Selecting an appropriate randomization procedure for a small population group trial on the basis of a linked optimization criterion

David Schindler, Ralf-Dieter Hilgers

Department of Medical Statistics
RWTH Aachen University

24. September 2015
Motivation

- No scientific evaluation of randomization procedures in the presence of several types of bias found in literature.
- Several demands on randomization procedures have been well studied independently of each other, but not simultaneously.
- Urgent need for a score that unifies several issues for measuring the different demands on the randomization process.

⇒ Propose a new framework for the selection of an appropriate randomization procedure based on desirability functions.
Motivation

Exemplary application of the new framework on

- **Selection bias:**
  - Assessed issue: correct guesses.

- **Chronological bias**
  - Assessed issue: type-I-error, power.

- **Balancing behavior**
  - Assessed issue: power loss due to differences in group sizes.

⇒ Propose a new framework for the selection of an appropriate randomization procedure based on desirability functions.
Two-armed clinical trial with parallel group design with continuous endpoint and total sample size $N$.

Experimental treatment $E$ and control treatment $C$.

Let $\mathbf{T} = (T_1, \ldots, T_N)' \in \{E, C\}^N$ be a randomization sequence and $T_i$ be the $i$th element of $\mathbf{T}$.

Let $N_s(i, \mathbf{T})$ be the number of patients assigned to $s \in \{E, C\}$ after $i$ allocations.
Selection bias

Assuming a balanced trial it is opportune for the experimenter to guess the $i$th allocation according to the convergence strategy:

(Blackwell and Hodges Jr., 1957)

$$g_{CS}(i, \mathbf{T}) = \begin{cases} 
E, & \text{if } N_E(i - 1, \mathbf{T}) < N_C(i - 1, \mathbf{T}) \\
\text{random guess}, & \text{if } N_E(i - 1, \mathbf{T}) = N_C(i - 1, \mathbf{T}) \\
C, & \text{if } N_E(i - 1, \mathbf{T}) > N_C(i - 1, \mathbf{T})
\end{cases}$$

Expected proportion of Correct Guesses (CG) of $\mathbf{T}$ is defined as:

$$CG(\mathbf{T}) = \frac{\mathbb{E} \left( \sum_{i=1}^{N} 1\{T_i=g_{CS}(i,\mathbf{T})\} \right)}{N}$$
Chronological bias

Model for chronological bias: (Tamm and Hilgers, 2014; Rosenkranz, 2011)

\[
Y = \begin{pmatrix}
1 & \tilde{T}_1 & 1 \\
1 & \tilde{T}_2 & 2 \\
\vdots & \vdots & \vdots \\
1 & \tilde{T}_N & N \\
\end{pmatrix}
\begin{pmatrix}
\mu \\
\xi \\
\vartheta \\
\end{pmatrix} + \epsilon,
\]

with \( \epsilon \sim \mathcal{N}(0, I_{N \times N}) \) and \( \tilde{T}_i := t(T_i) = \begin{cases}
1, & \text{if } T_i = E \\
-1, & \text{if } T_i = C
\end{cases} \)

- The trial is evaluated with a model including the effects \( \mu \) and \( \xi \), although the time effect \( \vartheta \neq 0 \) is present (misspecification).

\[ \Rightarrow \text{The type-I-error } \alpha \text{ and the power } (1 - \beta) \text{ when testing } \xi = 0 \text{ using a t-test is biased, due to not adjusting for } \vartheta. \]
Balancing behavior

Due to differences in group sizes $N_E(N, T) - N_C(N, T)$ arising at the end of a clinical trial a **loss in the power** when conducting Student’s t-test occurs.

**Example:**
Assuming a total sample size of $N = 50$, an effect size of $\Delta = 0.81$, and a type-I-error probability of $\alpha = 0.05$ it follows:

<table>
<thead>
<tr>
<th>$N_E(N, T)$</th>
<th>25</th>
<th>24</th>
<th>23</th>
<th>20</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_C(N, T)$</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>$1 - \beta_0(T)$</td>
<td>0.800</td>
<td>0.799</td>
<td>0.797</td>
<td>0.784</td>
<td>0.728</td>
</tr>
</tbody>
</table>
Right-sided Derringer-Suich desirability function

**Definition:** (Derringer and Suich, 1980)

\[
d_i(T) := d(c_i(T)) = \begin{cases} 
1, & \text{if } c_i(T) \leq TV_i \\
\frac{USL_i - c_i(T)}{USL_i - TV_i}, & \text{if } TV_i < c_i(T) < USL_i \\
0, & \text{if } c_i(T) \geq USL_i 
\end{cases}
\]

- \(c_i(T)\): value of the \(i\)-th issue for \(T\).
- \(TV_i\): Target Value of the \(i\)-th issue.
- \(USL_i\): Upper Specification Limit of the \(i\)-th issue.
Right-sided Derringer-Suich desirability function

**Definition:** (Derringer and Suich, 1980)

\[
d_i(T) := d(c_i(T)) = \begin{cases} 
1, & \text{if } c_i(T) \leq TV_i \\
\frac{USL_i - c_i(T)}{USL_i - TV_i}, & \text{if } TV_i < c_i(T) < USL_i \\
0, & \text{if } c_i(T) \geq USL_i
\end{cases}
\]

- \(c_i(T)\): value of the \(i\)-th issue for \(T\).
- \(TV_i\): Target Value of the \(i\)-th issue.
- \(USL_i\): Upper Specification Limit of the \(i\)-th issue.

\(\Rightarrow\) Need a meaningful TV and USL dependent on the practical need.
Right-sided Derringer-Suich desirability function

**Definition:** (Derringer and Suich, 1980)

\[ d_i(T) := d(c_i(T)) = \begin{cases} 
1, & \text{if } c_i(T) \leq TV_i \\
\frac{USL_i - c_i(T)}{USL_i - TV_i}, & \text{if } TV_i < c_i(T) < USL_i \\
0, & \text{if } c_i(T) \geq USL_i 
\end{cases} \]

Investigated standard setting:

<table>
<thead>
<tr>
<th>i</th>
<th>( c_i(T) )</th>
<th>( TV_i )</th>
<th>( USL_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( CG(T) )</td>
<td>0.50</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>( \alpha_{TT}(T) )</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>( \beta_{TT}(T) )</td>
<td>0.20</td>
<td>0.40</td>
</tr>
<tr>
<td>4</td>
<td>( \beta_0(T) )</td>
<td>0.20</td>
<td>0.21</td>
</tr>
</tbody>
</table>
Properties of desirability scores

- Desirability scores are dimensionless and $\in [0, 1]$.
- Desirability scores are summarizeable with the geometric mean:

$$\bar{d}(T) := \prod_{i=1}^{4} d_i(T)^{\omega_i} \text{ with } \sum_{i=1}^{4} \omega_i = 1.$$ 

- $T$ with $\bar{d}(T) = 0$ is called undesired randomization sequence.
Properties of desirability scores

- Desirability scores are dimensionless and $\in [0, 1]$.
- Desirability scores are summarizeable with the geometric mean:
  \[ \bar{d}(T) := \prod_{i=1}^{4} d_i(T)^{\omega_i} \text{ with } \sum_{i=1}^{4} \omega_i = 1. \]
- $T$ with $\bar{d}(T) = 0$ is called undesired randomization sequence.
- Weights should be chosen dependent on the planned trial.
  - Heuristical approach: Put one third of the weights on each demand.
    \[ \Rightarrow \omega_1 = 1/3, \omega_2 = \omega_3 = 1/6, \text{ and } \omega_4 = 1/3 \]
Correct guesses of PBR(4) for \( N = 4 \)

- randomizeR was used for the evaluation (Schindler and Uschner, 2015).

<table>
<thead>
<tr>
<th></th>
<th>( T'_j )</th>
<th>( P(T_j) )</th>
<th>( CG(T_j) )</th>
<th>( d_1(T_j) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EECC</td>
<td>1/6</td>
<td>0.625</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ECEC</td>
<td>1/6</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CEEC</td>
<td>1/6</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ECCE</td>
<td>1/6</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CECE</td>
<td>1/6</td>
<td>0.750</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CCEE</td>
<td>1/6</td>
<td>0.625</td>
<td></td>
</tr>
</tbody>
</table>

average value: 0.708

PBR(\( k \)) (Permuted Block Randomization with block length \( k \))
Within each block half of the patients are assigned to \( E \) and \( C \).
Correct guesses of PBR(4) for \( N = 4 \)

- randomizr was used for the evaluation (Schindler and Uschner, 2015).

<table>
<thead>
<tr>
<th>( j )</th>
<th>( T'_j )</th>
<th>( P(T'_j) )</th>
<th>( CG(T'_j) )</th>
<th>( d_1(T'_j) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EECC</td>
<td>( \frac{1}{6} )</td>
<td>0.625</td>
<td>0.500</td>
</tr>
<tr>
<td>2</td>
<td>ECEC</td>
<td>( \frac{1}{6} )</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>CEEC</td>
<td>( \frac{1}{6} )</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>ECCE</td>
<td>( \frac{1}{6} )</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>CECE</td>
<td>( \frac{1}{6} )</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>CCEE</td>
<td>( \frac{1}{6} )</td>
<td>0.625</td>
<td>0.500</td>
</tr>
</tbody>
</table>

Average value: 0.708 0.167

\[
d_1(T_1) = d(CG(T_1)) = \frac{USL_1 - CG(T_1)}{USL_1 - TV_1} = \frac{0.75 - 0.625}{0.75 - 0.5} = 0.5
\]
Correct guesses of PBR(4) for $N = 4$

randomizer was used for the evaluation (Schindler and Uschner, 2015).

<table>
<thead>
<tr>
<th>$j$</th>
<th>$T'_j$</th>
<th>$P(T_j)$</th>
<th>$CG(T_j)$</th>
<th>$d_1(T_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EECC</td>
<td>$\frac{1}{6}$</td>
<td>0.625</td>
<td>0.500</td>
</tr>
<tr>
<td>2</td>
<td>ECEC</td>
<td>$\frac{1}{6}$</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>CEEC</td>
<td>$\frac{1}{6}$</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>ECCE</td>
<td>$\frac{1}{6}$</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>CECE</td>
<td>$\frac{1}{6}$</td>
<td>0.750</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>CCEE</td>
<td>$\frac{1}{6}$</td>
<td>0.625</td>
<td>0.500</td>
</tr>
</tbody>
</table>

average value: 0.708 0.167

$$\bar{d}_1(T) = \frac{1}{6} (0.5 + 0 + 0 + 0 + 0 + 0 + 0.5) = 0.167$$
Assessment of PBR(4) for $N = 4$

<table>
<thead>
<tr>
<th>j</th>
<th>$T'_j$</th>
<th>$P(T_j)$</th>
<th>$d_1(T_j)$</th>
<th>$d_2(T_j)$</th>
<th>$d_3(T_j)$</th>
<th>$d_4(T_j)$</th>
<th>$\bar{d}(T_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EECC</td>
<td>$\frac{1}{6}$</td>
<td>0.500</td>
<td>0.804</td>
<td>0.649</td>
<td>1.000</td>
<td>0.712</td>
</tr>
<tr>
<td>2</td>
<td>ECEC</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.668</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>CEEC</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.776</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>ECCE</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.776</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>CECE</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.961</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>CCEE</td>
<td>$\frac{1}{6}$</td>
<td>0.500</td>
<td>0.804</td>
<td>1.000</td>
<td>1.000</td>
<td>0.765</td>
</tr>
</tbody>
</table>

average value: 0.167 0.935 0.805 1.000 0.246

Settings: $\vartheta = \frac{1}{4}$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.

\[ d_1(T) = d(CG(T)) \quad d_2(T) = d(\alpha_{TT}(T)) \quad d_3(T) = d(1 - \beta_{TT}(T)) \quad d_4(T) = d(1 - \beta_0(T)) \]

\[
\bar{d}(T_1) = \sqrt[3]{d_1(T_1)} \cdot \sqrt[6]{d_2(T_1)} \cdot \sqrt[6]{d_3(T_1)} \cdot \sqrt[3]{d_4(T_1)}
\]

\[
= \sqrt[3]{0.500} \cdot \sqrt[6]{0.804} \cdot \sqrt[6]{0.649} \cdot \sqrt[3]{1.000}
\]

\[
= 0.712
\]
Assessment of PBR(4) for $N = 4$

<table>
<thead>
<tr>
<th>$j$</th>
<th>$T'_j$</th>
<th>$P(T_j)$</th>
<th>$d_1(T_j)$</th>
<th>$d_2(T_j)$</th>
<th>$d_3(T_j)$</th>
<th>$d_4(T_j)$</th>
<th>$\bar{d}(T_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EECC</td>
<td>$1/6$</td>
<td>0.500</td>
<td>0.804</td>
<td>0.649</td>
<td>1.000</td>
<td>0.712</td>
</tr>
<tr>
<td>2</td>
<td>ECEC</td>
<td>$1/6$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.668</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>CEEC</td>
<td>$1/6$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.776</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>ECCE</td>
<td>$1/6$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.776</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>CECE</td>
<td>$1/6$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.961</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>CCEE</td>
<td>$1/6$</td>
<td>0.500</td>
<td>0.804</td>
<td>1.000</td>
<td>1.000</td>
<td>0.765</td>
</tr>
</tbody>
</table>

average value: 0.167 0.935 0.805 1.000 0.246

Settings: $\vartheta = \frac{1}{4}$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.

$d_1(T) = d(CG(T))$  $d_2(T) = d(\alpha_T(T))$  $d_3(T) = d(1 - \beta_T(T))$  $d_4(T) = d(1 - \beta_0(T))$

$\varnothing \bar{d}(T) = \frac{1}{6} (0.712 + 0 + 0 + 0 + 0 + 0 + 0.765)
= 0.246$
Assessment of PBR(4) for $N = 4$

<table>
<thead>
<tr>
<th>$j$</th>
<th>$T'_j$</th>
<th>$P(T_j)$</th>
<th>$d_1(T_j)$</th>
<th>$d_2(T_j)$</th>
<th>$d_3(T_j)$</th>
<th>$d_4(T_j)$</th>
<th>$\bar{d}(T_j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EECC</td>
<td>$\frac{1}{6}$</td>
<td>0.500</td>
<td>0.804</td>
<td>0.649</td>
<td>1.000</td>
<td>0.712</td>
</tr>
<tr>
<td>2</td>
<td>ECEC</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.668</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>CEEC</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.776</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>ECCE</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.776</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>CECE</td>
<td>$\frac{1}{6}$</td>
<td>0.000</td>
<td>1.000</td>
<td>0.961</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>CCEE</td>
<td>$\frac{1}{6}$</td>
<td>0.500</td>
<td>0.804</td>
<td>1.000</td>
<td>1.000</td>
<td>0.765</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>average value:</td>
</tr>
</tbody>
</table>

Settings: $\vartheta = \frac{1}{4}$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.

$d_1(T) = d(CG(T)) \quad d_2(T) = d(\alpha_{TT}(T)) \quad d_3(T) = d(1 - \beta_{TT}(T)) \quad d_4(T) = d(1 - \beta_0(T))$

$\frac{2}{3}$ of the randomization sequences are undesired.

$\Rightarrow$ Approach for $N = 4$ not usefull.
Investigated randomization procedures

**PBR(k)** (Permuted Block Randomization with block length \( k \))
Within each block half of the patients are assigned to \( E \) and \( C \).

**RPBR(k)** (Randomized Permuted Block Randomization with maximal block length \( k \)) PBR with random block lengths 2, 4, \( \ldots \), \( k \).

**CR** Complete randomization is accomplished by tossing a fair coin.

**BSD(\( a \))** (Big Stick Design) CR allow for imbalance within the limit \( a \).
Comparison for $N = 50$

Settings: $\psi = \frac{1}{50}$, $\xi = 0.40$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.
Results based on 100,000 simulations.

<table>
<thead>
<tr>
<th>Design</th>
<th>$\varnothing \bar{d}(T)$ (sd)</th>
<th>$P(\bar{d}(T) = 0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CR</td>
<td>0.5131 (0.388)</td>
<td>0.3534</td>
</tr>
<tr>
<td>RPBR(8)</td>
<td>0.6088 (0.081)</td>
<td>0.0011</td>
</tr>
<tr>
<td>PBR(8)</td>
<td>0.6759 (0.07)</td>
<td>0.0001</td>
</tr>
<tr>
<td>PBR(50)</td>
<td>0.7797 (0.181)</td>
<td>0.0408</td>
</tr>
<tr>
<td>BSD(4)</td>
<td>0.8400 (0.084)</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

- **BSD(4)** has low probability of generating undesired randomization sequences.
- **BSD(4)** seems to be the best compromise between handling a time trend, the proportion of correct guesses, and the loss in power.
Analysis of the used USLs for BSD(4)

- Change of the desirability scores, when a specification limit convergences against the TV$_i$. 

![Graphs showing desirability scores](image)
Analysis of the used weights for BSD(4)

Change of the weight of an fixed issue. The other weights are splitted equally.
Conclusions

- Presented a framework for the scientific evaluation of randomization procedures dependent on arising demands.

- Evaluation should be part of the statistical trial and analysis plan.

- Other TVs, USLs, and weights for the investigated issues lead to different recommendations.

- Other randomization procedures can be implemented easily.

- Include other issues for measuring (further) demands.

This project has received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under Grant Agreement no 602552.
References


