Clinical Research Methodology I: Introduction to Randomized Trials

Lillian S Kao, MD, MS, FACS, **Jon E Tyson, MD, MPH**, **Martin L Blakely, MD, MS, FACS**, and **Kevin P Lally, MD, MS, FACS**
From the Department of Surgery (Kao), Department of Pediatrics, Division of Neonatology (Tyson), and Department of Pediatric Surgery (Lally), The University of Texas Health Science Center at Houston, Houston, TX, and the Department of Surgery, Division of Pediatric Surgery, The University of Tennessee Health Science Center at Memphis, Memphis, TN (Blakely).

With increasing initiatives to improve the effectiveness and safety of patient care, there is a growing emphasis on evidence-based medicine and incorporation of high-quality evidence into clinical practice. The cornerstone of evidence-based medicine is the randomized controlled trial (RCT). The World Health Organization defines a clinical trial as "any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes."[1] Randomization refers to the method of assignment of the intervention or comparison(s). Fewer than 10% of clinical studies reported in surgical journals are RCTs,[2–4] and treatments in surgery are only half as likely to be based on RCTs as treatments in internal medicine.[5]

Multiple factors impede surgeons performing definitive RCTs, including the inability to blind health care providers and patients, small sample sizes, variations in procedural competence, and strong surgeon or patient preferences.[5–8] Not all questions can be addressed in an RCT; Solomon and colleagues[8] estimated that only 40% of treatment questions involving surgical procedures are amenable to evaluation by an RCT, even in an ideal clinical setting. In surgical oncology, trials evaluating survival after operations for a rare malignancy can require an unreasonably large sample size. Pawlik and colleagues[9] estimated that only 0.3% of patients with pancreatic adenocarcinoma could benefit from pancreaticoduodenectomy with extended lymphadenectomy; a randomized trial of 202,000 patients per arm would be necessary to detect a difference in survival.

These reasons should not dissuade surgeons from performing RCTs. Even for rare diseases, randomized trials remain the best method to obtain unbiased estimates of treatment effect.[10,11] Rigorously conducted RCTs minimize bias by controlling for known and unknown factors (confounders) that affect outcomes and distort the apparent treatment effect. Observational studies, including those with the most sophisticated design and analysis,[12,13] can control only for known confounders and might not adequately control for those. Many surgical and medical interventions recommended based on observational studies have later been demonstrated to be ineffective or even harmful. These have included blood transfusions to maintain a hemoglobin

>10 mg/dL in critically ill patients,[14,15] bone marrow transplantation for breast cancer,[16–19] and extracranial-intracranial bypass for carotid artery stenosis.[20,21]

Another major reason for RCTs to be of interest to surgeons is that patients enrolled in trials can have improved short-term outcomes, even if the intervention is ineffective.[22–25] Potential sources of this benefit include enrollment of lower-risk patients, use of standardized protocols and improved supportive care, and greater effort to prevent and address treatment hazards. Different outcomes can also be observed in trial participants because of either the Hawthorne or placebo effect, both of which can distort the apparent treatment effect and threaten the validity of the trial. The Hawthorne effect occurs when changes in clinicians' or patients' behavior, because of being observed, results in improved outcomes. For example, a prospective observational study evaluating operating room efficiency after an intervention can demonstrate improvement over historic performance, in part because the staff is aware of being observed rather than as a result of the intervention. The placebo effect occurs when the patient derives benefit not from the treatment itself, but from the patient's expectations for benefit. In a randomized trial of arthroscopic surgery versus sham surgery for osteoarthritis of the knee, the placebo procedure had equivalent results to debridement and lavage, despite lack of any therapeutic intervention.[26]

Despite the advantages of well-conducted RCTs, poorly conducted trials or inadequately reported results can yield misleading information.[27,28] Recently, Chang and colleagues[29] demonstrated the continued paucity of high-level evidence in surgical journals and called for articles on clinical research methodology to educate surgeons. The purpose of this article is to serve as an introduction to RCTs, focusing on procedures for assigning treatment groups that serve to minimize bias and error in estimating treatment effects. Common threats to validity and potential solutions to difficulties in randomizing patients in surgical trials will also be discussed.

OBSERVATIONAL COHORT STUDIES

RCTs are the gold standard for evaluating the effectiveness of an intervention. Many therapies have historically been evaluated in surgery using observational cohort studies where groups of patients are followed for a period of time, either consecutively or concurrently. These studies can be conducted retrospectively or prospectively. The fundamental criticism of observational cohort studies is that confounding can result in biased estimates of treatment effect.[30,31]

A confounder is a known or unknown factor that is related to the variable of interest (eg, an intervention) and is a cause of the outcomes. For example, suppose a study finds that patients undergoing a procedure by surgeon A have increased mortality when compared with surgeon B. The outcomes difference might not be a result of inferior operative technique of surgeon A, but rather confounders, such as patients' comorbidities or severity of disease (eg, if surgeon A is referred the more complicated patients).

Observational studies cannot account for unknown confounders. Novel statistical methods can improve estimates of treatment effect because of known and unknown confounders in nonrandomized trials, but are still subject to limitations.[32,33] Traditionally, nonrandomized or observational studies adjust for known confounders in the statistical analysis. Adjustment refers to the mathematic modeling of the relationship between one or more predictor variables and the outcomes to estimate the isolated effect of each variable. Even with advanced statistical analyses, such as propensity scoring, these models cannot completely adjust for all of the confounders.[12] Although observational cohort studies have a role in clinical research, such as in answering questions about harm, well-designed RCTs are the gold standard for evaluating an intervention because they minimize bias from known and unknown confounders.

OVERVIEW OF RANDOMIZATION AND ALLOCATION CONCEALMENT

Properly designed RCTs minimize imbalances in baseline characteristics between groups and could distort the apparent effect of the difference in treatment on patient outcomes. The randomization procedure used to assign treatment and prevent prediction of treatment assignment resulting in allocation of intervention bias is especially important. With random assignment, each patient has the same chance of being assigned to a specific treatment group. Equal (1:1) allocation results in the same likelihood of assignment to either group (50:50) and the greatest power to detect a difference in outcomes between the groups. Unequal or weighted randomization allows the investigators to maintain a balance between groups in their baseline characteristics, but allocate more patients to one group (eg, with a 2:1 allocation, two-thirds of patients will be assigned the first treatment and one-third to the second). Unequal randomization can be used to decrease costs when one treatment is considerably more expensive to provide than the other.[34,35]

Valid methods of randomization include flipping a coin, rolling a die, using a table of random numbers, or running a computerized random allocation generator (eg, http://www.random.org). Randomization should be performed in such a way that the investigator should not be able to anticipate treatment group.[36–38] Not all published trials reported as "randomized" are truly randomized. In fact, only 33% to 58% of published surgical trials describe a valid randomization process where the treatment assignment cannot be predicted.[39,40] For example, in a randomized trial evaluating screening mammography, participants were assigned based on which day of the month they were born (patients born between the 1st and 10th of the month or 21st and 31st were assigned mammography and patients born between the 11th and the 20th were assigned the control).[41] Results were questioned because anticipation of treatment group can inadvertently influence whether a patient is considered to meet eligibility criteria or how much effort is devoted to securing informed consent[42] and so can cause selection biases with baseline differences between groups that can influence the results. Other "pseudorandom" or "quasirandom" schemes include use of medical record number or date of enrollment.

Allocation concealment prevents the investigator or trial participant from consciously or subconsciously influencing the treatment assignment and causing selection bias. Allocation concealment, which occurs before randomization, should not be confused with blinding (also known as masking), which occurs after randomization. Where a valid randomization scheme has been used, allocation can still be inadequately concealed (eg, use of translucent envelopes containing treatment assignments). Methods of allocation concealment include use of sequentially numbered, sealed, opaque envelopes or allocation by a central office. Allocation concealment is always possible, although blinding is not.[38] Yet, between 1999 and 2003, only 29% of published surgical trials reported allocation concealment.[40] Although allocation concealment can be used but not noted in published reports, inadequate allocation concealment appears to be a large and generally unrecognized source of bias. Schulz and colleagues[43] found that treatment effect was overestimated by 41% when allocation concealment was inadequate or unclear.

VARIATIONS IN RANDOMIZATION SCHEMES

Simple randomization

The most straightforward scheme for allocating patients is simple randomization (Fig. 1A), with treatment assigned using one of the methods mentioned previously (eg, computer-generated random sequence). Simple randomization can result, by chance alone, in unequal numbers in each group—the smaller the sample size, the larger the likelihood of a major imbalance in the number of patients or the baseline characteristics in each group.[44] An example

of simple randomization would be the sequence of 20 random numbers generated using a computer program (Table 1).

Treatment could be allocated as evens receiving treatment A and odds receiving treatment B. In this case, 9 patients would receive treatment A and 11 patients would receive treatment B (Table 1).

Alternatively, patients assigned to a number between 1 and 50 could receive treatment A and patients assigned to a number between 51 and 100 could receive treatment B. In this case, 6 patients would receive treatment A and 14 would receive treatment B. In small trials or in larger trials with planned interim analyses, simple randomization can result in imbalanced group numbers. Even if the groups have equal numbers, there can be important differences in baseline characteristics that would distort the apparent treatment effect.

Blocked (restricted randomization)

Simple randomization can result not only in imbalanced groups, but also chronological bias in which one treatment is predominantly assigned earlier and the other later in the trial. Chronological bias is important if outcomes changes with time, as when surgeons become more adept at the procedure under investigation or increasing referrals for the procedures changes the patient population.[45] Chronological bias results in an inability to differentiate between the effects of temporally related factors, such as surgeon experience and treatment. For these reasons, blocked (restricted) randomization schemes can be used (Figs. 1B, 1C).

For example, a uniform block size of 4 can be used with 1:1 allocation and two treatment arms. The two arms will never differ at any time by more than two patients, or half of the block length. There are six possible assignment orders for each block (called a permuted block) of four patients: AABB, ABAB, ABBA, BAAB, BABA, and BBAA. Although blocked randomization will maintain equal or nearly equal group sizes across time, selection bias can occur if the investigators are not blinded to block size and treatment assignment. If the first three patients in the trial received treatments A, A, and B, then the unblinded investigator might anticipate that the fourth patient will receive treatment B. The decision whether to enroll the next study candidate could be inadvertently affected, as a result, by the investigator's treatment preference.[42] This problem can generally be avoided by randomly or systematically varying the block sizes (Fig. 1C).

Stratified randomization

Imbalances in prognostic factors between treatment arms can occur because of chance alone, even if randomization and allocation concealment are properly performed. Important imbalances are most likely to occur by chance in small trials or during interim analyses of large RCTs.[46] Prognostic stratification can be used to avoid such imbalances. Patients can be first categorized based on several prognostic factors into strata and then randomized within each stratum, guaranteeing no major imbalance between groups in these factors (Fig. 1D). For example, Fitzgibbons and colleagues[47] performed an RCT to evaluate whether watchful waiting is an acceptable alternative to tension-free repair for inguinal hernias in minimally symptomatic or asymptomatic patients. Eligible patients were stratified by center (six total), whether the hernia was primary or recurrent, and whether the hernia was unilateral or bilateral. The total number of strata for this study was 24, or the product of the number of levels of each factor ($6 \times 2 \times 2$). Once assigned to a stratum, patients were then randomized to either watchful waiting or hernia repair.

Because important prognostic factors will be balanced, stratified randomization can decrease the chance of a type I error (finding a difference between treatment arms because of chance

alone) and can increase the power (the chance of finding a difference if one exists) of small studies, where the stratified factors have a large effect on outcomes.[46] Additionally, stratification can increase the validity of subgroup or interim analyses.[46] If too many strata are used, some strata might not be filled with equal patients in both groups, leading to imbalances in other prognostic factors.[46] Excessive stratification also unduly increases the complexity of trial administration, randomization, and analysis. Stratification is usually performed using only a small number of carefully selected variables likely to have a large impact on outcomes.

Adaptive randomization

Another strategy to minimize imbalances in prognostic factors is to use an adaptive randomization scheme when randomization is influenced by analysis of either the baseline characteristics or outcomes of previous patients. When treatment assignment is based on patient characteristics, the adaptive randomization procedure known as minimization assigns the next treatment to minimize any imbalance in prognostic factors among previously enrolled patients. For the computer algorithm to run, minimization should be limited to larger trials.[48]

One response-adaptive randomization procedure used in trials examining a dichotomous outcomes (eg, yes/no or survival/death) involves the "play-the-winner" strategy to allocate treatment based on outcomes of the last patient enrolled (Fig. 2). The more successful a treatment, the more likely that the next patient will be randomized to that treatment.[49] For short-term trials where the treatments have been well-evaluated for safety, play-the-winner trials can reduce the likelihood that a patient is assigned to an ineffective or harmful treatment. [49] Adaptive trials are being increasingly used in phase 1 or 2 cancer trials.[50–52] The downside is that these trials are complex to plan and analyze, quite susceptible to chronological bias, and might not be persuasive.

The most well-known and controversial play-the-winner randomized trial was the Michigan Extracorporeal Membrane Oxygenation (ECMO) trial for neonatal respiratory failure by Bartlett and colleagues.[53] Neonates with respiratory failure and a predicted ≥ 80% chance of mortality were randomized to either ECMO or conventional treatment. The investigators' intent was to perform a methodologically sound, randomized trial that minimized the number of critically ill infants given the inferior treatment. The study was designed to end after 10 patients had received ECMO or 10 patients had received the control. The first patient enrolled received ECMO and survived; the second patient received the control and died. The trial was terminated after 11 patients, only 1 of whom received the control. The investigators concluded that ECMO improved survival when compared with conventional treatment in neonates with respiratory failure. The main criticism of the trial was that the control group included only one patient.[54] Widespread acceptance of ECMO for neonates with respiratory failure did not occur until after larger and more conventional trials were performed.[55]

RANDOMIZED DESIGNS INCORPORATING PATIENT AND SURGEON PREFERENCES

Patient preference trials

Strong patient preferences can result in failure to enroll patients into surgical RCTs or serve as a theoretical threat to validity.[6,35,56] Patients with strong preferences for one treatment can differ from those without, resulting in selection bias or a systematic difference in patients enrolled in trials from those not enrolled.[57] Patient preference can also be an independent prognostic factor or can interact with the treatment to affect outcomes, particularly in unblended RCTs.[58] For example, patients randomized to their preferred treatment can perform better because of increased compliance or a placebo effect, and patients randomized to their

nonpreferred treatment can perform worse.[57] One potential solution is to measure baseline patient preferences and to mathematically adjust for the interaction between preference and treatment, but this approach increases sample size.[58]

Another solution is to modify the trial design to incorporate patient or physician preferences using a comprehensive cohort design, Zelen's design, or Wennberg's design.[35,56,59–61] These trial designs have not been commonly used. In the comprehensive cohort design, patients without strong preferences are randomized, and patients with strong preferences are assigned to their treatment of choice.[35] It is often used in trials where participation in an RCT might be low because of strong patient preferences. Even if the proportion of patients randomized is low, this design should be less susceptible to bias than a nonrandomized cohort study, which is often encouraged by statisticians when problems in accrual can limit the power of a conventional randomized trial. Because of the lower randomization rates, these trials can be more expensive, require more total patients, and be more difficult to interpret than conventional trials.[57]

The National Institute of Child Health and Human Development Neonatal Research Network is currently planning a comprehensive cohort trial comparing laparotomy and peritoneal drainage for extremely low birth weight infants with severe necrotizing enterocolitis or isolated intestinal perforation (personal communication, Blakely). These conditions in this patient population are associated with a 50% mortality rate and a 72% rate of death or neurodevelopmental impairment. Surgical practice varies widely; caregivers often have strong treatment preferences,[62,63] and the consent rate can be no higher than 50% because of problems obtaining consent for emergency therapies. In this trial, the same risk and outcomes data will be collected for nonrandomized patients (observational or preference cohort) and for randomized patients. In the primary analysis, treatment effect will be assessed as the relative risk of death or neurodevelopmental impairment with laparotomy (relative to drainage) among randomized patients. The relative risk for death or impairment with laparotomy will also be assessed in the observational cohort after adjusting for important known risk factors. If the relative risk for observational cohort is similar to that for randomized patients, all patients can be combined in a supplemental analysis to increase the power, precision, and generalizability of the study in assessing treatment effect. An analysis with all patients combined would not be performed if the relative risk is not comparable for the randomized patients and the observational cohort. In this circumstance, the difference might well be because of an inability to adjust for unknown or unmeasured confounders among patients treated according to physician or parent preference

Zelen's design, also known as the postrandomization consent design, has two variants. In the single-consent design, patients randomized to the standard therapy are not informed of the trial or offered alternative therapy. Consent is sought only for patients randomized to the intervention. If consent is refused, they are administered standard therapy but analyzed with the intervention group. The single-consent design raises ethical concerns because patients are randomized before consent and because patients receiving standard therapy are included without informed consent of their participation in the trial. In the double-consent design, consent is sought for patients randomized to a standard therapy and those randomized to the intervention, and both groups are informed of the trial and both groups are allowed to receive the opposite treatment if consent is refused for the treatment to which they were randomized. Zelen's design has been used to evaluate screening tools such as fecal occult blood testing for colorectal cancer.[64]

With Wennberg's design, patients are randomized to either a preference group or a randomization group. Patients in the preference group are offered their treatment of choice, and patients in the other group are assigned treatment based on randomization. All groups are

analyzed to assess the impact of patient preference on outcomes.[35,59] Although patient preference trials are an alternative to RCTs, downsides include potential for additional differences between treatment groups other than preference and increased sample size requirements or cost to complete a trial.[34,57]

EXPERTISE-BASED TRIALS

A proposed solution to the problem of variation between surgeons in skill and preference is the expertise-based RCT. In a conventional RCT evaluating two surgical procedures (eg, open versus laparoscopic hernia repair), a surgeon can be asked to perform both procedures, even though he or she might be adept with only one. Differential expertise bias can result from favoring the less technically challenging or more familiar procedure if a higher percentage of experienced surgeons performed that procedure.[45] Additionally, differential expertise can result in increased cross-over from one procedure to another (eg, conversion from laparoscopic to open hernia repair), or bias resulting from use of different co-interventions.[45]

An expertise-based trial differs from a conventional RCT because surgeons perform only the procedure at which they believe they are most skilled. Proponents argue that expertise-based trials minimize bias resulting from differences in technical competency and surgeon preference, decrease crossover from one intervention to the other, and can be more ethical than conventional RCTs.[45] On the other hand, expertise-based RCTs present challenges in coordinating trials in which there are few experts for one or both procedures; changing surgeons after the initial patient contact, or generalizing the results to surgeons with less expertise.[45]

For example, in a trial comparing open with endovascular aortic aneurysm repair, the investigators required each participating surgeon to have performed 20 endovascular aortic aneurysm repair procedures to control for expertise bias.[65] The trial demonstrated no difference in all-cause mortality between the groups.[66] Performance of 60 endovascular repairs, or 40 more than the minimum requirement for surgeon participation in this study, appears to be necessary to achieve an acceptable failure rate of < 10%. The minimum number of procedures required to participate in a trial is often less than the number needed to reach the plateau of the learning curve, biasing the results.[45] An expertise-based RCT, NExT ERA: National Expertise Based Trial of Elective Repair of Abdominal Aortic Aneurysms: A Pilot Study, is planned to prevent problems in interpreting the trial because of differential surgical expertise from affecting outcomes after aneurysm repair (www.clinicaltrials.gov; NCT00358085).

INTERNAL AND EXTERNAL VALIDITY IN RANDOMIZED TRIALS

Before applying the results of RCTs to individual patients, the internal and external validity of the trial must be examined. Internal validity refers to the adequacy of the trial design to provide a true estimate of association between an exposure and outcomes in patients studied, and external validity assesses the generalizability of the results to other patients. Threats to internal validity can result from either random or systematic error. Random errors result in errors in either direction, but systematic errors are a result of bias, resulting in consistent variation in the same direction. An example of random error is the up and down variability in blood pressure measurements based on the precision of an automatic cuff. A systematic error occurs when all of the blood pressure measurements are high because the cuff is too small. Bias can occur at any point in a trial, including during design, selection of the participants, execution of the intervention, outcomes measurement, data analysis, results interpretation, or publication. Specific types of bias include selection bias that results from systematic differences between treatment groups, confounding, ascertainment bias that results from lack of blinding of outcomes assessors, compliance bias because of differential adherence to the study protocols, and bias because of losses or withdrawals to followup.[67]

External validity is dependent on multiple factors, including the characteristics of the participants, the intervention, and the setting of the trial.[68] Enrolled patients can differ substantially from eligible patients and ineligible patients with the condition of interest, or both, representing only a select population. An analysis of RCTs in high-impact medical journals found that only 47% of exclusion criteria were well-justified, and large subpopulations, such as women, children, the elderly, and patients with common medical conditions, were often excluded from RCTs.[69] In evaluating external validity, the difference between efficacy (explanatory) and effectiveness (management or pragmatic) trials must also be considered. Efficacy trials test whether therapies work under ideal conditions (eg, highly protocolized interventions and small number of homogeneous patients), and effectiveness trials test whether therapies work under routine or "real-world" circumstances (eg, large number of diverse patients and broad range of clinically acceptable co-interventions). Efficacy trials maximize internal validity and effectiveness trials emphasize external validity.[70]

Well-designed RCTs reduce systematic errors from selection bias, biased treatment assignment, ascertainment bias, and confounding. Even with adequate randomization and allocation concealment, as described here, both random and systematic errors can still occur. Larger sample sizes can decrease the risk of imbalances because of chance and can increase the external validity of trials as well by including more diverse patients (eg, pragmatic trials). A description of all potential threats to validity is beyond the scope of this article.

Despite the perceived barriers to performing randomized clinical trials in surgery, they remain the gold standard for evaluating an intervention. Surgeons must be aware of the potential methodologic flaws that can invalidate results, both in interpreting and applying the literature and in designing future trials. To promote rigorous, high-quality studies, surgeons should be aware of variations in trial design, and increase use of alternative designs when conventional trials would not be feasible or suitable.

REFERENCES

1. World Health Organization. International Clinical Trials Registry Platform (ICTRP). 2007 [Accessed August 30]. Available at: http://www.who.int/ictrp/en/

2. McLeod RS. Issues in surgical randomized controlled trials. World J Surg 1999;23:1210–1214. [PubMed: 10552108]

3. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. Surgery 1994;115:707–712. [PubMed: 8197563]

4. Hardin WD Jr, Stylianos S, Lally KP. Evidence-based practice in pediatric surgery. J Pediatr Surg 1999;34:908–912. [PubMed: 10359204]discussion 912–913

5. McCulloch P, Taylor I, Sasako M, et al. Randomised trials in surgery: problems and possible solutions. BMJ 2002;324:1448–1451. [PubMed: 12065273]

6. Abraham NS, Young JM, Solomon MJ. A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. Surgery 2006;139:469–483. [PubMed: 16627056]

7. Fung EK, Lore JM Jr. Randomized controlled trials for evaluating surgical questions. Arch Otolaryngol Head Neck Surg 2002;128:631–634. [PubMed: 12049555]

8. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? Surgery 1995;118:459–467. [PubMed: 7652679]

9. Pawlik TM, Abdalla EK, Barnett CC, et al. Feasibility of a randomized trial of extended lymphadenectomy for pancreatic cancer. Arch Surg 2005;140:584–589. [PubMed: 15967906] discussion 589–591

10. Lilford RJ, Thornton JG, Braunholtz D. Clinical trials and rare diseases: a way out of a conundrum. BMJ 1995;311:1621–1625. [PubMed: 8555809]

11. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. Lancet 2005;365:1348–1353. [PubMed: 15823387]

12. Adamina M, Guller U, Weber WP, Oertli D. Propensity scores and the surgeon. Br J Surg 2006;93:389–394. [PubMed: 16400708]

13. Provonost P, Weast B, Rosenstein B, et al. Implementing and validating a comprehensive unit-based safety program. J Patient Saf 2005;1:33–40.

14. Hebert PC, Wells G, Blajchman MA, et al. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. N Engl J Med 1999;340:409–417. [PubMed: 9971864]

15. Hill SR, Carless PA, Henry DA, et al. Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion. Cochrane Database Syst Rev. 2002CD002042

16. Antman K, Ayash L, Elias A, et al. A phase II study of high-dose cyclophosphamide, thiotepa, and carboplatin with autologous marrow support in women with measurable advanced breast cancer responding to standard-dose therapy. J Clin Oncol 1992;10:102–110. [PubMed: 1727912]

17. Farquhar C, Marjoribanks J, Basser R, et al. High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy for women with metastatic breast cancer. Cochrane Database Syst Rev. 2005CD003142

18. Peters WP, Shpall EJ, Jones RB, et al. High-dose combination alkylating agents with bone marrow support as initial treatment for metastatic breast cancer. J Clin Oncol 1988;6:1368–1376. [PubMed: 3047332]

19. Thomas EJ, Taggart B, Crandell S, et al. Teaching teamwork during the Neonatal Resuscitation Program: a randomized trial. J Perinatol 2007;27:409–414. [PubMed: 17538634]

20. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. Results of an international randomized trial. The EC/IC Bypass Study Group. N Engl J Med 1985;313:1191–1200. [PubMed: 2865674]

21. Weinstein PR, Rodriguez y, Baena R, Chater NL. Results of extracranial-intracranial arterial bypass for intracranial internal carotid artery stenosis: review of 105 cases. Neurosurgery 1984;15:787–794. [PubMed: 6514151]

22. Braunholtz DA, Edwards SJ, Lilford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". J Clin Epidemiol 2001;54:217–224. [PubMed: 11223318]

23. Janni W, Kiechle M, Sommer H, et al. Study participation improves treatment strategies and individual patient care in participating centers. Anticancer Res 2006;26:3661–3667. [PubMed: 17094382]

24. Hallstrom A, Friedman L, Denes P, et al. Do arrhythmia patients improve survival by participating in randomized clinical trials? Observations from the Cardiac Arrhythmia Suppression Trial (CAST) and the Antiarrhythmics Versus Implantable Defibrillators Trial (AVID). Control Clin Trials 2003;24:341–352. [PubMed: 12757998]

25. Weijer C, Freedman B, Fuks A, et al. What difference does it make to be treated in a clinical trial? A pilot study. Clin Invest Med 1996;19:179–183. [PubMed: 8724821]

26. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 2002;34:81–88. [PubMed: 12110735]

27. Balasubramanian SP, Wiener M, Alshameeri Z, et al. Standards of reporting of randomized controlled trials in general surgery: can we do better? Ann Surg 2006 244::663–667. [PubMed: 17060756]

28. Jacquier I, Boutron I, Moher D, et al. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. Ann Surg 2006;244:677–683. [PubMed: 17060758]

29. Chang DC, Matsen SL, Simpkins CE. Why should surgeons care about clinical research methodology? J Am Coll Surg 2006;203:827–830. [PubMed: 17116550]

30. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. Am J Med 1982;72:233–240. [PubMed: 7058834]

31. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev. 2007MR000012

32. Baggs JG, Schmitt MH, Mushlin AI, et al. Association between nurse-physician collaboration and patient outcomes in three intensive care units. Crit Care Med 1999;27:1991–1998. [PubMed: 10507630]

33. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. Am J Epidemiol 2005;162:279–289. [PubMed: 15987725]

34. Jadad, A. Randomised controlled trials: a user's guide. London: BMJ Books; 1998.

35. King M, Nazareth I, Lampe F, et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. JAMA 2005;293:1089–1099. [PubMed: 15741531]

36. Altman DG, Bland JM. How to randomise. BMJ 1999;319:703–704. [PubMed: 10480833]

37. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? BMJ 1999;318:1209. [PubMed: 10221955]

38. Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. J Crit Care 2005;20:187–191. [PubMed: 16139163]discussion 191–193

39. Hall JC, Hall JL. Randomization in surgical trials. Surgery 2002;132:513–518. [PubMed: 12324767]

40. Ellis C, Hall JL, Khalil A, Hall JC. Evolution of methodological standards in surgical trials. ANZ J Surg 2005;75:874–877. [PubMed: 16176230]

41. Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Followup after 11 years—update of mortality results in the Stockholm mammographic screening trial. Breast Cancer Res Treat 1997;45:263–270. [PubMed: 9386870]

42. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. N Engl J Med 1983;309:1358–1361. [PubMed: 6633598]

43. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–412. [PubMed: 7823387]

44. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. Lancet 2002;359:515–519. [PubMed: 11853818]

45. Devereaux PJ, Bhandari M, Clarke M, et al. Need for expertise based randomised controlled trials. BMJ 2005;330:88. [PubMed: 15637373]

46. Kernan WN, Viscoli CM, Makuch RW, et al. Stratified randomization for clinical trials. J Clin Epidemiol 1999;52:19–26. [PubMed: 9973070]

47. Fitzgibbons RJ, Jonasson O, Gibbs J, et al. The development of a clinical trial to determine if watchful waiting is an acceptable alternative to routine herniorrhaphy for patients with minimal or no hernia symptoms. J Am Coll Surg 2003;196:737–742. [PubMed: 12742207]

48. Bracken MB. On stratification, minimization and protection against types 1 and 2 error. J Clin Epidemiol 2001;54:104–105. [PubMed: 11246516]

49. Rosenberger WF. Randomized play-the-winner clinical trials: review and recommendations. Control Clin Trials 1999;20:328–342. [PubMed: 10440560]

50. Kuehn BM. Industry, FDA warm to "adaptive" trials. JAMA 2006;296:1955–1957. [PubMed: 17062852]

51. Schmidt C. Adaptive design may hasten clinical trials. J Natl Cancer Inst 2007;99:108–109. [PubMed: 17227993]

52. Morita S, Sakamoto J. Application of an adaptive design to a randomized phase II selection trial in gastric cancer: a report of the study design. Pharm Stat 2006;5:109–118. [PubMed: 17080767]

53. Bartlett RH, Roloff DW, Cornell RG, et al. Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study. Pediatrics 1985;76:479–487. [PubMed: 3900904]

54. Paneth N, Wallenstein S. Extracorporeal membrane oxygenation and the play the winner rule. Pediatrics 1985;76:622–623. [PubMed: 4047807]

55. Elbourne D, Field D, Mugford M. Extracorporeal membrane oxygenation for severe respiratory failure in newborn infants. Cochrane Database Syst Rev. 2002CD001340

56. Torgerson DJ, Sibbald B. Understanding controlled trials. What is a patient preference trial? BMJ 1998;316:360. [PubMed: 9487173]

57. Torgerson DJ, Klaber-Moffett J, Russell IT. Patient preferences in randomised trials: threat or opportunity? J Health Serv Res Policy 1996;1:194–197. [PubMed: 10180870]

58. Halpern SD. Evaluating preference effects in partially unblinded, randomized clinical trials. J Clin Epidemiol 2003;56:109–115. [PubMed: 12654404]

59. Jadad, A. Randomised controlled trials: a user's guide. 24. London: BMJ Books; 1998. Types of randomised controlled trials.

60. Adamson J, Cockayne S, Puffer S, Torgerson DJ. Review of randomised trials using the post-randomised consent (Zelen's) design. Contemp Clin Trials 2006;27:305–319. [PubMed: 16455306]

61. Torgerson DJ, Roland M. What is Zelen's design? BMJ 1998;316:606. [PubMed: 9518917]

62. Blakely ML, Lally KP, McDonald S, et al. Postoperative outcomes of extremely low birth-weight infants with necrotizing enterocolitis or isolated intestinal perforation: a prospective cohort study by the NICHD Neonatal Research Network. Ann Surg 2005;241:984–989. [PubMed: 15912048] discussion 989–994

63. Blakely ML, Tyson JE, Lally KP, et al. Laparotomy versus peritoneal drainage for necrotizing enterocolitis or isolated intestinal perforation in extremely low birth weight infants: outcomes through 18 months adjusted age. Pediatrics 2006;117:e680–e687. [PubMed: 16549503]

64. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 1996;348:1472–1477. [PubMed: 8942775]

65. Greenhalgh RM, Brown LC, Kwong GP, et al. Comparison of endovascular aneurysm repair with open repair in patients with abdominal aortic aneurysm (EVAR trial 1), 30-day operative mortality results: randomised controlled trial. Lancet 2004;364:843–848. [PubMed: 15351191]

66. Endovascular aneurysm repair versus open repair in patients with abdominal aortic aneurysm (EVAR trial 1): randomised controlled trial. Lancet 2005;365:2179–2186. [PubMed: 15978925]

67. Delgado-Rodriguez M, Llorca J. Bias. J Epidemiol Community Health 2004;58:635–641.

68. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. PLoS Clin Trials 2006;1

69. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA 2007;297:1233–1240. [PubMed: 17374817]

70. Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC Med Res Methodol 2003;3:28. [PubMed: 14690550]
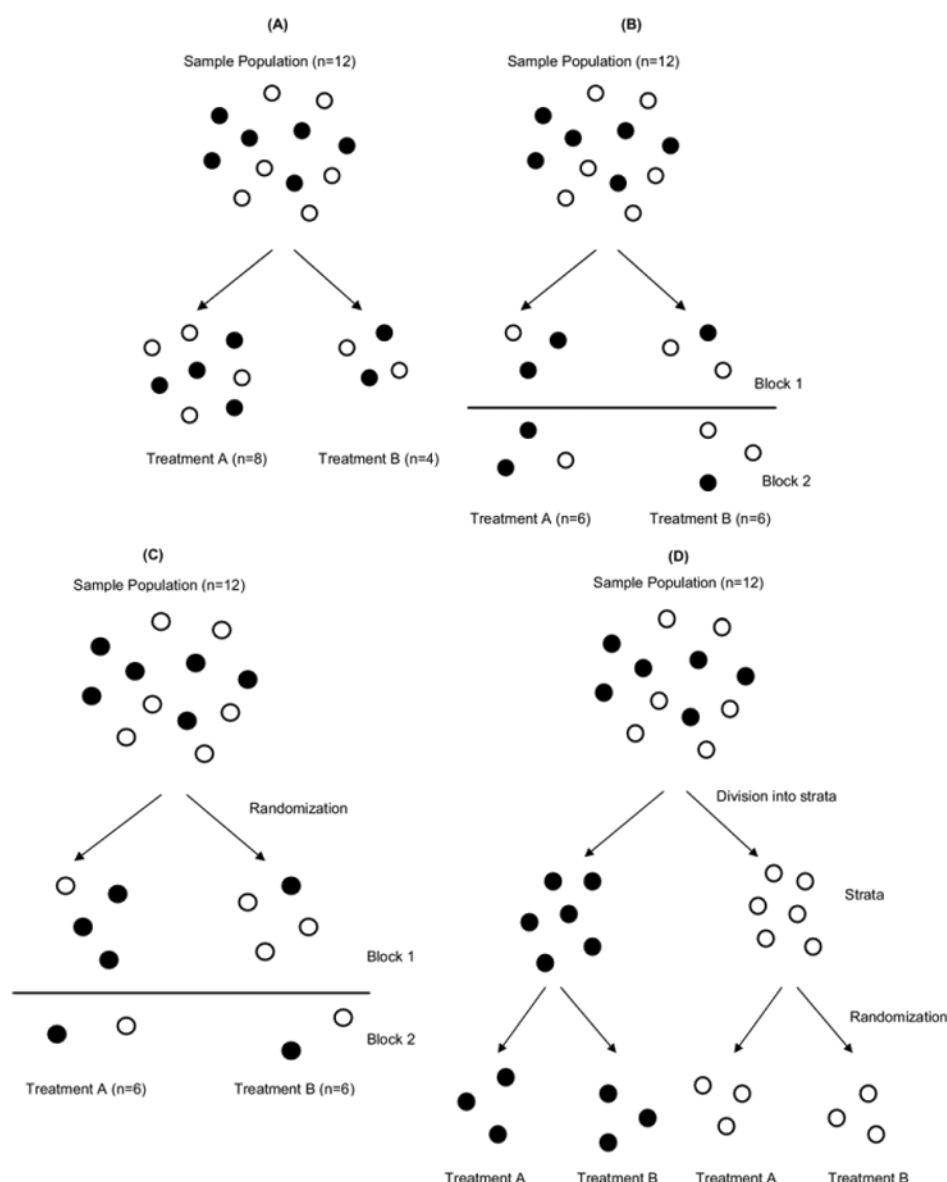
Figure 1.
Randomization procedures. In this example, balls represent patients and the color represents a prognostic factor (eg, ethnicity). (A) In simple randomization, group assignment can be determined by a flip of a coin, roll of a die, random number table, or computer program. With small sample sizes, there can be an unequal number in each treatment arm or an unequal distribution of prognostic factors, or both. Note that the numbers in the treatment arms are unequal and the prognostic factor is unevenly distributed between the two. (B) In blocked randomization with uniform or equal blocks, randomization occurs in groups (blocks), and the total sample size is a multiple of the block size. In the figure, balls were randomized in blocks of six. Note that the number in each treatment arm is equal, but the prognostic factor is not equally balanced between the two. (C) In blocked randomization with varied blocks, the size of the blocks changes either systematically or randomly to avoid predictability of treatment assignment. In the figure, the first block has eight balls and the second block has four balls. Again, the number in each treatment arm is equal, but the prognostic factor is not balanced

between them. (D) In stratified blocked randomization, the total sample is divided into one or more subgroups (strata) and then randomized within each stratum. In this example, the sample population is divided into black and white balls and then randomized in blocks of six. There are an equal number of balls in both arms, and the distribution of the black and white balls is balanced as well.

Sample Population (n=12)
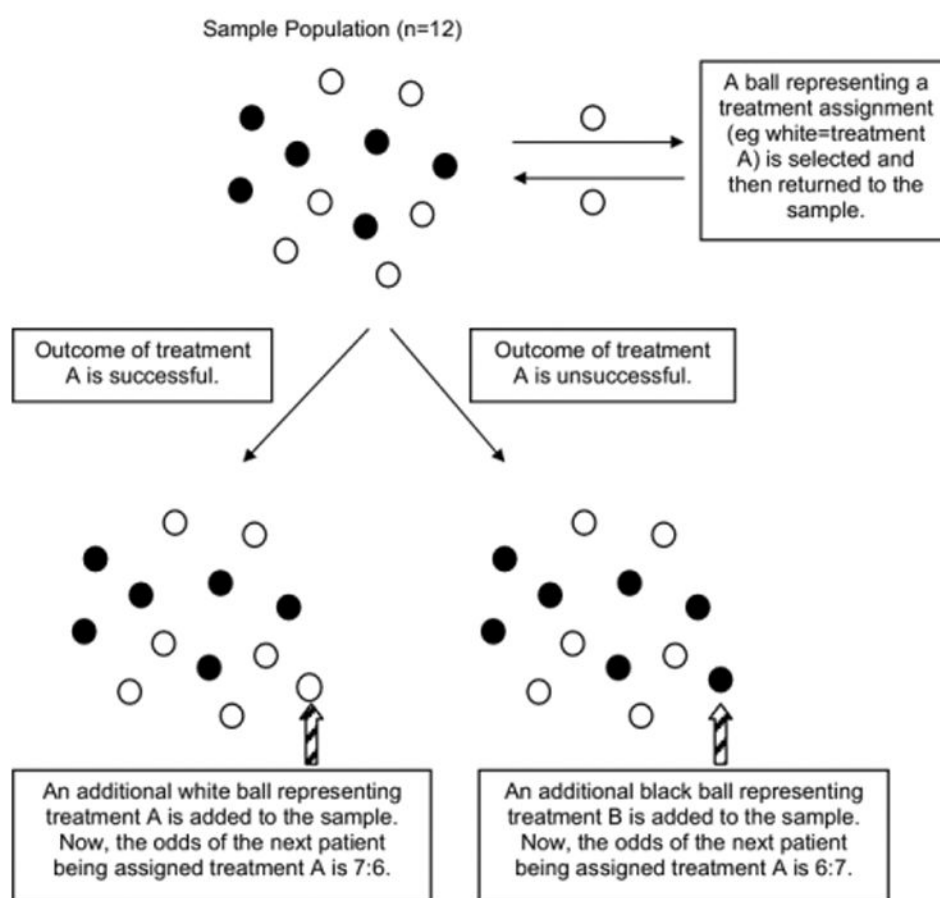
Figure 2.
Adaptive randomization: play the winner randomization rule. In this example, the color of the ball represents the treatment assignment (white = treatment A, black = treatment B). A ball is selected and then replaced. Based on the outcomes of the treatment selected, a ball representing the same or opposite treatment is added to the sample. The rule repeats itself. If the two treatments have similar outcomes, then there will be an equal distribution of balls at the end of the trial. If one treatment has substantially better outcomes, then there will be more balls representing that treatment. Patients entering the trial have a better chance of having an effective treatment and less of a chance of having an ineffective or harmful treatment.

Table 1

Examples of Randomization

Simple randomization				
21	46	67	64	29
79	64	60	55	67
86	72	36	100	95
43	61	25	50	53
Random allocation of treatment				
B	A	B	A	B
B	A	A	B	B
A	A	A	A	B
B	B	B	A	B