Desirability of restricted randomization procedures in small population group trials in case of selection and chronological bias

David Schindler, Ralf-Dieter Hilgers

Department of Medical Statistics
RWTH Aachen University

24. August 2015

FP7 HEALTH 2013 - 602552

Motivation

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

$\Rightarrow$ Propose a new framework for the selection of an appropriate randomization procedure for a small population group trial based on desirability functions.

FP7 HEALTH 2013 - 602552

Motivation

Examplary application of the new framework on two types of bias

- Selection bias:
  - Assessed issue: correct guesses.
- Chronological bias
  - Assessd issue: type-I-error, power.

$\Rightarrow$ Propose a new framework for the selection of an appropriate randomization procedure for a small population group trial based on desirability functions.

FP7 HEALTH 2013 - 602552

Terminology

- Two-armed clinical trial with parallel group design with continuous endpoint and total sample size N.
- Experimental treatment E and control treatment C.
- Let $\boldsymbol{T} = (T_1, \ldots, T_N)' \in \{E, C\}^N$ be a randomization sequence and T_i be the ith element of $\boldsymbol{T}$.
- Let $N_s(i, \boldsymbol{T})$ 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)

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

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

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

FP7 HEALTH 2013 - 602552

Chronological bias

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

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

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

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

  $\Rightarrow$ The type-I-error α and the power $(1-\beta)$ when testing $\xi = 0$ using a t-test is biased, due to not adjusting for ϑ.

FP7 HEALTH 2013 - 602552

Right-sided Derringer-Suich desirability function

Definition: (Derringer and Suich, 1980)

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

- $c_i(\boldsymbol{T})$: value of the i-th issue for $\boldsymbol{T}$.
- 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)

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

- $c_i(\boldsymbol{T})$: value of the i-th issue for $\boldsymbol{T}$.
- 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 issue and practical need.

FP7 HEALTH 2013 - 602552

Right-sided Derringer-Suich desirability function

Definition: (Derringer and Suich, 1980)

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

Example for settings:

i	$c_i(\boldsymbol{T})$	TV_i	USL_i
1	$CG(\boldsymbol{T})$	0.50	0.75
2	$\alpha_{TT}(\boldsymbol{T})$	0.05	0.10
3	$\beta_{TT}(\boldsymbol{T})$	0.20	0.40

FP7 HEALTH 2013 - 602552

Properties of desirability scores

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

$$\bar{d}(\boldsymbol{T}) := \prod_{i=1}^{3} d_i(\boldsymbol{T})^{\omega_i} \text{ with } \sum_{i=1}^{3} \omega_i = 1.$$

- $\boldsymbol{T}$ with $\bar{d}(\boldsymbol{T}) = 0$ is called undesired randomization sequence.

FP7 HEALTH 2013 - 602552

Properties of desirability scores

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

$$\bar{d}(\boldsymbol{T}) := \prod_{i=1}^{3} d_i(\boldsymbol{T})^{\omega_i} \text{ with } \sum_{i=1}^{3} \omega_i = 1.$$

- $\boldsymbol{T}$ with $\bar{d}(\boldsymbol{T}) = 0$ is called undesired randomization sequence.

- Weights should be chosen dependent on the planned trial.
- Heuristical approach: Put half of the weights on selection bias and half of the weights on chronological bias.

$\Rightarrow \omega_1 = 1/2$ and $\omega_2 = \omega_3 = 1/4$

FP7 HEALTH 2013 - 602552

Assessment of PBR(4) with $N = 4$

j	$\boldsymbol{T}_j'$	$P(\boldsymbol{T}_j)$	$CG(\boldsymbol{T}_j)$	$d_1(\boldsymbol{T}_j)$	$\alpha_{TT}(\boldsymbol{T}_j)$	$d_2(\boldsymbol{T}_j)$	$1-\beta_{TT}(\boldsymbol{T}_j)$	$d_3(\boldsymbol{T}_j)$	$\bar{d}(\boldsymbol{T}_j)$
1	EECC	$1/6$	0.625						
2	ECEC	$1/6$	0.750						
3	CEEC	$1/6$	0.750						
4	ECCE	$1/6$	0.750						
5	CECE	$1/6$	0.750						
6	CCEE	$1/6$	0.625						
	average value:		0.708						

Settings: $\vartheta = 1/4$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.
$\alpha_{TT}(\boldsymbol{T}_j)/1 - \beta_{TT}(\boldsymbol{T}_j)$:= type-I-error/power of $\boldsymbol{T}_j$ in case of the assumed linear time trend

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

FP7 HEALTH 2013 - 602552

Assessment of PBR(4) with $N = 4$

j	T'_j	$P(T_j)$	$CG(T_j)$	$d_1(T_j)$	$\alpha_{TT}(T_j)$	$d_2(T_j)$	$1-\beta_{TT}(T_j)$	$d_3(T_j)$	$\bar{d}(T_j)$
1	EECC	1/6	0.625	0.500					
2	ECEC	1/6	0.750	0.000					
3	CEEC	1/6	0.750	0.000					
4	ECCE	1/6	0.750	0.000					
5	CECE	1/6	0.750	0.000					
6	CCEE	1/6	0.625	0.500					
	average value:		0.708	0.167					

Settings: $\vartheta = 1/4$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.
$\alpha_{TT}(T_j)/1 - \beta_{TT}(T_j)$:= type-I-error/power of T_j in case of the assumed linear time trend

$$d_1(T_1) = d(CG(T_1)) = \frac{CG(T_1) - TV_1}{USL_1 - TV_1} = \frac{0.625 - 0.5}{0.5 - 0.75} = 0.5$$

Assessment of PBR(4) with $N = 4$

j	T'_j	$P(T_j)$	$CG(T_j)$	$d_1(T_j)$	$\alpha_{TT}(T_j)$	$d_2(T_j)$	$1-\beta_{TT}(T_j)$	$d_3(T_j)$	$\bar{d}(T_j)$
1	EECC	1/6	0.625	0.500	0.060	0.804	0.730	0.649	0.601
2	ECEC	1/6	0.750	0.000	0.047	1.000	0.734	0.668	0.000
3	CEEC	1/6	0.750	0.000	0.043	1.000	0.755	0.776	0.000
4	ECCE	1/6	0.750	0.000	0.043	1.000	0.755	0.776	0.000
5	CECE	1/6	0.750	0.000	0.047	1.000	0.792	0.961	0.000
6	CCEE	1/6	0.625	0.500	0.060	0.804	0.842	1.000	0.670
	average value:		0.708	0.167	0.050	0.935	0.768	0.805	0.212

Settings: $\vartheta = 1/4$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.
$\alpha_{TT}(T_j)/1 - \beta_{TT}(T_j)$:= type-I-error/power of T_j in case of the assumed linear time trend

$$\begin{aligned}\bar{d}(T_1) &= \sqrt{d_1(T_1)} \cdot \sqrt[4]{d_2(T_1)} \cdot \sqrt[4]{d_3(T_1)} \\ &= \sqrt{0.500} \cdot \sqrt[4]{d_2(0.804)} \cdot \sqrt[4]{d_3(0.649)} \\ &= 0.601\end{aligned}$$

FP7 HEALTH 2013 - 602552

Assessment of PBR(4) with $N = 4$

j	T_j'	$P(T_j)$	$CG(T_j)$	$d_1(T_j)$	$\alpha_{TT}(T_j)$	$d_2(T_j)$	$1-\beta_{TT}(T_j)$	$d_3(T_j)$	$\bar{d}(T_j)$
1	EECC	$1/6$	0.625	0.500	0.060	0.804	0.730	0.649	0.601
2	ECEC	$1/6$	0.750	0.000	0.047	1.000	0.734	0.668	0.000
3	CEEC	$1/6$	0.750	0.000	0.043	1.000	0.755	0.776	0.000
4	ECCE	$1/6$	0.750	0.000	0.043	1.000	0.755	0.776	0.000
5	CECE	$1/6$	0.750	0.000	0.047	1.000	0.792	0.961	0.000
6	CCEE	$1/6$	0.625	0.500	0.060	0.804	0.842	1.000	0.670
	average value:		0.708	0.167	0.050	0.935	0.768	0.805	0.212

Settings: $\vartheta = 1/4$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.
$\alpha_{TT}(T_j)/1 - \beta_{TT}(T_j)$:= type-I-error/power of T_j in case of the assumed linear time trend

$$\varnothing \bar{d}(T) = 1/6\,(0.601 + 0 + 0 + 0 + 0 + 0.670)$$
$$= 0.212$$

FP7 HEALTH 2013 - 602552

Assessment of PBR(4) with $N = 4$

j	T'_j	$P(T_j)$	$CG(T_j)$	$d_1(T_j)$	$\alpha_{TT}(T_j)$	$d_2(T_j)$	$1-\beta_{TT}(T_j)$	$d_3(T_j)$	$\bar{d}(T_j)$
1	EECC	$1/6$	0.625	0.500	0.060	0.804	0.730	0.649	0.601
2	ECEC	$1/6$	0.750	0.000	0.047	1.000	0.734	0.668	0.000
3	CEEC	$1/6$	0.750	0.000	0.043	1.000	0.755	0.776	0.000
4	ECCE	$1/6$	0.750	0.000	0.043	1.000	0.755	0.776	0.000
5	CECE	$1/6$	0.750	0.000	0.047	1.000	0.792	0.961	0.000
6	CCEE	$1/6$	0.625	0.500	0.060	0.804	0.842	1.000	0.670
	average value:		0.708	0.167	0.050	0.935	0.768	0.805	0.212

Settings: $\vartheta = 1/4$, $\xi = 2.83$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.
$\alpha_{TT}(T_j)/1 - \beta_{TT}(T_j)$:= type-I-error/power of T_j in case of the assumed linear time trend

$2/3$ 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 assigend to E and C.

CR Complete randomization is accomplished by tossing a fair coin.

BSD(a) (Big Stick design) CR allow for imbalance within the limit a.

FP7 HEALTH 2013 - 602552

Comparison for $N = 12$

Settings: $\vartheta = 1/12$, $\xi = 0.90$, $\alpha_0 = 0.05$, and $1 - \beta_0 = 0.8$.

Design	$\varnothing \bar{d}(T)$ (sd)	$P(\bar{d}(T) = 0)$
PBR(4)	0.3199 (0.222)	0.2963
PBR(12)	0.5199 (0.211)	0.0942
CR	0.6503 (0.302)	0.1331
BSD(3)	0.7287 (0.211)	0.0291

- BSD(3) has low probability of generating undesired randomization sequences.
- BSD(3) seems to be the best compromise between handling a time trend and the proportion of correct guesses.

FP7 HEALTH 2013 - 602552

Conclusions

- Presented a framework for the scientific evaluation of randomization procedures dependent on arising bias.
- 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 further issues (e.g. punishment of high imbalances in the end of a trial).

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

FP7 HEALTH 2013 - 602552

References

Blackwell, D. and J. L. Hodges Jr. (1957). Design for the control of selection bias. *Annals of Mathematical Statistics 25*, 449–460.

Derringer, G. and R. Suich (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology 12*, 214–219.

ICH E9 (1998). Statistical principles for clinical trials. *Current Step 4 version dated 5 Februrary 1998. Last access in Mai 2015. Available from:* `http://www.ich.org`.

Rosenkranz, G. K. (2011). The impact of randomization on the analysis of clinical trials. *Statistics in Medicine 30*, 3475–3487.

Tamm, M. and R.-D. Hilgers (2014). Chronological bias in randomized clinical trials under different types of unobserved time trends. *Methods of Information in Medicine 53*, 501–510.

FP7 HEALTH 2013 - 602552