Assessment of randomization procedures with respect to multiple objectives

David Schindler, Ralf-Dieter Hilgers

Department of Medical Statistics
RWTH Aachen University

August 22, 2016

Challenges in small population group trials

- Total number of eligible patients may be very limited, which impacts the choice of **study design** and the statistical methodology (see O'Connor and Hemmings, 2014)
- Choice of a randomization procedure does not follow scientific arguments up to now.
- Unequal performance of randomization procedures in the presence of
  - Selection bias
  - Chronological bias
- Treatment comparisons should involve consideration of the potential contribution of bias to the *p*-value (ICH E9, 1998).

FP7 HEALTH 2013 - 602552

Challenges in small population group trials

- Total number of eligible patients may be very limited, which impacts the choice of **study design** and the statistical methodology (see O'Connor and Hemmings, 2014)
- Choice of a randomization procedure does not follow scientific arguments up to now.
- Unequal performance of randomization procedures in the presence of
  - Selection bias
  - Chronological bias
- Treatment comparisons should involve consideration of the potential contribution of bias to the *p*-value (ICH E9, 1998).

FP7 HEALTH 2013 - 602552

Model

Assuming a (random) bias vector $\mathbf{b} = (b_1, b_2, \ldots, b_N)^T$ the ith patient's response with $i \in \{1, 2, \ldots, N\}$ can be expressed as:

$$y_i = \mu_E \, T_i + \mu_C \, (1 - T_i) + b_i + \epsilon_i \, . \tag{1}$$

- The ith allocation is done as follows:

$$T_i = \begin{cases} 1, & \text{if patient } i \text{ is allocated to group } E \\ 0, & \text{if patient } i \text{ is allocated to group } C \end{cases} .$$

- Expected response μ_j under treatment $j \in \{E, C\}$.
- Errors $\epsilon_i \underset{iid}{\sim} \mathcal{N}(0, 1)$.

FP7 HEALTH 2013 - 602552

Test Statistic

We test the hypotheses

$$H_0 : \mu_E = \mu_C \text{ vs. } H_1 : \mu_E \neq \mu_C$$

with Student's t-test (under misspecification) and test statistic

$$W := \sqrt{\frac{N_E\, N_C}{N_E + N_C}}\, \frac{\bar{y}_E - \bar{y}_C}{S_{\text{pooled}}} \sim t_{N-2,\delta,\lambda}$$

$$\text{with } \bar{y}_E = \frac{1}{N_E} \sum_{i=1}^{N} y_i\, T_i \text{ and } \bar{y}_C = \frac{1}{N_C} \sum_{i=1}^{N} y_i\, (1 - T_i),$$

where N_E and N_C are the final numbers of patients assigned to the corresponding treatment group.

Types of bias

For **chronological bias** according to Tamm and Hilgers (2014) b_i is assumed to be increasing/decreasing in N. For a linear time trend we define:

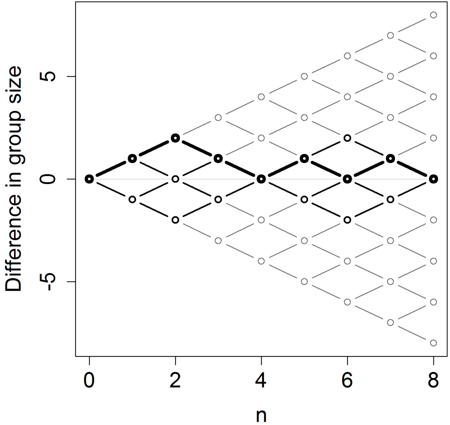
$$b_i = \frac{(i-1)\,\vartheta}{N} \quad \text{with } \vartheta \in \mathbb{R} \text{ and } i \in \{1, 2, \ldots, N\}\ .$$

In the situation of **selection bias** b_i is dependent on the patients assigned to the corresponding treatment groups (Proschan, 1994):

$$b_i = \begin{cases} \eta, & \text{if } N_E(i-1) < N_C(i-1) \\ -\eta, & \text{if } N_E(i-1) > N_C(i-1) \\ 0, & \text{if } N_E(i-1) = N_C(i-1) \end{cases} \quad \text{with } \eta \in \mathbb{R}_+\ .$$

FP7 HEALTH 2013 - 602552

Permuted Block Randomization



- At the end of each block there is no difference in patient numbers.
- All sequences are equiprobable.

PBR(4): Permuted Block Randomization with block length 4

FP7 HEALTH 2013 - 602552

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

Investigated settings for selection bias:

- $\alpha = 0.05$
- $\eta = 1.42$ (one quarter of the effect size)
- $\alpha_{SB}(\mathbf{T}_j) :=$ Type-I-error probability in case of selection bias

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	overall
1	CCEE	$1/6$	0.047			
2	CECE	$1/6$	0.138			
3	ECCE	$1/6$	0.060			
4	CEEC	$1/6$	0.060			
5	ECEC	$1/6$	0.138			
6	EECC	$1/6$	0.047			
	average value:		0.081			

FP7 HEALTH 2013 - 602552

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

Investigated settings for chronological bias:

- $\alpha = 0.05$, $(1-\beta) = 0.8$, $\mu_E - \mu_C = 5.65$
- $\vartheta = 1$
- $\alpha_{TT}(\mathbf{T}_j)$:= Type-I-error probability in case of a linear time trend
- $1-\beta_{TT}(\mathbf{T}_j)$:= Power in case of a linear time trend

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	overall
1	CCEE	$1/6$	0.047	0.060	0.842	
2	CECE	$1/6$	0.138	0.047	0.792	
3	ECCE	$1/6$	0.060	0.043	0.755	
4	CEEC	$1/6$	0.060	0.043	0.755	
5	ECEC	$1/6$	0.138	0.047	0.734	
6	EECC	$1/6$	0.047	0.060	0.730	
	average value:		0.081	0.050	0.768	

FP7 HEALTH 2013 - 602552

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

- No linked assessment score available
  $\Rightarrow$ How is the performance of PBR(4) in comparison to other randomization procedures?

j	$\mathbf{T}'_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	overall
1	CCEE	$1/6$	0.047	0.060	0.842	?
2	CECE	$1/6$	0.138	0.047	0.792	?
3	ECCE	$1/6$	0.060	0.043	0.755	?
4	CEEC	$1/6$	0.060	0.043	0.755	?
5	ECEC	$1/6$	0.138	0.047	0.734	?
6	EECC	$1/6$	0.047	0.060	0.730	?
	average value:		0.081	0.050	0.768	?

FP7 HEALTH 2013 - 602552

Right-sided Derringer-Suich desirability function

Definition (Derringer and Suich (1980)):

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

TV: Target Value USL: Upper Specification Limit

i	Criterion$_i$ (c_i)	TV$_i$	USL$_i$
1	$\alpha_{SB}(\mathbf{T})$	0.05	0.10
2	$\alpha_{TT}(\mathbf{T})$	0.05	0.10
3	$\beta_{TT}(\mathbf{T})$	0.20	0.40

FP7 HEALTH 2013 - 602552

Multi-objective combination criterion

- Desirability scores are in the interval $[0, 1]$.
- Desirability scores can be combined with the geometric mean:

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

- The geometric mean is a multi-objective combination criterion.

FP7 HEALTH 2013 - 602552

Multi-objective combination criterion

- Desirability scores are in the interval $[0, 1]$.
- Desirability scores can be combined with the geometric mean:

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

- The geometric mean is a multi-objective combination criterion.
- Weights should be chosen dependent on the planned trial.
- To give an example:
  Distribute the weight uniformly on selection bias and 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	$\mathbf{T}'_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
1	EECC	$1/6$	0.047	1.000					
2	ECEC	$1/6$	0.138	0.000					
3	CEEC	$1/6$	0.060	0.809					
4	ECCE	$1/6$	0.060	0.809					
5	CECE	$1/6$	0.138	0.000					
6	CCEE	$1/6$	0.047	1.000					
	average value:		0.081	0.603					

$$d_1(\mathbf{T}_1) = d(\alpha_{SB}(\mathbf{T}_1)) = 1, \text{because } 0.047 < 0.05$$

FP7 HEALTH 2013 - 602552

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

j	$\mathbf{T}'_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
1	EECC	$1/6$	0.047	1.000	0.060	0.804	0.842	1.000	0.947
2	ECEC	$1/6$	0.138	0.000	0.047	1.000	0.792	0.961	0.000
3	CEEC	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.844
4	ECCE	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.844
5	CECE	$1/6$	0.138	0.000	0.047	1.000	0.734	0.668	0.000
6	CCEE	$1/6$	0.047	1.000	0.060	0.804	0.730	0.649	0.850
	average value:		0.081	0.603	0.050	0.935	0.768	0.805	0.581

$$
\begin{aligned}
\bar{d}(\mathbf{T}_1) &= \sqrt{d_1(\mathbf{T}_1)} \cdot \sqrt[4]{d_2(\mathbf{T}_1)} \cdot \sqrt[4]{d_3(\mathbf{T}_1)} \\
&= \sqrt{1} \cdot \sqrt[4]{0.804} \cdot \sqrt[4]{d_3(1)} \\
&= 0.947
\end{aligned}
$$

FP7 HEALTH 2013 - 602552

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

j	$\mathbf{T}'_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
1	EECC	$1/6$	0.047	1.000	0.060	0.804	0.842	1.000	0.947
2	ECEC	$1/6$	0.138	0.000	0.047	1.000	0.792	0.961	0.000
3	CEEC	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.844
4	ECCE	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.844
5	CECE	$1/6$	0.138	0.000	0.047	1.000	0.734	0.668	0.000
6	CCEE	$1/6$	0.047	1.000	0.060	0.804	0.730	0.649	0.850
	average value:		0.081	0.603	0.050	0.935	0.768	0.805	0.581

$$\varnothing \bar{d}(\mathbf{T}) = 1/6\,(0.947 + 0 + 0.844 + 0.844 + 0 + 0.850)$$
$$= 0.581$$

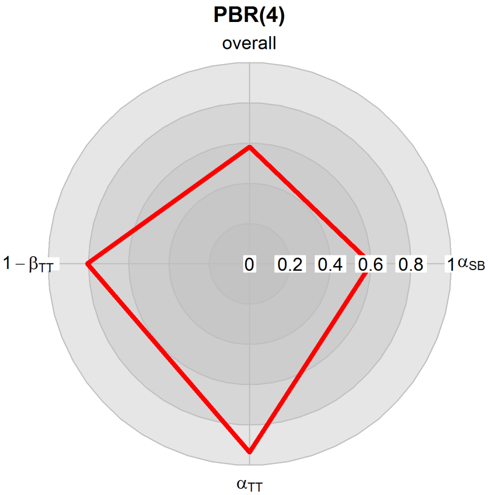
FP7 HEALTH 2013 - 602552

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

j	$\mathbf{T}'_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
1	EECC	$1/6$	0.047	1.000	0.060	0.804	0.842	1.000	0.947
2	ECEC	$1/6$	0.138	0.000	0.047	1.000	0.792	0.961	0.000
3	CEEC	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.844
4	ECCE	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.844
5	CECE	$1/6$	0.138	0.000	0.047	1.000	0.734	0.668	0.000
6	CCEE	$1/6$	0.047	1.000	0.060	0.804	0.730	0.649	0.850
	average value:		0.081	0.603	0.050	0.935	0.768	0.805	0.581

- Average desirability scores can be visualized in a radar plot, which is available in the `randomizeR` package (Schindler et al., 2015).

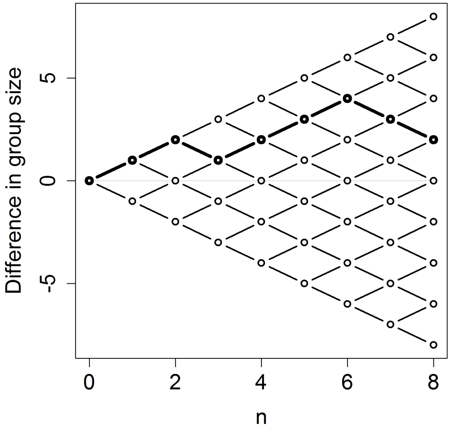
Radar plot



- PBR(4) seems to be good in handling the assumed linear time trend.
- PBR(4) seems to be susceptible to the convergence strategy.

FP7 HEALTH 2013 - 602552

Complete Randomization

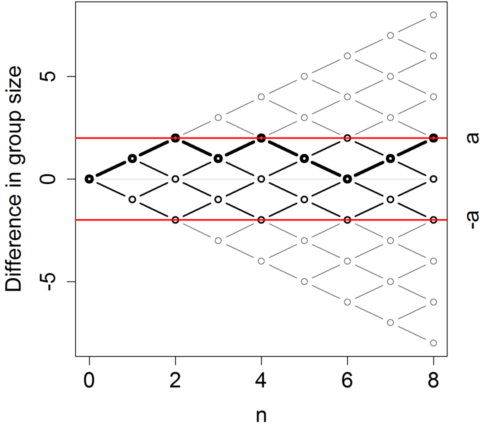


- Fair coin toss for each patient allocation.

CR: Complete Randomization

FP7 HEALTH 2013 - 602552

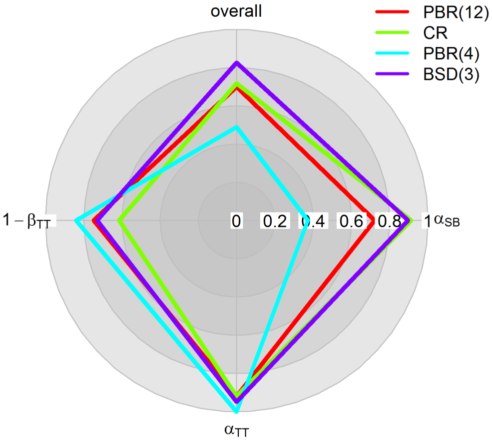
Big Stick Design (Soares and Wu, 1983)



- Fair toin toss with imbalance boundary a.

BSD(2): Big Stick Design with imbalance boundary $a = 2$

Comparison for $N = 12$



- PBR(4) seems to be very susceptible to selection bias.
- BSD(3) manages the investigated criteria the best.

FP7 HEALTH 2013 - 602552

Conclusion

- Randomization procedures differ in terms of their susceptibility to selection bias and chronological bias.
- The linked assessment criterion makes a fair comparison of different randomization procedures possible.
- The radar plot compares the behavior of randomization procedures at a glance.
- We developed `randomizeR` (Schindler et al., 2015) for making fair comparisons of randomization procedures concerning different types of bias and their balancing behavior.

The IDeAl 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 I

Atkinson, A. C. (2001). The comparison of designs for sequential clinical trials with covariate information. *Journal of the Royal Statistical Socitey 165*, 349–373.

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

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

O'Connor, D. J. and R. J. Hemmings (2014). Coping with small of patients in clinical trials. *Expert Opinion on Orphan Drugs 2*, 765–768.

Proschan, M. (1994). Influence of selection bias on type 1 error rate under random permuted block designs. *Statistica Sinica 4*, 219–231.

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

References II

Schindler, D., D. Uschner, R.-D. Hilgers, and N. Heussen (2015). *randomizeR: Randomization for clinical trials*. R package version 1.2.

Soares, J. F. and C. Wu (1983). Some restricted randomization rules in sequential designs. *Communications in Statiscs - Theory and Methods 12*, 2017–2034.

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

Flexibility of the approach

- The linked assessment criterion summarizes all imaginable criteria to one unified score and takes their importance into account.
- Other suggested criteria in the literature are:
  - Correct Guesses (Blackwell and Hodges Jr., 1957)
  - Loss in treatment estimation (Atkinson, 2001)
- Other randomization procedures can be easily assessed such as:
  - Efron's Biased Coin Design
  - Truncated Binomial Design
  - Randomized Permuted Block Randomization
  - Maximal Procedure

FP7 HEALTH 2013 - 602552

Comparison for $N = 12$

RP	$\bar{d}(1-\beta_{TT}(\mathbf{T}_j))$	$\bar{d}(\alpha_{TT}(\mathbf{T}_j))$	$\bar{d}(\alpha_{SB}(\mathbf{T}_j))$	$\varnothing\bar{d}(\mathbf{T}_j)$
PBR(4)	0.840	1.000	0.371	0.489
PBR(12)	0.747	0.919	0.721	0.699
CR	0.615	0.919	0.911	0.717
BSD(3)	0.729	0.947	0.895	0.825

- PBR(4) seems to be very susceptible to selection bias.
- BSD(3) manages the investigated criteria the best.

FP7 HEALTH 2013 - 602552