Considerations of Parallel Groups Design (PGD)
Randomization Schemes
Significance Test based on PGD data
Multiple Comparison

ST4241 — Design and Analysis of Clinical Trials
Lecture 2: Parallel groups design and analysis

Chen Zehua

Department of Statistics & Applied Probability

8:00-10:00 am, Tuesday, August 16, 2016
Outline

Considerations of Parallel Groups Design (PGD)

Randomization Schemes

Significance Test based on PGD data

Multiple Comparison
When to use PGD?

There must be available eligible patients who are homogeneous in the sense:

- Similar in medical conditions;
- Any personal differences do not affect the effect of the treatments;

An example

- A clinic trial is to be conducted to compare four anesthetics.
- The purpose is to find the differences among the anesthetics.
- It is possible to recruit patients who are of the same gender, same age group and going through the same kind of surgery.
- Other aspects of the patients do not affect the effect of the anesthetics.
Elements of PGD

- $g$ treatments are to be compared.
- $n$ (a multiple of $g$) homogeneous patients need to be recruited.
- The $n$ patients are randomly divided into groups of about equal size.
- The $g$ treatments are randomly assigned to the $g$ groups.

Why is randomization necessary?

- Randomization is to make the groups assigned to different treatments as homogeneous as possible so as to avoid bias.
- Bias might be caused by assigning a particular treatment to a particular group of patients because of other considerations, such as race, severity of the disease, etc.
- When bias presents the treatment effects will be confounded with the effects of other (unknown) factors.
A simple randomization scheme is obtained by randomly permute the $n$ patients and then divide the permute patients in order into $g$ groups.

An example: 45 patients are randomly permuted and divided into 3 groups.

<table>
<thead>
<tr>
<th>44</th>
<th>42</th>
<th>10</th>
<th>21</th>
<th>28</th>
<th>38</th>
<th>6</th>
<th>31</th>
<th>27</th>
<th>30</th>
<th>17</th>
<th>35</th>
<th>29</th>
<th>41</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>25</td>
<td>14</td>
<td>16</td>
<td>37</td>
<td>45</td>
<td>40</td>
<td>36</td>
<td>8</td>
<td>24</td>
<td>32</td>
<td>5</td>
<td>20</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>43</td>
<td>13</td>
<td>23</td>
<td>12</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>18</td>
<td>22</td>
<td>11</td>
<td>26</td>
<td>19</td>
<td>7</td>
<td>39</td>
</tr>
</tbody>
</table>

The drawback of the simple randomization scheme is that, if the trials stops earlier than planned, it might result in an imbalance of patients among the groups, some bias might occur.
Randomly permuted blocks scheme

Instead of permuting all patients together, patients are permuted in batches with size as a smaller multiple of $g$.

- Permute the patients in batches of size $g$: This scheme is easier to implement, but might still cause bias if the trial cannot be completely double-blinded.
- Permute the patients in batches of size $gr$ with $r \leq 3$ or 4 chosen randomly for each batch: This scheme is a good compromise between simplicity and unbiasedness.

45 patients are assigned into 3 groups with randomly permuted blocks of size 9, 12, 12, 12:

6 4 1 14 10 13 11 24 32 25 33 34 43 40 36
3 5 7 16 15 17 21 30 23 27 28 37 38 35 42
9 2 8 20 19 12 18 31 26 29 22 44 39 45 41
Considerations of Parallel Groups Design (PGD)
Randomization Schemes
Significance Test based on PGD data
Multiple Comparison

Summary of data

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample Size</th>
<th>Mean</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$n_1$</td>
<td>$\bar{X}_1$</td>
<td>$s_1^2$</td>
</tr>
<tr>
<td>2</td>
<td>$n_2$</td>
<td>$\bar{X}_2$</td>
<td>$s_2^2$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$g$</td>
<td>$n_g$</td>
<td>$\bar{X}_g$</td>
<td>$s_g^2$</td>
</tr>
<tr>
<td>Total</td>
<td>$n = \sum_i n_i$</td>
<td>$\bar{X}$</td>
<td>$S^2$</td>
</tr>
</tbody>
</table>

where

- $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$, $X_{ij}$: the end-point measurement of patient $j$ in group $i$.
- $\bar{X} = \frac{1}{n} \sum_{i=1}^{g} n_i \bar{X}_i$.
- $s_i^2 = \frac{1}{n_i-1} \sum_{i=1}^{n_i} (X_{ij} - \bar{X}_i)^2$.
- $S^2 = \frac{1}{n-1} \sum_{i=1}^{g} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$. 
Analysis of Variance (ANOVA)

ANOVA table for PGD data

<table>
<thead>
<tr>
<th>Source</th>
<th>Between</th>
<th>Within</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>$g - 1$</td>
<td>$n. - g$</td>
<td>$n. - 1$</td>
</tr>
<tr>
<td>SS</td>
<td>$\sum n_i(\bar{X}_i - \bar{X}.)^2$</td>
<td>$\sum(n_i - 1)s_i^2$</td>
<td>$(n. - 1)S^2$</td>
</tr>
<tr>
<td>MS</td>
<td>BSS/$(g - 1)$</td>
<td>WSS/$(n. - g)$</td>
<td></td>
</tr>
<tr>
<td>F-ratio</td>
<td>BMS/WMS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that $TSS = BSS + WSS$. BSS measures the difference between the $g$ groups. WSS measures the difference between the individuals within groups. Under the hypothesis of no difference between the $g$ groups,

$$\frac{BMS}{WMS} \sim F_{g-1,n.-g}.$$
Model assumption and significant test

The model for $X_{ij}$'s is

$$X_{ij} = \mu_i + \epsilon_{ij}, \ i = 1, \ldots, g; j = 1, \ldots, n_i,$$

where $\epsilon_{ij}$ are i.i.d. $N(0, \sigma^2)$.

The significant test is to answer the question: Are there any difference among the treatments?

- **Hypotheses:** $H_0 : \mu_1 = \mu_2 = \cdots = \mu_g, H_1 : \text{not } H_0$.
- **Test statistic:** $F = \frac{\text{BMS}}{\text{WMS}}$.
- **p-value:** $p = P(F_{g-1, n-g} > \frac{\text{BMS}}{\text{WMS}})$.
- **Decision rule:** At significance level $\alpha$, $H_0$ is rejected if $p < \alpha$ or $F > F_{g-1, n-g}(\alpha)$, the upper $\alpha$-quantile of the $F$-distribution with d.f. $g - 1$ and $n - g$. 
The anesthetics trial example

In the trial for the comparison of anesthetics, four anesthetic methods: Ether, Cyclopropane, Thiopental and Spinal, are to be compared. A simple randomization scheme was used to assign patients to the anesthetics, it resulted in group sizes: 5, 7, 9, 8. The summary data of the trial is as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample size</th>
<th>Mean</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ether</td>
<td>5</td>
<td>4.64</td>
<td>1.2080</td>
</tr>
<tr>
<td>Cyclopropane</td>
<td>7</td>
<td>4.63</td>
<td>0.7390</td>
</tr>
<tr>
<td>Thiopental</td>
<td>9</td>
<td>3.53</td>
<td>0.2025</td>
</tr>
<tr>
<td>Spinal</td>
<td>8</td>
<td>3.08</td>
<td>0.5479</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>3.86</strong></td>
<td><strong>0.9915</strong></td>
</tr>
</tbody>
</table>
Considerations of Parallel Groups Design (PGD)
Randomization Schemes
Significance Test based on PGD data
Multiple Comparison

The anesthetics trial example (cont.)

Computation of ANOVA table:

- **BSS:** \[ \sum n_i (\bar{X}_i - \bar{X})^2 = 4(4.67 - 3.86)^2 + 6(4.63 - 3.86)^2 + 8(3.53 - 3.86)^2 + 7(3.08 - 3.86)^2 = 13.04. \]

- **WSS:** \[ \sum (n_i - 1)s_i^2 = 4 \times 1.208 + 6 \times 0.739 + 8 \times 0.2025 + 7 \times 0.5479 = 14.72. \]

- **TSS:** \[ (n - 1)S^2 = 28 \times 0.9915 = 27.76 \]

- **BMS** = **BSS**/3 = 4.35, \[ **WMS** = **WSS**/25 = 0.59. \]

- **F-ratio:** \[ **BMS**/**WMS** = 7.37 \]
Considerations of Parallel Groups Design (PGD)
Randomization Schemes
Significance Test based on PGD data
Multiple Comparison

The anesthetics trial example (cont.)

ANOVA table for the anesthetics trial example

<table>
<thead>
<tr>
<th>Source</th>
<th>Between</th>
<th>Within</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>3</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>SS</td>
<td>13.04</td>
<td>14.72</td>
<td>27.76</td>
</tr>
<tr>
<td>MS</td>
<td>4.35</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>F-ratio</td>
<td>7.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ p = P(F_{3,25} > 7.37) = 0.00106. \]

Conclusion: The null hypothesis of no difference of four anesthetics is rejected at level \( \alpha > 0.00106 \). For example, \( \alpha = 0.002, F_{3,25}(0.002) = 6.5645 < 7.37 = F \), the null hypothesis of no difference of anesthetics is rejected.
General problems of multiple comparison

The multiple comparison is to find the concrete differences among the treatments. It answers the question: *what are the differences?*

There are two situations:

- (a) There are no pre-hypothesized specific differences to investigate and a general exploring needs to be made to discover the differences, the significance test must precede the multiple comparison. Multiple comparison is meaningful only when the significant test rejects the null hypothesis.

- (b) There are certain scientifically important specific differences which are pre-hypothesized, the significance test is not necessary, multiple comparison can be done directly.

Usual pre-specified hypotheses: pairwise differences, the differences between a particular treatment and the others, some specific contrasts.
Contrasts

A general structure of differences are expressed by contrasts. Let \( \mu_i \) denote the effect of treatment \( i \). A contrast is defined as

\[
C = \sum_{i=1}^{g} c_i \mu_i \quad \text{with} \quad \sum_{i=1}^{g} c_i = 0.
\]

All contrasts together describe the totality of the differences among the treatments, because of the following equivalence:

\[
\mu_1 = \mu_2 = \cdots = \mu_g \iff \sum_{i=1}^{g} c_i \mu_i = 0 \quad \text{for all } C.
\]

Examples of contrasts:

\[
\begin{align*}
\mu_i - \mu_j, & \quad 1 \leq i < j \leq g; \\
\frac{1}{2} (\mu_1 + \mu_2) - \frac{1}{3} (\mu_3 + \mu_4 + \mu_6); \\
& \text{etc.}
\end{align*}
\]
Scheffe’s procedure

- Scheffe’s procedure is for a general exploring.
- Scheffe’s procedure tests the hypothesis:
  \[ H_0 : \sum_{i=1}^{g} c_i \mu_i = 0 \text{ for all } C. \]

- In order to prevent artifacts, Scheffe’s procedure controls the family-wise type I error rate
  \[ P\left( \frac{\left| \sum_{i=1}^{g} c_i \bar{X}_i \right|}{\sqrt{WMS \sum_{i=1}^{g} \frac{c_i^2}{n_i}}} \geq c_\alpha, \text{ for at least one } C \right) \]
  at level \( \alpha \).
- Critical value \( c_\alpha = \sqrt{(g - 1)F_{g-1, n-g}(\alpha)} \) is referred to as the Scheffe's criterion.
Scheffe’s procedure (cont.)

- In the application of Scheffe’s procedure, the contrasts to be tested are identified from the summary of the data.
- Let $L_C$ denote 
  \[ \frac{\sum_{i=1}^{g} c_i \bar{X}_i}{\sqrt{\text{WMS} \sum_{i=1}^{g} \frac{c_i^2}{n_i}}} \]
  The contrast $C = \sum_{i=1}^{g} c_i \mu_i$ is significant, if 
  \[ |L_C| \geq \sqrt{(g-1)F_{g-1,n.-g}(\alpha)} \]. The $p$-value of $L_C$ is given by
  \[ P(F_{g-1,n.-g} \geq L_C^2/(g-1)). \]
- **Simultaneous confidence intervals** for the contrasts $\sum_{i=1}^{g} c_i \mu_i$ are constructed as
  \[ \sum_{i=1}^{g} c_i \bar{X}_i \pm \sqrt{\text{WMS} \sum_{i=1}^{g} \frac{c_i^2}{n_i}} \sqrt{(g-1)F_{g-1,n.-g}(\alpha)}. \]
The anesthetics trial example (cont.)

From the summary of the data:

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample size</th>
<th>Mean</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ether</td>
<td>5</td>
<td>4.64</td>
<td>1.2080</td>
</tr>
<tr>
<td>Cyclopropane</td>
<td>7</td>
<td>4.63</td>
<td>0.7390</td>
</tr>
<tr>
<td>Thiopental</td>
<td>9</td>
<td>3.53</td>
<td>0.2025</td>
</tr>
<tr>
<td>Spinal</td>
<td>8</td>
<td>3.08</td>
<td>0.5479</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>29</strong></td>
<td><strong>3.86</strong></td>
<td><strong>0.9915</strong></td>
</tr>
</tbody>
</table>

the following contrasts are suggested:

- $C_1 : (1, 0, -1, 0)$
- $C_2 : (0, 1, 0, -1)$
- $C_3 : (0.5, 0.5, 0, -1)$
- $C_4 : (0.5, 0.5, -0.5, -0.5)$
### Illustration of computation:
Recall $WMS = 0.59$.

\[
\hat{C}_1 = \bar{X}_1 + 0\bar{X}_2 - \bar{X}_3 + 0\bar{X}_4 = \bar{X}_1 - \bar{X}_3 = 4.64 - 3.53 = 1.11.
\]

\[
se(\hat{C}_1) = \sqrt{0.59 \times (1^2/5 + (-1)^2/9)} = 0.428.
\]

\[
L_{C_1} = \hat{C}_1/se(\hat{C}_1) = 1.11/0.428 = 2.59.
\]

\[
p-value = P(F_{3,25} \geq 2.59^2/3) = 0.1089.
\]

The computed values for all the four contrasts are given below:

<table>
<thead>
<tr>
<th></th>
<th>$\hat{C}$</th>
<th>$se(\hat{C})$</th>
<th>$L$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.11</td>
<td>0.428</td>
<td>2.59</td>
<td>0.1089</td>
</tr>
<tr>
<td>2</td>
<td>1.55</td>
<td>0.398</td>
<td>3.89$^a$</td>
<td>0.0072</td>
</tr>
<tr>
<td>3</td>
<td>1.56</td>
<td>0.353</td>
<td>4.42$^a$</td>
<td>0.0021</td>
</tr>
<tr>
<td>4</td>
<td>1.33</td>
<td>0.292</td>
<td>4.55$^a$</td>
<td>0.0015</td>
</tr>
</tbody>
</table>
Remarks

(i) It is possible that the null hypothesis of the significance $F$-test is rejected but none of the contrasts considered is significant.

(ii) For any single contrast, \( \frac{\hat{C}}{\text{se}(\hat{C})} \) follows a $t$-distribution with df \( n. - g \).

(iii) If the test for each contrast is based on the $t$-distribution, it is possible that even if the overall $F$-test is not significant at level $\alpha$, some of the $t$-tests are significant at level $\alpha$.

Takehome question

Why?