Published in final edited form as:
J Soc Integr Oncol. 2006 ; 4(4): 194–198.

How to randomize

Andrew J. Vickers[1,2]

1 *Department of Medicine, Memorial Sloan-Kettering Cancer Center*

2 *Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center*

Abstract

Randomized trials are an important method for deciding whether integrative oncology therapies do more good than harm. Many investigators do not pay sufficient attention to randomization procedures and several studies have shown that only a fraction of trial reports describe randomization adequately. The purpose of randomization is to prevent selection bias: randomization procedures must therefore ensure that researchers are unable to predict the group to which a patient will be randomized until the patient is unambiguously registered on study; moreover, researchers must be unable to change a patient's allocation, after they are registered. The use of telephone randomization, and opaque envelopes have been suggested as good randomization method, but both can be subverted. Randomization should be conducted either by a pharmaceutical company, which sends blinded medication to the hospital pharmacy, or by a secure, password protected database system. Computer randomization can easily incorporate extensions of randomization, such as blocking, stratification and minimization, that can help ensure balance between groups.

Introduction

Given the importance and prevalence of randomized trials, it is a continual surprise to me how little most researchers understand about their distinguishing feature: randomization. In my experience, it is rare indeed that a trial I review either as grant applications or as submitted papers includes an adequate description of randomization in sufficient detail. This is not a problem specific to integrative oncology: in one systematic review, Pildal et al. reported that just 16% of randomized trials adequately report an appropriate randomization procedure(1).

In this didactic paper, I will outline the basics for effective randomization and reporting thereof. My focus will be on practicalities and I will therefore not include mention of trials of oral or injected agents, where randomization is typically conducted by a pharmaceutical company, with blinded medication sent to the pharmacy at each study site. I will start this paper by discussing the "why" of randomization, as a preface to explaining the "how".

Why randomize?

To understand why we undertake randomized trials, it is useful to consider the alternative, the observational study. As a simple example, imagine a study to compare death rates in patients receiving surgery for coronary artery disease with those receiving medication alone. In an observational study, in which we simply compare a group of patients who received surgery with a group that did not, the two groups will typically differ in several important ways. A surgeon is unlikely to treat a patient who has very severe disease, has other serious medical problems or who is obese. Moreover, a surgeon may choose to forgo treating a patient who is

Corresponding author: Andrew J. Vickers, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, NY, NY 10021, vickersa@mskcc.org, T: 212 639 6556, F: 212 794 5851.

not compliant with medication or advice on smoking cessation. The likelihood that a patient is treated surgically may also be related to financial status, which in turn is a predictor of survival. In sum, even if surgery had no effect, we might expect patients in the surgery group to do better due to factors known at the time of surgery (such as severity of disease) and events that happen after surgery (such as death due to other causes). The difference between patients is related to patient *selection:* the surgeons get to choose their patients and so surgery treated patients are different from patients not receiving surgical treatment.

The aim of randomization is to ensure a fair comparison between like groups. Randomly selecting patients for treatment means that groups will typically be similar for important prognostic factors, and so would likely have similar outcomes if they received identical treatment. Thus any observed differences in outcome are likely to result from differences in treatment.

Simple methods of randomization (and how to subvert them)

A first-time researcher once told me that she had gone to unusual lengths to ensure good randomization for her study: she had asked a statistician to generate a randomization list using sophisticated software. She'd then tacked the list to the wall in the nurses' office so that when anyone entered a patient to trial they would know which group to put them in. The obvious downside to this method is that the list on the wall allows anyone to know a patient's allocation *before* they are registered on trial. Let's go back to the surgery example. Say that, on a given day, the surgeon has seen the randomization list and knows that the next patient will be randomly assigned to the surgery group. In walks a patient who meets the eligibility criteria for the trial but who the surgeon feels, on balance, is probably not going to do that well. Accordingly, the surgeon advises against surgery and does not raise the study with the patient; the next patient, however, is a great candidate for surgery, and although he is rather wary of research, the surgeon pressures him to consent. In other words, the surgeon is able to subvert randomization and select which patients get which treatment, the very problem randomization was designed to avoid.

This is not merely a theoretical problem: Ken Schulz has documented various ways in which clinical researchers have attempted to subvert randomization, including bright light sources to reveal allocations in sealed envelopes and even ransacking the principal investigator's office to find the randomization list(2,3). My own favorite anecdote illustrates another aspect of randomization that must be protected. In a trial of social support for women with at-risk pregnancies, randomization was implemented by having a researcher pick a marble from an urn, blue for treatment, white for control; in practice, however, if the research nurse picked out a white marble for a patient she felt really needed social support she simply replaced it and selected another one. As a result, baseline distress scores were higher in women in the treatment group: this would make the trial more likely to found a difference between groups, even if none existed, due to regression to the mean.

In sum, the two fundamental characteristics of randomization are:

1. Researchers must be unable to predict the group to which a patient will be randomized until the patient is unambiguously registered on study
2. Researchers must be unable to change a patient's allocation after they are randomized

A trial meeting these criteria is said to have adequate "allocation concealment". This term is often mistakenly confused with blinding, which concerns knowing a patient's allocation during the study. Not all trials can or should be blinded – a trial of radical prostatectomy is an obvious example – however, all trials can and should have full allocation concealment.

Inadequate solutions to allocation concealment

The two most widely discussed practical methods for ensuring allocation concealment are telephone randomization, and opaque envelopes. Both of these methods seem suspect to me. The idea behind telephone randomization is to separate the researcher consenting the patient from the researcher conducting the randomization. This is on the grounds that having the same individual consent and randomize can introduce selection bias, such as in the surgery example above. But use of a telephone does not of itself protect against selection bias. There is nothing to stop the surgeon becoming friendly with the researcher in charge of randomization and getting information on future randomizations: “Great, this one is randomized to surgery. So … what will the next patient get?”

Use of opaque envelopes is similarly vulnerable to subversion. Several authors (see, for example, Doig and Simpson(4)) have described in great detail the practicalities of creating what is known as SNOSE schemes (Sequentially Numbered Opaque Sealed Envelopes). This involves numerous complexities such as putting silver foil in the envelope, so that a researcher cannot use a light source to see numbers inside, and carbon paper, so that patient identifiers written on the outside of the envelope will show up inside. There are also complicated rules such as requiring that trialists complete an audit sheet before opening the envelope. Nonetheless, there is little to stop an investigator from subverting randomization. The simplest ruse is simply to open the envelope and look inside. If the allocation is the one the researcher wants, he or she continues and registers the patient to trial. If not, they simply wait for another patient to come along.

Practicalities of randomization

In my view, there is only one way to randomize patients and that is to use a computer. Now some researchers have described this as problematic because, unlike envelope randomization, computer randomization “require[s] the use of specialized technology”(4). This rather makes me fear for the future of clinical research: my 11-year old neighbor is programming a website; my clinical research colleagues are wrapping bits of paper in silver foil. Anyone who considers a computer to be “specialized technology” should seriously reconsider whether they really want a career in scientific research.

Programming a randomization scheme is relatively straightforward: I have no formal programming training but I have created randomization software for several trials with no real difficulty using Filemaker Pro, a commercially available database package. One simple approach to randomization requires two separate databases: a “patients” database that lists basic information such as patient name, hospital number, contact details and so on, and a “randomization” database that holds data on which patients have been registered on trial along with their treatment allocations. This database also includes programming code for randomization. The key point is that the “patients” database is accessible to any researcher whereas the randomization database is password protected so that it is accessible only by the principal investigator and a nominated, anonymous individual, such as a computer technician. To use the surgery trial as an example, the surgeon would access the “patients” database, type in the patient’s name and details and then hit an icon to randomize the patient. This would lead to a dialog box asking the surgeon to confirm that the patient is eligible. After the “okay” button is pressed, the patient’s information is automatically sent through to the randomization database, where randomization takes place. The result of randomization – patient allocated to surgery or medicine alone – is stored and sent back to the “patients” database where the surgeon can see it. As the surgeon cannot access or modify the secure, password protected randomization database, there is no way to predict a patient’s allocation before registration to trial, or change it afterwards.

"Programming code for randomization" sounds very complicated, but it is actually pretty straightforward for a computer technician or statistician. The simplest form of randomization is the coin flip: the computer randomly generates a number between 0 and 1, and assigns the patient to treatment if the number if less than 0.5 and to control otherwise. There are also several more complex randomization schemes.

Blocking and stratification

Simple randomization has the disadvantage that it is quite possible to have important differences between groups simply by chance. For example, in a 40 patient trial, there is about a 15% chance that there will be an imbalance greater than 24 patients in one group and only 16 in the other. Simple randomization can also lead to imbalances on important prognostic factors. Imagine that patients with bone pain tended to do better than those with neuropathic pain and that each type of pain was equally represented in the patient population. In a 40 patient trial, simple randomization would lead to one group having at least a 50% higher or lower incidence of bone pain about 20% of the time.

An alternative to simple randomization is to use what is known as "blocked randomization". Instead of randomizing each patient individually, this scheme randomizes several patients at a time in such a way as to ensure that equal numbers are allocated to each group. For example, if the block size is four, we randomize four patients at a time ensuring that two patients are allocated to the treatment group and two patients to control. As it happens, there are six different possible ways we could randomize four patients equally to two treatments (see table 1). So we ask the computer to randomly select one of the six types of block at the time when the first patient is randomized and this determines the treatments received by the first four patients. A second block is randomly selected when patient 5 is registered in order to create allocations for patients 5–8.

One obvious problem with blocked randomization is that it can lead to allocation becoming unconcealed. For example, if our cardiac surgeon knows that the block size is four, or figured this out halfway through the trial, it is a simple matter to work out that if, say, two patients have just been allocated to surgery, the next patient will be in the control group. Hence many blocked randomization schemes use a two-stage process: the computer first randomly selects whether the block will contain 4, 6, 8 or 12 patients and then randomly chooses one of the possible different blocks for the selected size.

To ensure that groups are similar for an important prognostic feature, such as type of pain, blocking should be "stratified". As a simple example, imagine that the first patient on the pain trial had bone pain. The computer randomly selects the number 2 and so this patient is allocated to the treatment group (see table 1). The next two patients also have bone pain and so are allocated to control and treatment respectively according to their block (table 1). The fourth patient presents with neuropathic pain and so a new block needs to be selected. The computer randomly selects block 1 so this patient is randomized to the treatment group. Extending stratification to more than one prognostic factor is straightforward: where randomization is stratified by both type of pain and gender, separate blocks are selected for each possible "cell": men with bone pain, men with nerve pain, women with bone pain and women with nerve pain.

Minimization

One difficulty with stratified randomization comes when the investigators have identified a large number of different prognostic factors. In the pain trial, such factors might include type of pain (neuropathic / bone), gender, baseline intensity of pain (mild / moderate / severe), current treatment (opiates / no opiates). This is a total of $2 \times 2 \times 3 = 12$ different "cells". Now in a 40 patient trial, this is an average of only 3 patients a cell and by chance there will be

several cells with only a single patient. This can lead to chance imbalances between groups, the very problem stratification is meant to avoid.

The solution is to use what is called a "minimization" or "biased coin" design. Using the pain trial as a simple example, let us imagine that the trial has been running for several months; 8 of the 12 (67%) patients in the treatment group had neuropathic pain compared to 7 of the 14 (50%) patients randomized to control. Now a new patient is randomized and this patient has neuropathic pain. In an ideal world, you would want this patient to be randomized to the control group to ensure balance between groups as to type of pain. So instead of giving the patient a 50:50 chance of being in treatment or control, randomization uses the "biased coin": the computer selects a random number between 0 and 1 and if this is less than, say, 0.6 (rather than 0.5 as usual), the patient is assigned to control. But note that the groups are not just imbalanced in terms of proportion of patients with neuropathic pain, but in the total assigned to each group: you really want the new patient assigned to the treatment group so that there would be more equal numbers of patients in the two arms.

The solution is to use a scoring system. This is easiest to explain by illustration. Imagine that the current distribution of patients halfway through a pain trial was as shown in table 2 and that the next patient registered eligible on trial was an opiate-using woman with moderate neuropathic pain. Using the scoring system in table 3, she would be scored −0.5 for type of pain (there are a higher proportion of patients with neuropathic pain in the treatment group), +0.5 for baseline pain severity (fewer patients with moderate pain in the treatment group) and +0.25 for opiates (there are fewer opiate users in the treatment group). Men and women are distributed equally, so the patient is scored 0 for gender. As for total number of patients, there are 2 fewer in the treatment group so she scores + 0.25 × 2 = 0.5 (if there had been, say, 3 more patients in the treatment group, should would have scored −0.25 × 3). The total score for this woman is −0.5 + 0.5 + 0.25 + 0.5 = 0.75. To convert this to a probability, use the formula: $e^{score} \div (e^{score} + 1)$. In this case, e^{score} is 2.12 and so the probability is 0.679. So for the woman, the computer would randomly select a number between 0 and 1, and assign to treatment group if it was less than 0.679 and to control otherwise. Note that this is just one approach to a scoring system for minimization, there are numerous others.

One general rule of thumb about both minimization and stratification is to keep the number of variables to the minimum possible. A common mistake is to want a randomization scheme to ensure that groups are balanced for all possible baseline characteristics. This might make the trial's "table 1" look good but it can causes a number of problems. First, stratifying on a very large number of characteristics can lead to balance on trivial variables but imbalance on more important ones. Moreover, for some complex statistical reasons(5) any variable used to stratify randomization has to be included as a covariate in analysis. It is generally undesirable to have a statistical model with a very large number of covariates. Accordingly, trialists are advised to include only those variables known, or highly likely, to have an important impact on outcome. In a trial of pain, this will undoubtedly include baseline pain score; inclusion of gender would be harder to argue; use of age as a stratifying variable is unlikely to have an important impact.

Conclusion

The preceding two sections of this paper have gone into some detail about the underlying mathematics of randomization. But randomization is not primarily about complicated statistics. At its heart randomization is very simple: make sure that no-one can predict a patient's assignment before they are registered on trial and make sure that no-one can change it afterwards. In randomized drug trials, this is generally straightforward, as a pharmaceutical company sends blinded medication to the hospital pharmacy. Where researchers need to implement randomization themselves, there is really no alternative to use of a secure, password-

protected database: envelope randomization should really be sent the way of the slide rule and the Rolodex.

On a final point, “justice not only needs to be done, justice needs to be seen to be done”. It isn’t enough just to implement a good randomization system, you have to make sure that you explain it carefully in any grant application and published report.

Acknowledgements

This research was funded, in part, by a P50-CA92629 SPORE from the National Cancer Institute.

References

1. Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman DG, Gotzsche PC. Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study. Bmj 2005 May 7;330(7499):1049. [PubMed: 15817527]
2. Schulz KF. Subverting randomization in controlled trials. Jama 1995 Nov 8;274(18):1456–8. [PubMed: 7474192]
3. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. Lancet 2002 Feb 16;359(9306):614–8. [PubMed: 11867132]
4. Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. J Crit Care 2005 Jun;20(2):187–91. [PubMed: 16139163]discussion 91–3
5. Kalish LA, Begg CB. The impact of treatment allocation procedures on nominal significance levels and bias. Control Clin Trials 1987 Jun;8(2):121–35. [PubMed: 3608506]

Table 1

Block randomization with a block size of four

There are six different ways in which four patients can be assigned equally to two different treatments.

	Allocation			
	Patient 1	**Patient 2**	**Patient 3**	**Patient 4**
1	Treatment	Treatment	Control	Control
2	Treatment	Control	Treatment	Control
3	Treatment	Control	Control	Treatment
4	Control	Control	Treatment	Treatment
5	Control	Treatment	Control	Treatment
6	Control	Treatment	Treatment	Control

Table 2

Baseline features of accrued patients during a hypothetical pain trial

The table gives the characteristics of each group early in the trial.

Variable	Treatment group	Control group
Total number of patients	12	14
Neuropathic pain	8 (67%)	7 (50%)
Bone pain	4 (33%)	7 (50%)
Men	6 (50%)	7 (50%)
Women	6 (50%)	7 (50%)
Severe pain at baseline	7 (58%)	8 (57%)
Moderate pain at baseline	5 (42%)	6 (43%)
Use opiates	10 (83%)	12 (86%)
No use of opiates	2 (17%)	2 (14%)

Table 3

Example of a scoring system for minimization

The choice of scores is at the discretion of the investigators and depends on a judgment as to which factors are most strongly prognostic and therefore the most important to keep balanced between groups. For example, baseline pain is likely to affect pain at the end of the trial more than gender, and so baseline pain is given a higher score.

Variable	Score
Type of pain	0.5
Gender	0.25
Use of opiates	0.25
Baseline pain	0.5
Total number of patients	0.25 × difference