A software tool for the design and analysis of small population group trials

Diane Uschner David Schindler Ralf-Dieter Hilgers

September 24th, 2015

FP7 HEALTH 2013 - 602552

Motivation

- Situation: *Randomized* controlled clinical trial, two treatment arms.
- Aim: Estimate the treatment effect.
- Randomization is used in order to
  - balance (unknown) prognostic factors
  - to control selection bias (in combination with blinding)
- **But: No sound scientific basis exists for choosing a randomization procedure**
- Incorporate the constraints that peril the estimation of the treatment effect in the design stage.
- We present a software tool that provides a basis for choosing a randomization procedure.

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.

Integrated DEsign and AnaLysis of small population group trials

There exist more than 7000 rare diseases worldwide and the European Society of Paediatric Oncology stated that 75% of rare diseases affect children and 30% of rare disease patients die before the age of five.

Usual statistical methods for proving efficacy and safety of therapies fail to provide cost-efficient and reliable results in small populations.

There is a pressing need to integrate a broad range of innovative methodologies improving clinical trials in the setting of small sample population groups (SPG).

The objective of this research is to produce methods of general applicability irrespective of indication by Integrated DEsign and AnaLysis of clinical trials in SPG (IDeAl) through a multidisciplinary closely collaborating consortium of researchers from European universities, research institutes and industry.

The consortium will work in 11 WPs, focussed on assessment of randomization procedures, extrapolating dose-response information, investigation of adaptive designs, optimal designs in mixed models, pharmacogenetic designs, simulation of clinical trials, genetic factors influencing the response, decision analysis and biomarker surrogate endpoints as well as WPs on project management and dissemination of results. Relevant stakeholder concerns (patient needs, regulatory issues, reimbursement, clinical feasibility) will be monitored by a Clinical Scientific Advisory Board. Because of its integrative structure, this research program extends previous approaches, which focus on a certain methodology only. In its totality, the WPs constitute a logically coherent set of methodologies that is of sufficient breadth to tackle theseimportant, multidisciplinary challenges. By combining, enhancing and developing different statistical methodologies and assessment methods, this research program will impact the scientific discussion in promoting efficient statistical methodology for clinical trials in SPG, also in view of existing regulatory guidance in the EU.

Learn about our Work

Work Packages

NEWS

IDeAl researchers discuss frontiers of confirmatory inference in small populations at MCP 2015
August 28, 2015 - 12:35

IDeAl researchers disseminate latest findings at ISCB 2015
August 24, 2015 - 09:16

Gerald Hlavin receives Arthur Linder Award
August 10, 2015 - 16:18

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

DOCUMENTS FOR DOWNLOAD

In this section, you can download the files that have been produced as official output in the project.

- Articles in peer-reviewed journals
- Presentations
- Input to Regulatory Documents
- Short Courses
- Conference Posters
- Comments
- Conferences
- Statistical Software Programs

  - **R package for testing similarity of dose response curves** [published 2015-09-16]
  - **R package on Randomization for clinical trials** [published 2015-09-09]
  - **R package "MIXFIM" for the evaluation and optimization of the Fisher Information Matrix in NonLinear Mixed Effect Models using Markov Chains Monte Carlo for both discrete and continuous data** [published 2015-09-01]
  - **R package for the estimation of within subject correlations based on linear mixed effects models** [published 2015-04-29]
  - **R package on the Prediction of Therapeutic Success** [published 2015-04-27]
  - **R package for dimensionality reduction via variables clustering** [published 2014-12-06]
  - **R-Code to calculate worst case type I error inflation in multiarmed clinical trials** [published 2014-04-23]
  - **R package on Surrogate Markers** [published 2014-03-18]

How to get randomizeR on my computer? - it's easy!

```
install.packages("randomizeR")
library("randomizeR")
```

randomizeR: Randomization for Clinical Trials

This tool enables the user to choose a randomization procedure based on sound scientific criteria. It comprises the generation of randomization sequences as well the assessment of randomization procedures based on carefully selected criteria. Furthermore, randomizeR provides a function for the comparison of randomization procedures.

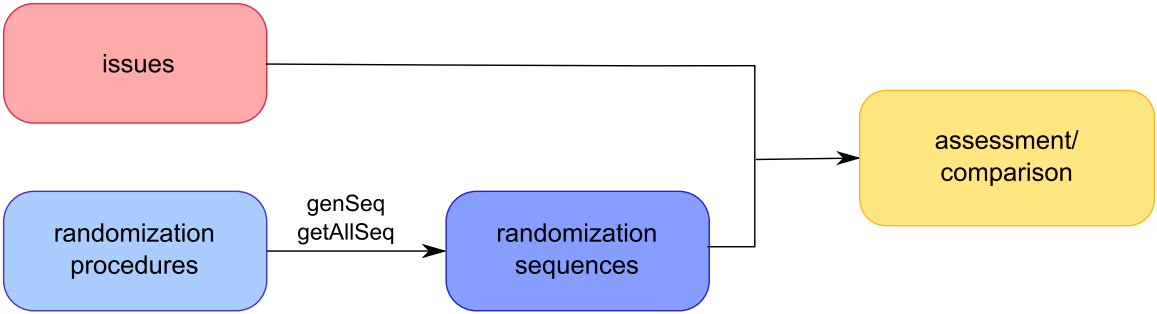
Version:	1.0
Depends:	R (≥ 3.2.2), methods, ggplot2
Suggests:	testthat, knitr, quantreg
Published:	2015-09-09
Author:	Thi Mui Pham [ctb], David Schindler [aut], Diane Uschner [aut, cre]
Maintainer:	Diane Uschner <duschner at ukaachen.de>
License:	GPL (≥ 3)
NeedsCompilation:	no
Materials:	README
CRAN checks:	randomizeR results

Downloads:

Reference manual:	randomizeR.pdf
Vignettes:	Comparing randomizaton procedures
Package source:	randomizeR_1.0.tar.gz
Windows binaries:	r-devel: not available, r-release: not available, r-oldrel: not available
OS X Snow Leopard binaries:	r-release: not available, r-oldrel: not available
OS X Mavericks binaries:	r-release: not available

Schindler and Uschner (2015)

Structure of the package



Randomization Sequence

Let N be the total sample size.

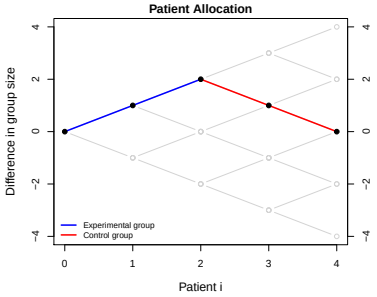
Definition

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

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

A *randomization sequence* is a vector $t \in \Gamma$ with

$$t_i = \begin{cases} 0 & \text{if subject } i \text{ is assigned to the control group,} \\ 1 & \text{if subject } i \text{ is assigned to the experimental group.} \end{cases}$$



Rosenberger and Lachin (2016)

Several randomization procedures exist.

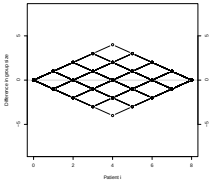
Formal definition:

(a) Random allocation rule:
$\mathbb{P}_{RAR}(t) = \binom{N}{N/2}^{-1}, t \in \Gamma_{RAR}$

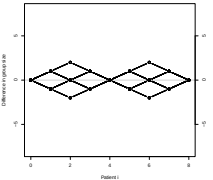
(b) Permuted block randomization:
$\mathbb{P}_{PBR}(t) = \binom{k}{k/2}^{-N/k}, t \in \Gamma_{PBR}$

(c) Maximal procedure:
$\mathbb{P}_{MP}(t) = |\Gamma_{MP(N,b)}|^{-1}, t \in \Gamma_{MP(N,b)}$

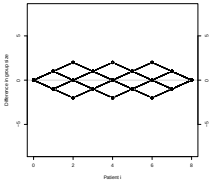
(d) Big Stick design: $\mathbb{P}_{BSD}(t) = 0.5^{N-da(t)}$,
$t \in \Gamma_{BSD}$, $da(t)$ = number of deterministic allocations of t



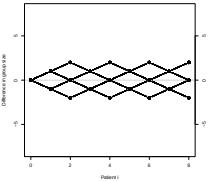
(a) Γ_{RAR}



(b) Γ_{PBR}



(c) Γ_{MP}



(d) Γ_{BSD}

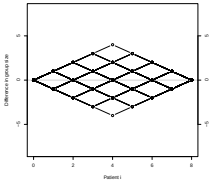
FP7 HEALTH 2013 - 602552

Several randomization procedures exist.

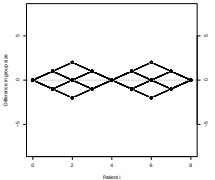
Sampling algorithms model the probability
$\Rightarrow$ sampled relative frequency is equal to theoretical probability.

Representation in `randomizeR`:

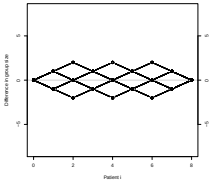
(a) Random allocation rule: `rarPar(N)`
(b) Permuted block randomization: `pbrPar(bc)`
(c) Maximal procedure: `mpPar(N,b)`
(d) Big Stick design: `bsdPar(N,b)`



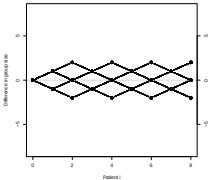
(a) Γ_{RAR}



(b) Γ_{PBR}



(c) Γ_{MP}



(d) Γ_{BSD}

FP7 HEALTH 2013 - 602552

Randomization procedures implemented in randomizeR

- Complete Randomization
- Random Allocation Rule
- Permuted Block Randomization
- Permuted Block Randomization with random block lengths
- Truncated Binomial Design (in blocks)
- Truncated Binomial Design with random block lengths
- Efron's Biased Coin Design
- Big Stick Design
- Maximal Procedure
- Hadamard Randomization
- Wei's Urn Design

Rosenberger and Lachin (2016)

FP7 HEALTH 2013 - 602552

Representation of randomization procedures

```
bsdPar(8,2, groups= c("E","C"))
```

```
Object of class "bsdPar"

design = BSD(2)
mti = 2
N = 8
groups = E C
```

Generate randomization sequences

```
bsd <- bsdPar(8,2, groups= c("E","C"))
genSeq(bsd)
```

```
Object of class "rBsdSeq"

design = BSD(2)
seed = 1425574311
N = 8
groups = E C
mti = 2

The sequence M:

1 E E C E C C E C
```

Generate randomization sequences

```
genSeq(bsd, 4)
```

```
Object of class "rBsdSeq"

design = BSD(2)
seed = 1425574311
N = 8
groups = E C
mti = 2

The first 3 of 4 sequences of M:

1 E E C E C C E C
2 C E E C C C E E
3 C E E C E C C E
...
```

Access the randomization sequences

```
getRandList(genSeq(bsd, 4))
```

```
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] "E"  "E"  "C"  "E"  "C"  "C"  "E"  "C"
[2,] "C"  "E"  "E"  "C"  "C"  "C"  "E"  "E"
[3,] "C"  "E"  "E"  "C"  "E"  "C"  "C"  "E"
[4,] "C"  "C"  "E"  "C"  "E"  "C"  "E"  "C"
```

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
   0.08    0.00    0.08

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

   user  system elapsed
   0.70    0.03    0.73
```

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

   user  system elapsed
   6.88    0.06    7.06

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

   user  system elapsed
  67.85    0.44   68.79
```

More functionality for randomization

`randomizeR` supports several functions for randomization procedures:

`createParam()`	Creates a `<.>Par` object according to user input.
`createSeq()`	Generates a random sequence according to user input.
`getAllSeq(myPar)`	Compute $\Gamma_{\mathcal{M}}$ for $N \leq 24$.
`getProb(seqs)`	Compute the theoretical probabilities for an object `seqs` of type `randSeq`.
`saveRand(seqs)`	Save the randomization protocol including the randomization sequence(s) to `.csv`.

Choose a **suitable** randomization procedure!

- Provide sound scientific basis for the selection of a tailored randomization procedure.
- Assumptions concerning the distribution of the responses may not be fulfilled.
- Biases may compromise the estimation of the treatment effect.
  - Selection bias
  - Chronological bias
  - Imbalances

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.

FP7 HEALTH 2013 - 602552

Choose a **suitable** randomization procedure!

- Provide sound scientific basis for the selection of a tailored randomization procedure.
- Assumptions concerning the distribution of the responses may not be fulfilled.
- Biases may compromise the estimation of the treatment effect.
  - Selection bias
  - Chronological bias
  - Imbalances

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.

FP7 HEALTH 2013 - 602552

randomizeR implements several issues.

Simple issues:

`corGuess`	Represent the proportion of correct guesses (i.e. predictability).
`imbal`	Represent the imbalance in allocation numbers (i.e. balance).

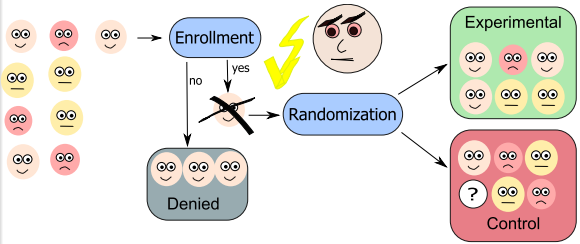
Sophisticated issues:

`selBias`	Represent exact rejection probability (size/ power) in case the responses are influenced by third order **selection bias**. Hilgers et al (2016),Proschan (1994)
`chronBias`	Represent exact rejection probability (size/ power) in case the responses are influenced by **chronological bias**. Tamm and Hilgers (2014), Rosenkranz (2011)
`setPower`	Represent the power for a given detectable effect and size. Lachin (1988)

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 (due to side effects).
  - restrictions of the randomization procedure.
- Investigator can deny enrollment due to soft inclusion criteria.



Berger (2005)

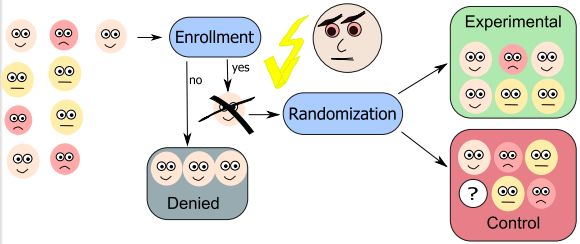
FP7 HEALTH 2013 - 602552

Selection bias under convergence strategy

Convergence strategy

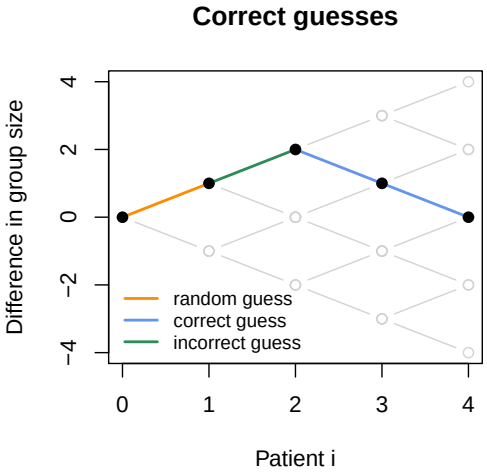
In the situation of third order selection bias, the investigator can count $N_E(i)$ and $N_C(i)$ respectively. $\Rightarrow$ Under the assumption that $N_E(N) = N_C(N)$ for all $t \in \Gamma$ it is opportune for the investigator to guess

$$g(t_{i+1}) = \begin{cases} 0 & N_E(i) > N_C(i) \\ Ber(0.5) & N_E(i) = N_C(i) \\ 1 & N_E(i) < N_C(i), \end{cases}$$



Blackwell and Hodges Jr. (1957)

Correct guesses - a measure for selection bias



Expected number of correct guesses

The *expected number of correct guesses* of a randomization sequence t is given by

$$CG(t) := \mathbb{E}(\#\{i = 1, \dots N : g(t_{i+1}) = t_{i+1}\}).$$

The expected proportion of correct guesses of t is

$$propCG(t) := \frac{CG(t)}{N}.$$

Blackwell and Hodges Jr. (1957)

FP7 HEALTH 2013 - 602552

Use all the information! Take the **sequence based approach**!

- Model selection bias, e.g. as the proportion of correct guesses.
- For each randomization sequence t, compute the proportion of correct guesses.
- Compute descriptive measures, s.a. mean and quantiles, weighting the proportion of correct guesses of each sequence with its probability of occurrence.

	Sequence	Probability	propCG(CS)
1	CCEE	0.1666667	0.625
2	CECE	0.1666667	0.750
3	ECCE	0.1666667	0.750
4	CEEC	0.1666667	0.750
5	ECEC	0.1666667	0.750
6	EECC	0.1666667	0.625

mean = 0.71

With randomizeR this is easy.

First, represent the issue of correct guesses:

```
cg <- corGuess(type = "CS")
```

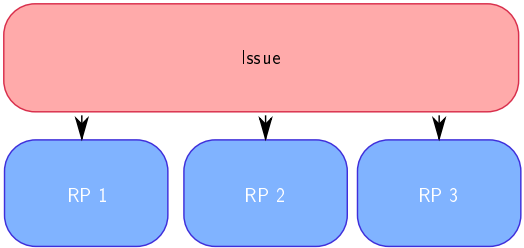
Then assess the sequences:

```
rarS <- genSeq(rarPar(6),10)
assess(rarS, cg)
```

```
Object of class "assessment"

design = RAR
N = 6
K = 2
groups = A B


The first 3 rows of 10 rows of D:

  Sequence Relative_Frequency propCG(CS)
1   BAABAB                0.1  0.7500000
2   BABBAA                0.1  0.6666667
3   AAABBB                0.1  0.5833333
...
```

FP7 HEALTH 2013 - 602552

Putting things together: compare



- Compare a number of randomization procedures according to one issue.

FP7 HEALTH 2013 - 602552

First, set the parameters:

```
bsd <- genSeq(bsdPar(100,5),1000)
mp <- genSeq(mpPar(100,5),1000)
rar <- genSeq(rarPar(100),1000)
pbr <- genSeq(pbrPar(rep(10,10)),1000)

cg <- corGuess(type = "CS")
```

Next, compare the randomization procedures according to selection bias:

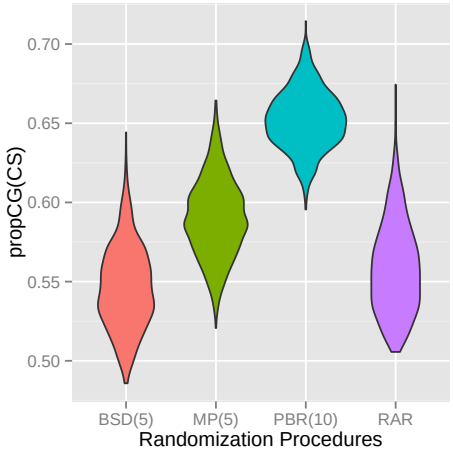
```
compare(cg, bsd, mp, rar, pbr)
```

```
Comparison for propCG(CS)

     BSD.5. MP.5.   RAR PBR.10.
mean  0.546 0.590 0.558   0.652
sd    0.026 0.025 0.028   0.019
max   0.645 0.665 0.675   0.715
min   0.485 0.520 0.505   0.595
x05   0.505 0.550 0.515   0.620
x25   0.525 0.575 0.535   0.640
x50   0.545 0.590 0.555   0.650
x75   0.565 0.605 0.575   0.665
x95   0.595 0.635 0.610   0.685
```

One picture is worth ten thousand words

Finally, visualize the comparison with

```
C <- compare(cg, bsd, mp, rar, pbr)
plot(C)
```



And it's fast, too!

Input:

```
bsd = genSeq(bsdPar(100,5),1000)
mp  = genSeq(mpPar(100,5),1000)
rar = genSeq(rarPar(100),1000)
pbr = genSeq(pbrPar(rep(10,10)),1000)
cg <- corGuess(type = "CS")

comp <- compare(cg, bsd, mp, rar, pbr)
```

Performance [sec]:

	test	elapsed	replications	average
5	CG	0.02	100	0.0002
3	RAR	6.15	100	0.0615
4	PBR	15.65	100	0.1565
2	MP	67.85	100	0.6785
1	BSD	68.56	100	0.6856
6	COMP	159.08	100	1.5908

Summary & Outlook

`randomizeR`

- establishes an open source, easily extendable, strongly structured framework for randomization.
- incorporates various randomization procedures and issues.
- provides a sound scientific basis for the assessment of randomization procedures according to pre-specified criteria.
- enables the user to chose a tailored randomization procedure.
- will lead to more reliable treatment effect estimates.

Summary & Outlook

`randomizeR`

- establishes an open source, easily extendable, strongly structured framework for randomization.
- incorporates various randomization procedures and issues.
- provides a sound scientific basis for the assessment of randomization procedures according to pre-specified criteria.
- enables the user to chose a tailored randomization procedure.
- will lead to more reliable treatment effect estimates.

Next Propose a unified criterion that comprises several issues

⇒ David Schindler: "Selecting an appropriate randomization procedure for a small population group trial on the basis of a linked optimization criterion".

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.

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.

Hilgers, R.-D. et al (2016). A general framework to assess randomization procedures under selection and chronological bias with respect to type I error probability. *(in preparation)*

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.

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

FP7 HEALTH 2013 - 602552

References II

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.

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.