Covariates Adaptive Randomization Designs in Clinical Trials: A Comparative Study

Montasir Ahmed Osman

Mathematics Department, Faculty of Education, University of Kassala, Kassala, Sudan

Email address: a.montasir@yahoo.com


Received: September 29, 2016; Accepted: November 24, 2016; Published: December 30, 2016

Abstract: This paper investigated in covariate adaptive randomization designs, which are used to reduce covariate variables imbalance between treatments in clinical trials. Critical percentage and imbalance minimization methods are compared each one to another, and both are compared with pure randomization method in term of imbalance. The comparison is intended to show which method has minimum imbalance at three covariate variables with twelve single layers and three sample sizes 10, 20 and 100. The results which carried out from the simulation experiment clearly shown that the performance of critical percentage approach is closely similar to imbalance minimization method in full balance case as well as maximum imbalance at all sample sizes. And pure randomization method has the maximum imbalance compared to others at each sample size.

Keywords: Clinical Trials, Imbalance, Minimization, Randomization

1. Introduction

Adaptive design is a design which allows modifications to the trial or statistical procedures of the trial after its initiation without undermining the validity and integrity. The purpose is to make clinical trials more flexibility, effectively and fast. Due to the level of flexibility involved, these trial designs are termed as flexible design [1].

Liu and Pledger have addressed the following definition. The adaptive design methods are usually developed based on observed treatments effects to allow wider flexibility, and adaptations in clinical investigation of treatment. These may include changes of sample size, inclusion or exclusion criteria, study dose, study endpoints or methods of analysis [2].

According to Chow and Chang adaptive design is design which allows for changing or modifications the characteristics of a trial based on cumulating information to increase the probability of success, reduce the cost, reduce the time or preserve the and validity of the trial [3].

Food and Drug Administration (FDA) is an American organization interested on statistical procedures in biological industries. In 2010 FDA addressed guidance for that industry titled in “adaptive design clinical trials for drugs and biologics”. That guidance has defined adaptive design as “a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study”. Analysis of the accumulating study data are performed at pre-planned endpoints within the study, not important if with or without formal statistical hypotheses testing [4].

The covariate adaptive randomization (CAR) is usually used instead of pure randomization to reduce the covariate imbalance between treatment groups in clinical trials. Allocation probability for the covariate adaptive randomization is adapted over time during the trial based on the cumulative information about baseline covariates and treatment assignments. The following methods are addressed by biostatisticians to reduce the covariate imbalance in experiments. “Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial” [5] is one of the famous papers that discussed (CAR). The paper by Zelen “The randomization and stratification of patients to clinical trial” [6] is one of the most important designs in (CAR), too. Wei has mentioned two designs. The first one aimed to decrease the imbalance between treatments, by depending on marginal urn design. The second one used “Play the Winner Rule” for achieving the desired balance [7]. “Atkinson Optimal Model” is a model which considered a linear regression model to obtain the same
above advantage [8]. “Imbalance minimization method” has been addressed by Birkett [9]. The minimization method (MIN) has been widely used in clinical trials. The using of this method achieves minimum imbalance in the number of patients and their characteristics also, in each treatment.

“Critical percentage method” (CPM) is the last method in CAR methods which suggested by Osman [10]. In critical percentage method, all previous data is used to assign a new patient to treatments. It is designed to bridge the gap between the goal of covariate adaptive randomization designs and the current methods which are used to achieve this purpose.

This paper aims to compare two of CAR methods with pure randomization method (RM). The comparison is between imbalance minimization (MIN), critical percentage methods (CPM) and RM. Which one of these methods has the minimum imbalance is our purpose.

2. Adaptive Randomization Methods

Covariate adaptive randomization methods which the comparison will be between are addressed in this paragraph.

2.1. Imbalance Minimization Method

The minimization method (MIN) has been widely used in clinical trials. The using of this method achieves minimum imbalance in the number of patients and their characteristics also, in each treatment [9]. If there are T treatments and C covariates. Let \( l_1, l_2, l_3, ..., l_c \) be the levels of covariates 1, 2, 3, ..., c respectively. Then the number of strata here are \( l_1 \ast l_2 \ast l_3 \ast ... \ast l_c \).

Let \( n_{ijk} \) be the number of patients who were assigned with covariate \( i \) in level \( j \) to treatment \( k \).

Where \( i = 1, 2, 3, ..., c; j = 1, 2, 3, ..., l_i \) and \( k = 1, 2, 3, ..., T \), the next step is to assign \( n_{ijk} \) at each level. Let \( r_1, r_2, r_3, ..., r_c \) be the levels of new a patient covariates.

The assigning of this patient is as follows:

Step 1:
Add the new patient to the first treatment, treatment 1 say, temporarily.

Then compute the amount of imbalance:

\[
d_i = n_{ir_{1j}} - n_{ir_{2j}}
\]

where, \( n_{ir_{1j}} \) is the number of patients with covariate \( i \) in level \( j \) that are assigned to treatment 1.

\( n_{ir_{2j}} \) is the number of patients with covariate \( i \) in level \( j \) that are assigned to treatment 2.

\[
G = (d_{11}, d_{12}, ..., d_{1T}) = \sum_{i=1}^{c} |d_i|
\]

Step 2:
Add the new patient to treatment 2 temporarily.

Then compute the amount of imbalance:

\[
d_i = n_{ir_{1j}} - n_{ir_{2j}}
\]

where, \( n_{ir_{1j}} \) be the number of patients with covariate \( i \) in level \( j \) who have been assigned to treatment \( T_1 \).

\( n_{ir_{2j}} \) be the number of patients with covariate \( i \) in level \( j \) have been assigned to treatment \( T_2 \).

2.2. Critical Percentage Method

In critical percentage method, all previous data is used when assigning a new patient to treatments. It is designed to bridge the gap between the goal of covariate adaptive randomization designs and the current methods which are used to achieve this purpose [8].

As mentioned earlier, adaptive randomization designs are used in clinical trials to avoid the imbalance in the number of patients and their characteristics which could happen in pure randomization. The earliest method of adaptive randomization worked to reduce the imbalance by making more balance in each single layer in the experiment, but ignored the total of layers. This problem is solved in MIN method which focuses on total randomization imbalance. But the imbalance increases in single layers in this method. So, the purpose of CPM is to make more balance in the single layers and in the total randomization at the same time.

In the following paragraphs, assumptions and steps of CPM are explained for two treatments, and it is easy to generalize it for more than two treatments.

It is assumed in CPM that, patients are entered to the trial sequentially.

Suppose that there are two treatments \( T_1 \) and \( T_2 \), and \( C \) covariate variables. The \( i^{th} \) covariate has \( l_i \) levels, where \( C \geq 1, l_1 \geq 2, i = 1, 2, 3, ..., C \).

There are thus \( l_1 \ast l_2 \ast l_3 \ast ... \ast l_c = S \) single layers (strata) in the trial.

Step 1:
In this step the desirable percentage (critical percentage) to divide each part of each covariate variable between treatments is determined. That means, if we choose critical percentage equal 50% for \( l_{ij} \) (\( j^{th} \) level of \( i^{th} \) covariate) the number of patients who have the \( j^{th} \) level of the \( i^{th} \) covariate must be such that half of them in treatment \( T_1 \), and the other half in \( T_2 \). And if we choose 60% as a critical percentage for \( l_{ij} \), that means the number of patients who have \( l_{ij} \) in \( T_1 \) or \( T_2 \) is \( \leq 60\% \) from the total patients in this layer.

Let \( \lambda_{ij} \) be the critical percentage for level \( j \) of covariate \( i \).

Where \( 0 < \lambda_{ij} < 1 \).

The value of \( \lambda_{ij} \) would increase or decrease according to the importance of the covariate or the covariate level. And this flexibility in \( \lambda_{ij} \) value is considered as an of advantage of CPM.

Step 2:
1) The first patient in the trial would be assigned randomly to treatment $T_1$ or $T_2$ with probability equal to $\frac{1}{2}$ for each.

2) To assign the $(k+1)$th patient, where $k = 1, 2, 3, \ldots, n-1$ with $n$ the number of patients in the experiment.

a. Determine the covariates levels of the patient, let this $(i = 1, \ldots, c; j = 1, \ldots, l)$. This specifies the stratum to be $k_{ij}$ which the patient belongs.

b. Letting: $n_{k_{ij}1}$ The number of patients in level $j$ of covariate $i$ who are assigned to treatment $T_1$ after $k$ assignments.

$n_{k_{ij}2}$ The number of patients in level $j$ of covariate $i$ who are assigned to treatment $T_2$ after $k$ assignments.

Compute:

\[ p_{ij1} = \frac{n_{k_{ij}1}}{n_{k_{ij}1} + n_{k_{ij}2}} \]  
\[ p_{ij2} = \frac{n_{k_{ij}2}}{n_{k_{ij}1} + n_{k_{ij}2}} \]

(1) (2)

c. Compute the $r_1$ values defined as:

\[ r_1 = \begin{cases} 
1 \text{ if } p_{ij1} \leq \lambda_{ij}, & \forall ij \\
0 \text{ otherwise} & 
\end{cases} \]

\[ r_2 = \begin{cases} 
1 \text{ if } n_{k_{ij}1} < n_{k_{ij}2} \\
0 \text{ otherwise} & 
\end{cases} \]

\[ r_3 = \begin{cases} 
1 \text{ if } p_{ij2} > \lambda_{ij}, & \forall ij \\
0 \text{ otherwise} & 
\end{cases} \]

\[ r_4 = \begin{cases} 
1 \text{ if } p_{ij2} \geq \lambda_{ij}, & \forall ij \\
0 \text{ otherwise} & 
\end{cases} \]

\[ r_5 = \begin{cases} 
1 \text{ if } n_{k_{ij}1} = n_{k_{ij}2} \\
0 \text{ otherwise} & 
\end{cases} \]

\[ r_6 = \begin{cases} 
1 \text{ if } p_{ij1} = \lambda_{ij}, & \forall ij \\
0 \text{ otherwise} & 
\end{cases} \]

\[ r_7 = \begin{cases} 
1 \text{ if } p_{ij1} > \lambda_{ij}, & \exists ij \\
0 \text{ otherwise} & 
\end{cases} \]

$\lambda = \begin{cases} 
1 & \text{ if } p_{ij2} > \lambda_{ij}, & \exists ij \\
0 & \text{ otherwise} & 
\end{cases}$

\[ r_8 = \begin{cases} 
1 & \text{ if } p_{ij2} > \lambda_{ij}, & \exists ij \\
0 & \text{ otherwise} & 
\end{cases} \]

Patient number $(k+1)$ will then be assigned to treatment $T_1$ with probability $p_{k+1,1}$ where:

\[ p_{k+1,1} = \begin{cases} 
1 & \text{ if } r_8 = 1 \text{ or } (r_2 = 1 \text{ and } (r_5 = 1 \text{ and } r_7 = 1)) \\
\frac{1}{2} & \text{ if } r_5 = 1 \text{ and } (r_6 = 1 \text{ or } (r_7 = 1 \text{ and } r_8 = 1)) \\
0 & \text{ otherwise} & 
\end{cases} \]

3. The Simulation Experiment

In this section a description is given to the simulation experiment conducted to examine the performance of the CPM relative to that of RM and the MIN. To conduct this simulation, we assumed that, there are two treatments $T_1$ and $T_2$, and three covariate variables $c_1, c_2$ and $c_3$. The first covariate $c_1$ has two levels $l_{11}$ and $l_{12}$, $c_2$ has three levels $l_{21}, l_{22}$ and $l_{23}$ and $c_3$ has two levels $l_{31}$ and $l_{32}$. This yields $2 \times 3 \times 2 = 12$ single layers in the trial. The data will be generated with above covariates then allocated to the treatments by MIN method and CPM with $\lambda = 50\%$ and by RM with equal probabilities. When the sample is allocated, the imbalance would be calculated as the absolute difference between numbers of units which are assigned to treatments $T_1$ and $T_2$ for each method.

Different sample sizes are chosen to make reliable simulation. The first sample size is 10. This helps to test the methods performance in case, when the sample size is less than the size of single layers in the trial. Other sample sizes are 20 and 100. The simulation would be repeated 1000 times for each sample size. Imbalance calculated as the absolute value between numbers of patients who are assigned to treatments $T_1$ and $T_2$ at each single layer and total assigning. The software STATA12 is used to conduct this simulation.

4. Summary Results

The following table and figures show the frequencies and percentages of the imbalance for MIN, CPM and RM from the simulation study.

<table>
<thead>
<tr>
<th>Imbalance</th>
<th>Sample size 10</th>
<th>Sample size 20</th>
<th>Sample size 100</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CPM%</td>
<td>MIN%</td>
<td>RM%</td>
</tr>
<tr>
<td>0</td>
<td>72.8</td>
<td>82.1</td>
<td>69.3</td>
</tr>
<tr>
<td>2</td>
<td>27.2</td>
<td>17.9</td>
<td>30.4</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>0.0</td>
<td>8.8</td>
</tr>
<tr>
<td>6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
</tr>
<tr>
<td>8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>12</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>14</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>16</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>18</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>22</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>24</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 1. Imbalance and its percent for the three randomization methods when sample size 10, 20 and 100.
Table 1 above shows the imbalance values and its percent. The first column is the imbalance amount (the absolute difference between units which assigned to $T_1$ and $T_2$) according to randomization method and sample size. When the sample size 10, CPM has 72.8% full balance case (728 of 1000) compared to 82.1% in MIN and 23.2% in RM. And the maximum imbalance is 2 in both CPM and MIN, while is 10 in RM. In case of sample size 20 we find zero imbalance cases are 69.3% for CPM, 81.5% for MIN and 17.8% for RM. The maximum imbalance is large in RM (14), but is small for each of CPM (2) and MIN (4). Performances of CPM and MIN are similar where sample size 100. Maximum imbalance is 4 for both, compared to 24 for RM. And full balance is 62.9% in CPM and 79.6% in MIN compared to 7.2% in RM.

From Figure 1 (a), it is clear that the cumulative percent equal 100 when the imbalance is just 2 in each of CPM and MIN where is 10 in case of RM with sample size 10. Figure 1 (b) shows the cumulative percent is 100 where imbalances are 4, 2 and 14 for CPM, MIN and RM respectively when sample size 20. Cumulative percent is converged to a large extent in CPM and MIN and both are different from RM in Figure 1 (c).

5. Conclusion

From above results, we can conclude:

1) CPM has minimum imbalance compared with RM at all sample sizes.
2) The performance of CPM is similar to a large extent to MIN in terms of full balance and maximum imbalance.
3) The similarity between methods performance is unaffected by sample size.

References