, which guides the researcher on the sample size needed to demonstrate a treatment effect.	No power analysis was performed in advance to specify the sample size needed to detect a moderate difference. This author used G-Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) to calculate power on the basis of the findings of the study and found the study to be adequately powered for the number of voids per hour (effect size = 0.66, power = 0.92) and the average bladder scan amount (effect size = 0.516, power = 0.78) and underpowered for the average amount voided (effect size 0.280, power = 0.373) and the need for catheterization (effect size 0.3, total sample size suggested = 220).
4 *Methods for assuring internal validity of the study*	
a. Rigorous approaches to control confounding variables are used. Assess: Were participants randomized to the experimental and control groups? Was group assignment concealed from those enrolling subjects into the groups?	Random sampling was not used. The 97 participants in the original sample were randomly assigned using a computer-generated random-numbers table, thus concealing allocation. This resulted in subgroup sizes of: Control group: 45 Intervention group: 52

(*continued*)

TABLE 1. KEY COMPONENTS TO THE CRITIQUE OF RANDOMIZED CONTROLLED TRIALS (*Continued*)

Critique components	Review of Gallo et al.'s study
The goal of this approach is to have equivalence across groups on all factors but the intervention variable itself. It does not mean that the groups must be equal in number.	
b. Blinding strategies are used to decrease detection bias. Assess: Were key groups blinded to treatment allocation? Specifically determine whether blinding occurred with: Patients Those administering intervention(s) Those scoring outcome(s) Data analysts	No blinding was used even though blinding would have been possible. It is important to assess to what degree detection bias could be present. No subjective patient measures were used; therefore, patient detection bias is not likely. However, scorer bias could be present, as the scorer is responsible for measuring urinary output and bladder distension and these reports could be skewed toward the observer's bias.
c. Were follow-up assessments are conducted long enough to study the effects of the intervention?	Follow-up assessments were effectively planned until normal urinary function returned Every 4 hr: Total morphine delivered No. of patient voids Total voiding amount Every 8 hr: Subjective feelings of bladder fullness Palpation for bladder distention Bladder ultrasound As indicated: Need for catheterization Amount of residual urine
d. Did all subjects complete the study? (The goal is for at least 80% of subjects to complete the study.)	A total of 97 participants consented to study and 96 completed the study (98%). One participant from the experimental group was lost because of increased pain after administration of naloxone. Data analysis was completed on 45 participants from the control group and 51 from the intervention group. No significant bias results from this. Two subjects from the control group and four from the experimental group did not have ongoing monitoring (no. of voids, amount voided, and bladder retention) because of need for indwelling Foley catheter placement. These subjects were included in the need for catheterization analysis.
e. Were participants analyzed in the groups to which they were randomized? Participants should be analyzed within the group to which they were allocated, irrespective of whether they experienced the intended intervention (intention to treat analysis).	There was one dropout in the intervention group that was not included in the final data analysis according to the assigned group. This one case should not significantly impact the results; however, preference would be to continue with data collection and included the subject in the experimental group even though they did not receive ongoing naloxone administration.
f. Were patients in the treatment and control groups similar with respect to known prognostic factors (key demographic and clinical factors)?	There were no significant differences across the intervention and control groups on gender, procedure type, age, hours on PCA, or hours in study.
5 *What are the results of the study and are they important?*	
a. Analysis is focused on estimating the size of the difference in predefined outcomes between intervention groups—How large was the treatment effect? This is accomplished by determining the effect size or level of risk reduction.	Authors reported *p* values rather than risk reduction or effect size. Significance was found for average number of voids per hour (higher in the experimental group), average bladder scan amount per hour (lower in the experimental group), and percentage of patients who were catheterized.
Effect size is an estimate of the strength of the treatment. It is calculated by taking the mean of the control group minus the mean of the experimental group and dividing this amount by the pooled standard deviation.	Calculation of effect size for number of voids, bladder scan amount, and amount voided was not reported by Gallo but was calculated by this author using the formula given on the left. Results showed a medium effect size for number of voids (0.66) and bladder scan amount (–0.516) and a small effect size for amount voided (0.28) and need for catheterization (0.30).

(*continued*)

TABLE 1. KEY COMPONENTS TO THE CRITIQUE OF RANDOMIZED CONTROLLED TRIALS (*Continued*)

Critique components	Review of Gallo et al.'s study
Relative risk reduction measures how much the risk is reduced in the experimental group compared to the control group. Relative risk reduction would be calculated as Control Event Rate minus Experimental Event Rate divided by the Control Event Rate.	Calculation of risk was not reported by Gallo but was calculated for the variable "need for catheterization" by this author using the formula given on the left. *Relative risk reduction*: (0.244 – 0.115)/0.244 = 0.528. This means that the need for catheterization was reduced by 52.8% in the treatment group (PCA morphine with naloxone) compared with the control group (PCA morphine only).
Absolute risk reduction is the absolute difference in outcome rates between the control and treatment groups.	*Absolute risk reduction*: (0.244 – 0.115) = 0.129 or 12.9%. This means that for every 100 patients enrolled in the treatment group, about 13 catheterizations would be averted. These calculations support the clinical significance of the findings.
b. Having established the clinical importance of the findings, one must then determine whether the probable treatment benefits are worth the effort. This is calculated by the "number needed to treat" (NNT) and is responsive to the question: Are the likely treatment benefits worth the potential harm and costs? This is calculated by taking the inverse of the absolute risk reduction.	Absolute risk reduction = 12.9%. NNT = 1/12.9% = 7.75. This means that for every eight patients treated with naloxone, one case of catheterization would be prevented.
c. How precise is the treatment effect? This is determined by confidence intervals of the outcome values. The narrower the confidence interval, the more confident one can be that the study result is the true result.	Confidence intervals were not calculated by the researchers.
6 *Relevance to practice*	
a. The applicability of the results is assessed by the question: Can the results be applied to my patient?	If the patients in your practice meet the inclusion criteria and do not violate any of the exclusion criteria, then you can apply the results to their care with confidence. In this case, exclusion criteria consisted of patients older than 65 years. So, the results may not be applicable for older patients.
Is the treatment feasible in my clinical setting?	Factors relevant to feasibility include costs of naloxone therapy, which has not been reported.
b. What are the patient's values and expectations regarding the outcome that is trying to be prevented and the treatment itself?	This is not known; however, it is likely that patients would prefer not to be catheterized because of the accompanying risks of such a procedure. The naxolone treatment is not accompanied by any significant discomfort.

Note. PCA = patient-controlled analgesia.

Confidence intervals are used to express the precision and uncertainty of the findings. When taking the findings from a sample, the ultimate goal is to be able to generalize to the larger population itself. No matter how carefully the sample has been selected to be a fair and unbiased representation of the population, relying on information from a sample will always lead to some level of uncertainty. A confidence interval is a range of values that tries to quantify this uncertainty. It can be viewed as a range of plausible values. A narrow confidence interval implies high precision; we can specify plausible values within a tiny range. A wide interval implies poor precision; we can only specify plausible values to a broad and uninformative range.

An important question needed to interpret the results of a clinical trial is whether the measured effect size is clinically important. Three commonly used measures of effect size are relative risk reduction, absolute risk reduction, and the number needed to treat to prevent one bad outcome. These findings put the results into a meaningful interpretation of clinical relevance.

Many studies report only the statistical significance of the findings, which is largely dependent on the sample size and statistical power. The larger the sample, the greater the power and probability of detecting significant differences between groups even when the effect size of the treatment is small. In the study by Gallo et al., the effect size of catheterization was small at 0.30 and statistical significance was not found. When risk and number needed to treat was calculated, it was found that even with a small effect size, the need for catheterization was reduced by 52.8% in the treatment group; or when looking at the absolute risk reduction, the findings indicated that for every 100 patients enrolled in the treatment group, about 13 catheterizations could be averted. With the inherent risks of catheterization, averting 13 catheterizations is clinically meaningful. If one considered only the *p* value, one would conclude that naloxone was not effective in reducing the number of complications. However, looking at the effect size, risk reduction, and number needed to treat, it is apparent that these findings are clinically meaningful.

Relevance to Practice

Evidence-based practice requires the integration of individual clinical expertise with the best available external clinical evidence from systematic research, available resources, and our patient's unique values and circumstances. Thus, the final area to assess when critically appraising an intervention study is its relevance to one's own clinical population. Answering questions of clinical relevance assists the practitioner in determining whether the proposed intervention should be adopted in their setting with their specific population. To determine this, compare your practice population with the population in the study. Determine whether the treatment is feasible in your own clinical setting. Finally, determine your patient's values and expectations and whether they are congruent with the proposed treatment and outcome effect.

Summary

Gallo and colleagues have studied an important topic to orthopaedic nurses—urinary retention in the postoperative orthopaedic surgical patient receiving morphine PCA. A randomized controlled design was the appropriate selection for an intervention study. The research design was clearly described and the outcome measures were objective and clinically relevant. Random assignment and allocation concealment were effectively accomplished and the intervention and control groups were similar on important demographic and clinical characteristics. Blinding was not done, which decreases the validity of the findings.

The authors reported only on statistical significance with findings that naloxone was effective in influencing the number of voids per hour and the average bladder scan amount. Average amount voided and need for catheterization showed differences; however, they were not statistically significant. As the sample size was small, a post hoc power analysis showed the study to be underpowered which would account for the lack of statistical significance. Further analysis by this author in calculating clinical meaningfulness found a moderate effect size for number of voids per hour and the average bladder scan amount, and a small effect size for average amount voided and the need for catheterization. Calculation of risk reduction and number needed to treat revealed the findings to be clinically meaningful in reducing need for catheterization,

Study findings showed that low-dose naloxone can be safely used without compromising pain management and that it may be efficacious in reducing the negative outcome of urinary retention and need for catheterization. Replication of this study should consider three methodological changes: (1) using a preimplementation power analysis to determine necessary sample size; (2) blinding the outcome assessor; (3) and using statistics associated with clinical relevance. Moreover, patients over the age of 65 should be included, as this is the group that is most vulnerable to urinary retention and it is important to establish the effectiveness of naloxone in this age group.

References

Bhandari, M., & Haynes, R. B. (2005). How to appraise the effectiveness of treatment. *World Journal of Surgery, 29*(5), 570–575.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191.

Fernandes, B., Carmo, M., Costa, V., & Saraiva, R. A. (2007). Postoperative urinary retention: Evaluation of patients using opioid analgesic. *Revista Latino-Americana de Enfermagem, 15*(2), 318–322.

Gallo, S., DuRand, J., & Pshon, N. (2008). A study of naloxone effect on urinary retention in the patient receiving morphine PCA. *Journal of Orthopaedic Nursing, 27*(2), 111–115.

Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research.* Belmont, CA: Wadsworth.

Koch, Grinberg, Gary, & Farley. (2000).

Licciardone, J. C., & Russo, D. P. (2006). Blinding protocols, treatment credibility, and expectancy: Methodological issues in clinical trials of osteopathic manipulative treatment. *Journal of the American Osteopathic Association, 106*(8), 457–463.

Melnyk, B. M., & Fineout-Overholt, E. (2005). Rapid critical appraisal of randomized controlled trials (RCTs): An essential skill for evidence-based practice (EBP). *Pediatric Nursing, 31*(1), 50–52.

O'Riordan, J. A., Hopkins, P. M., Ravenscroft, A., & Stevens, J. D. (2000). Patient-controlled analgesia and urinary retention following lower limb joint replacement: Prospective audit and logistic regression analysis. *European Journal of Anaesthesiology, 17,* 431–435.

Poolman, R. W., Struijs, A. A., Krips, R., Sierevelt, I. N., Marti, R. K., Farrokhyar, F., et al. (2007). Reporting of outcomes in orthopaedic randomized trials: Does blinding of outcome assessors matter? *Journal of Bone & Joint Surgery, 89-A*(3), 550–558.

Sarasin, S. M., Walton, M. J., Singh, H. P., & Clark, D. I. (2006). Can a urinary tract symptom score predict the development of postoperative urinary retention in patients undergoing lower limb arthroplasty under spinal anaesthesia? A prospective study. *Annals of the Royal College of Surgeons of England, 88*(4), 394–398.

Sibbaid, B., & Roland, M. (1998). Understanding controlled trials: Why are randomised controlled trials important? *British Medical Journal, 316,* 201.

Stanley, K. (2007). Design of randomized controlled trials. *Circulation, 115*(3), 1164–1169.

Stommel, M., & Wills, C. E. (2004). *Clinical research: Concepts and principles for advanced practice nurses.* Philadelphia: Lippincott, Williams & Wilkins.

Wynd, C. A., Wallace, M., & Smith, K. M. (1996). Factors influencing postoperative urinary retention following orthopaedic surgical procedures. *Orthopaedic Nursing, 15*(1), 43–50.

AOGS
Acta Obstetricia et Gynecologica Scandinavica

AOGS REVIEW

A simplified guide to randomized controlled trials

AMAR BHIDE[1], PRAKESH S. SHAH[2,3] & GANESH ACHARYA[4,5,6]

[1]Fetal Medicine Unit, St. Georges University Hospital, London, UK, [2]Division of Neonatology, Department of Pediatrics, Mount Sinai Hospital, Toronto, ON, [3]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada, [4]Department of Clinical Science, Intervention and Technology, Karolinska Institute and Center for Fetal Medicine, Karolinska University Hospital, Stockholm, Sweden, [5]Women's Health and Perinatology Research Group, Department of Clinical Medicine, UiT-The Arctic University of Norway, Tromsø, and [6]Department of Obstetrics and Gynecology, University Hospital of North Norway, Tromsø, Norway

Key words
Clinical trial, good clinical practice, random allocation, randomized controlled trial, research methods, study design

Correspondence
Ganesh Acharya, Department of Clinical Science, Intervention and Technology, Karolinska Institute and Center for Fetal Medicine, Karolinska University Hospital – K57, 141 86 Stockholm, Sweden.
E-mail: ganesh.acharya@ki.se

Conflict of interest
The authors have stated explicitly that there are no conflicts of interest in connection with this article.

Please cite this article as: Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand 2018; 97:380–387.

Received: 12 January 2018
Accepted: 22 January 2018

DOI: 10.1111/aogs.13309

Abstract

A randomized controlled trial is a prospective, comparative, quantitative study/experiment performed under controlled conditions with random allocation of interventions to comparison groups. The randomized controlled trial is the most rigorous and robust research method of determining whether a cause–effect relation exists between an intervention and an outcome. High-quality evidence can be generated by performing an randomized controlled trial when evaluating the effectiveness and safety of an intervention. Furthermore, randomized controlled trials yield themselves well to systematic review and meta-analysis providing a solid base for synthesizing evidence generated by such studies. Evidence-based clinical practice improves patient outcomes and safety, and is generally cost-effective. Therefore, randomized controlled trials are becoming increasingly popular in all areas of clinical medicine including perinatology. However, designing and conducting an randomized controlled trial, analyzing data, interpreting findings and disseminating results can be challenging as there are several practicalities to be considered. In this review, we provide simple descriptive guidance on planning, conducting, analyzing and reporting randomized controlled trials.

Introduction

Evidence from randomized controlled trials (RCTs) is considered to be at the top of the evidence pyramid. It is recommended that clinical practice decisions are based on evidence emanating from well-conducted RCTs when available (1). Following the pioneering works by James Lind for scurvy (2) and Sir Bradford-Hill for the treatment of tuberculosis (3,4), gathering of evidence for clinical practice based on RCTs has become increasingly common. This document will review the need and scope for RCT evidence in perinatology. It will provide a

Key Message

Appropriately planned and rigorously conducted randomized controlled trials (RCTs) remain the most robust research method available to find the real effect of an intervention, but a biased RCT can lead to the adoption of a wasteful intervention and may even harm patients. This manuscript provides a step-by-step guide to planning, conducting, analyzing and reporting RCTs.

practical guide to researchers wanting to pursue this exciting field of research.

Why RCTs?

A fundamental question that any researcher may ask is, why is it that evidence based on RCTs is considered to be of the highest quality. The main reason for this is that evidence based on observational data is prone to bias. Bias is defined as the systematic tendency of any factors associated with the design, conduct, analysis, evaluation and interpretation of the results of a study to make the estimate of the effect of a treatment or intervention deviate from its true value. If two or more groups are being compared in an observational study, there are often systematic differences between the groups, so much so that the outcome of the groups may be different because of these differences rather than actual exposure or intervention. This is known as confounding. The only way to eliminate these differences is to allocate each individual to one or the other intervention at random. Therefore, the probability of any individual receiving one intervention or the other is decided solely by chance. In this scenario, if sample size is sufficient, all the factors influential in the outcome are likely to be distributed equally between the groups, because the allocation was at random. Therefore, any difference in the observed outcome between the groups is likely to be due to the intervention rather than any other factors. Although randomization is the best way to minimize the risk of confounding by unmeasured factors, RCTs could still be confounded. A recent article by Howards provides several suggestions on how to address confounding (5).

What questions are suitable for an RCT?

An RCT is a study design that is generally used in experiments testing the effectiveness and/or safety of one or more interventions. The intervention being tested is allocated to two or more study groups that are followed prospectively, outcomes of interest are recorded, and comparisons are made between intervention and control groups. The control group may receive no intervention, a standard treatment, or a placebo. The intervention can be therapeutic or preventive and does not necessarily have to be a pharmaceutical agent or a surgical intervention.

Randomized controlled trials are suitable both for preclinical and clinical research. There should be sufficient uncertainty about the utility of an intervention. This is referred to as equipoise. For clinical trials, the proposed intervention is sometimes based on logic, but mostly on data obtained from *in vitro* laboratory studies, animal experiments or preliminary serendipitous/planned observation in an uncontrolled setting. Observational (case–control or cohort) studies may suggest the benefit of an intervention, but they are prone to bias. Important and relevant gaps in the scientific knowledge sometimes come to light in the process of developing guidelines, and such gaps need to be addressed by producing robust evidence.

The question to be answered by RCT design should also be safe from participants' point of view. Investigators cannot subject participants to any undue risk/harm. For clinical trials, it can be argued that it is unethical to expose participants to any risk or burden if the research lacks social value (6). There are conditions/situations or statuses for which there is unlikely to ever be an RCT; for example, the effectiveness of a parachute in preventing death while sky-diving. One needs to be practical in understanding such dilemmas and not be blind-sided by proponents of RCTs (7). Finally, the research question also needs to be ethically appropriate to be answered in an RCT setting. Experiments conducted on human subjects before or during World War II provide striking reminders of the importance of obtaining voluntary informed consent from participants and ensuring their well-being in the design of RCTs (8). What is considered ethical may change with time and accumulation of knowledge. As an example, clinical trials of intravenous administration of alcohol to pregnant women as a tocolytic agent to halt preterm labor were performed from the 1960s to the late 1980s (9,10). However, administering alcohol to pregnant women is no longer ethical in light of its harmful effects on the mother and developing fetus.

Randomized controlled trials are equally useful to study effectiveness and/or safety of diagnostic and screening tests. However, RCTs are not suitable for investigating etiology or natural history of disease. Outcomes that are extremely rare or take a very long time to develop are also impractical to study using an RCT.

How should an RCT be designed?

Detailed guidance on planning of clinical trials is available (11), and should be followed. Here, we will only consider the basic principles. The first step is to assess if an RCT is the best research design for the research question. Then, it is necessary to confirm that the question has not already been answered by an appropriately powered (please see later) RCT. This requires a systematic literature search.

Research question formulation

A hypothesis has to be formed for a research question to be answered using an RCT. The hypothesis has to be precise. The key components of a sound research question

should include: P (population of interest), I (Intervention to be studied), C (comparator intervention), O (outcomes to be evaluated) and T (is there a time duration for intervention/outcome ascertainment time). Adequate time needs to be devoted to converting a "free form" question arising from a clinical or nonclinical context to convert it into a properly answerable "PICOT" format question. This is best demonstrated using an example of tranexamic acid that is shown to reduce blood loss during cesarean section (12,13). To address the free flowing research question "Does administration of tranexamic acid prevent postpartum hemorrhage?", investigators need to convert this into an answerable and specific question. The population needs to be defined clearly, such as whether one would investigate women delivering vaginally, by cesarean section (which could also be emergency or elective), or both. Next, the intervention requires specific stipulation as to how much, how frequently, what dose, what time and what route would be used for administering tranexamic acid. For a comparator, it needs to be clearly outlined whether this intervention is compared with placebo or no treatment or any other currently used measures to prevent postpartum hemorrhage. The outcome also needs to be clearly defined with regard to what definition of postpartum hemorrhage will be used for the trial (for example, blood loss >500 mL or >1000 mL). As the outcome of interest here occurs within a fairly short time frame of completion of intervention, the time component may not be very relevant to this question. Hence, a refined form of the question may look like this: "Does 1.0 g of intravenous tranexamic acid given 4 h prior to elective cesarean section compared with placebo reduce postpartum hemorrhage diagnosed as estimated blood loss of >500 mL within 24 h of birth?"

As evident from the above exercise, an ample amount of time must be devoted to carefully honing the research question. This should result in a very well-defined, clear, feasible, specific, measurable, ethical and clinically important question before one initiates the experiment.

How should an RCT be conducted?

Once a research question is generated, the next step in the conduct of an RCT will be to clearly define the target population, inclusion and exclusion criteria, process of randomization, allocation, blinding of intervention, treatment and control delivery, outcomes assessment, definitions of outcomes, sample size required, ethical requirements, consent process and, finally, data management. These topics should be written as a well-defined protocol. The protocol should be reviewed and approved by an independent ethics committee before starting the trial. For clinical trials, it is now obligatory that protocols are registered with a publicly available trial registry before recruiting any participants to the trial. Apart from promoting transparency, most journals have this as a mandatory requirement and would not publish the results of an RCT unless trial registration details were provided. Retrospective registration may not be acceptable. Some medical journals also accept RCT protocols for publication. This is to make sure that protocols are not altered as a result of trial results. It also helps to ensure that negative trials are not left unpublished. Nevertheless, only two-thirds of RCTs in women's health published in 2015 were prospectively registered, and more than half did not achieve the planned sample size (14). However, this is expected to change in future. Moreover, the International Committee of Medical Journal Editors (ICMJE) has recently suggested that clinical trials that will begin enrolling participants on or after 1 January 2019 must include a data-sharing plan when registering the trial (15).

An independent data monitoring committee is usually established to periodically oversee the progress of a clinical trial, safety data and critical efficacy variables, and to recommend to the sponsor (funder) whether to continue, modify or terminate a trial.

The key components of design of an RCT are highlighted below.

Random allocation

Each of the eligible participants should have an equal chance to be allocated the intervention or not. The simplest way of achieving this is by parallel group design, in which each group of participants is exposed to only one of the study interventions. In a crossover design, all the trial participants receive both interventions in a sequential manner and only the order of intervention is randomly assigned. In this way, each participant serves as his/her own control, thereby eliminating individual participant differences. However, this design is more vulnerable to drop out and attrition. If a particular baseline characteristic is of such fundamental importance as to have a big influence on the outcome, it can be taken into account at randomization. Participants with or without that baseline characteristic are randomized separately (stratified randomization). Block randomization is used to maintain a balance between the intervention group and control, so that the numbers are not too dissimilar, which could rarely happen by chance. Cluster randomization can be used when randomization of individual participants is not feasible/practical, in which case hospitals, clinics, geographic areas etc. can be used as units for the allocation of intervention or control groups. Generation of random sequence should be done by some independent personnel, usually a statistician, who is not going to be involved in

the conduct of the RCT. The access to this sequence should be restricted to only a few individuals who absolutely need to have access (such as the pharmacist who will be preparing the medication) and not the investigators or personnel involved in ascertaining outcome. The sequence should be opened by this individual only on a case-by-case basis and specified sequence should be followed.

Allocation concealment

One of the key components of an RCT is allocation concealment. This means that neither front-line care providers, investigators or participants are aware of whether the next eligible participant will be receiving treatment or control intervention. This should be masked until the time when participants are ready to receive intervention. By this virtue, unnecessary adjustments as to whether to enroll a participant or not (such as after knowing that the prognosis is not good and the patient is randomized to an experimental treatment, the investigator changes her/his mind and decides not to include the participant in the study) can be avoided. This is very important in situations when blinding of intervention is not possible.

Blinding

The focus of conducting an RCT is elimination of bias. Unconscious information bias may be introduced if the investigators or participants are aware of who is getting the intervention and who is not. The procedure of blinding the participants (single blind) or both investigators and participants (double blind) helps to eliminate this unconscious information bias. Whenever possible, blinding should be used in an RCT. It is not always possible to blind either the participants or investigators due to the nature of the RCT. For example, Jozwiak et al. (16) conducted an RCT comparing vaginal prostaglandins with trans-cervical balloon catheter for induction of labor (PROBAAT study). It was possible to randomize participants to the two types of interventions, but it was not possible to blind the participant or the investigator. Therefore, this trial was conducted as an open-label RCT.

Conduct

An RCT can be conducted at a single site or at multiple sites. RCTs conducted according to a single protocol but at more than one site are referred to as a multi-center trials. Including several sites has the advantage of reaching the required sample size within a shorter time and may also improve generalizability of findings.

The main premise of conducting an RCT is that the participants should be treated exactly the same way in both arms except for the intervention/control treatment. All other procedures of treatment, diagnosis, investigations, alterations etc. should follow the routine process and no undue advantage or testing should be performed on patients in the trial. These data should be collected to identify issues of contaminations, crossover of intervention and co-interventions.

Outcome ascertainment

The prespecified primary and secondary outcomes should be collected by independent observers who are unaware of the allocation and treatment arms of participants. As far as possible, it is advisable that objective measures are used for ascertaining outcome so that personal bias on the part of the collector does not come into play. This is particularly important when the intervention cannot be masked (such as surgical scars). It is also important that the outcome is collected in all randomized patients. The number of patients with missing outcome data should be minimized as far as possible. A high rate of attrition will lead to reduced confidence in the results and may lead to biased estimates.

Sample size

One would always like to conduct a study that has adequate sample size and power so that the conclusions generated from the experiment can be applied to the broader population with ample confidence. The required sample size to test a hypothesis is governed by the effect size (17). In a superiority trial, one would like to detect difference in the effect of intervention vs. placebo, which is a minimum clinically important difference, that is required to detect between two groups and convince users of the information to utilize the intervention. This number is usually derived from previous experiments/observations, previous trials or by consensus opinion. In general, the more widely the two groups are separated from each other and the smaller the variability in each group, the fewer participants are necessary in each group to show that the difference is unlikely to be due to chance and more likely to be due to the intervention. The inclusion of too few patients in a study increases the risk that a significant treatment benefit will not be shown, even if such an effect exists (a type 2 error). The details of actual calculations are beyond the scope of this manuscript but published literature on this topic is abundant and standard formulae are available to calculate sample size (18,19). Expert statistical help should be sought when needed. However, some key information should be

gathered before seeking statistical help. First, one needs to know the baseline estimate of outcome rate in the placebo/control arm, i.e. how many patients are expected to benefit from the control intervention. Second, it is important to have an expected estimate as to what percentage of patients are expected to benefit from the intervention. This is where one needs to be careful to not overestimate the benefit (as this will need lower sample size) or to underestimate the benefit (as one may end up experimenting on more patients than necessary). Two more aspects that must be kept in mind are: how many type 1 and type 2 errors are we willing to accept before rejecting the null hypothesis (as described below).

Power of a study

If there are significant differences in the primary outcome between the two groups, one can conclude that the difference is likely to be due to the intervention. Typically, the difference is thought to be 'significant' if the probability of this difference arising solely due to chance is <0.05. This is the well-known probability (p)-value. Therefore, the chance that a difference will be found even if there is no real difference is 1:20 (0.05). This is known as a type 1 error, and is usually fixed at 0.05 due to convention. However, it is not necessary to fix this at 0.05, and other levels of significance (0.01 or 0.005) may be chosen. It is possible that a significant difference may not be observed even if this is present. The chance that the study will be able to demonstrate a significant difference if it is present, is known as the power of the study. By convention it is fixed at 0.8 to 0.95 level (80–95%). Inability to demonstrate a significant difference even when one does exist is known as a type 2 error. The probability of type 2 error is conventionally set at 0.2 to 0.05. With these four values, it will be easy for a statistician to calculate sample size in an experiment where effect size is measured in proportion. If the effect size is measured as a continuous variable (such as difference in blood pressure) then one would need the mean and standard deviation of the variable in each group in addition to type 1 and 2 error values to calculate the sample size. This provides adequate information for most routine types of RCTs. For different and complicated statistical parameters to be evaluated in an RCT, it is advisable to consult an experienced statistician before designing the experiment, as prohibitive sample size may jeopardize research efforts. RCTs can be designed and conducted to evaluate the superiority, equivalence or noninferiority of an intervention compared with another, and power calculations are different for these different types of RCTs. Power calculations may be based on simulations performed by a statistician, the details of which are beyond the scope of this review.

Trial phases

Clinical trials evaluating safety and efficacy of pharmaceutical agents generally have to go through a series of studies before they can be used in clinical practice. There are four sequential phases of clinical trial that have the objective of: studying human pharmacology of the agent (phase I); exploring therapeutic potential (phase II); confirming therapeutic effect (phase III); and evaluating it for therapeutic use (phase IV). Phase I trials are conducted in a small number of healthy participants (20–80) to determine the absorption, distribution, metabolism and toxicity of a new drug in humans for the first time. Phase II trials are designed to estimate dose and test the safety and therapeutic efficacy in a slightly larger population (100–300) afflicted with the condition for which the drug was developed. Phase III is a definitive study of efficacy of the drug after sample size estimation for proper evaluation. Data on side effects are collected meticulously. Phase IV trials are post marketing studies after a drug has been approved by a regulatory body such as the US Food and Drug Administration in the USA or the European Medicines Agency in Europe. Such trials provide additional information including the benefits, optimal dose, effectiveness and adverse events of the drug in different patient populations.

Improving generalizability

Generalizability should be considered as a very important criterion in designing an RCT. The answer provided by an RCT is applicable to the patient population similar to that used in the trial. Extrapolating it to other patients is not strictly valid. Therefore, the inclusion criteria should be as wide as is practically possible, while at the same time maintaining scientific rigor. None of the aforementioned steps should be so strict that once the RCT is over, replication of the intervention is impossible in a practical setting. The process of administering intervention to a group of subjects should be relatively easy, and collection of outcome data should also not be too onerous. For example, the investigators of WOMAN trial (13) used a clinical definition of postpartum hemorrhage, and supplemented it with a quantitative estimation of blood loss: >500 mL for vaginal births, and >1000 mL for operative deliveries. This implied that the trial result was valid both for spontaneous and operative delivery.

Interim analysis

Randomized controlled trials are designed with an anticipated incidence of the primary outcome in the control arm. The observed incidence may be lower, making the

trial underpowered, or higher, making the trial unnecessarily prolonged. Interim analysis is a useful way to make sure that the observed incidence is not too different from the expected incidence. However, interim analyses should be preplanned and stated in the protocol. Analysis should be performed by an independent statistician blinded to the identity of either group. Interim analysis may sometimes show that differences in the two groups are large and show a clear advantage of the intervention. In this case, continuing the trial is unethical because the control group will be denied the clearly superior alternative. On the other hand, early discontinuation may be advisable if the incidence of primary outcome in the control is far too low and the revised sample size is deemed to be unfeasible. Upward revision of the required sample size may result from the interim analysis as was the case in the WOMAN trial (13). When event rates are lower than anticipated or variability is larger than expected, methods for sample size re-estimation are available without unblinding. The observed incidence in the placebo arm of the study is often lower than anticipated. In the WOMAN trial, the anticipated rate of death was 3.0% in the placebo arm, but the observed incidence was 1.9%. An independent data monitoring committee usually oversees the interim analysis.

Ethical considerations

Rigorous ethical principles must be applied to all RCTs involving experiments on animals, humans or human biological material (20). Evaluation of risk and benefit to the participants and society, obtaining ethical approval, and informed consent are crucial. Before planning and conducting an RCT, it must be considered and evaluated whether it is ethical to use randomization to allocate participants to an intervention group. Where there is previous evidence showing superiority of an intervention over that of doing nothing, an RCT using a placebo (or doing nothing) is unethical. For example, before the development of anti-retroviral therapy, untreated human immunodeficiency virus infection was associated with near certain death. Today it is a treatable condition although newer, more effective treatments continue to be invented. Therefore, RCTs in this field comparing one drug or treatment regimen against another would be ethical, but the use of placebo would not.

Safety concerns

As is true of medicine, participants must not be harmed by the experiment. Therefore, predefined, serious, adverse events need to be reported to the sponsor and to the trial monitoring committee. Any complications or side effects of drugs should be reported to regulatory bodies. Unacceptably high frequency of adverse events in the intervention group may lead to early discontinuation of the trial, usually recommended by the data monitoring committee.

How should an RCT be reported?

There are guidelines for reporting on RCTs (http://www.consort-statement.org), which should be followed. Many RCTs report baseline characteristics of the two (intervention and control) groups. If the allocation was random, any differences between the baseline characteristics of the two groups must be by chance. Therefore, a comparison with statistical testing and reporting *p*-values is superfluous.

The type of comparison made between intervention and control groups is important, especially when efficacy of an intervention is being evaluated. Therefore, whether it is a superiority, noninferiority or equivalence trial should be reported.

It is not uncommon that some participants do not receive the intervention allocated by the randomization process. The gold standard of reporting is 'intention–to-treat' analysis. Outcomes of all participants randomized to the intervention arm should be reported in that group even if some of the participants may not have received the intervention. The analysis by intention-to-treat, and according to the intervention that they actually received (per-protocol analysis) can rarely lead to differing results. Reporting per-protocol analysis rather than intention-to-treat analysis often results in overestimation of the effect of intervention. Grouping participants according to the actual treatment received can introduce a bias, and is discouraged. The best strategy to guard against such a possibility is to maintain protocol violations to a minimum. While reporting the primary outcome, it is becoming increasingly customary to report the effect size (and its 95% confidence interval) rather than just the *p*-value, as this provides meaningful information about the magnitude of change. It is best to always use two-sided *p* values (21). It is also a good practice to give the actual *p*-values rather than $p < 0.05$ or "significant," and $p > 0.05$ or "not significant" (21).

How should RCTs be interpreted?

An RCT is an experiment. If the difference in the primary outcome is significant at the customary level of $p < 0.05$, chances are that the observed difference is real. The magnitude of the observed difference is also important. The magnitude may be small, but still statistically significant. In this case, the clinical significance is most often limited. The magnitude may be large, but still statistically

nonsignificant. In this case, the study remains underpowered. Appropriate sample size calculations before embarking on the study should prevent this situation. In many cases, the observed difference is small, and is not statistically significant. This constitutes a negative trial. It does not mean that the trial has failed. When a trial is too small to detect modest treatment effects, it is appropriate to describe the findings as inconclusive rather than negative. It is a matter of great joy for the investigators when the anticipated difference is found between the two groups and is statistically significant. As great care has been observed in the conduct of the trial, the results are likely to be reproducible. However, the scientific community was recently shaken by reports that a troubling proportion of peer-reviewed preclinical studies are not reproducible. New initiatives have been proposed to increase confidence in the published studies (22). Generalizability should also be taken into account when interpreting results. When applying the evidence gathered from an RCT to a clinical situation, the question to ask is "is my patient so different compared with the participants of the RCT so as to make the results of the RCT inapplicable?" The evidence is valid as long as the answer to this question is "no." Interpretation of any trial should depend not only on the primary outcome, but on the totality of the evidence (i.e. the primary, secondary and safety outcomes).

Conclusion

The RCT is the most rigorous and robust research method for determining whether a cause–effect relation exists between an intervention and an outcome. Therefore, it is important to perform RCTs to generate evidence in basic, translational and clinical research and improve the management of our patients. However, an RCT should be conducted only if it is ethically feasible, economically viable and clinically worthwhile.

Funding

No specific funding.

References

1. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21:125–7.
2. Milne I. Who was James Lind, and what exactly did he achieve. J R Soc Med. 2012;105:503–8.
3. Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. Br Med J. 1948;2:769–82.
4. Hill AB. The clinical trial. N Engl J Med. 1952;247:113–9.
5. Howards PP. An overview of confounding part 1: the concept and how to address it. Acta Obstet Gynecol Scand. 2018;97:394–399.
6. Wendler D, Rid A. In defense of a social value requirement for clinical research. Bioethics. 2017;31:77–86.
7. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ. 2003;327:1459–61.
8. Annas GJ. Beyond Nazi war crimes experiments: the voluntary consent requirement of the Nuremberg Code at 70. Am J Public Health. 2018;108:42–6.
9. Haas DM, Morgan AM, Deans SJ, Schubert FP. Ethanol for preventing preterm birth in threatened preterm labor. Cochrane Database Syst Rev. 2015;11:CD011445.
10. Schröck A, Fidi C, Löw M, Baumgarten K. Low-dose ethanol for inhibition of preterm uterine activity. Am J Perinatol. 1989;6:191–5.
11. ICH. Guideline for good clinical practice E6(R2). EMA/CHMP/ICH/135/1995. London: European Medicines Agency, 2017. Available online at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf (accessed December 11, 2018).
12. Simonazzi G, Bisulli M, Saccone G, Moro E, Marshall A, Berghella V. Tranexamic acid for preventing postpartum blood loss after cesarean delivery: a systematic review and meta-analysis of randomized controlled trials. Acta Obstet Gynecol Scand. 2016;95:28–37.
13. WOMAN Trail Collaborators. Effect of early tranexamic acid administration on mortality, hysterectomy, and other morbidities in women with post-partum haemorrhage (WOMAN): an international, randomised, double-blind, placebo-controlled trial. Lancet. 2017;389:2105–16.
14. Nijjar SK, Khan KS. Threats to reliability risk erroneous conclusions: a survey of prospective registration and sample sizes of randomised trials in women's health. BJOG. 2017;124:1057–61.
15. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. JAMA. 2017;317:2491–2.
16. Jozwiak M, Oude Rengerink K, Benthem M, van Beek E, Dijksterhuis MG, de Graaf IM, et al. Foley catheter versus vaginal prostaglandin E2 gel for induction of labour at term (PROBAAT trial): an open-label, randomised controlled trial. PROBAAT Study Group. Lancet. 2011;378:2095–103.
17. Ialongo C. Understanding the effect size and its measures. Biochem Med (Zagreb). 2016;26:150–63.
18. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen. 2012;141:2–18.

19. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA. 1994;22:122–4.
20. WMA Declaration of Helsinki. Ethical principles for medical research involving human subjects. Available online at: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (accessed January 11, 2018).
21. Pocock SJ, McMurray JJ, Collier TJ. Making Sense of statistics in clinical trial reports: part 1 of a 4-part series on statistics for clinical trials. J Am Coll Cardiol. 2015;66:2536–49.
22. McNutt M. Reproducibility. Science. 2014;343:229.

KJA
Korean Journal of Anesthesiology

Statistical Round
pISSN 2005-6419 • eISSN 2005-7563

Randomization in clinical studies

Chi-Yeon Lim[1] and Junyong In[2]

[1]*Department of Biostatistics, Dongguk University College of Medicine,* [2]*Department of Anesthesiology and Pain Medicine, Dongguk University Ilsan Hospital, Goyang, Korea*

Randomized controlled trial is widely accepted as the best design for evaluating the efficacy of a new treatment because of the advantages of randomization (random allocation). Randomization eliminates accidental bias, including selection bias, and provides a base for allowing the use of probability theory. Despite its importance, randomization has not been properly understood. This article introduces the different randomization methods with examples: simple randomization; block randomization; adaptive randomization, including minimization; and response-adaptive randomization. Ethics related to randomization are also discussed. The study is helpful in understanding the basic concepts of randomization and how to use R software.

Keywords: Adaptive randomization; Minimization; Random allocation; Randomization; Randomized controlled trial; Restrictive randomization; Simple randomization; Stratified randomization.

Introduction

Statistical inference in clinical trials is a mandatory process to verify the efficacy and safety of drugs, medical devices, and procedures. It allows for generalizing the results observed through sample, so the sample by random sampling is very important. A randomized controlled trial (RCT) comparing the effects among study groups carry out to avoid any bias at the stage of the planning a study protocol. Randomization (or random allocation of subjects) can mitigate these biases with its randomness, which implies no rule or predictability for allocating subjects to treatment and control groups.

Another property of randomization is that it promotes comparability of the study groups and serves as a basis for statistical inference for quantitative evaluation of the treatment effect. Randomization can be used to create similarity of groups. In other words, all factors, whether known or unknown, that may affect the outcome can be similarly distributed among groups. This similarity is very important and allows for statistical inferences on the treatment effects. Also, it ensures that other factors except treatment do not affect the outcome. If the outcomes of the treatment group and control group show differences, this will be the only difference between the groups, leading to the conclusion that the difference is treatment induced [1].

CONSORT[1)], a set of guidelines proposed to improve completeness of the clinical study report, also includes randomization. Randomization plays a crucial role in increasing the quality of evidence-based studies by minimizing the selection bias that could affect the outcomes. In general, randomization places programming for random number generation, random allocation concealment for security, and a separate random code manager. After then, the generated randomization is implemented to the study [2]. Randomization is based on probability theory and hence difficult to understand. Moreover, its reproducibility problem requires the use of computer programming language. This study tries to alleviate these difficulties by enabling even a

Corresponding author: Junyong In, M.D., Ph.D.
Department of Anesthesiology and Pain Medicine, Dongguk University Ilsan Hospital, 27 Dongguk-ro, Ilsandong-gu, Goyang 10326, Korea
Tel: 82-31-961-7875, Fax: 82-31-961-7864
Email: dragona1@dumc.or.kr
ORCID: https://orcid.org/0000-0001-7403-4287

Received: February 8, 2019.
Revised: March 12, 2019.
Accepted: March 18, 2019.

Korean J Anesthesiol 2019 June 72(3): 221-232
https://doi.org/10.4097/kja.19049

[1)]http://www.consort-statement.org/

non-statistician to understand randomization for a comparative RCT design.

Methods of Randomization

The method of randomization applied must be determined at the planning stage of a study. "Randomness" cannot be predicted because it involves no rule, constraint, or characteristic. Randomization can minimize the predictability of which treatment will be performed. The method described here is called simple randomization (or complete randomization). However, the absence of rules, constraints, or characteristics does not completely eliminate imbalances by chance. For example, assume that in a multicenter study, all subjects are randomly allocated to treatment or control groups. If subjects from center A are mainly allocated to the control group and lots of subjects from center B are allocated to the treatment group, even though this is allocated with simple randomization, can we ignore the imbalance of the randomization rate in each center?

For another example, if the majority of subjects in the control group were recruited early in the study and/or the majority of those in the treatment group were recruited later in the study, can the chronological bias be ignored? The imbalance in simple randomization is often resolved through restrictive randomization, which is a slightly restricted method [3,4]. Furthermore, adaptive randomization can change the allocation of subjects to reflect the prognostic factors or the response to therapy during the study. The use of adaptive randomization has been increasing in recent times, but simple or restrictive randomization continues to be widely used [4]. In the Appendix, the R commands are prepared for the various randomization methods described below.

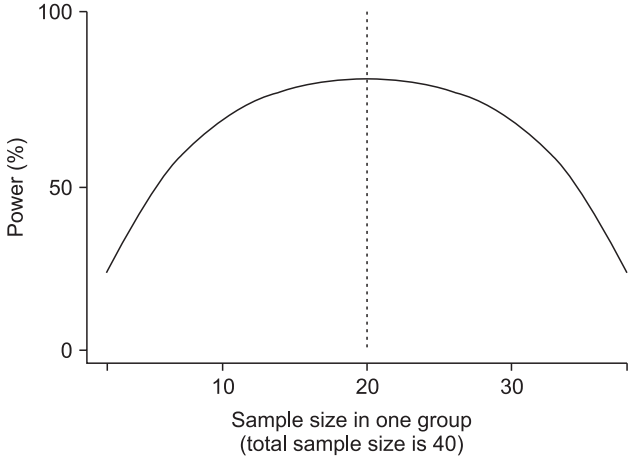


Fig. 1. Influence of sample size ratio in two groups on power (difference (d) = 0.9, two-tailed, significant level = 0.05). The dashed line indicates the same sample size in two groups (n = 20) and maximized power.

Simple randomization

In simple randomization, a coin or a die roll, for example, may be used to allocate subjects to a group. The best part of simple randomization is that it minimizes any bias by eliminating predictability. Furthermore, each subject can maintain complete randomness and independence with regard to the treatment administered [5]. This method is easy to understand and apply,[2] but it cannot prevent the imbalances in the sample size or prognostic factors that are likely to occur as the number of subjects participating in the study decreases. If the ratio of number of subjects shows an imbalance, that is, it is not 1 : 1, even with the same number of subjects participating, the power of the study will fall. In a study involving a total of 40 subjects in two groups, if 20 subjects are allocated to each group, the power is 80%; this will be 77% for a 25/15 subject allocation and 67% for a 30/10 subject allocation (Fig. 1).[3] In addition, it would be difficult to consider a 25/15 or 30/10 subject allocation as aesthetically balanced.[4] In other words, the balancing of subjects seems plausible to both researchers and readers. Unfortunately, the nature of simple randomization rarely lets the number of subjects in both groups to be equal [6]. Therefore, if it is not out of the range of the assignment ratio (e.g., 45%–55%),[5] it is balanced. As the total number of subjects increases, the probability of departing from the assignment ratio, that is, the probability of imbalance, decreases. In the following, the total number of subjects and the probability of imbalance were examined in the two-group study with an assignment ratio of 45%–55% (Fig. 2). If the total number of subjects is 40, the probability of the imbalance is 52.7% (Fig. 2, point A), but this decreases to 15.7% for 200 subjects (Fig. 2, point B) and 4.6% for 400 subjects (Fig. 2, point C). This is the randomization method recommended for large-scale clinical trials, because the likelihood of imbalance in trials with a small number of subjects is high [6–8].[6] However, as the number of subjects does not always increase, other solutions need to be considered. A block randomization is helpful to resolve the imbalance in number of subjects, while a stratified randomization and an adaptive randomization can help resolve the imbalance in prognostic factors.

---

[2] However, since the results and process of randomization cannot be easily recorded, the audit of randomization is difficult.
[3] Two-tailed test with difference (d) = 0.91 and type 1 error of 0.05.
[4] "Cosmetic credibility" is often used.
[5] The difference in number of subjects does not exceed 10% of the total number of subjects. This range is determined by a researcher, who is also able to choose 20% instead of 10%.
[6] These references recommend 200 or more subjects, but it is not possible to determine the exact number.

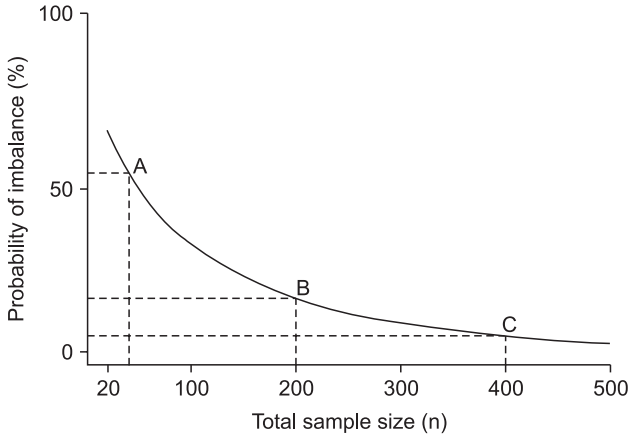


Fig. 2. Probability curves of imbalance between two groups for complete randomization as a function of total sample size (n). When n = 40, there is a 52.7% chance of imbalance beyond 10% (allocation ratio 45%–55%) (point A). When n = 200, there is a 15.7% chance of imbalance (point B), but n = 400 results in only 4.6% chance of imbalance (point C).

Block randomization

If we consider only the balance in number of subjects in a study involving two treatment groups A and B, then A and B can be repeatedly allocated in a randomized block design with predefined block size. Here, a selection bias is inevitable because a researcher or subject can easily predict the allocation of the group. For a small number of subjects, their number in the treatment groups will not remain the same as the study progresses, and the statistical analysis may show the problem of poor power. To avoid this, we set blocks for randomization and balance the number of subjects in each block.[7] When using blocks, we need to apply multiple blocks and randomize within each block. At the end of block randomization, the number of subjects can easily be balanced, and the maximum imbalance in the study can be limited to an appropriate level. That is, block randomization has the advantage of increasing the comparability between groups by keeping the ratio of the number of subjects between groups almost the same. However, if the block size is 2, the allocation result of the second subject in the block can be easily predicted with a high risk of observation bias.[8] Therefore, the block size used should preferably be 4 or more. However, note that even when the block size is large, if the block size is known to the researcher, the risk of selection bias will increase because the treatment of the last subject in the block will be revealed. To reduce the risk of predictability from the use of one block size, the size may be varied.[9]

Restricted randomization for unbalanced allocation

Sometimes unbalanced allocation becomes necessary for ethical or cost reasons [9]. Furthermore, if you expect a high dropout rate in a particular group, you have to allocate more subjects. For example, for patients with terminal cancer who are not treated with conventional anticancer agents, it would be both ethical and helpful to recruit those who would be more likely to receive a newly developed anticancer drug [10] (of course, contrary to expectations, the drug could be harmful).

As for simple randomization, the probability is first determined according to the ratio between the groups, and then the subjects are allocated. If the ratio between group A and group B is 2 : 1, the probability of group A is 2/3 and that of group B is 1/3. Block randomization often uses a jar model with a random allocation rule. To consider the method, first drop as many balls as the number of subjects into the jar according to the group allocation ratio (of course, the balls have different colors depending on the group). Whenever you allocate a subject, take out one ball randomly and confirm it, and do not place the ball back into the jar (random sampling without replacement). Repeat this allocation for each block.

Stratified randomization

Some studies have prognostic factors or covariates affecting the study outcome as well as treatment. Researchers hope to balance the prognostic factors between the study groups, but randomization does not eliminate all the imbalances in prognostic factors. Stratified randomization refers to the situation where the strata are based on level of prognostic factors or covariates. For example, if "sex" is the chosen prognostic factor, the number of strata is two (male and female), and randomization is applied to each stratum. When a male subject participates, the subject is first allocated to the male strata, and the group (treatment group, control group, etc.) is determined through randomization applied to the male strata. In a multicenter study, one typical prognostic factor is the "site." This may be due to the differences in characteristics between the subjects and the manner and procedure in which the patients are treated in each hospital.

Stratification can reduce imbalances and increase statistical power, but it has certain problems. If several important prognostic factors affect the outcome, the number of strata would

[7] Random allocation rule, truncated binomial randomization, Hadamard randomization, and the maximal procedure are forced balance randomization methods within blocks, and one of them is applied to the block. The details are beyond the scope of this study, and are therefore not covered.

[8] The block size of 2 applies mainly to a study of allocating a pair at the same time.

[9] Strictly speaking, the block size is randomly selected from a discrete uniform distribution, and so the use of a random block design rather than a "varying" block size would be a more formal procedure.

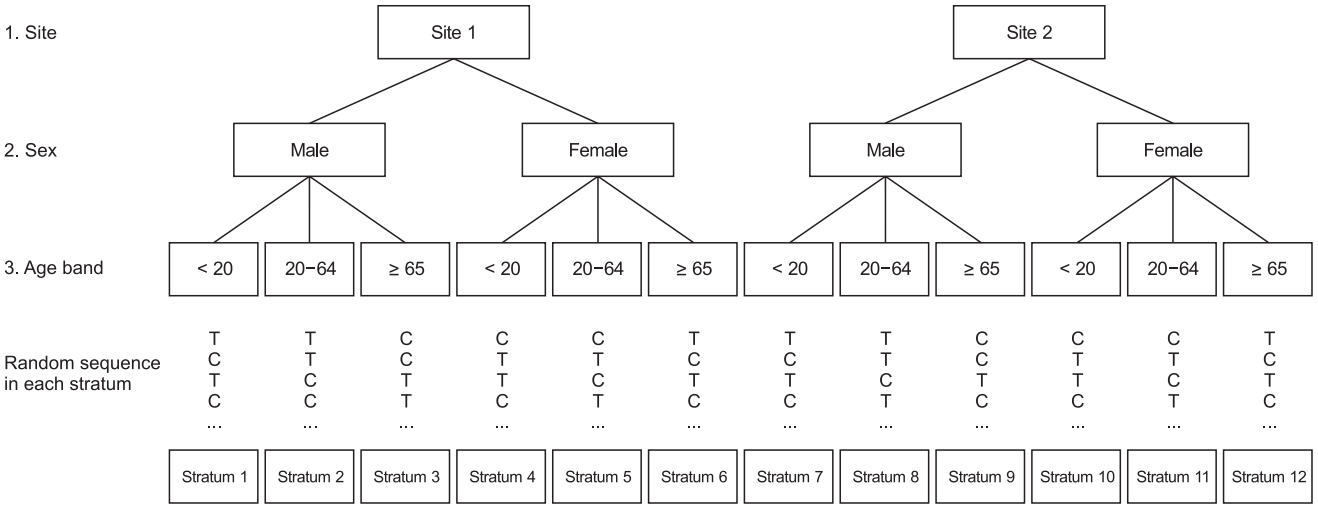


Fig. 3. Example of stratification with three prognostic factors (site, sex, and age band). Eventually, randomization with 12 strata should be accomplished using 12 separate randomization processes. C: control group, T: treatment group.

increase [11]. For example, 12 (2 × 2 × 3) strata are formed solely from recruitment hospitals (sites 1 and 2), sex (male and female), and age group (under 20 years, 20–64 years, and 65 years and older) (Fig. 3). In case of several strata in relation to the target sample size, the number of subjects allocated to a few strata may be empty or sparse. This causes an imbalance[10] in the number of subjects allocated to the treatment group. To reduce this risk, the prognostic factors should be carefully selected. These prognostic factors should be considered again during the statistical analysis and at the end of the study.

Adaptive randomization

Adaptive randomization is a method of changing the allocation probability according to the progress and position of the study. It may be used to minimize the imbalance between treatment groups as well as to change the allocation probability based on the therapeutic effect. Covariate-adaptive randomization adjusts the allocation of each subject to reduce the imbalance, taking into account the imbalance of the prognostic factors. One example is the "minimization technique of randomization (minimization)" to develop indicators that collectively determine the distributional imbalance of various prognostic factors and allocates them to minimize the imbalance.

Minimization[11]

Minimization was first introduced as a covariate adaptive method to balance the prognostic factors [12,13]. The first subject is allocated through simple randomization, and the subsequent ones are allocated to balance the prognostic factors. In other words, the information of the subjects who have already participated in the study is used to allocate the newly recruited subjects and minimize the imbalance of the prognostic factors [14].

Several methods have emerged following Taves [13]. Pocock and Simon define a more general method [12].[12] First, the total number of imbalances is calculated after virtually allocating a newly recruited subject to all groups, respectively. Then, each group has its own the total number of imbalances. Here, this subject will be allocated to the group with lowest total number of imbalances.

We next proceed with a virtual allocation to the recruitment hospitals (Sites 1 and 2), sex (male and female), and age band (under 20 years, 20–64 years, and 65 years or older) as prognostic factors. This study has two groups: a treatment group and a control group.

Assume that the first subject (male, 52-years-old) was recruited from Site 2. Because this subject is the first one, the allocation is determined by simple randomization.

Further, assume that the subject is allocated to a treatment group. In this group, scores are added to Site 2 of the recruiting hospital, sex Male, and the 20–64 age band (Table 1). Next, assume that the second subject (female, 25-years-old) was recruited through Site 2. Calculate the total number of imbalances when this subject is allocated to the treatment group and to the

[10] As the number of strata increases, the imbalance increases due to various factors. The details are beyond the scope of this study.
[11] This paragraph introduces how to allocate "two" groups.
[12] We can set the weights on the variables or the allowable range for the total number of imbalance, but in this study, we did not set any weights or allowable range for the total number of imbalances.

control group. Add the appropriate scores to the area within each group, and sum the differences between the areas.

First, the total number of imbalances when the subject is allocated to the control group is

$$[(1 - 1) + (1 - 0) + (1 - 1)] = 1.$$

The total number of imbalances when the subject is allocated to the treatment group is

$$[(2 - 0) + (1 - 0) + (2 - 0)] = 5.$$

Since the total number of imbalances when the subject is allocated to the control group has 1 point (< 5), the second subject is allocated to the control group, and the score is added to Site 2 of the recruiting hospital, Sex female, and the 20–64 age band in the control group (Table 2). Next, the third subject (Site 1, Sex male, 17-years-old) is recruited.

Now, the total number of imbalances when the subject is allocated to the control group is

$$[(1 - 0) + (1 - 1) + (1 - 0)] = 2.$$

The total number of imbalances when the subject is allocated to the treatment group is

$$[(1 - 0) + (2 - 0) + (1 - 0)] = 4.$$

The total number of imbalances when the subject is allocated to the control group is 2 point (< 4). Therefore, the third subject is allocated to the control group, and the score is added to Site 1 of the recruiting hospital, sex male, and the < 20 age band (Table 3). The subjects are allocated and scores added in this manner.

Table 1. How Adaptive Randomization Using Minimization Works

Prognostic factor	Control group	Treatment group
Site		
Site 1	0	0
Site 2	0	1
Sex		
Male	0	1
Female	0	0
Age band		
< 20	0	0
20–64	0	1
≥ 65	0	0

The score in each factor is 0. The first patient (sex male, 52 yr, from site 2) is allocated to the treatment group through simple randomization. Therefore, site 2, sex male, and the 20–64 years age band in the treatment group receive the score.

Now, assume that the study continues, and the 15th subject (female, 74-years-old) is recruited from Site 2.

Here, the total number of imbalances when the subject is allocated to the control group is

$$[(5 - 4) + (4 - 3) + (4 - 3)] = 3.$$

The total number of imbalances when the subject is allocated to the treatment group is

$$[(6 - 3) + (4 - 3) + (4 - 3)] = 5.$$

The total number of imbalances when the subject is allocated to the control group is lower than that when the allocation is to the treatment group (3 < 5). Therefore, the 15th subject is allocated to the control group, and the score is added to Site 2 of the recruiting hospital, female sex, and the ≥ 65 age band (Table 4). If the total number of imbalances during the minimization technique is the same, the allocation is determined by simple

Table 2. How Adaptive Randomization Using Minimization Works

	Prognostic factor	Control group	Treatment group
If allocated to control group	Site		
	Site 1	0	0
	Site 2	1	1
	Sex		
	Male	0	1
	Female	1	0
	Age band		
	< 20	0	0
	20–64	1	1
	≥ 65	0	0
	Total number of imbalances	[(1 − 1) + (1 − 0) + (1 − 1)] = 1	
If allocated to treatment group	Site		
	Site 1	0	0
	Site 2	0	2
	Sex		
	Male	0	1
	Female	0	1
	Age band		
	< 20	0	0
	20–64	0	2
	≥ 65	0	0
	Total number of imbalances	[(2 − 0) + (1 − 0) + (2 − 0)] = 5	

The second patient has factors sex female, 25 yr, and site 2. If this patient is allocated to the control group, the total imbalance is 1. If this patient is allocated to the treatment group, the total imbalance is 5. Therefore, this patient is allocated to the control group, and site 2, sex female, and the 20–64 years age band in the control group receive the score.

Table 3. How Adaptive Randomization Using Minimization Works

	Prognostic factor	Control group	Treatment group
If allocated to control group	Site		
	Site 1	1	0
	Site 2	1	1
	Sex		
	Male	1	1
	Female	1	0
	Age band		
	< 20	1	0
	20–64	1	1
	≥ 65	0	0
	Total number of imbalances	[(1 − 0) + (1 − 1) + (1 − 0)] = 2	
If allocated to treatment group	Site		
	Site 1	0	1
	Site 2	1	1
	Sex		
	Male	0	2
	Female	1	0
	Age band		
	< 20	0	1
	20–64	1	1
	≥ 65	0	0
	Total number of imbalances	[(1 − 0) + (2 − 0) + (1 − 0)] = 4	

The third patient has factors sex male, 17 yr, and site 1. If this patient is allocated to the control group, the total imbalance is 2. If this patient is allocated to the treatment group, the total imbalance is 4. Therefore, this patient is allocated to the control group, and then site 1, sex male, and the < 20 age band in the control group receive the score.

Table 4. How Adaptive Randomization Using Minimization Works

	Prognostic factor	Control group	Treatment group
If allocated to control group	Site		
	Site 1	4	2
	Site 2	4	5
	Sex		
	Male	4	4
	Female	4	3
	Age band		
	< 20	2	2
	20–64	2	2
	≥ 65	4	3
	Total number of imbalances	[(5 − 4) + (4 − 3) + (4 − 3)] = 3	
If allocated to treatment group	Site		
	Site 1	4	2
	Site 2	3	6
	Sex		
	Male	4	4
	Female	3	4
	Age band		
	< 20	2	2
	20–64	2	2
	≥ 65	3	4
	Total number of imbalances	[(6 − 3) + (4 − 3) + (4 − 3)] = 5	

The 15th patient has factors sex female, 74 yr, and site 2. If this patient is allocated to the control group, the total imbalance is 3. If this patient is allocated to the treatment group, the total imbalance is 5. Therefore, this patient is allocated to the control group, and site 2, sex female, and the ≥ 65 age band in the control group receive the score.

randomization.

Although minimization is designed to overcome the disadvantages of stratified randomization, this method also has drawbacks. A concern from a statistical point of view is that it does not satisfy randomness, which is the basic assumption of statistical inference [15,16]. For this reason, the analysis of covariance or permutation test are proposed [13]. Furthermore, exposure of the subjects' information can lead to a certain degree of allocation prediction for the next subjects. The calculation process is complicated, but can be carried out through various programs.

Response-adaptive randomization

So far, the randomization methods is assumed that the variances of treatment effects are equal in each group. Thus, the number of subjects in both groups is determined under this assumption. However, when analyzing the data accruing as the study progresses, what happens if the variance in treatment effects is not the same? In this case, would it not reduce the number of subjects initially determined rather than the statistical power? In other words, should the allocation probabilities determined prior to the study remain constant throughout the study? Alternatively, is it possible to change the allocation probability during the study by using the data accruing as the study progresses? If the treatment effects turn out to be inferior during the study, would it be advisable to reduce the number of subjects allocated to this group [17,18]?

An example of response-adaptive randomization is the randomized play-the-winner rule. Here, the first subject is allocated by predefined randomization, and if this patient's response is "success," the next patient will be allocated to the same treatment group; otherwise, the patient will be allocated to another treatment. That is, this method is based on statistical reasoning that is not possible under a fixed allocation probability and on the ethics of allowing more patients to be allocated to treatments that benefit the patients. However, the method can lead to imbalances between the treatment groups. In addition, if clinical studies take a very long time to obtain the results of patient responses, this method cannot be recommended.

Ethics of Randomization

As noted earlier, RCT is a scientific study design based on the probability of allocating subjects to treatment groups in order to ensure comparability, form the basis of statistical inference, and identify the effects of treatment. However, an ethical debate needs to examine whether the treatment method for the subjects, especially for patients, should be determined by probability rather than by the physician. Nonetheless, the decisions should preferably be made by probability because clinical trials have the distinct goals of investigating the efficacy and safety of new medicines, medical devices, and procedures, rather than merely reach therapeutic conclusions. The purpose of the study is therefore to maintain objectivity, which is why prejudice and bias should be excluded. That is, only an unconstrained attitude during the study can confirm that a particular medicine, medical device, or procedure is effective or safe.

Consider this from another perspective. If the researcher maintains an unconstrained attitude, and the subject receives all the information, understands it, and decides to voluntarily participate, is the clinical study ethical? Unfortunately, this is not so easy to answer. Participation in a clinical study may provide the subject with the benefit of treatment, but it could be risky. Furthermore, the subjects may be given a placebo, and not treatment. Eventually, the subject may be forced to make personal sacrifices for ambiguous benefit. In other words, some subjects have to undergo further treatment, representing the cost that society has to pay for the benefit of future subjects or for a larger number of subjects [4,19]. This ethical dilemma on the balance between individual ethics and collective ethics [20] is still spawning much controversy. If, additionally, the researcher is biased, the controversy over this dilemma will obviously become more confused and the reliability of the study will be lowered. Therefore, randomization is a key factor in a study having to clarify causality through comparison.

Conclusions

Studies have described a random table with subsequent randomization. However, if accurate information on randomization is not provided, it would be difficult to gain enough confidence to proceed with the study and arrive at conclusions. Furthermore, probability-based treatment is allowed with the hope that the trial will be conducted through proper processes, and that the outcome will ultimately benefit the medical profession. Concurrently, it should be fully appreciated that the contribution of the subjects involved in this process is a social cost.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Author Contributions

Chi-Yeon Lim (Software; Supervision; Validation; Writing – review & editing)
Junyong In (Conceptualization; Software; Visualization; Writing – original draft; Writing – review & editing)

ORCID

Chi-Yeon Lim, https://orcid.org/0000-0003-0178-6976
Junyong In, https://orcid.org/0000-0001-7403-4287

References

1. Altman DG. Randomisation. BMJ 1991; 302: 1481-2.
2. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c869.
3. Kang H. Random allocation and dynamic allocation randomization. Anesth Pain Med 2017; 12: 201-12.
4. Rosenberger WF, Lachin JM. Randomization in Clinical Trials. 2nd ed. Hoboken, Wiley. 2016, pp 1-14.
5. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? BMJ 1999; 318: 1209.
6. Lachin JM, Matts JP, Wei LJ. Randomization in clinical trials: conclusions and recommendations. Control Clin Trials 1988; 9: 365-74.
7. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. Lancet 2002; 359: 966-70.
8. Schulz KF. Randomized controlled trials. Clin Obstet Gynecol 1998; 41: 245-56.
9. Rosenberger WF, Lachin JM. Randomization in Clinical Trials. 2nd ed. Hoboken, Wiley. 2016, pp 37-66.
10. Cocconi G, Bella M, Zironi S, Algeri R, Di Costanzo F, De Lisi V, et al. Fluorouracil, doxorubicin, and mitomycin combination versus PELF chemotherapy in advanced gastric cancer: a prospective randomized trial of the Italian Oncology Group for Clinical Research. J Clin Oncol 1994; 12: 2687-93.
11. Rosenberger WF, Lachin JM. Randomization in Clinical Trials. 2nd ed. Hoboken, Wiley. 2016, pp 133-46.
12. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975;

31: 103-15.
13. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. Clin Pharmacol Ther 1974; 15: 443-53.
14. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. J Clin Epidemiol 1999; 52: 19-26.
15. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. Lancet 2002; 359: 515-9.
16. Rosenberger WF, Sverdlov O. Handling Covariates in the Design of Clinical Trials. Statistical Science 2008; 23: 404-19.
17. Rosenberger WF, Lachin JM. The use of response-adaptive designs in clinical trials. Control Clin Trials 1993; 14: 471-84.
18. Rosenberger WF, Lachin JM. Randomization in Clinical Trials. 2nd ed. Hoboken, Wiley. 2016, pp 189-215.
19. Nardini C. The ethics of clinical trials. Ecancermedicalscience 2014; 8: 387.
20. Palmer CR, Rosenberger WF. Ethics and practice: alternative designs for phase III randomized clinical trials. Control Clin Trials 1999; 20: 172-86.

Appendix

To utilize this appendix, the statistical program R version 3.6.0 (R Core Team (2019). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. R-project.org/) is needed. It is also necessary to understand the installation of its package, data path specification, input and execution of the command, and so on.

There are a few packages that make it easier to do randomization. In this appendix, we use "randomizeR" package [1]. This package contains functions and commands for creating various randomization sequences.

The following command invokes the randomizeR package.

```
> library(randomizeR)
```

Let's try some of the randomization methods introduced in this article.

1. Simple randomization

The function for simple randomization is crPar, which defines and includes the basic elements needed for a random sequence. The genSeq function then generates a randomization sequence. Let's look at simple randomization for a study in which the total number of subjects is 1000 (N) and the number of groups is 3 (K).

First, specify rand with basic information for full randomization. If you open this rand, you can check the stored information.

```
> N ← 1000
> K ← 3
> (rand ← crPar(N, K))

Object of class "crPar"
design = CR
N = 500
K = 3
groups = A B C
```

Next, run the genSeq function with rand to generate a random sequence, and specify it as crs.

```
> (crs ← genSeq(rand))
Object of class "rCrSeq"

design = CR
seed = 476000002
N = 500
K = 3
groups = A B C

The sequence M:

1 B C B B B C B C A B ...
```

Here you can see only some of the entire random sequence, so to see the whole,

```
> getRandList(crs)
```

Note that a new random sequence is generated each time you run it. If you want to see the random sequence already generated, you need to keep the seed[1] value. The seed generated here is 476000002. If you generate crs.2 with the seed, crs.2 can be equal to crs.

```
> (crs.2 ← genSeq(rand, 476000002))
> getRandList(crs.2)
```

The generated random sequence can be saved as a 'simpleRS.csv' file with the following command.

```
> saveRand(crs, file = "simpleRS.csv")
```

2. Block randomization

This randomization method is performed using one block size. For example, suppose there is a study in which a total of 90 subjects are allocated to three groups (K = 3), each with 30 subjects. If six is selected as the block size (bc), 15 iterations (R = 15) are needed for allocation of 90 subjects (if the block size is 9, you will need 10 iterations). First, basic information is assigned to rand by using pbrPar, a function for block randomization. Next, the allocation order generated by the genSeq function with 'rand' is stored in brs.

```
> bc ← 90
> K ← 3
> R ← 15
> (rand ← pbrPar(bc, K))
> (brs ← genSeq(rand, R))

Object of class "rPbrSeq"

design = PBR(6)
seed = 408740403
N = 6
K = 3
groups = A B C
bc = 6

The first 3 of 15 sequences of M:

1 C C B A B A
2 B C C A A B
3 C B A C A B
...

> getRandList(brs)
```

To save the generated random sequence (brs) as a csv file

```
> saveRand(brs, file = "blockRS.csv")
```

A seed value is needed to regenerate and confirm the same random sequence. If you apply the seed value given during brs genera-

[1]Whenever a new random sequence is generated, the seed value will change.

tion to generate brs.2, you can see the same random sequence as brs.

```
> (brs.2 ← genSeq(rand, 408740403))
> getRandList(brs.2)
```

3. Randomized block randomization

The default function is rpbrPar. The researcher determines the total number of participants (N = 80) and the randomly chosen block size (rb, size of 3, 6, 9). The basic function repeats the selection of any block size (one of 3, 6, or 9) and the allocation of subjects within the selected block until all subjects are assigned.

If the last block size is larger than the remaining number of subjects, the last block is not filled. If the last block should be filled, set filledBlock = FALSE. Otherwise, set filledBlock = TRUE. The random sequence is generated and stored in rbrs. BlockConst shows that blocks of various sizes are used.

```
> N ← 60
> rb ← c(3, 6, 9)
> K ← 3

> rand ← rpbrPar(N, rb, K, ratio = rep(1, K), groups = LETTERS[1:K], filledBlock = FALSE)
> (rbrs ← genSeq(rand))

Object of class "rRpbrSeq"

design = RPBR(3,6,9)
rb = 3 6 9
filledBlock = FALSE
seed = 939989023
N = 60
K = 3
ratio = 1 1 1
groups = A B C

RandomizationSeqs BlockConst
B C A C B C B A ... 3 9 9 6 3 ...
```

You can view or save the random sequence generated by the following command.

```
> getRandList(rbrs)
> saveRand(rbrs, file = "RBRS.csv")
```

4. Random allocation rule

Te jar model is often used. Put as many balls as the number of subjects (the color of the ball varies according to the group) in the jar according to the allocation ratio. Whenever the subject is assigned, take out the ball and confirm it, and do not put the ball back into the jar (random sampling without replacement). This allocation is repeated on a block-by-block basis. The following is the order to apply the random allocation rule to compare the three groups (K) with the total number of subjects (N) of 60.

```
> N ← 60
> K ← 3
> rand ← rarPar(N, K)
> (rar ← genSeq(rand))

Object of class "rRarSeq"

design = RAR
seed = 144561
N = 60
K = 3
groups = A B C

The sequence M:

1 A B B C A B A B C C ...
```

You can view or save the random sequence generated by the following command.

```
> getRandList(rar)
> saveRand(rar, file = "RAR.csv")
```

5. Stratified randomization

Once the prognostic factors and strata have been determined, randomization for each stratum can be performed using the methods previously described. However, random sequences are required as many as the number of strata. In the previous example, the prognostic factors were the recruitment hospitals (site 1, 2), sex (male, female), and age band (under 20 years old, 20-64 years, 65 years old or older). Therefore, 12 random sequences are required.

Reference

1. Uschner D, Schindler D, Hilgers RD, Heussen N. randomizeR: an R package for the assessment and implementation of randomization in clinical trials. J Stat Softw 2018; 85: 1-22.