Choosing a suitable randomization procedure with randomizeR

Diane Uschner David Schindler Ralf-Dieter Hilgers

RWTH Aachen University

November 2, 2015

FP7 HEALTH 2013 - 602552

Acknowledgment

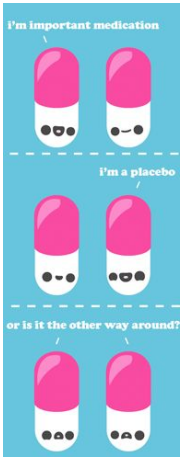
This research is part of the **IDeAI project** which has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under Grant Agreement no 602552.
Please visit

`http://www.ideal.rwth-aachen.de/`

to get the latest news.

Introduction

- Randomized Controlled Clinical Trial with $K \geq 2$ treatment arms
- Restricted Randomization is used for the allocation of treatments to patients.
- Aim: Choose suitable randomization procedure according to problems that might occur during the trial
- Propose a tool for the design of a clinical trial to
  - Assess and compare randomization procedures **sequence wise** wrt *issues* (e.g. selection bias)
  - Calculate the exact distribution of the issue (e.g. distribution of the type-I-error for the sequences).



FP7 HEALTH 2013 - 602552

Aim: Exact distribution of the issue

Example: Exact distribution of the type-I-error in case the responses are influenced by selection bias (Convergence Strategy).

	Sequence	Probability	P(rej)(CS)
1	BBAA	0.1666667	0.04229902
2	BABA	0.1666667	0.18880215
3	ABBA	0.1666667	0.04972876
4	BAAB	0.1666667	0.04972876
5	ABAB	0.1666667	0.18880215
6	AABB	0.1666667	0.04229902

Suitable Randomization Procedure

Process

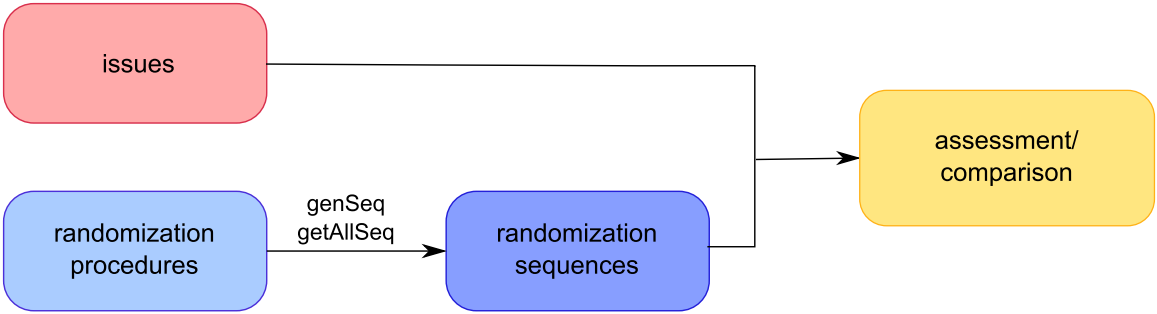
1. Identify constraints that impact the validity of the trial.
2. Define **issues** that measure the constraint.
3. Assess randomization procedures according to the issues.
4. Select appropriate randomization procedure on the basis of the assessment.

Definition of issue

An *issue* is a criterion for the assessment of randomization procedures that can be measured for each randomization sequence.

FP7 HEALTH 2013 - 602552

Structure of the package

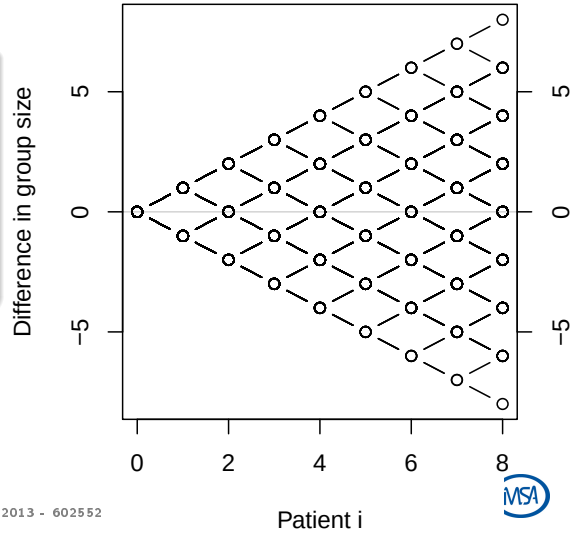


Randomization procedure

A randomization procedure $\mathcal{M}$ is a probability distribution on $\Gamma = \{0,1\}^N$. $t \in \Gamma$ is called randomization sequence. $\mathcal{M}$ *produces* the sequences

$$\Gamma_{\mathcal{M}} = \{t \in \Gamma \mid \mathbb{P}_{\mathcal{M}}(t) \neq 0\}$$

```
install.packages('randomizeR')
library(randomizeR)
N<-8
```



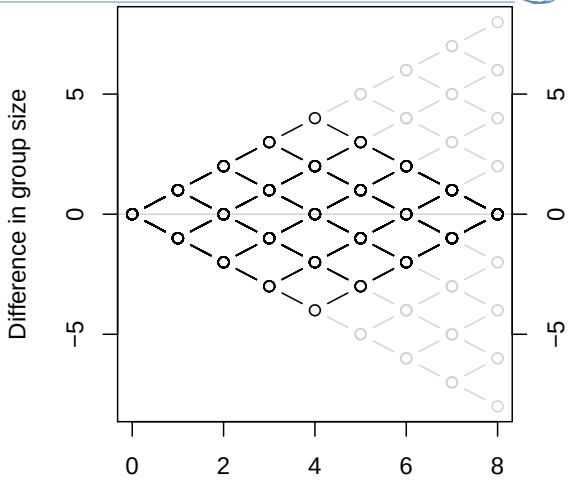
FP7 HEALTH 2013 - 602552

Random Allocation Rule

Equally probable final balance sequences:

$$\mathbb{P}_{RAR}(t) = \begin{cases} \binom{N}{N/2}^{-1} & \sum_{i=1}^{N}(2 \cdot t_i - 1) = 0 \\ 0 & \text{else.} \end{cases}$$

```
rarPar(N)
```



FP7 HEALTH 2013 - 602552

Permuted Block Randomization

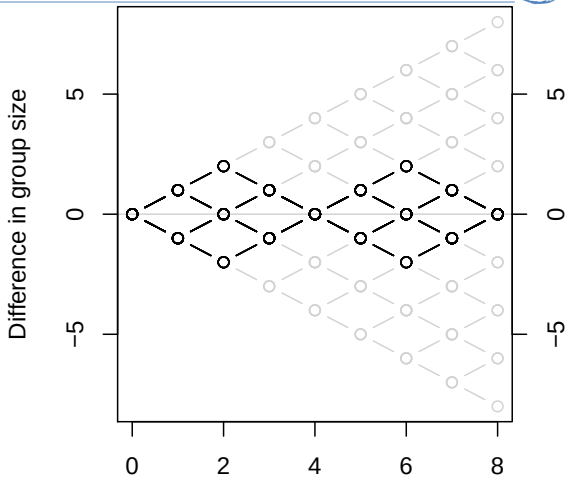
Equally probable balance sequences that attain balance after each block:

$$\mathbb{P}_{PBR}(t) = \begin{cases} \binom{k}{k/2}^{-N/k} & \sum_{i=1}^{j \cdot k}(2 \cdot t_i - 1) = 0 \\ 0 & \text{else.} \end{cases}$$

for $j = 1, \ldots, N/k$.

```
k <- 4 #block length

bc <- rep(k, N/k) #block constellation
pbrPar(bc)
```



FP7 HEALTH 2013 - 602552

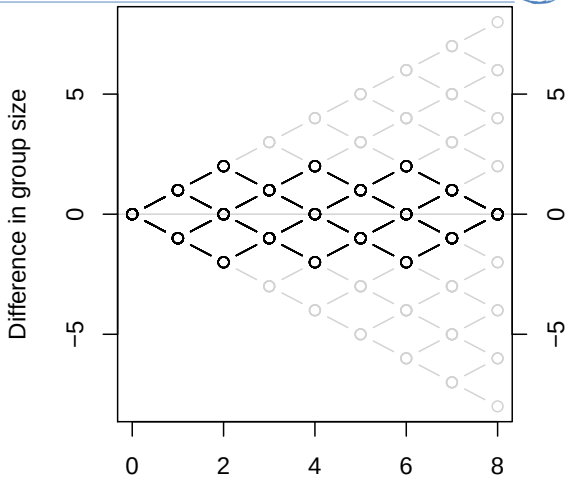
Maximal Procedure (Berger (2005))

Equally probable final balance sequences that do not exceed an imbalance boundary b.

$$\mathbb{P}_{MP}(t) = \begin{cases} \frac{1}{|\Gamma_{MP}|} & D_N = 0, \forall i: \ |D_i| \leq b \\ 0 & \text{else.} \end{cases}$$

```
b <- 2

mpPar(N,b)
```



FP7 HEALTH 2013 - 602552

Big Stick Design

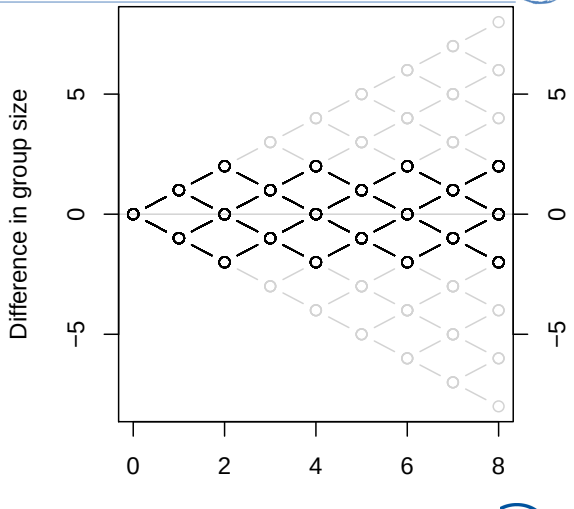
Toss a fair coin in until you hit the imbalance boundary. Then make a deterministic allocation.

$$\mathbb{P}_{BSD}(t) = \begin{cases} 0.5^{N-da} & \sum_{i=1}^{N} |2 \cdot t_i - 1| \leq b \\ 0 & \text{else.} \end{cases}$$

with imbalance boundary b and number of deterministic allocations $da := |\{j : \sum_{i=1}^{j} t_i = b\}|$.

```
b <- 2

bsdPar(N,b)
```



FP7 HEALTH 2013 - 602552

Let's get random!

`createParam()`	Creates a `<.>Par` object according to user input.
`createSeq()`	Generates a random sequence according to user input.
`genSeq()`	Generate a random sequence from a `<.>Par object`.
`getAllSeq(myPar)`	Compute $\Gamma_{\mathcal{M}}$ for $N < 20$.
`getProb(seqs)`	Compute the theoretical probabilities for an object `seqs` of type `randSeq`.
`saveRand(seqs)`	Save the randomization protocol inluding a the randomization sequence(s) to `.csv`.

FP7 HEALTH 2013 - 602552

Performance of genSeq

Performance of generating 10^x RAR sequences, $x \in \{3, 4, 5, 6\}$.

```
system.time(genSeq(rarPar(100),10^3))

   user  system elapsed
   0.06    0.00    0.06

system.time(genSeq(rarPar(100),10^4))

   user  system elapsed
   0.70    0.00    0.71
```

```
system.time(genSeq(rarPar(100),10^5))

   user  system elapsed
   6.16    0.05    6.23

system.time(genSeq(rarPar(100),10^6))

   user  system elapsed
  62.95    0.44   63.48
```

Definition

An issue is a criterion for the assessment of randomization procedures that can be measured ***for each randomization sequence****.*

`selBias`	Represent exact rejection probability (size/ power) in case the responses are influenced by selection bias.
`corGuess`	Represent the proportion of correct guesses.
`chronBias`	Represent exact rejection probability (size/ power) in case the responses are influenced by chronological bias.
`setPower`	Represent the power for a given detectable effect and size.
`imbal`	Represent the imbalance in allocation numbers.

Table: Issues implemented in randomizeR

FP7 HEALTH 2013 - 602552

Model for the responses (unbiased

Response

Let E and C be treatments that influence a continuous outcome Y. For $i = 1, \ldots, N$, we write

$$Y_i \sim \mathcal{N}(\mu + d \cdot T_i, \sigma^2) \qquad (1)$$

where $d \in \mathbb{R}$ denotes the *treatment effect*, $\mu > 0$ the overall mean and $\sigma^2 > 0$ the equal but unknown variance. Y_i is called *response* of patient i. Higher values of Y are regarded as better.

Represent normal endpoints in randomizeR

```
normEndp(mu=c(0,0), sigma=c(1,1))
```

Hypothesis of no treatment effect

Test Model:

$$Y_i \sim \mathcal{N}(\mu + d \cdot T_i, \sigma^2)$$

Test the hypothesis **under model miss-specification**!

True Model:

$$Y_i \sim \mathcal{N}(\mu + d \cdot T_i + g(\theta, i), \sigma^2)$$

Null hypothesis

We test the null hypothesis that the expected effect of the experimental treatment does not differ from the expected effect of the control treatment

$$H_0 : d = 0$$

against the two-sided alternative that the expected treatment effects differ

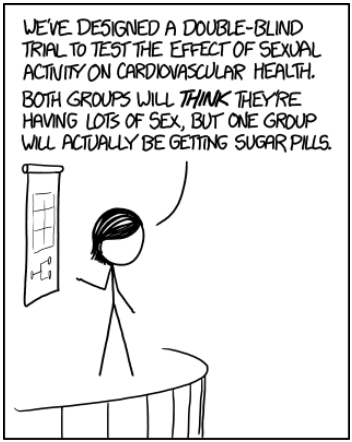
$$H_1 : d \neq 0$$

Selection bias under convergence strategy

Third order selection bias

- Trial is randomized.
- Allocation list is concealed.
- But: the investigator can guess the next treatment assignment due to
  - unmasking of past assignments (e.g. due to side effects).
  - restrictions of the randomization procedure.
- Investigator can deny enrollment due to soft inclusion criteria.

Berger (2005)

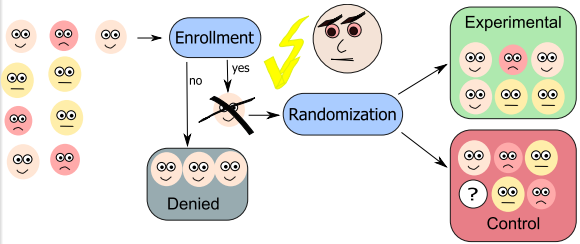


FP7 HEALTH 2013 - 602552

Selection bias under convergence strategy

Third order selection bias

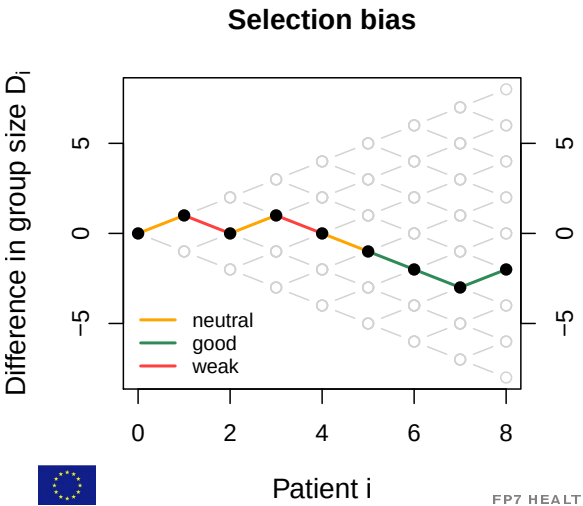
- Trial is randomized.
- Allocation list is concealed.
- But: the investigator can guess the next treatment assignment due to
  - unmasking of past assignments (e.g. due to side effects).
  - restrictions of the randomization procedure.
- Investigator can deny enrollment due to soft inclusion criteria.



Berger (2005)

FP7 HEALTH 2013 - 602552

Biasing Policy



Choose patient $i+1$ with expected response

$$E(Y_{i+1}) = \begin{cases} \mu - \eta & N_E(i) > N_C(i) \\ \mu & N_E(i) = N_C(i) \\ \mu + \eta & N_E(i) < N_C(i) \end{cases}$$

with *selection effect* $\eta > 0$.

Proschan (1994)

FP7 HEALTH 2013 - 602552

Given the randomization sequence $t \in \Gamma$ and using Student's t-test in order to test the hypothesis $H_0 : d = 0$ of no treatment effect, the test statistic

$$S = \frac{\sqrt{\frac{N_E N_C}{N_E+N_C}}\,(\tilde{y}_E - \tilde{y}_C)}{\frac{1}{N_E+N_C-2}\left(\sum_{i=1}^{N} t_i(y_i - \tilde{y}_E)^2 + \sum_{i=1}^{N}(1-t_i)(y_i - \tilde{y}_C)^2\right)}$$

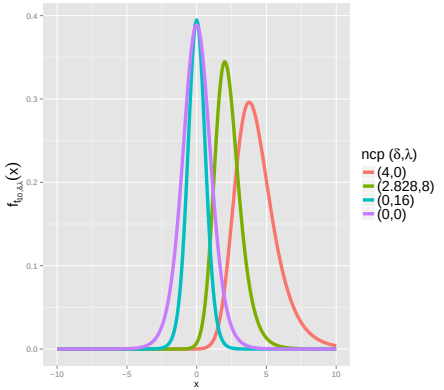
with $\tilde{y}_E = \frac{1}{N_E}\sum_{i=1}^{N} y_i t_i$, $\tilde{y}_C = \frac{1}{N_C}\sum_{i=1}^{N} y_i(1-t_i)$ and $N = N_E + N_C$ is doubly noncentrally t-distributed with parameters δ and λ.

Exact rejection probability in case of selection bias(2)

The noncentrality parameters can be determined as follows

$$\delta = \eta \sqrt{\frac{1}{\sigma^2 N}} \sum_{i=1}^{N} 2 \cdot (t_i - \frac{1}{2}) \cdot \text{sign}(D_{i-1})$$

$$\lambda = \frac{\eta^2}{\sigma^2} \left(\sum_{i=1}^{N} \text{sign}(D_{i-1})^2 - \frac{2}{N} \Big(\sum_{i=1}^{N} t_i \cdot \text{sign}(D_{i-1}) \Big)^2 - \frac{2}{N} \Big(\sum_{i=1}^{N} (1 - t_i) \cdot \text{sign}(D_{i-1}) \Big)^2 \right)$$



Langer (2014) Figure: Doubly noncentral t-distribution, $N = 12$

FP7 HEALTH 2013 - 602552

Assess randomization procedure with randomizeR

```
pbr <- getAllSeq(pbrPar(bc))

sb <- selBias("CS",eta = 0.6, method = "exact")

endp <- normEndp(mu=c(0,0), sigma = c(1,1))

assess(pbr, sb, endp = endp)
```

Comparison of randomization procedures

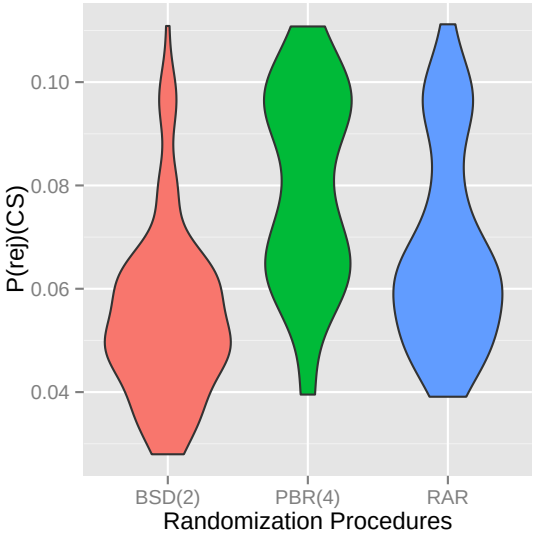
```
pbr <- getAllSeq(pbrPar(bc))

mp <- getAllSeq(mpPar(N,2))

bsd <- getAllSeq(bsdPar(N,2))

C <- compare(sb, pbr, mp, bsd,
                          endp = endp)

plot(C)
```



Conclusions

- randomizeR makes it easy to generate randomization sequences for a large number of randomization procedures.
- Easy to assess and compare randomization procedures for a large number of issues.
- Assessment should be done before conducting a clinical trial.

FP7 HEALTH 2013 - 602552

Want some more?

Try it yourself! Just type

```
install.packages("randomizeR")
library("randomizeR")
vignette("comparison-example")
```

in your R command line.

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.

Berger, V. W. (2005). *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. Wiley.

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

ICH E9. Statistical principles for clinical trials, 1998. *Current version dated 5 Februrary 1998. Last access in September 2014. Available from:* `http://www.ich.org`.

S. Langer. The modified distribution of the t-test statistic under the influence of selection bias based on random allocation rule. Master's thesis, RWTH Aachen, 2014.

Lehmann (1975) Nonparametrics: Statistical Methods Based on Ranks

References II

Lachin, J. M. (1988). Statistical properties of randomization in clinical trials. *Controlled Clinical Trials 9*, 289–311.

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

Rosenberger, W.F., and Lachin, J.M. (2016). Randomization in clinical trials - Theory and practice. Wiley.

Tamm, M., and Hilgers, R.-D. (2014). Chronolgical bias in randomized clinical trials under different types of unobserved time trends *Meth. Inf. Med. 6*, 501–510.

Schindler, D., and Uschner, D. (2015). randomizeR: Randomization for Clinical Trials R package version 1.0.

FP7 HEALTH 2013 - 602552