Adaptive Randomization: 
Institutional Balancing Using SAS® Macro

Rita Tsang, Aptiv Solutions, Southborough, Massachusetts
Katherine Kacena, Aptiv Solutions, Southborough, Massachusetts

ABSTRACT

Adaptive designs in clinical trials have become increasingly popular and well received in drug development. This approach needs to be prospectively planned and includes such trial designs as adaptive dose finding, adaptive group sequential, sample size reassessment, seamless phase II/III, and population enrichment. We also consider institutional balancing (IB) in the randomization algorithm (Zelen, 1974) to be an adaptive feature which has the potential of making the trial more efficient. This kind of adaption optimizes the likelihood of treatment balance across risk strata and at the site (institution) level.

This paper will introduce programs that perform adaptive randomization based on an IB rule of a predefined treatment and control ratio within a site by utilizing SAS® macro. These programs will dynamically assign the first available treatment or control to a patient that enters the study based on a risk stratum when the IB rule is satisfied.

The SAS Macro facility is an excellent tool for dynamic randomization for its capacity to perform conditional iteration based on data-driven statistical input. In addition, simulations are used to verify the operating characteristics of the randomization. These operating characteristics may include but are not limited to the prevalence across risk strata and the patient recruitment across sites. This paper will show you how to program the IB design and will present scenarios that will benefit most from the IB design through simulation.

INTRODUCTION

The benefit of institutional balancing applied to a central randomization scheme is that the benefit of treatment balance within a stratum does not come at the detriment of an imbalance at the institutional level.

In our current design, we have 8 different risk strata (three dichotomous variables such that 23). The allocation ratio within each stratum is 2:1 (treatment: control) based on a block randomization scheme of size 3.

In all, there are expected to be 375 patients coming from 25 sites. Patients will be randomized into the study in sequence of their patient identifier based on the first available randomization number for their risk strata.

The rule created for institutional balancing requires that before a treatment from the central randomization can be assigned, the overall balance between treatment and control (including the assignment about to be made) must not violate the following rule: \( P = |\text{Treatment Count}/2 - \text{Control Count}| \leq 2. \)

For example, if Site 1 has 5 treatment and 1 control already assigned, and the next patient for that site, according to the central randomization for that risk strata happens to be a treatment, we have \(|6/2-1|=2\), and the assignment will be in compliance with institutional balancing rule.

However, if the very next patient for this institution is slated to receive treatment, the rule would be violated, i.e. \(|7/2-1|=2.5\) which is greater than 2. In this case, the next available control in the randomization list for that risk strata will be assigned instead.

For the next patient in this risk stratum, the central randomization would attempt to allocate any assignment skipped, the institutional balancing rule will be checked, and the process continues until all the patients have been randomized.

This METHODS/RESULTS section that follows is organized in the following fashion:
1) The Central Randomization List
2) Data from the First Simulation (Equal Patients per Site),
3) Macro Programs (IB, Fill, Chkfreq macros required to perform the institutional balancing algorithm)
4) Results from the First Simulation (Equal Patients per Site)
5) Data from 100 Simulations (Variable Patients per Site)
6) Results from the 100 Simulations (Variable Patients per Site)
METHODS/RESULTS

1) The Central Randomization List
For each stratum, 210 randomization numbers were created ($8 \times 210 = 1680$). Here is the frequency distribution of the treatment assignment within each risk stratum of the central randomization list:

<table>
<thead>
<tr>
<th>Risk Strata</th>
<th>Treatment</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>140</td>
<td>70</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>140</td>
<td>70</td>
<td>210</td>
</tr>
</tbody>
</table>

A sample printout of 10 observations from the central randomization data (block3.sas7bdat) is shown below:

<table>
<thead>
<tr>
<th>STRATA</th>
<th>TRT</th>
<th>RECID</th>
<th>RANDOMID</th>
<th>AVAILABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1001</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1002</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1003</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1004</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>1005</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>6</td>
<td>1006</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>7</td>
<td>1007</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>8</td>
<td>1008</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>9</td>
<td>1009</td>
<td>.</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>10</td>
<td>1010</td>
<td>.</td>
</tr>
</tbody>
</table>

STRATA (risk strata), TRT (treatment assignment, 2 = Treatment, 1 = Control), RECID (a unique record id within each stratum), RANDOMID (a unique randomization number), and AVAILABLE (an indicator for the availability of the randomization number, i.e. whether the number has been taken via randomization)

2) Data from the First Simulation (Equal Patients per Site)
We considered two distributions of patient recruitment by site. The first distribution is described in the table below, in which the total patients per site is uniformly distributed (15 patients for each of the 25 sites such that $n=375$):

<table>
<thead>
<tr>
<th>Site</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13</td>
<td>42</td>
<td>32</td>
<td>6</td>
<td>21</td>
<td>51</td>
<td>105</td>
<td>105</td>
<td>13</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

Posters
NESUG 2011
A sample printout of 10 observations assuming the first distribution of uniform patients per site in the cohort data (cohort.sas7bdat) is shown below:

<table>
<thead>
<tr>
<th>SITE</th>
<th>STRATA</th>
<th>PATID</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

SITE (site identifier), STRATA (risk strata), PATID (a unique patient identifier)

The second distribution uses an S+2 algorithm, where S is the site number, such that site 1 will have 3 patients (1+2) and site 25 will have 27 patients (25+2). In total, this assigns 375 patients to all 25 sites. This assumption is addressed within section 5) Data from 100 Simulations (Variable Patients per Site).

3) Macro Programs (IB, Fill, Chkfreq)
The IB macro will go through each patient (PATID) in the COHORT data in sequence. The first available treatment assignment from the BLOCK3 data will be obtained based on the risk strata. After the treatment assignment, the macro will check for the balance between the treatment and control within the site to determine if the IB is maintained. It will skip to the next available desired treatment or control to maintain the IB.

Two other macros, FILL and CHKFREQ, will be called within the IB macro. The FILL macro performs the task of adding patients from the COHORT data to the BLOCK3 data based on the first available randomization number for that risk strata. The CHKFREQ macro checks the IB rule within a site for the current assigned patient and determines whether the treatment assignment needs to be switched.

libname perm 'H:\My Documents\backup\papers\NESUG11\programs';
%let path=H:\My Documents\backup\papers\NESUG11\programs;

The MAUTOSOURCE option enables the macro processor to search the autocall libraries when resolving macro program references. The SASAUTOS option identifies the location of the autocall libraries to the macro processor. In our case, the user defined macros are stored in H:\My Documents\backup\papers\NESUG11\programs. The autocall macros supplied by various SAS products will be made available by the sasautos reference. Note that the library definition for sasautos is defined in the SAS configuration file.

options mautosource sasautos="&path", sasautos;
The IB Macro

%macro ib;

A macro variable LAST for the last patient in the COHORT data is created. In our current data, &last will be resolved to 375.

data cohort;
  set cohort end=eof;
  if eof then call symput('last',patid);
run;
%put &last;

The macro variable &i is equivalent to the patient number in the COHORT data. The do-loop will create one patient per data set in the COHORT&i data based on the PATID.

%do i=1 %to &last;
  data cohort&i;
  set cohort;
  if patid=&i;
  run;
%end;

The stratification factor (STRATA) and site (SITE) associated with PATID will be saved as macro variables.

proc sql noprint;
  select strata into: strata from cohort&i;
  select site into: site from cohort&i;
quit;

Three macro variables RECID_, CHECK, and NEWTRT are initialized.
&RECID_ is the RECID (a unique record id within each risk stratum) value to be created in the COHORT data for mapping with the BLOCK3 data.
&CHECK is an indicator for treatment balance such that 1=balanced, and missing (.) = imbalanced.
&NEWTRT is an indicator for the new treatment that needs to be switched based on the fulfillment of the IB rule such that 1=control and 2=treatment.

%let recid_=.;
%let check=.;
%let newtrt=.;

For the first assigned patient (&i=1), there is no need to check for IB. We will subset the BLOCK3 data based on the STRATA associated with PATID and the availability of the randomization number (available=.). The FILL macro is called to map the patient in the COHORT data to BLOCK3.

%if &i=1 %then %do;
  data checktrt;
  set block3;
  if strata=&strata and available=.;
  run;
  %fill;
%end;
The IB rule only applies to the second patient and onwards (i>1).

%if i>1 %then %do;

Note that for the second patient and onwards, the treatment assignment is performed before the IB rule is checked. We will initialize the macro variable CHECK to 1 (IB is balanced). When &CHECK=1, the FILL macro will be called to assign a treatment assignment to the patient based on the first available randomization number for that risk strata.

Once the assignment is performed, the CHKFREQ macro is called to check for IB. The CHKFREQ macro will implement the IB rule by checking the proportion of treatment and control for the assigned patients (available=1) within the site associated with the PATID.

%let check=1;
%if &check=1 %then %do;
  data checktrt;
  set block3;
  if strata=&strata and available=.;
  run;
  %fill;
%end;
%chkfreq;

After the CHKFREQ macro is called, the macro variables CHECK and NEWTRT will be generated based on the IB rule of P<=2. If the balance is not maintained (&CHECK ne 1), the current assignment will be skipped for the first alternative treatment assignment (&NEWTRT=1 or 2). The variables PATID, AVAILABLE and SITE will then be reset to blank to make the randomization number available for later use.

%if &check ne 1 %then %do;
  data block3;
  set block3;
  if site=&site and patid=&i then do;
    patid=.; available=.; site=.;
  end;
  run;
%end;

To improve the IB, it will get the next available randomization number of the same risk strata based on the first available alternative treatment assignment from the NEWTRT macro variable. The FILL macro will be called again to assign the patient to the right treatment in accordance with the IB rule.

data checktrt;
set block3;
if strata=&strata and trt=&newtrt and available=.;
run;
%fill;
%end;

This is the end of the &i>1 do-loop.

%end;
An IB data set will be saved with the information of the IB rule.

```sas
%if &i=2 %then %do;
data ib;
  set h;
  run;
%end;
%if &i>2 %then %do;
data ib;
  set ib h;
  run;
%end;
```

This is the end of the patient do-loop. The patient do-loop will repeat 375 times till all patients are randomized.

```sas
%mend ib;
```

The FILL Macro

The fill macro will identify the first available RECID value from the BLOCK3 data and save it as a macro variable RECID_. Note that the work data set CHECKTRT is created from BLOCK3 based on the STRATA associated with the PATID created in the IB macro. Note that the COHORT&I data set is also created in the IB macro.

```sas
%macro fill;
data checktrt;
  set checktrt;
  if _n_=1 then do;
    call symput('recid_',recid);
  end;
  run;
%put &recid_;
data cohort&i;
  set cohort&i;
  recid=&recid_;
  run;
%mend fill;
```

After COHORT&i is merged with BLOCK3 by STRATA and RECID, the available flag is turned on (available=1) to indicate that the treatment assignment has been taken.

```sas
data block3;
  merge block3 cohort&i (in=a);
  by strata recid;
  if a then available=1;
  run;
%mend fill;
```
The CHKFREQ Macro

```sas
%macro chkfreq;
   proc freq data=block3 noprint;
      where available=1 and site=&site;
      tables trt/out=h;
   run;
```

P is the IB rule, \( P = \text{abs}(\text{trt2}/2 - \text{trt1}) \), where \( \text{trt2} \) is the treatment count and \( \text{trt1} \) is the control count. Note that these counts include the proposed new treatment assignment.

When \( P \leq 2 \), the IB rule is maintained (CHECK=1). \( P1 \) is an indicator for whether there are too many controls or treatments within the site. \( P1 \) has the same value as \( P \) except that it is not an absolute value.

If \( P1 \) is negative (\( P1 < 0 \)), there are too many controls and we need to switch the current assignment for the first available treatment (NEWTRT=2). Conversely, if \( P1 \) is positive (\( P1 > 0 \)) then there are too many treatments. We will need to switch the current assignment for the first available control (NEWTRT=1).

```sas
data h;
   set h end=eof;
   retain trt1 trt2 .;
   if trt=1 then trt1=count;
   if trt=2 then trt2=count;
   if trt1=. then trt1=0;
   if trt2=. then trt2=0;
   p=abs(trt2/2-trt1);
   p1=(trt2/2-trt1);
   if p<=2 then check=1;
   if p1<0 and check ne 1 then newtrt=2;
   if p1>0 and check ne 1 then newtrt=1;
   if eof then output;
   run;

   proc sql noprint;
      select(check) into: check from h;
      select(newtrt) into: newtrt from h;
   quit;

   %put &check &newtrt;
```

The CHKFREQ macro will save all the IB information in a work data set \( h \), with added STRATA, SITE and PATID information.

```sas
data h;
   set h;
   strata=&strata;
   site=&site;
   patid=&i;
   run;

%mend chkfreq;
```
Calling the IB macro and save the final work data sets into permanent data sets.

```sas
%ib;

data perm.matched;
  set block3;
run;

data perm.ib;
  set ib;
run;
```

4) Results of the First Simulation (Equal Patients per Site)

IB data: There are 9 out of 375 patients (2.4%) who have had their treatment assignment switched due to the violation of the IB rule in the current data. There are 7 out of 25 sites (28%) with an imbalance.

<table>
<thead>
<tr>
<th>SITE</th>
<th>PATID</th>
<th>STRATA</th>
<th>P</th>
<th>P1</th>
<th>CHECK</th>
<th>NEWTRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>64</td>
<td>3</td>
<td>2.5</td>
<td>-2.5</td>
<td>.</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>157</td>
<td>6</td>
<td>2.5</td>
<td>2.5</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>182</td>
<td>7</td>
<td>2.5</td>
<td>2.5</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>215</td>
<td>7</td>
<td>2.5</td>
<td>2.5</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>286</td>
<td>8</td>
<td>2.5</td>
<td>2.5</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>297</td>
<td>8</td>
<td>2.5</td>
<td>2.5</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>305</td>
<td>8</td>
<td>2.5</td>
<td>-2.5</td>
<td>.</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>352</td>
<td>8</td>
<td>2.5</td>
<td>2.5</td>
<td>.</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>355</td>
<td>8</td>
<td>2.5</td>
<td>2.5</td>
<td>.</td>
<td>1</td>
</tr>
</tbody>
</table>

MATCHED data: Note that the first patient got their treatment assignment switched due to imbalance within the site is PATID 64 (STRATA 3, SITE 7, NEWTRT=2). The patient would have originally taken a treatment assignment of 1 at RECID 9 (indicated in red).

However due to the imbalance within the site, a treatment assignment of 2 is required. PATID 64 skipped to RECID 11 for the first available alternative treatment 2 (indicated in blue). Instead PATID 65 took the previously skipped treatment (TRT=1).

<table>
<thead>
<tr>
<th>STRATA</th>
<th>TRT</th>
<th>RECID</th>
<th>RANDOMID</th>
<th>AVAILABLE</th>
<th>SITE</th>
<th>PATID</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3001</td>
<td>1</td>
<td>1</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3002</td>
<td>1</td>
<td>1</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3003</td>
<td>1</td>
<td>2</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>3004</td>
<td>1</td>
<td>5</td>
<td>59</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3005</td>
<td>1</td>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>6</td>
<td>3006</td>
<td>1</td>
<td>6</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>7</td>
<td>3007</td>
<td>1</td>
<td>6</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>8</td>
<td>3008</td>
<td>1</td>
<td>6</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>9</td>
<td>3009</td>
<td>1</td>
<td>7</td>
<td>65</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>10</td>
<td>3010</td>
<td>1</td>
<td>8</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>11</td>
<td>3011</td>
<td>1</td>
<td>7</td>
<td>64</td>
</tr>
</tbody>
</table>
5) Data from 100 Simulations (Variable Patients per Site)

To assess the value of institutional balancing, we simulated 100 trials in which the number of patients within each site reflect real world scenarios. In each simulation, we created \((S+2)\) patients for each site, where \(S\) = site number, such that site 1 will have 3 patients, and site 25 will have 27 patients. In total, this assigns 375 patients across 25 sites. The 8 risk strata are based on three dichotomous risk characteristics such that \(2^3\). These three risk characteristics followed a Bernoulli probability distribution of Risk 1 at 0.5, Risk 2 at 0.3, and Risk 3 at 0.2.

<table>
<thead>
<tr>
<th>Risk Strata</th>
<th>Site</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>10</td>
<td>43</td>
<td>18</td>
<td>10</td>
<td>22</td>
<td>33</td>
<td>83</td>
<td>81</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
<td>41</td>
<td>27</td>
<td>18</td>
<td>35</td>
<td>45</td>
<td>105</td>
<td>114</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10</td>
<td>70</td>
<td>36</td>
<td>13</td>
<td>36</td>
<td>53</td>
<td>146</td>
<td>136</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>18</td>
<td>61</td>
<td>42</td>
<td>14</td>
<td>45</td>
<td>75</td>
<td>175</td>
<td>170</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24</td>
<td>78</td>
<td>56</td>
<td>19</td>
<td>45</td>
<td>78</td>
<td>200</td>
<td>200</td>
<td>700</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>19</td>
<td>89</td>
<td>46</td>
<td>26</td>
<td>54</td>
<td>92</td>
<td>245</td>
<td>229</td>
<td>800</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>31</td>
<td>83</td>
<td>61</td>
<td>21</td>
<td>60</td>
<td>99</td>
<td>272</td>
<td>273</td>
<td>900</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>32</td>
<td>125</td>
<td>56</td>
<td>33</td>
<td>74</td>
<td>113</td>
<td>303</td>
<td>264</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>30</td>
<td>118</td>
<td>78</td>
<td>24</td>
<td>76</td>
<td>132</td>
<td>325</td>
<td>317</td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>36</td>
<td>156</td>
<td>80</td>
<td>35</td>
<td>91</td>
<td>147</td>
<td>291</td>
<td>364</td>
<td>1200</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>43</td>
<td>174</td>
<td>91</td>
<td>45</td>
<td>93</td>
<td>140</td>
<td>355</td>
<td>359</td>
<td>1300</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>53</td>
<td>157</td>
<td>89</td>
<td>45</td>
<td>108</td>
<td>154</td>
<td>375</td>
<td>419</td>
<td>1400</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>51</td>
<td>170</td>
<td>113</td>
<td>33</td>
<td>92</td>
<td>186</td>
<td>410</td>
<td>445</td>
<td>1500</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>52</td>
<td>189</td>
<td>99</td>
<td>36</td>
<td>119</td>
<td>175</td>
<td>462</td>
<td>468</td>
<td>1600</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>51</td>
<td>216</td>
<td>124</td>
<td>65</td>
<td>112</td>
<td>202</td>
<td>461</td>
<td>469</td>
<td>1700</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>43</td>
<td>212</td>
<td>123</td>
<td>60</td>
<td>136</td>
<td>198</td>
<td>516</td>
<td>512</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>66</td>
<td>217</td>
<td>128</td>
<td>52</td>
<td>135</td>
<td>220</td>
<td>551</td>
<td>531</td>
<td>1900</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>63</td>
<td>253</td>
<td>134</td>
<td>64</td>
<td>146</td>
<td>240</td>
<td>573</td>
<td>527</td>
<td>2000</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>57</td>
<td>255</td>
<td>135</td>
<td>56</td>
<td>148</td>
<td>262</td>
<td>596</td>
<td>591</td>
<td>2100</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>69</td>
<td>266</td>
<td>147</td>
<td>62</td>
<td>147</td>
<td>254</td>
<td>629</td>
<td>626</td>
<td>2200</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>61</td>
<td>276</td>
<td>174</td>
<td>72</td>
<td>182</td>
<td>271</td>
<td>611</td>
<td>653</td>
<td>2300</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>71</td>
<td>268</td>
<td>172</td>
<td>56</td>
<td>196</td>
<td>298</td>
<td>655</td>
<td>684</td>
<td>2400</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>69</td>
<td>302</td>
<td>157</td>
<td>82</td>
<td>178</td>
<td>304</td>
<td>680</td>
<td>728</td>
<td>2500</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>68</td>
<td>337</td>
<td>190</td>
<td>78</td>
<td>187</td>
<td>283</td>
<td>772</td>
<td>685</td>
<td>2600</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>64</td>
<td>314</td>
<td>206</td>
<td>75</td>
<td>211</td>
<td>326</td>
<td>722</td>
<td>782</td>
<td>2700</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1106</td>
<td>4470</td>
<td>2582</td>
<td>1094</td>
<td>2728</td>
<td>4380</td>
<td>10513</td>
<td>10627</td>
<td>37500</td>
</tr>
</tbody>
</table>

The new cohort data set will have 375 patients and 100 simulated stratification factors (SIM1 – SIM100). Here is a sample printout of 10 observations from the COHORT data:

<table>
<thead>
<tr>
<th>SITE</th>
<th>PATID</th>
<th>SIM1</th>
<th>SIM2</th>
<th>SIM3</th>
<th>SIM4</th>
<th>SIM5</th>
<th>SIM6</th>
<th>...</th>
<th>SIM100</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>24</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>2</td>
</tr>
</tbody>
</table>

The program used to create this data is presented in the APPENDIX.
We repeated the same institutional balancing rule \((P = |\text{Treatment Count}/2 - \text{Control Count}| \leq 2)\) for each 100 simulated cohorts of size \(n=375\).

6) Results of 100 Simulations (Variable Patients per Site)

Our findings of the 100 simulated cohorts show that an average of 28 out of 375 patients (7.5%) had their treatment switched due to imbalance within the site. On average, 15 out of the 25 sites (60%) had at least one imbalance during the course of randomization (treatment assignments).

CONCLUSION

The benefit of institutional balancing (IB) applied to a central randomization scheme allows the benefit of treatment balance within a risk stratum to be achieved without the risk of an imbalance at the institutional level.

The IB decision rule programming was applied so that for each patient entry, the IB rule determined whether the patient would be randomized according to the first available randomization number for that stratum, or for the first alternative treatment assignment to prevent an institutional imbalance.

This type of IB randomization algorithm can be easily implemented. Because of the ease of implementation, IB perhaps should become a standard addition to central randomization. The relative importance of adding IB can vary under certain assumptions of the prevalence across risk strata and patient recruitment across sites. The particular rule for IB should be study dependent and assessed with the use of simulation.

REFERENCES


Food and Drug Administration (February 2010), Guidance for the Industry: Adaptive Design Clinical Trials for Drugs and Biologics, Rockville, MD.


CONTACT INFORMATION

The authors may be contacted at:

Rita Tsang, MA
Principal Statistical Programmer, Aptiv Solutions
225 Turnpike Road, Southborough, MA 01772, USA
rita.tsang@aptivsolutions.com
aptivsolutions.com

Katherine A. Kacena, PhD
Director of Biostatistics, Aptiv Solutions
225 Turnpike Road, Southborough, MA 01772, USA
katherine.kacena@aptivsolutions.com
aptivsolutions.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.
APPENDIX: The Cohort Macro

The following cohort macro program creates 100 simulated cohorts with 375 patients in each cohort as shown below.

<table>
<thead>
<tr>
<th>Create a cohort data of 375 patients, with (S+2) patients for each site where S=site number, such that site 1 will have 3 patients, and site 25 will have 27 patients. In total, this assigns 375 patients to all 25 sites.</th>
</tr>
</thead>
</table>
%macro cohort;
   data cohort;
   do site=1 to 25;
      if site=1 then do;
         do i=1 to 3;
            output;
         end;
      end;
      if site >1 then do;
         do i=1 to (3+1*(site-1));
            output;
         end;
      end;
   end;
run;

Scramble the patient number.

| data cohort;
   set cohort;
   x=ranuni(0323);
run; |
| proc sort data=cohort;
   by x;
run; |
| data cohort;
   set cohort;
   patid=_n_;
   drop x;
run; |
| %do i=1 %to 100; |

Patients are assigned to one of the 8 stratification factors in each simulated cohort based on three Bernoulli distribution probabilities. With the probability fixed at 0.5, 0.3 and 0.2, the prevalence of the risk strata will be consistent if the data is being regenerated. The macro variable SIM&i variable is synonymous to the STRATA variable. For example, SIM1 denotes the stratification factor, ranging from 1 to 8, each patient is assigned to for the first simulation.

| data cohort&i;
   set cohort;
   risk1=rand('BERNOULLI',.5);
   risk2=rand('BERNOULLI',.3);
   risk3=rand('BERNOULLI',.2);
   if risk1=1 and risk2=1 and risk3=1 then sim&i=1; |
if risk1=1 and risk2=1 and risk3=0 then sim&i=2;
if risk1=1 and risk2=0 and risk3=1 then sim&i=3;
if risk1=0 and risk2=1 and risk3=1 then sim&i=4;
if risk1=0 and risk2=0 and risk3=1 then sim&i=5;
if risk1=0 and risk2=1 and risk3=0 then sim&i=6;
if risk1=1 and risk2=0 and risk3=0 then sim&i=7;
if risk1=0 and risk2=0 and risk3=0 then sim&i=8;
output;
drop risk: i;
run;

%if &i=1 %then %do;
data final;
set cohort&i;
run;
%end;
%if &i>1 %then %do;
data final;
merge final cohort&i;
by patid;
run;
%end;
%end;

%mend cohort;
%cohort;

Data perm.cohort;
Set final;
Run;