Linked assessment criterion for the choice of a randomization procedure

David Schindler, Ralf-Dieter Hilgers

Department of Medical Statistics
RWTH Aachen University

June 14th, 2016

Introduction

- Choice of a randomization procedure does not follow scientific arguments up to now.
- Treatment comparisons should involve consideration of the potential influence of bias on the p-value (ICH E9, 1998).
- Unequal performance of randomization procedures with respect to the following objectives:
  - Selection bias
  - Chronological bias
  - Balancing behavior

$\Rightarrow$ Presentation of a linked assessment criterion for the choice of a randomization procedure

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 treatment indicator takes the values:

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

- 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

$$W := \sqrt{\frac{N_E \, N_C}{N_E + N_C}} \, \frac{\bar{y}_E - \bar{y}_C}{S_{\text{pooled}}}$$

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

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

FP7 HEALTH 2013 - 602552

Distribution of the test statistic

Lemma (Generalization of Langer (2014)):
The test statistic W follows a doubly-noncentral t-distribution with noncentrality parameteres $\delta := \delta(\mathbf{T}, \mathbf{B})$ and $\lambda := \lambda(\mathbf{T}, \mathbf{B})$. Under $H_0 : \mu_E = \mu_C$ and $\mathbf{B} = (b_1, b_2, \ldots, b_N)^T$ it follows that the noncentrality parameters are given by:

$$\delta = \sqrt{\frac{N_E\, N_C}{N_E + N_C}}\left(\bar{B}_E - \bar{B}_C\right) \text{ and } \lambda = \sum_{i=1}^{N}\left(B_i^2 - N_E \bar{B}_E^2 - N_C \bar{B}_C^2\right)$$

$$\text{with } \bar{B}_E = \frac{1}{N_E}\sum_{i=1}^{N} b_i\, \mathbb{1}_{\{T_i=1\}} \text{ and } \bar{B}_C = \frac{1}{N_C}\sum_{i=1}^{N} b_i\, \mathbb{1}_{\{T_i=0\}}\,.$$

FP7 HEALTH 2013 - 602552

Rejection probability

Theorem:
Assuming $H_0 : \mu_E = \mu_C$, the rejection probability dependent on **T** and **B** can be calculated as follows:

$$\alpha(\mathbf{T}) := \alpha(\mathbf{T},\mathbf{B}) := P\left(|W| > t_{N-2,1-\alpha/2} \middle| \mathbf{T},\mathbf{B}\right)$$
$$= F_{N-2,\delta,\lambda}\left(t_{N-2,\alpha/2}\right) + 1 - F_{N-2,\delta,\lambda}\left(t_{N-2,1-\alpha/2}\right) ,$$

where $F_{N-2,\delta,\lambda}(x)$ is the distribution fuction of the doubly noncentral t-distribution with $N-2$ degrees of freedom and noncentrality parameteres δ and λ.

Corollary:
The noncentrality parameter λ is invariant with respect to a treatment difference, δ not.

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:

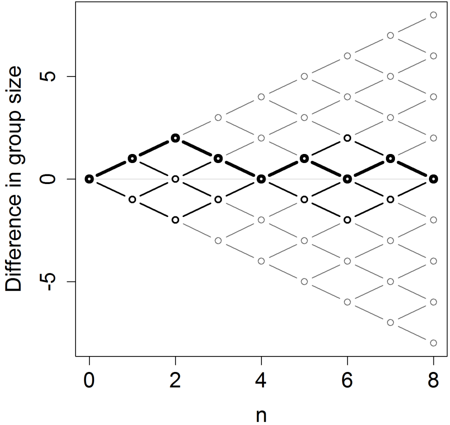
$$b_i = \frac{(i-1)\,\vartheta}{N} \quad \text{with } \vartheta \in \mathbb{R} \text{ and } i \in \{1, 2, \ldots, N\}\ .$$

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

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 of $\mathbf{T}_j$ in case of selection bias

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$1-\beta_0(\mathbf{T}_j)$	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 of $\mathbf{T}_j$ in case of a linear time trend
- $1 - \beta_{TT}(\mathbf{T}_j)$:= Power of $\mathbf{T}_j$ in case of a linear time trend

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$1 - \beta_{TT}(\mathbf{T}_j)$	$1 - \beta_0(\mathbf{T}_j)$	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 $N = 4$

Investigatd settings for the balancing behavior:

- $(1 - \beta) = 0.8$
- $\mu_E - \mu_C = 5.65$

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$1 - \beta_{TT}(\mathbf{T}_j)$	$1 - \beta_0(\mathbf{T}_j)$	overall
1	CCEE	$1/6$	0.047	0.060	0.842	0.800	
2	CECE	$1/6$	0.138	0.047	0.792	0.800	
3	ECCE	$1/6$	0.060	0.043	0.755	0.800	
4	CEEC	$1/6$	0.060	0.043	0.755	0.800	
5	ECEC	$1/6$	0.138	0.047	0.734	0.800	
6	EECC	$1/6$	0.047	0.060	0.730	0.800	
	average value:		0.081	0.050	0.768	0.800	

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 procedurs?

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$1-\beta_0(\mathbf{T}_j)$	overall
1	CCEE	$1/6$	0.047	0.060	0.842	0.800	?
2	CECE	$1/6$	0.138	0.047	0.792	0.800	?
3	ECCE	$1/6$	0.060	0.043	0.755	0.800	?
4	CEEC	$1/6$	0.060	0.043	0.755	0.800	?
5	ECEC	$1/6$	0.138	0.047	0.734	0.800	?
6	EECC	$1/6$	0.047	0.060	0.730	0.800	?
	average value:		0.081	0.050	0.768	0.800	?

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
4	$\beta_0(\mathbf{T})$	0.20	0.25

FP7 HEALTH 2013 - 602552

Properties of desirability scores

- Desirability scores take values in the interval $[0, 1]$.
- Desirability scores can be summarized with the geometric mean:

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

- The geometric means serves as linked assessment criterion.

FP7 HEALTH 2013 - 602552

Properties of desirability scores

- Desirability scores take values in the interval $[0, 1]$.
- Desirability scores can be summarized with the geometric mean:

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

- The geometric means serves as linked assessment criterion.
- Weights should be chosen dependent on the planned trial.
- Heuristical approach:
  Distribute the weight uniformly on selection bias, chronological bias, and balancing behavior.

  $\Rightarrow \omega_1 = 1/3$, $\omega_2 = \omega_3 = 1/6$, and $\omega_4 = 1/3$

FP7 HEALTH 2013 - 602552

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

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$1-\beta_0(\mathbf{T}_j)$	$d_4(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
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	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$1-\beta_0(\mathbf{T}_j)$	$d_4(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
1	EECC	1/6	0.047	1.000	0.060	0.804	0.842	1.000	0.8000	1.000	0.964
2	ECEC	1/6	0.138	0.000	0.047	1.000	0.792	0.961	0.8000	1.000	0.000
3	CEEC	1/6	0.060	0.809	0.043	1.000	0.755	0.776	0.8000	1.000	0.893
4	ECCE	1/6	0.060	0.809	0.043	1.000	0.755	0.776	0.8000	1.000	0.893
5	CECE	1/6	0.138	0.000	0.047	1.000	0.734	0.668	0.8000	1.000	0.000
6	CCEE	1/6	0.047	1.000	0.060	0.897	0.730	0.649	0.8000	1.000	0.850
	average value:		0.081	0.603	0.050	0.935	0.768	0.805	0.8000	1.000	0.608

$$\bar{d}(\mathbf{T}_1) = \sqrt[3]{d_1(\mathbf{T}_1)} \cdot \sqrt[6]{d_2(\mathbf{T}_1)} \cdot \sqrt[6]{d_3(\mathbf{T}_1)} \cdot \sqrt[3]{d_4(\mathbf{T}_1)}$$
$$= \sqrt[3]{1} \cdot \sqrt[6]{0.804} \cdot \sqrt[6]{1} \cdot \sqrt[3]{1}$$
$$= 0.964$$

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

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$1-\beta_0(\mathbf{T}_j)$	$d_4(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
1	EECC	$1/6$	0.047	1.000	0.060	0.804	0.842	1.000	0.8000	1.000	0.964
2	ECEC	$1/6$	0.138	0.000	0.047	1.000	0.792	0.961	0.8000	1.000	0.000
3	CEEC	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.8000	1.000	0.893
4	ECCE	$1/6$	0.060	0.809	0.043	1.000	0.755	0.776	0.8000	1.000	0.893
5	CECE	$1/6$	0.138	0.000	0.047	1.000	0.734	0.668	0.8000	1.000	0.000
6	CCEE	$1/6$	0.047	1.000	0.060	0.804	0.730	0.649	0.8000	1.000	0.897
	average value:		0.081	0.603	0.050	0.935	0.768	0.805	0.8000	1.000	0.608

$$\varnothing \bar{d}(\mathbf{T}) = 1/6\,(0.964 + 0 + 0.893 + 0.893 + 0 + 0.897)$$
$$= 0.608$$

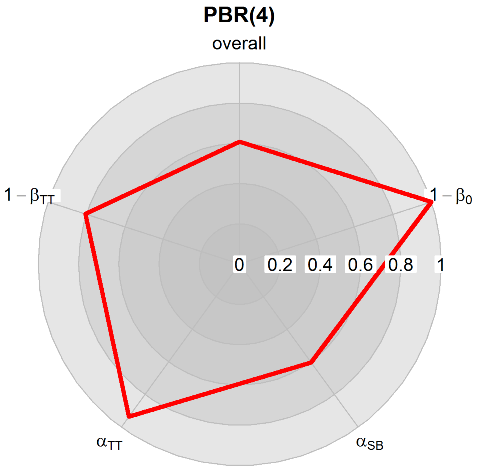
Assessment of PBR(4) with $N = 4$

j	$\mathbf{T}_j$	$P(\mathbf{T}_j)$	$\alpha_{SB}(\mathbf{T}_j)$	$d_1(\mathbf{T}_j)$	$\alpha_{TT}(\mathbf{T}_j)$	$d_2(\mathbf{T}_j)$	$1-\beta_{TT}(\mathbf{T}_j)$	$d_3(\mathbf{T}_j)$	$1-\beta_0(\mathbf{T}_j)$	$d_4(\mathbf{T}_j)$	$\bar{d}(\mathbf{T}_j)$
1	EECC	1/6	0.047	1.000	0.060	0.804	0.842	1.000	0.8000	1.000	0.964
2	ECEC	1/6	0.138	0.000	0.047	1.000	0.792	0.961	0.8000	1.000	0.000
3	CEEC	1/6	0.060	0.809	0.043	1.000	0.755	0.776	0.8000	1.000	0.893
4	ECCE	1/6	0.060	0.809	0.043	1.000	0.755	0.776	0.8000	1.000	0.893
5	CECE	1/6	0.138	0.000	0.047	1.000	0.734	0.668	0.8000	1.000	0.000
6	CCEE	1/6	0.047	1.000	0.060	0.804	0.730	0.649	0.8000	1.000	0.897
	average value:		0.081	0.603	0.050	0.935	0.768	0.805	0.8000	1.000	0.608

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

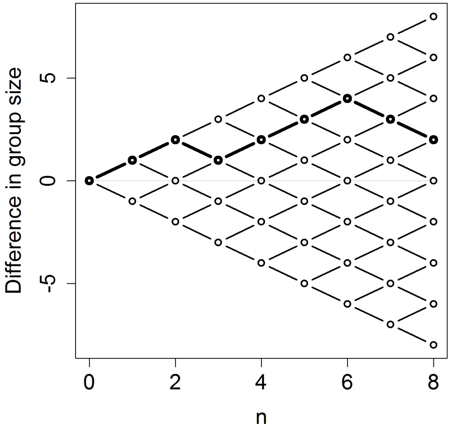
FP7 HEALTH 2013 - 602552

Radar plot



- PBR(4) is . . .
  - good in handling the assumed linear time trend
  - susceptible to selection bias
  - perfect with respect to its balancing behavior

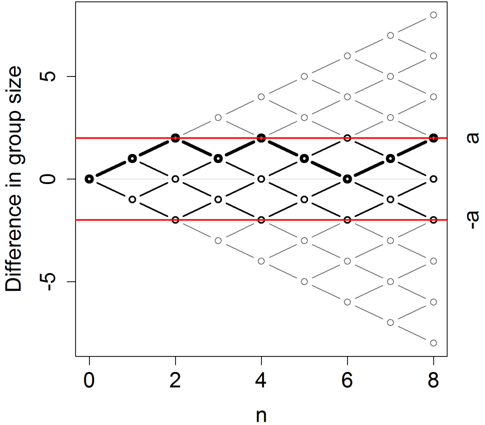
Complete Randomization



- Fair coin toss for each patient allocation

CR: Complete Randomization

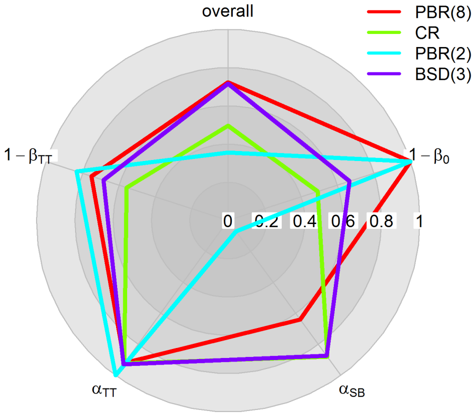
Big Stick Design (Soares and Wu, 1983)



- Fair toin toss with an imbalance boundary

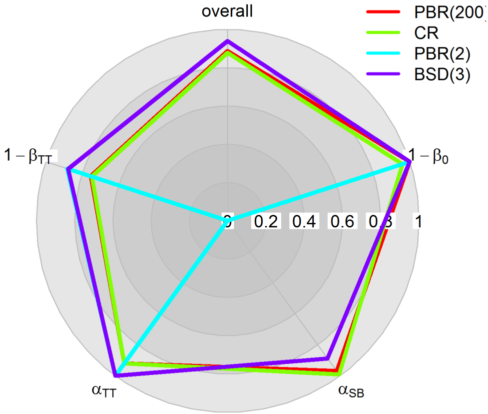
BSD(2): Big Stick Design with tolerated imbalance boundary 2

Comparison for $N = 8$



- PBR(2) seems to be very susceptible to selection bias
- CR has worse balancing behavior
- BSD(3) and PBR(8) manage the investigated criteria the best

Comparison for $N = 200$



- PBR(2) far too susceptible to selection bias
- CR and PBR(200) have nearly the same behavior
- BSD(3) manages the investigated criteria the best

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

Conclusion

- Randomization procedures differ in terms of their susceptibility to selection bias, chronological bias, and balancing behavior.
- 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.

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. Available from:* `http://www.ich.org`.

Kennes, L. N., E. Cramer, R.-D. Hilgers, and N. Heussen (2011). The impact of selection bias on test decision in randomized clinical trials. *Statistics in Medicine 30*, 2573–2581.

Langer, S. (2014). *The modified distribution of the t-test statistic under the influence of selection bias based on random allocation rule*. master thesis, RWTH Aachen University.

References II

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

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

Randomized Permuted Block Randomization

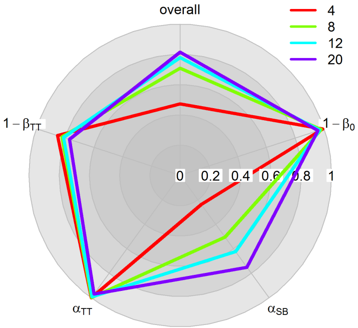


Figure: $N = 20$

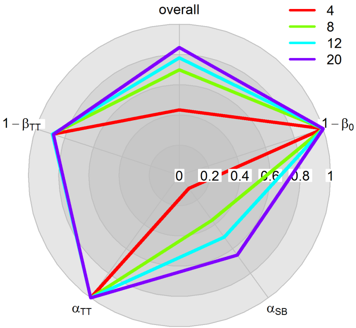


Figure: $N = 200$

FP7 HEALTH 2013 - 602552

Randomized Truncated Binomial Design

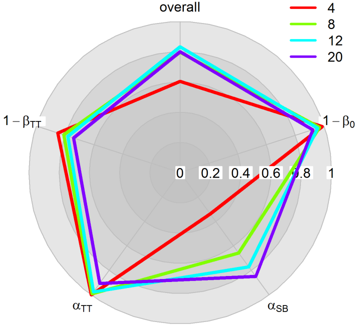


Figure: $N = 20$

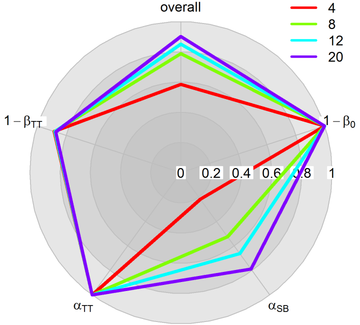


Figure: $N = 200$

FP7 HEALTH 2013 - 602552

Comparison

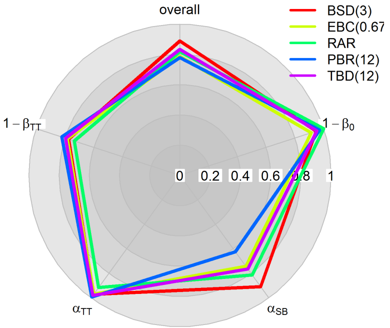


Figure: $N = 20$

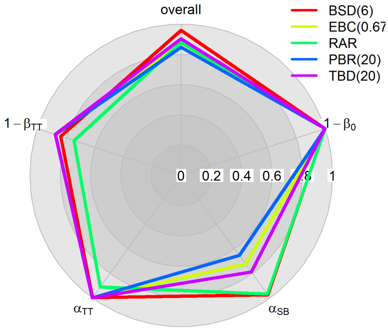


Figure: $N = 200$

FP7 HEALTH 2013 - 602552

Loss in treatment estimation (Atkinson, 2001)

Under the assumption of the following model

$$\mathbf{Y} = \begin{pmatrix} 1 & T_1 \\ 1 & T_2 \\ \vdots & \vdots \\ 1 & T_N \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

we can compute $\mathrm{Var}(\hat{\theta}_1)$ as follows:

$$\mathrm{Var}(\hat{\theta}_1) = \frac{\sigma^2}{N - \frac{D(N,\mathbf{T})^2}{N}} = \frac{\sigma^2}{N - L(\mathbf{T})},$$

where $D(N, \mathbf{T})$ defines the imbalance in group sizes at the end of the clinical study and $L(\mathbf{T})$ the loss.

FP7 HEALTH 2013 - 602552

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) :

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

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

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

FP7 HEALTH 2013 - 602552

Right-sided Derringer-Suich desirability function

Definition (Derringer and Suich (1980)):

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

TV: Target Value USL: Upper Specification Limit

i	Criterion$_i$ (c_i)	TV$_i$	USL$_i$
1	$L(\mathbf{T})$	0	1
2	$CG(\mathbf{T})$	0.5	0.75

FP7 HEALTH 2013 - 602552

Randomized Permuted Block Randomization

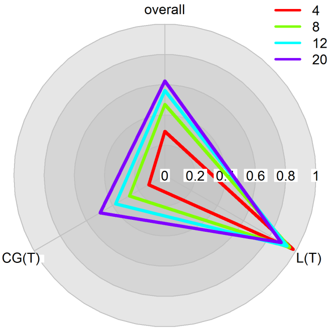


Figure: $N = 20$

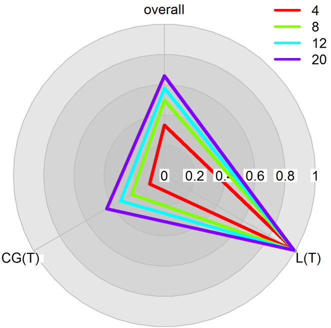


Figure: $N = 200$

Randomized Truncated Binomial Design

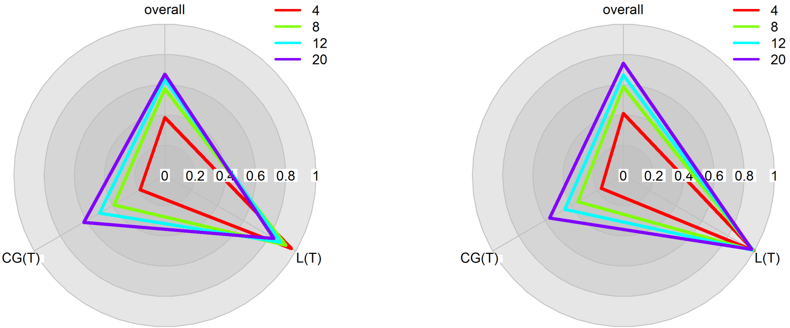


Figure: $N = 20$ Figure: $N = 200$

FP7 HEALTH 2013 - 602552

Comparison

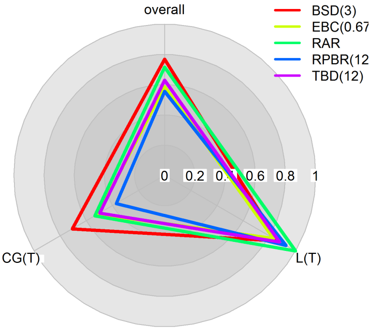


Figure: $N = 20$

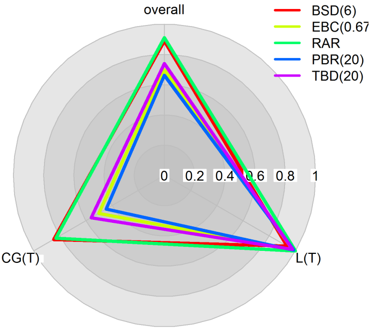
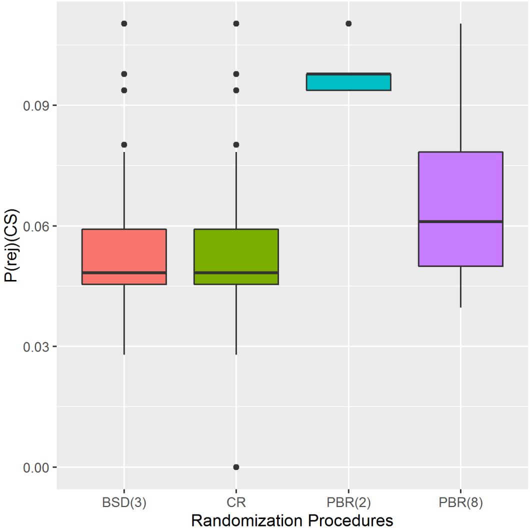


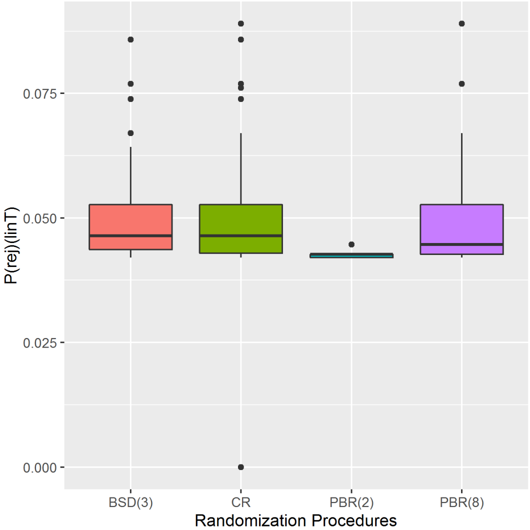
Figure: $N = 200$

Distribution of α_{SB} with $N = 8$



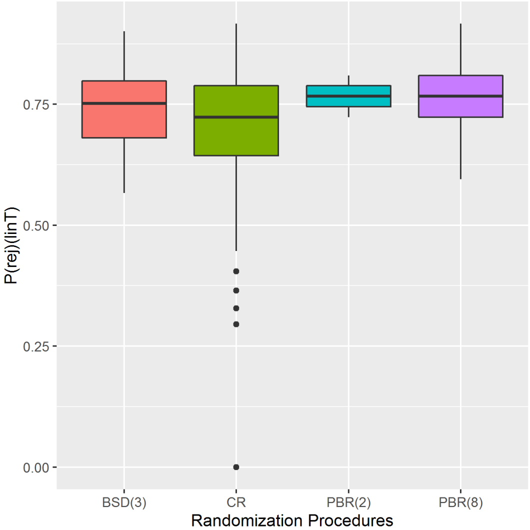
- Settings:
  - $\eta = \Delta_0/4 = 0.6$
  - $\alpha_0 = 0.05$
- Results:
  - PBR(2) most susceptible to selection bias
  - CR and BSD(3) have the best performance

Distribution of α_{CB} for $N = 8$



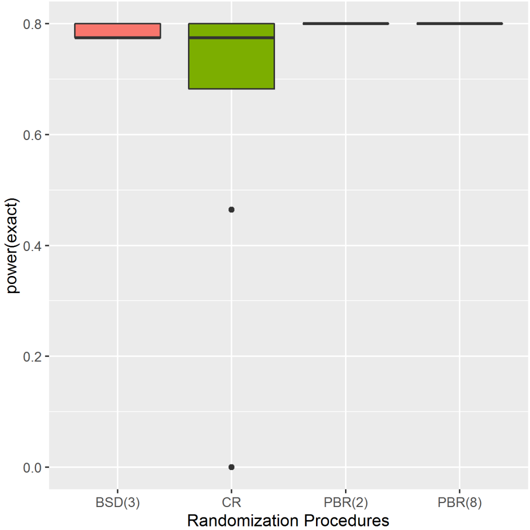
- Settings:
  - $\vartheta = 1$
  - $\alpha_0 = 0.05$
- Results:
  - All sequences of PBR(2) are conservative
  - CR, BSD(3), and PBR(8) have a similar performance

Distribution of $(1 - \beta_{SB})$ with $N = 8$



- Settings:
  - $\vartheta = 1$
  - $(1 - \beta_0) = 0.8$
- Results:
  - PBR(2) has lowest variance
  - CR has greatest variability and outliers are possible

Distribution of $(1 - \beta_0)$ with $N = 8$



- Settings:
  - $(1 - \beta_0) = 0.8$
- Results:
  - PBR maintains the planned power
  - BSD controlls the loss in power, CR does not

FP7 HEALTH 2013 - 602552