



# Randomization in clinical studies

Chi-Yeon Lim<sup>1</sup> and Junyong In<sup>2</sup>

<sup>1</sup>Department of Biostatistics, Dongguk University College of Medicine, <sup>2</sup>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 com-

parability 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<sup>1)</sup>, 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>

---

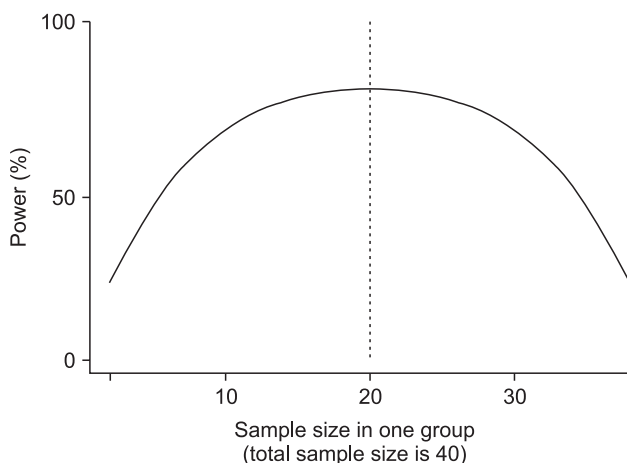
<sup>1)</sup><http://www.consort-statement.org/>

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,<sup>2)</sup> 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).<sup>3)</sup> In addition, it would be difficult to consider a 25/15 or 30/10 subject allocation as aesthetically balanced.<sup>4)</sup> 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%),<sup>5)</sup> 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].<sup>6)</sup> 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.

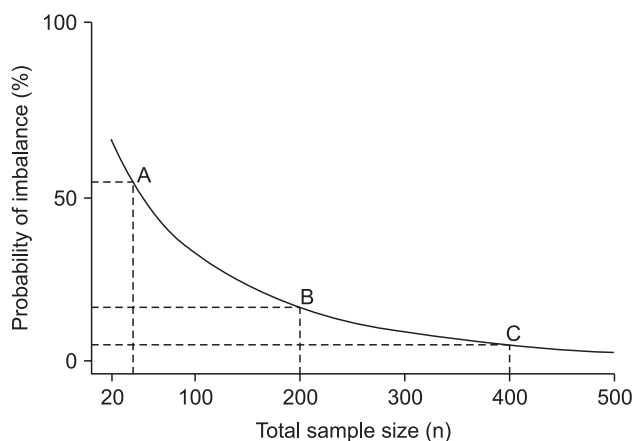
<sup>2)</sup>However, since the results and process of randomization cannot be easily recorded, the audit of randomization is difficult.

<sup>3)</sup>Two-tailed test with difference ( $d$ ) = 0.91 and type 1 error of 0.05.

<sup>4)</sup>“Cosmetic credibility” is often used.

<sup>5)</sup>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%.

<sup>6)</sup>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.<sup>7)</sup> 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.<sup>8)</sup> 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.<sup>9)</sup>

## Restricted randomization for unbalanced allocation

Sometimes unbalanced allocation becomes necessary for eth-

ical or cost reasons [9]. Furthermore, if you expect a high drop-out 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

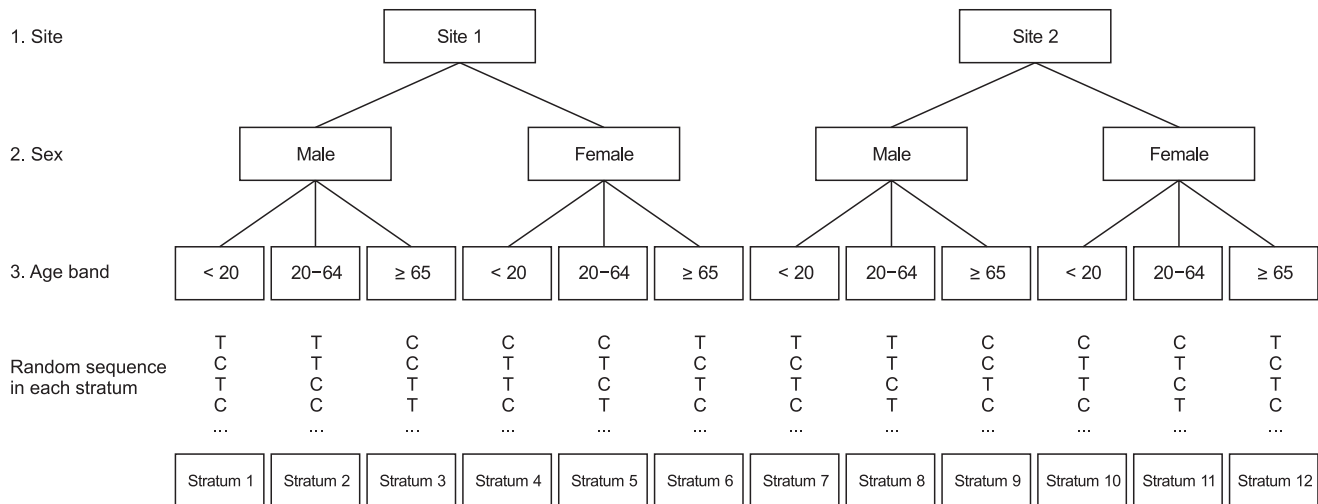
Some studies have prognostic factors or covariates affecting the study outcome as well as treatment. Researchers hope to balance the prognostic factors between the study groups, but randomization does not eliminate all the imbalances in prognostic factors. Stratified randomization refers to the situation 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

<sup>7)</sup>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.

<sup>8)</sup>The block size of 2 applies mainly to a study of allocating a pair at the same time.

<sup>9)</sup>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<sup>10</sup> 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<sup>11</sup>

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].<sup>12</sup> 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

<sup>10</sup>As the number of strata increases, the imbalance increases due to various factors. The details are beyond the scope of this study.

<sup>11</sup>This paragraph introduces how to allocate “two” groups.

<sup>12</sup>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.

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

**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.

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. *Ecancermedalscience* 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`<sup>1)</sup> 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-

---

<sup>1)</sup>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

The 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.