Methodology

Randomization: Techniques and Software-aided Implementation in Medical Studies

Mohammad Asghari-Jafarabadi¹, Homayoun Sadeghi-Bazargani³⁻⁶⁻*  
¹Department of Statistics and Epidemiology, Tabriz University of Medical Sciences, Tabriz, Iran,  
²Road Traffic Injury Research Center, Department of Statistics and Epidemiology, Tabriz University of Medical Sciences, Tabriz, Iran,  
³WHO Collaborating Center on Community Safety Promotion, PHS department, Karolinska Institute, Stockholm, Sweden

Abstract

In this methodology review we have briefly discussed both the theory and practice of randomization in clinical trials. Necessity, principals, definitions and classifications along with the current contraversies are presented. Practical notes on the methods and procedures of randomization in clinical research with with particular focus on software-aided randomization is provided and the review is ended by a section on reporting the randomization in clinical trials.

Introduction

Randomized controlled trial (RCT) benefits the most reliable research methodology for interventional studies and seeks to lead the scientific ideals onto clinical experimentation. It is the sole effective method known to be capable of controlling selection bias and controls confounding without adjustment [1]. Randomization and blinding are keystones of RCTs turning them into the gold standard of original research evidence [2]. Randomization, in general, is defined as the process of making something random. However, a random procedure is a series of random numbers describing a procedure whose products do not pursue a deterministic model but obey a progress described by probability distributions. For example, a random sample of individuals from a population refers to a sample where every individual has a known probability of being sampled. In interventional research, however, randomization has a specific meaning being a technical word different from random sampling.
**Random assignment versus random allocation**
These two words are usually used alternatively and nearly carry the same meaning.

**Randomization versus random allocation**
Randomization is a more general term that is defined to have three steps, the last of which is the act of random allocation often by the clinician. These two terms are many times used alternatively maybe due to the fact that randomization is finalized when allocation is done. However, from another point of view, random allocation and random selection could also be considered as two alternatives of applying randomization [6].

**Random sampling versus randomization**
Random sampling is used to define any procedure to select a sample of study units from a larger group or population based on chance. The purpose of random sampling is to provide a sample representative of the larger group or source population and improve generalizability of the study results. In clinical trials, however, the random sampling process is usually impractical or unethical to apply. Patients are, most of the times, not available in known groups of known size to be randomly sampled. We cannot wait for the patients to accumulate and form a group for random sampling. The process of most randomized clinical trials is such that patients are enrolled in a consecutive manner over time as they attend a clinic or are admitted to a hospital. Moreover, the selection of clinical settings, physicians and researchers engaged in process of allocating patients doesn’t follow a random procedure either. Therefore, usually the random allocation is done on a convenience sample of patients and is defined as a process that units are randomly assigned to conditions/ interventions but it doesn’t essentially mean randomly enrolling the units into the study.

**Random selection versus randomization or random sampling**
Random selection is a term with more general application sometimes used as a synonym both for random sampling and for random allocation in different contexts [7]. Nevertheless, some authors have also their own perception of the applicability of the term random selection. For instance, Shein-Chung and Jen-Pei,L state that randomization can be performed either by random selection or by random allocation for methods of complete and permuted-block randomization [6].

**Why randomization?**
Randomization is of essential importance in medical studies. In controlled clinical trials, randomization is often used to control conscious or unconscious bias in the assigning patients to treatment groups [3]. A good study diminishes inconsistency of the assessment and offers a balanced assessment of the intervention staying away from confounding by other factors. Randomization assures that each patient has a known, usually identical, probability of getting any of the treatments under study; producing analogous intervention groups which are similar in all central features apart from the intervention that each group receives. It also offers a foundation for the inferential statistical methods used to analyze the data. Benefits of randomization include; Minimizing the bias mainly selection bias; Balancing arms considering factor affecting identified or non-identified; Forms basis for statistical tests, a basis for an assumption-free statistical test of the equality of treatments.

Let’s assume that there are two treatments A and B needing to be compared for their efficacy in a clinical trial study. Further, assume that the patients for Treatment A are all to use a special type of medication and the patients for Treatment B are all not to use that type of medication. If there would be a difference in the primary outcome between the two groups in the analysis of the baseline measurements, might we then definitely interpret this difference as a treatment effect? Under this situation, the answer to the question would clearly be “No”. In this case, the effect of treatment might be confounded by medication, or both could affect the outcome and their pure effects cannot be determined. In any study due to confounding the (identifiable or unidentifiable) subject’s characteristics that may be related to the outcome under study, with treatment could cause a bias between the groups which is quite separate from any treatment effect.

It is expected that in the reasonable and long run, random allocation would balance the subjects characteristics among treatment groups and in doing so it is expected that irrelevant effects would be eliminated and permit the treatment effect to be recognized purely. In the exact words, in the random allocation it is expected that the effect of possible confounders would be minimized, results in to a fair comparison among treatments. Additionally, the assumption of randomness in the inferential statistical methods utilized to assess the treatment effect would be fulfilled for reliable findings. The random allocation will maintain this assumption and accordingly valid and defensible data analysis and results would be expected.

Since randomization is highly necessary based on above-mentioned points and its implementation needs full care to be taken by researchers, it is crucial for researchers to be familiar with the techniques of randomization. Nevertheless, it should be taken into account that randomization doesn’t always guarantee the balanced distribution of the potential confounders unless appropriately applied on large sample size. However, we could be optimistic of such balance to happen in phase III 2-arm clinical trials of the minimum size of 1000-2000 or above. Therefore, it is essential to assess the effectiveness of randomization in fulfilling balanced distribution of known confounders at the analysis phase.

**Principles for randomization**
1. Minimize predictability: Each participant has the same probability of participating in the study. Randomization is conducted by a probability system such that neither the contributors nor the researcher will know in advance which how the allocation would be.
2. Balance: Study groups are alike in size and formation and identical in all main relevant features.
3. Feasibility: Straightforward and feasible for investigator/staff to apply.
4. Avoid sequence effects: In interventional studies the
sequence of two treatments, A -> B may not have the same consequences than B -> A [8].

Techniques of random allocation
Regardless of the general classification of randomization into three categories of the complete randomization, the permuted-block randomization, and the adaptive randomization, for space limit and practical reasons here in this article, we just focus on commonly encountered terms as simple randomization, block randomization, stratified randomization and minimization. The latter term could hardly be considered as a real type of randomization. Each of these techniques will be briefly explained.

Simple randomization
Although very few clinicians may prefer to toss a coin in their clinics or hospitals, the simplest and traditional way of randomization would be to toss a coin to decide assignment of a patient to a study group. In simple randomization the only restriction enforced on randomization sequence is the number of patients to be allocated and the allocation ratio. Assume that there are three treatment groups and within each group 10 patients would be allocated using simple randomization procedure. For this purpose a sequence of random numbers between 1 and 30 would be generated using a random number generation system (such as a computerized random number generation system), for the three groups say the numbers in Table 1. In this Table, persons with id number 2, 4, 7 … and 30 would be allocated to group 1, for example.

This procedure is rather easy to arrange, preserves the prearranged design parameters. When the sample size is fairly large, simple randomization is expected to generate roughly balance sized treatment groups though this is not assured and such method may cause errors especially in small sized clinical trial studies [9]. Simple randomization decreases bias by balancing some features that have not been considered in the design of study such as a subgroup of subjects with special characteristics expected to influence treatment efficacy.

Although the simple randomization is too easy to implement, there are problems with that in some study situations. For any possible reason such as exhausting the budget or not finding enough eligible subjects, this type of randomization may encounter the problems. For example, as can be seen in the Table 1, if the study is going to end after 18 subjects get recruited, there would be unequal numbers of subjects in each group (7 persons in group 1 and 3 and 5 persons in group 2). This is because the subjects are usually recruited to the study sequentially. An additional case is that some characteristics such as sex could be an important prognostic variable. Possible inequality in allocation or a systematic bias may happen if gender distribution is not taken into account in some way.

Block randomization
Block randomization procedure is usually utilized in the situations where sample sizes for the study groups are to be identical or roughly identical. The procedure includes allocating the recruited subjects in the blocks of definite size. In this case it is assured that the subjects within blocks are allocated to defined study groups with random order of subjects inside each block.

The concept of block randomization will be illustrated using blocks of size 4 which can be used in the studies with two arms (A and B). There are six different permutations of the AABB letters is as follow:

1. AABB, 2. ABAB, 3. ABBA, 4. BAAB, 5. BABA, 6. BBAA.

To assign the participant in study groups, one of these six different blocks would be randomly selected for each group of four participants that are recruited. To randomly arrange the blocks, a list of random numbers is generated in the next step such as “263438181475359” which was generated by rand function in MS Excel software package. Since there are only six different permutations, numbers above 6 would be removed reaching the “26343114535” sequence.

At the next step, the block would be chosen based on this sequence. For example the first twenty subjects who would be randomly allocated into two groups, the random block allocation is as follows:

Therefore the allocation of subjects would be as in following Table:

<table>
<thead>
<tr>
<th>Study groups</th>
<th>2</th>
<th>6</th>
<th>3</th>
<th>4</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABBA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BAAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BABA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BBAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table 2, for the first permutation showed in the first line, the first, second, third and fourth subjects are randomly allocated to A, B, A and B study groups. As described in simple random sampling methodology, the group size may not equally proceed and the group imbalance would be a major concern when using simple random allocation. Simple sequential randomization
may be considered as a special case of block randomization with blocks of size 1. However, the block randomization almost generates the groups with identical size and without losing the rule of random selection. It is obvious in Table 2 that if the study stops before the determined sample size is reached, the maximum discrepancy between groups’ sample size would be 1. The length of the blocks used in block randomization is dependent on several factors such as study design and efficiency, study interruption likelihood and targeted allocation concealment rate. For instance, in clinical trials with blocks containing only one allocation for each of the two interventions in a 2-arm parallel design, the maximum allocation concealment rate cannot exceed 50% through the study exposing the study to noticeable risk of selection bias generally or even detection bias if the allocating clinician is the same as outcome assessor. At the same time if a study using long blocks interrupts too early, a higher rate of allocation imbalance could be expected leading at least to some loss of statistical power for the given resources used through the study. Additionally, the block size may randomly change over the study to maintain the investigator’s unbiased estimate of outcome measures.

Software applications

Microsoft Excel may be the most widely available software package for generating random allocation sequence. Here we provide a practical simple example for block randomization using Microsoft Excel 2010. The same process could also be run for its earlier and later versions.

Suppose you want to allocate 20 patients either to a new drug treatment or placebo through a one to one allocation weight. Open a blank Excel sheet. Fill in from 1 to 20 in the first column (column A). You may easily do this for larger sets using the drag options of Microsoft Excel. Leave the second column blank (Column B). Fill in the first 10 rows in column C with word “Drug” and the remaining rows with the word “Placebo”. Put the pointer in the first row cell of the column D (D1) and type in the function “=rand()”. This will create a random number in D1. Copy this command to the remaining 19 rows by using the mouse drag down from where shown by the red arrow in Figure 1. This will create a list of random numbers in column D. The last step would be to sort the random numbers in column D while expanding the sorting to column C. Now the random allocation list is generated. You may extend this practice to more than two interventions or do it with various allocation ratios. In this example, we created the random allocation sequence for one block. The process could be repeated to create several blocks of given sizes. However, for more complex randomization designs randomly permuted blocks and random sequences are recommend to be generated using statistical software packages such as Stata, R, SAS and other similar packages by means of available modules or through programming.

There are also some commercial software packages or web-based applications available for generating random sequences. The randomization could be helped using PROC PLAN in SAS, block random in R and RALLOC user defined module in Stata. To run the above example in SPSS following is a sample syntax in SPSS. Open your SPSS software and open a new syntax file through the menu path: File/new/syntax.

Type in the following code. You may change the parameters to fulfill your needs.

input program.
loop #I=1 to 20.
if (#I<11) group= 0.

Figure 1. Generating random allocation sequence using Microsoft Excel (Excel 2010 in this example).

A). You may easily do this for larger sets using the drag options of Microsoft Excel. Leave the second column blank (Column B). Fill in the first 10 rows in column C with word “Drug” and the remaining rows with the word “Placebo”. Put the pointer in the first row cell of the column D (D1) and type in the function “=rand()”). This will create a random number in D1. Copy this command to the remaining 19 rows by using the mouse drag down from where shown by the red arrow in Figure 1. This will create a list of random numbers in column D. The last step would be to sort the random numbers in column D while expanding the sorting to column C. Now the random allocation list is generated. You may extend this practice to more than two interventions or do it with various allocation ratios. In this example, we created the random allocation sequence for one block. The process could be repeated to create several blocks of given sizes. However, for more complex randomization designs randomly permuted blocks and random sequences are recommend to be generated using statistical software packages such as Stata, R, SAS and other similar packages by means of available modules or through programming.

There are also some commercial software packages or web-based applications available for generating random sequences. The randomization could be helped using PROC PLAN in SAS, block random in R and RALLOC user defined module in Stata. To run the above example in SPSS following is a sample syntax in SPSS. Open your SPSS software and open a new syntax file through the menu path: File/new/syntax.

Type in the following code. You may change the parameters to fulfill your needs.

input program.
loop #I=1 to 20.
if (#I<11) group= 0.
if (#I>10) group= 1.
comput x=normal(1).
end case.
end loop.
end file.
end input program.
sort cases by x.
string Intervention (A8).
recode group (0='Placebo') (1='Drug') into Intervention.
variable labels Intervention 'Intervention name'.
print table/casenum Intervention.
execute.

**Stratified randomization**

Stratification on random allocation is used when there is a source(s) of variation which may affect the results in a way not aimed by the researcher. The accuracy and reliability of the estimation of primary trial outcomes for assessing the effect of intervention can be affected by the heterogeneity caused by some covariates such as age, sex, geographical location, and underlying disease severity. The methodology is used to apply the results to some targeted patient population [6]. Hence, a stratification variable is a categorical (or discretized continuous) covariate which divides the participants in relation to its categories. For example, if in the allocation performed in the Table 2, the BMI categories (or any other prognostic factor) could affect the results, a solution would be to stratify on BMI categories. Using this procedure, the groups are allocated within each stratum using any of the aforementioned procedures. For example, the allocation presented in Table 3 could be constructed within various strata of the BMI categories. As can be seen, within for example the < 18.5 category, the block randomization results in allocation of first, second, third and fourth subjects in A, B, A and B groups respectively. In this case the number of subjects allocated within each block could be equal or unequal; for example the number of subjects within the <18.5 BMI category was 4 persons while it was 8 persons for the 25.1-30 BMI category. However, stratification assured that the number of participants within a given BMI category would be equal in study groups (this may clarify the difference between blocking and stratification).

**Minimization**

The rationale behind this procedure is to keep the balance of prognostic factors among groups. In minimization procedure, the first or early sequences of patients may really be randomly allocated to study groups, however, for the following patients, the allocation would be conducted in such a way that minimizes the imbalance among groups at that assignment through a non-probabilistic procedure.

For instance, consider a situation where there are two prognostic factors; BMI categories and sex which may confound the results if they distribute in an imbalance way among groups. Suppose that the 12 subjects recruited in the study are allocated as in Table 4.

In the table above there are imbalanced BMI categories and Sex distribution between groups and this may affect the results. Minimization procedure seeks to minimize this effect via subsequent allocation. For example if the next recruitment would be a male with 25.1 – 30 BMI category, to make closer the groups to balance, it will be better to allocate this subject to group A.

Therefore in this procedure, the new subjects would be allocated to the group that makes the group more balanced with respect to those prognostic factors. Minimization is not completely random and is a method with implementing limitations in randomizations. Actually, it is “a pseudorandom method of allocating intervention to subjects to try to balance the distribution of covariates across the treatment groups” [10].

Pure minimization even could be considered as a totally deterministic method [6]. Regardless of the current controversies in classification or application of the technique, minimization is considered a valuable method in clinical trial methodology [3,11-15].

**Allocation Concealment and Blinding**

In some cases of randomization when the researcher is able to guess the next subject’s group allocation, the study may be prone to a selection bias and this is because the randomization is inadequately executed. A convenient matter in conducting trials is allocation concealment, which means to conceal the next allocation group form the investigator. In the better words, this procedure shields those who admitted subjects to a trial from knowing the subsequent allocations. For example by coding the groups and not using their real name the allocation concealment and masking would occur and this is totally different from blinding. Additionally the allocation concealment seeks to prevent selection bias and it is always applicable while the blinding seeks to prevent observation-performance bias and it is not always applicable or ethical or relevant [16-18].
concealment minimizes the selection bias by shielding the randomization code before and until the interventions are given to trial participants but blinding helps in avoiding bias by protecting the randomization code after the interventions have been applied [18].

**Reporting the randomization**

Randomization is such an important issue in reporting clinical trials that several items in CONSORT checklist 2010 explores the reporting of it. The items attributed to randomization in CONSORT 2010 checklist include; Item 8a. Method used to generate the random allocation sequence; Item 8b. Type of randomization; details of any restriction (such as blocking and block size); Item 9. Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned; and Item 10 to clarify who generated the allocation sequence, who enrolled participants, and who assigned participants to interventions. Authors should give adequate information for the reader to assess the methods used randomization and provide the potential for the reader to assess risk of bias [19].

**References**