How to use randomizeR to investigate randomization tests in the presence of selection bias

Diane Uschner

joint work with David Schindler

RWTH Aachen University

October 28, 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.

Table of Contents

1. Randomization with randomizeR

2. Randomization tests in the presence of selection bias

3. References

FP7 HEALTH 2013 - 602552

Table of Contents

1. Randomization with randomizeR

2. Randomization tests in the presence of selection bias

3. References

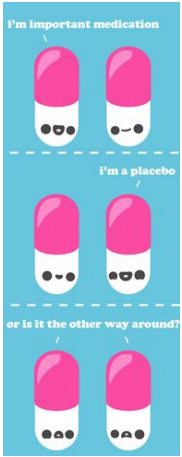
FP7 HEALTH 2013 - 602552

Controlled Clinical Trial

Experiment designed to determine whether the new treatment has a beneficial effect by comparing the patients receiving it with a group of patients receiving a control treatment.

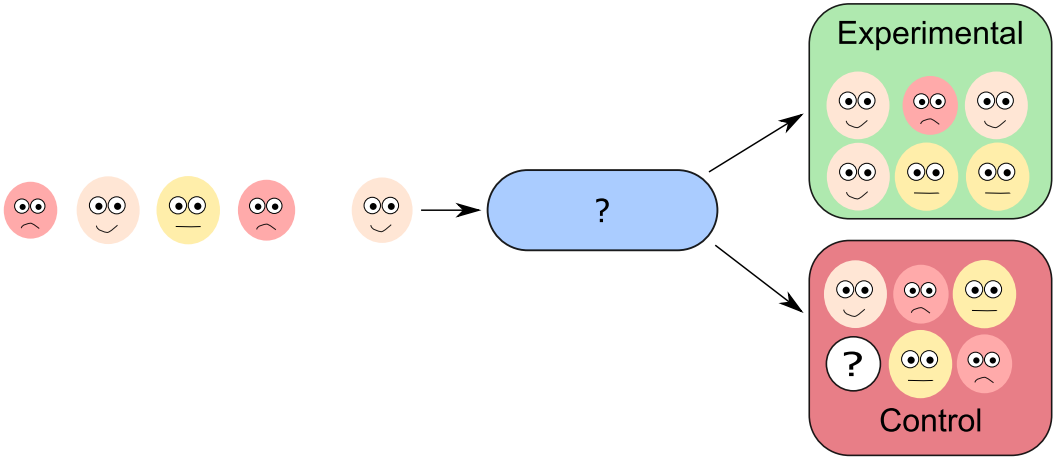
- Experimental and control group must be comparable.

Randomization is used to balance the variability of patients between treatment groups.



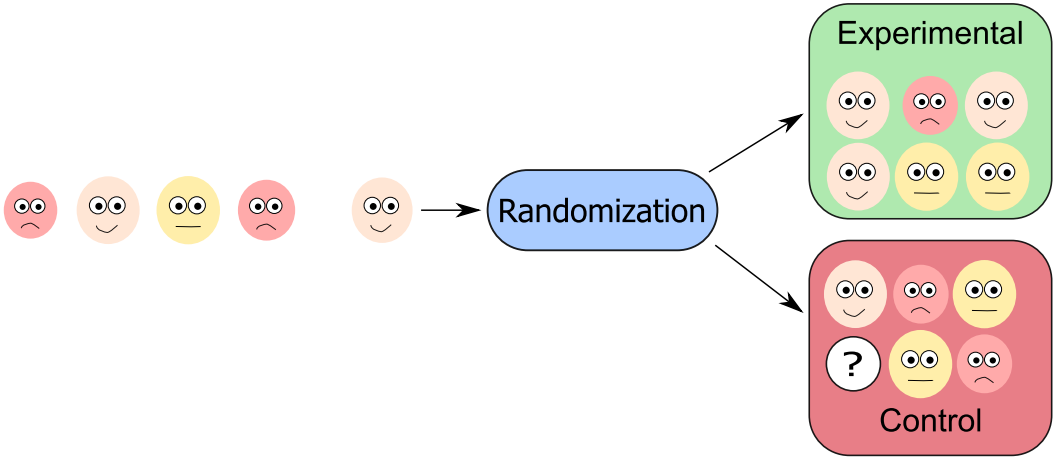
FP7 HEALTH 2013 - 602552

Allocation of patients to treatment groups



FP7 HEALTH 2013 - 602552

Allocation of patients to treatment groups



FP7 HEALTH 2013 - 602552

Randomization?

Which randomization procedure should we use?

FP7 HEALTH 2013 - 602552

Which randomization procedure should we use?

A suitable one!

FP7 HEALTH 2013 - 602552

Goal

Enable the practitioner to choose a suitable randomization procedure.

FP7 HEALTH 2013 - 602552

Goal

Enable the practitioner to choose a suitable randomization procedure.

Suitable Randomization Procedure

Process

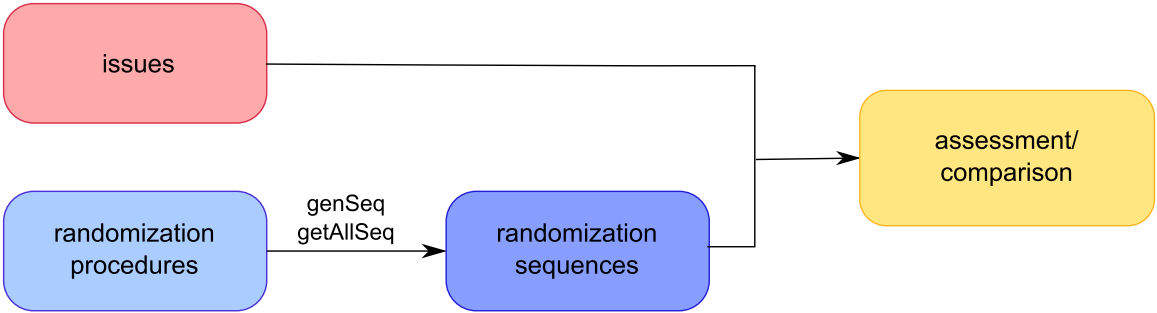
1. Identify constraints that impact the validity of the trial.
2. Define **issues** that measure the constraint.
3. Assess randomization procedures according to the issues.
4. Select appropriate randomization procedure on the basis of the assessment.

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



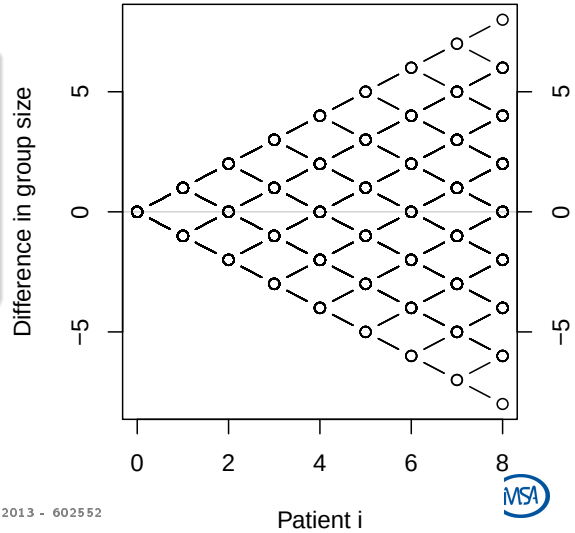
FP7 HEALTH 2013 - 602552

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
```



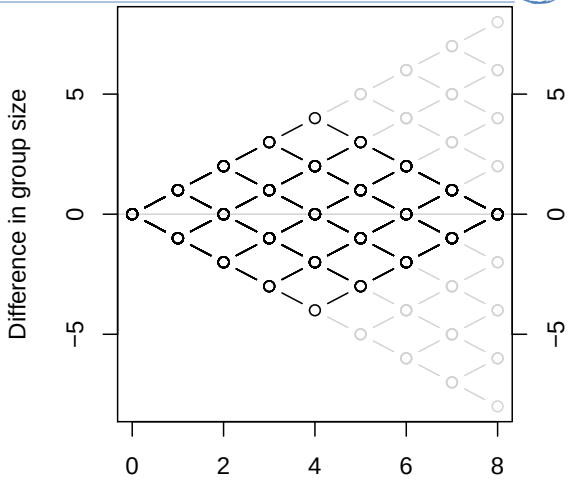
FP7 HEALTH 2013 - 602552

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

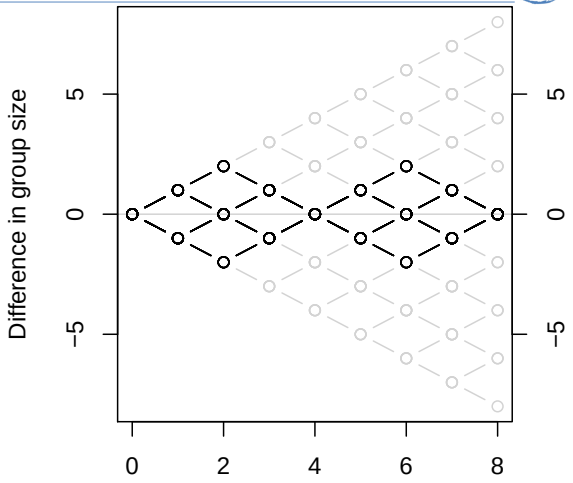
Equally probable balance sequences that attain balance after each block:

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

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

```
k <- 4 #block length

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



FP7 HEALTH 2013 - 602552

Big Stick Design

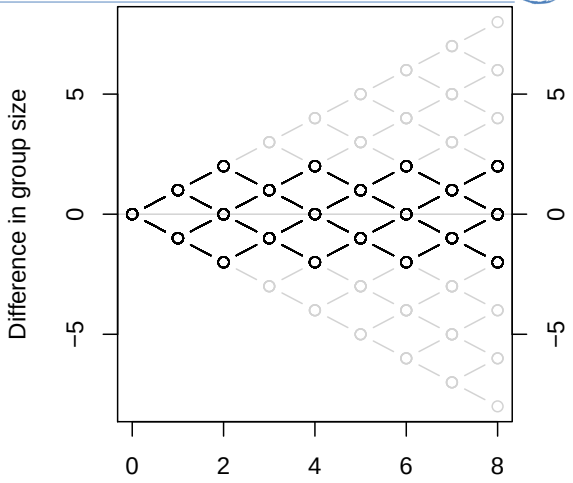
Equally probable final balance sequences:

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

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

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:

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

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

$$H_1 : d \neq 0$$

FP7 HEALTH 2013 - 602552

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

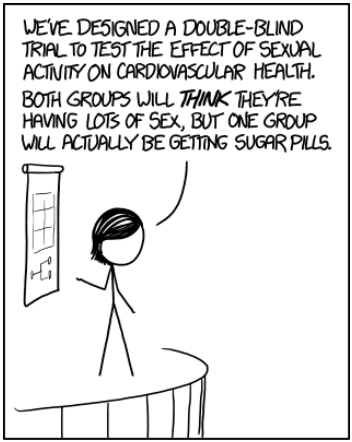
$$H_1 : d \neq 0$$

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)



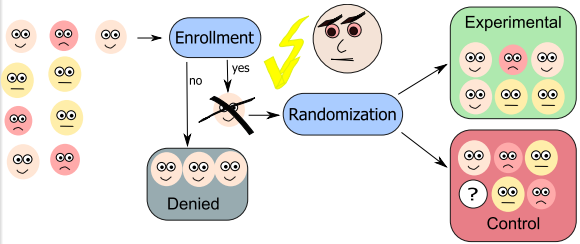
THE LIMITATIONS OF BLIND TRIALS

FP7 HEALTH 2013 - 602552

Selection bias under convergence strategy

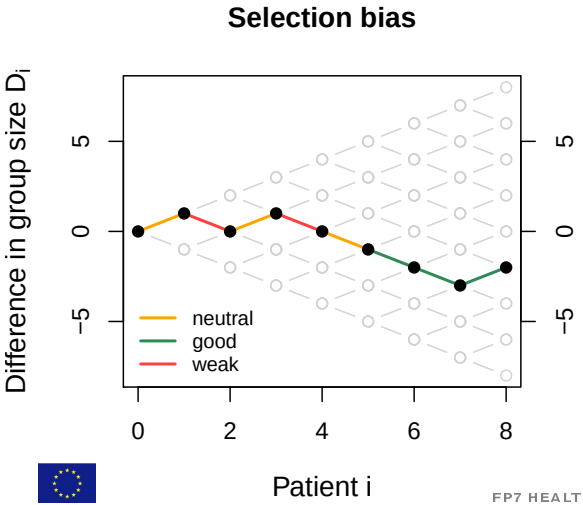
Third order selection bias

- 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



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

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

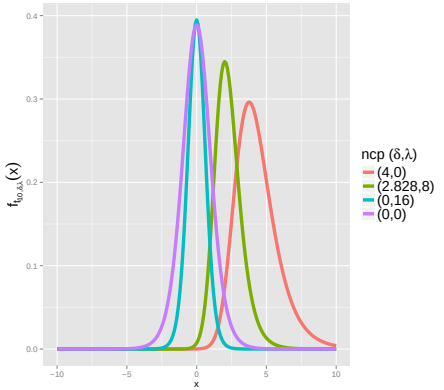
with $\tilde{y}_E = \frac{1}{N_E}\sum_{i=1}^{N} y_i t_i$, $\tilde{y}_C = \frac{1}{N_C}\sum_{i=1}^{N} y_i(1-t_i)$ and $N = N_E + N_C$ is doubly noncentrally t-distributed with parameters δ and λ.

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)
```

FP7 HEALTH 2013 - 602552

Comparison of randomization procedures

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

rar <- getAllSeq(rarPar(N))

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

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

plot(C)
```

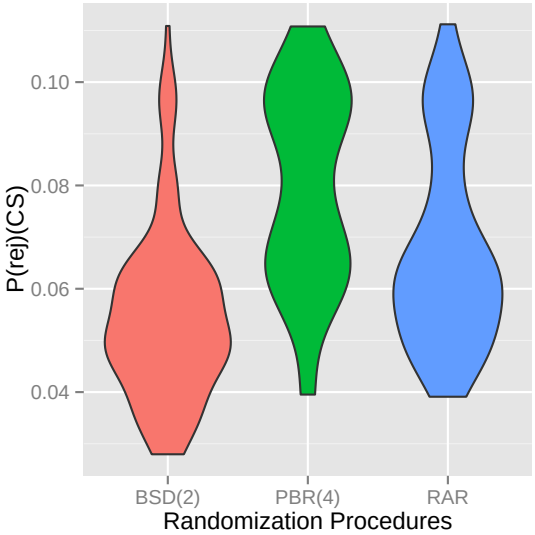


Table of Contents

1. Randomization with randomizeR

2. Randomization tests in the presence of selection bias

3. References

FP7 HEALTH 2013 - 602552

Randomization test

Null hypothesis

H_0: For each patient i the outcome y_i is the same disregarding of the treatment he receives.

1. Observe randomization sequence t_{obs}.
2. Observe the response $y_{obs} = (y_1, \ldots, y_N)$.
   $\Rightarrow$ Treat the response as fixed!
3. Calculate the randomization distribution of the test statistic:

$$\forall t \in \Omega : \text{Compute} \quad S(t, y_{obs}).$$

4. Then the p-value is $p = \sum_{t \in \Omega} \mathbb{P}_{\mathcal{M}}(t) \cdot I(|S(t, y_{obs})| \geq |S(t_{obs}, y_{obs})|)$

Lehmann (1975)

FP7 HEALTH 2013 - 602552

Test statistics

- Difference in means test statistic:

$$S(t, y_obs) = \sum_{i=1}^{N} y_i \cdot (2 \cdot t_i - 1)$$

where y_i denotes the responses of the ith patient.

- Logrank test statistic:

$$S(t, y_obs) = \sum_{i=1}^{N} (a_i - \bar{a}) \cdot (2 \cdot t_i - 1)$$

where a_i denotes the simple rank of the responses.

FP7 HEALTH 2013 - 602552

Model for selection bias

Assume now that the y_i are realizations of a random variable:

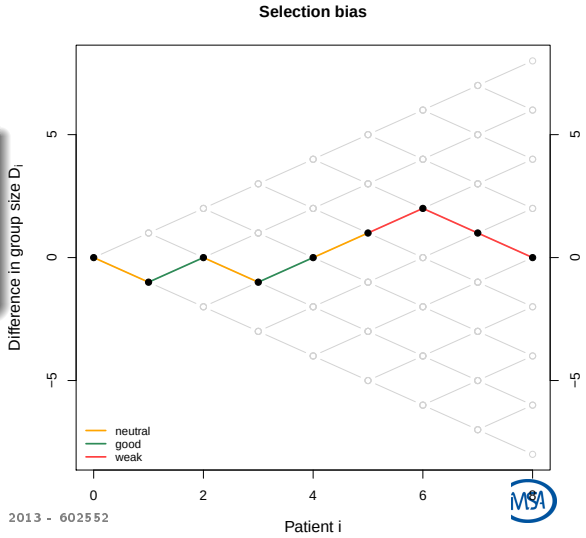
Convergence strategy

$$Y_i \sim \begin{cases} \mathcal{N}(\mu - \eta, \sigma^2) & D_{i-1} > 0 \\ \mathcal{N}(\mu, \sigma^2) & D_{i-1} = 0 \\ \mathcal{N}(\mu + \eta, \sigma^2) & D_{i-1} < 0 \end{cases}$$

Proschan (1994)

Blackwell and Hodges Jr. (1957)

→ Investigator wants to assign patients with higher values to the experimental group.



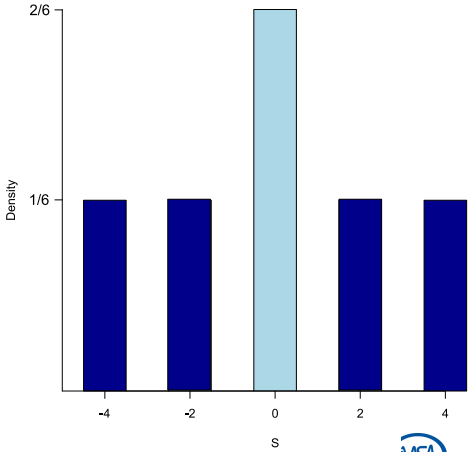
FP7 HEALTH 2013 - 602552

Example: Randomization test for RAR

Ranks of observed response: $y = (1, 2, 3, 4)$, test statistc $S_t = \sum_{i=1}^{4} y_i \cdot t_i - \sum_{i=1}^{4} y_i \cdot (1 - t_i)$

	t	$\mathbb{P}_{RAR}(t)$	S_t	$J_t = I(\lvert S_t \rvert \geq \lvert S_{obs} \rvert)$
1	E E C C	1/6	-4	1
*2	E C E C	1/6	-2	1
3	C E E C	1/6	0	0
4	E C C E	1/6	0	0
5	C E C E	1/6	2	1
6	C C E E	1/6	4	1

$$\Rightarrow p = \sum_t \mathbb{P}_{RAR}(t) \cdot J_t = 4/6$$



FP7 HEALTH 2013 - 602552

Two approach according to sample size

Exact Approach

Use the **complete set** $\Gamma_{\mathcal{M}}$ of sequences as the reference set Ω!

```
Omega<-getAllSeq(myPar)
```

Monte Carlo Approach

Sample L sequences from $\Gamma_{\mathcal{M}}$ and use this sample as the reference set Ω!

```
Omega<-genSeq(myPar, 1000000)
```

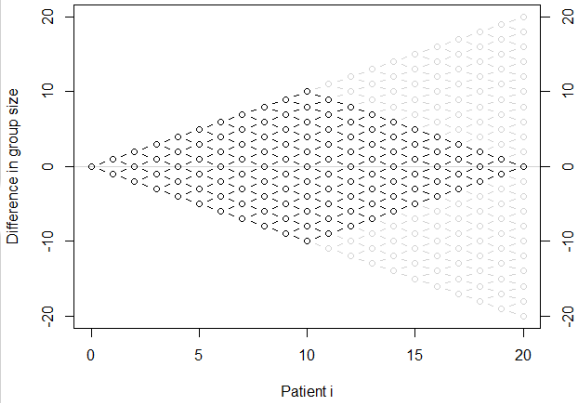
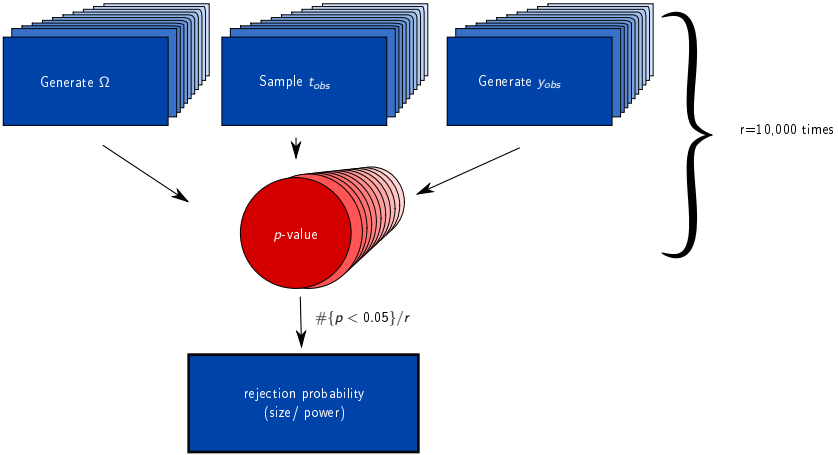


Figure: $|\Omega| = 184,756$

FP7 HEALTH 2013 - 602552

Simulations



FP7 HEALTH 2013 - 602552

Effect of selection bias on the test size ($N = 12$)

Table: Comparison of size and power of Student's t-test (TT) with the randomization test (RT) using the linear rank test statistic and the exact reference set, nominal significance level $\alpha = 0.05$. For the power, we assume a detectable effect $d = d(N, \alpha, \beta)$of the two-sided t-test where $\beta = 0.8$ denotes the nominal power.

	$\eta = 0$				$\eta = d/2$			
	RT		TT		RT		TT	
	size	power	size	power	size	power	size	power
TBD(4)	0.043	0.695	0.053	0.805	0.151	0.894	0.132	0.935
RAR	0.038	0.733	0.049	0.802	0.076	0.856	0.094	0.896
PBR(4)	0.040	0.697	0.048	0.801	0.135	0.898	0.158	0.949
MP(2)	0.041	0.733	0.051	0.803	0.115	0.887	0.128	0.923

FP7 HEALTH 2013 - 602552

Effect of selection bias on the test size ($N = 48$)

Table: Comparison of size and power of Student's t-test (TT) with the randomization test (RT) using the linear rank test statistic and a MC reference set of $L = 16,000$ sequences, and nominal significance level $\alpha = 0.05$. For the power, we assume the detectable effect $d = d(N, \alpha, \beta)$ of the two-sided t-test where $\beta = 0.8$ denotes the nominal power.

	$\eta = 0$				$\eta = d/2$			
	RT		TT		RT		TT	
	size	power	size	power	size	power	size	power
TBD(4)	0.048	0.768	0.049	0.803	0.163	0.953	0.170	0.965
RAR	0.048	0.769	0.049	0.796	0.060	0.845	0.065	0.886
PBR(4)	0.047	0.772	0.049	0.799	0.193	0.967	0.201	0.970
MP(2)	0.046	0.774	0.051	0.795	0.140	0.947	0.141	0.955

FP7 HEALTH 2013 - 602552

Comparison of RT different test statistics

Table: Comparison of the size of the randomization test using the linear rank test statistic (lr) and the difference of means test statistic (dm) with a MC reference set of $L = 16,000$ sequences, and nominal significance level $\alpha = 0.05$.

	$\eta = 0$		$\eta = d/2$	
	lr	dm	lr	dm
TBD(4)	0.048	0.036	0.163	0.161
RAR	0.048	0.052	0.060	0.069
PBR(4)	0.047	0.056	0.193	0.192
MP(2)	0.046	0.046	0.140	0.143

FP7 HEALTH 2013 - 602552

Conclusions

- randomizeR makes it easy to generate randomization sequences and compute reference sets for the randomization tests.
- The randomization test presented in this talk does not protect against selection bias (it is just as bad as Student's t-test).
- Aim: Develop a new randomization test (reference distribution + test statistic) that is not influenced by selection bias.

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.

Or just talk to me at lunch!

FP7 HEALTH 2013 - 602552

Table of Contents

1. Randomization with randomizeR

2. Randomization tests in the presence of selection bias

3. References

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.

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