Stratification and randomization in clinical trials

Armin Koch Ralf-Dieter Hilgers and Colleagues

August 23rd, 2015

FP7 HEALTH 2013 - 602552

Randomization in clinical trials

Ralf-Dieter Hilgers Diane Uschner David Schindler

Department for Medical Statistics RWTH Aachen University

August 23rd, 2015

FP7 HEALTH 2013 - 602552

Mission

Why? We believe that we can make a difference for patient health care by developing innovative methods for the design and analysis of clinical trials.

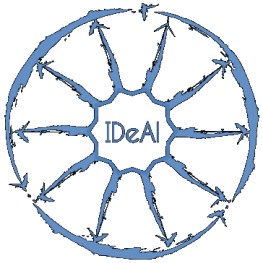
What? We believe that the design and analysis have to go hand in hand. The design stage must account for the problems that can occur in the analysis of the trial. We believe that a sound scientific evaluation of the randomization procedure has to be performed.

How? We propose a software tool for the selection of a tailored randomization procedure, taking into account possible biases that may arise throughout the trial.

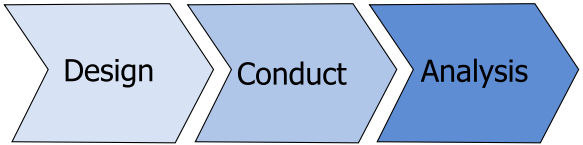
FP7 HEALTH 2013 - 602552

This project has received funding from the European Unions 7th Framework Programme for research, technological development and demonstration under Grant Agreement no 602552.

- Development of `randomizeR`.
- Please give us your feedback (in the break)!



Steps of a clinical trial



Three steps of a clinical trial:

1. Design including randomization
2. Recruitment and conduct
3. Analysis and reporting

FP7 HEALTH 2013 - 602552

Roots

Initially: Patient allocation by alternation (`ABABA...`)

Bradford Hill, 1933: It is necessary to conceal the allocation sequence!

Alternating allocation sequence is too predictable!
$\Rightarrow$ Replace alternation by *randomization*!

Aims of the randomization (ICH E9)

- Statistical basis for quantitative evaluation of the evidence relating to treatment effects
- Produce treatment groups with similar distribution of prognostic factors
- In combination with blinding: Help avoid *bias* in the selection and allocation of subjects arising from predictability of treatment assignments.

Bias

The systematic tendency of any factors associated with the design, conduct, analysis and evaluation of the results of a clinical trial to make the estimate of a treatment effect deviate from its true value.

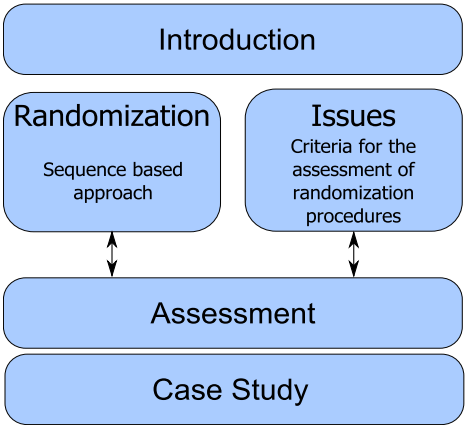
- Selection bias
- Chronological bias

ICH E9 (1998)

FP7 HEALTH 2013 - 602552

Structure of the workshop

1. Introduction
2. Issues
3. Randomization procedures
4. Assessment
5. Case study

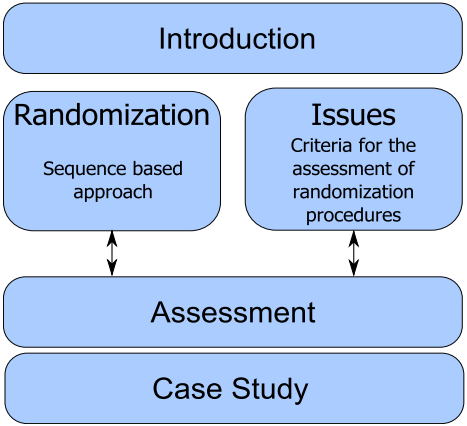


FP7 HEALTH 2013 - 602552

Structure of the workshop

1. Introduction
2. Issues
3. Randomization procedures
4. Assessment
5. Case study

At the same time: Get acquainted with the `randomizeR` package $\Rightarrow$ Give feedback!

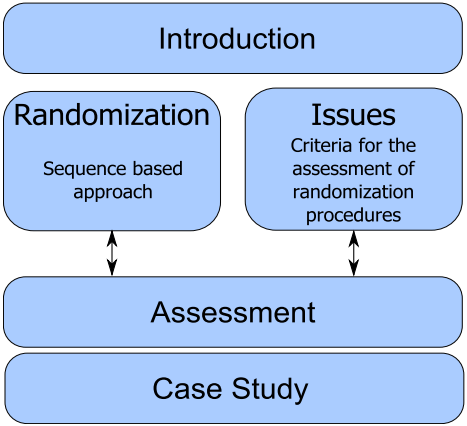


FP7 HEALTH 2013 - 602552

Structure of the workshop

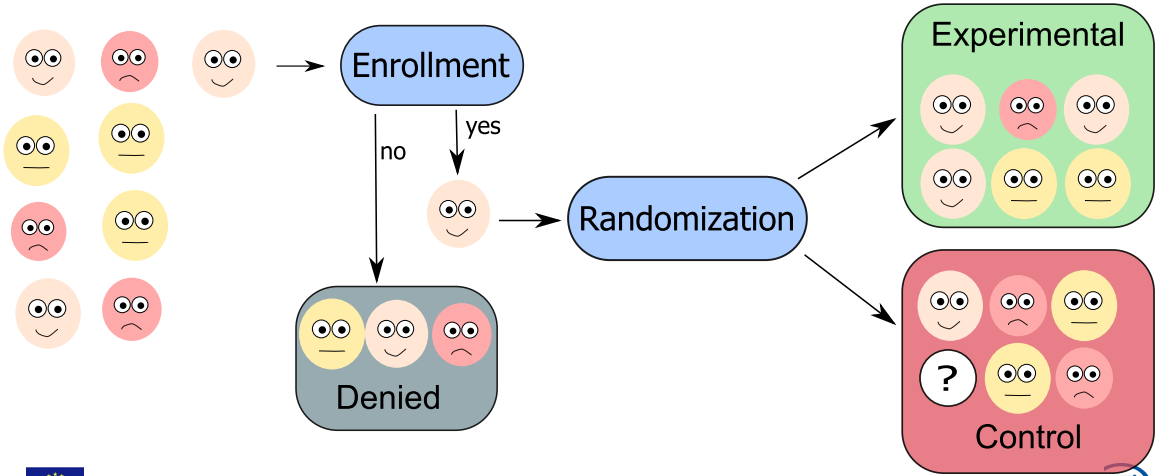
1. **Introduction**
2. Issues
3. Randomization procedures
4. Assessment
5. Case study

At the same time: Get acquainted with the `randomizeR` package $\Rightarrow$ Give feedback!



FP7 HEALTH 2013 - 602552

What is randomization?



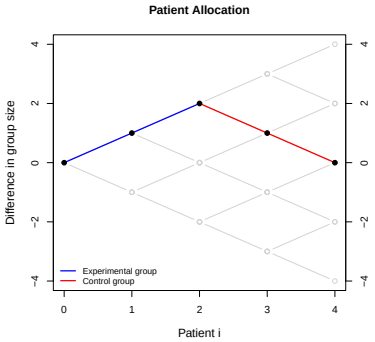
FP7 HEALTH 2013 - 602552

Randomization Sequence

Let N be the total sample size, i be the patient index. A randomization sequence is a vector $t \in \Gamma := \{0,1\}^N$ with

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

We call Γ the *set of all sequences*. t is the realization of a random vector T.



Rosenberger and Lachin (2016)

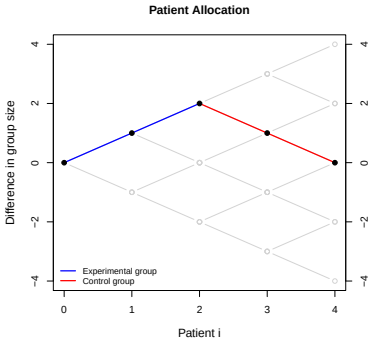
FP7 HEALTH 2013 - 602552

Group sizes

For the *number of patients in group E* (resp. C) at time i we write:

$$N_E(i) := N_E(T, i) := \sum_{j=1}^{i} T_j$$

$$N_C(i) := N_C(T, i) := i - \sum_{j=1}^{i} T_j$$



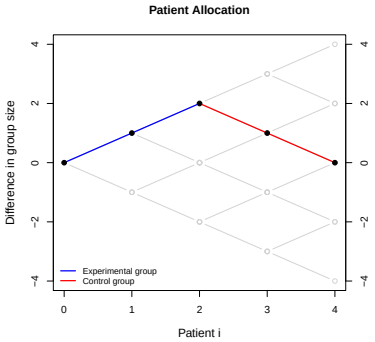
Rosenberger and Lachin (2016)

FP7 HEALTH 2013 - 602552

Imbalance

The *imbalance after the inclusion of i patients* is given by the difference in group sizes:

$$D_i := D_i(T) := N_C(T,i) - N_E(T,i)$$
$$= i - 2 \cdot \sum_{j=1}^{i} T_j$$



Rosenberger and Lachin (2016)

FP7 HEALTH 2013 - 602552

Randomization procedure

Definition

A randomization procedure $\mathcal{M}$ is a probability distribution on Γ. $\mathcal{M}$ *produces* the sequences

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

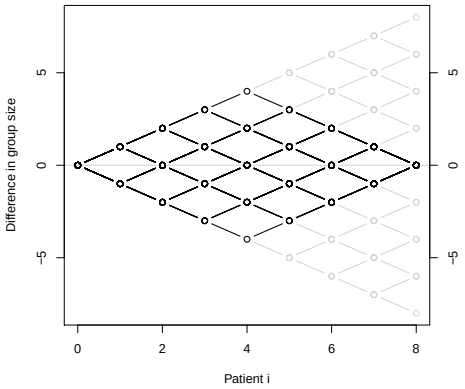


Figure: Γ_{RAR} for $N = 8$

Randomization procedure

Definition

A randomization procedure $\mathcal{M}$ is a probability distribution on Γ. $\mathcal{M}$ *produces* the sequences

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

Example: random allocation rule (RAR)

For $\mathcal{M} = RAR$ it is

$$P_{RAR}(T) = \begin{cases} \binom{N}{N/2}^{-1} & D_N(T) = 0 \\ 0 & D_N(T) \neq 0 \end{cases}$$

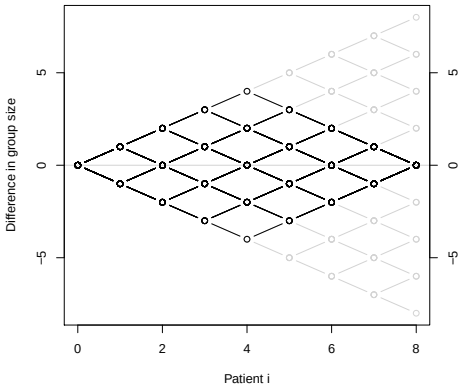


Figure: Γ_{RAR} for $N = 8$

FP7 HEALTH 2013 - 602552

Response

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

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

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.

- W.l.o.g. higher values of Y_i are regarded as better.

FP7 HEALTH 2013 - 602552

Hypothesis of no treatment effect

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

$$H_1 : d \neq 0$$

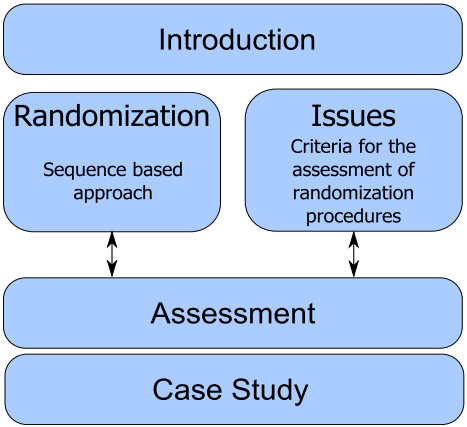
- Prove the hypothesis with Student's t-test.

FP7 HEALTH 2013 - 602552

Structure of the workshop

1. Introduction
2. **Issues**
3. Randomization procedures
4. Assessment
5. Case study

At the same time: Get acquainted with the `randomizeR` package $\Rightarrow$ Give feedback!



FP7 HEALTH 2013 - 602552

Issues

- Assumptions in Model 1 may not be fulfilled.
- Biases may compromise the estimation of the treatment effect.
- Provide sound scientific basis for the selection of a tailored randomization procedure.

Process

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

FP7 HEALTH 2013 - 602552

Sequence based approach

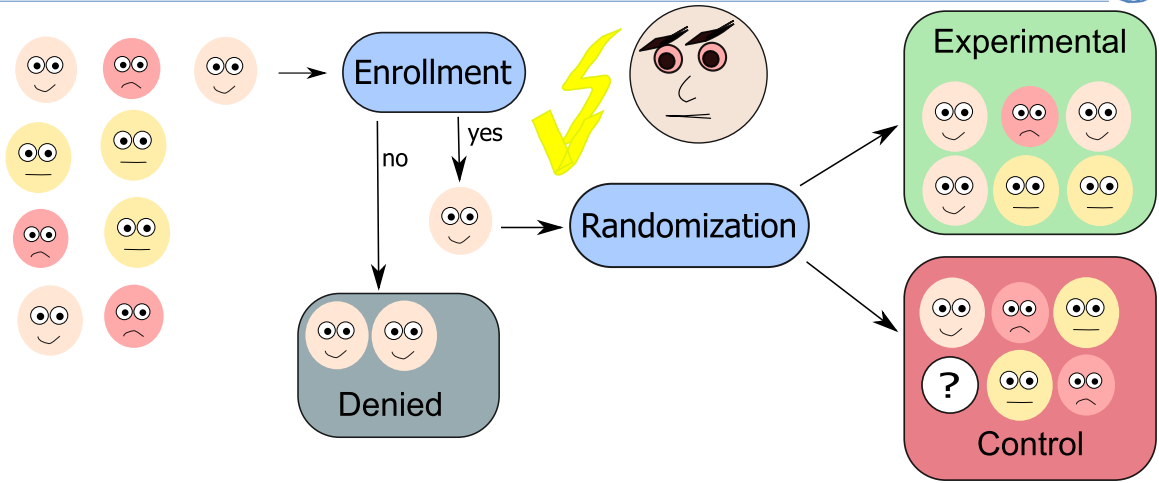
- Model each constraint via an issue.
- For each randomization sequence t, compute the value of the issue
- Weight the values of issue according to the probability of occurrence of the randomization sequence.
- Compute descriptive measures for the issue in the randomization procedure, s.a. mean and quantiles.

	Sequences	Probability	Issue
1	C C E E	0.167	0.463
2	C E C E	0.167	0.689
3	E C C E	0.167	-0.583
4	C E E C	0.167	-0.654
5	E C E C	0.167	2.005
6	E E C C	0.167	-0.029

mean = 0.315

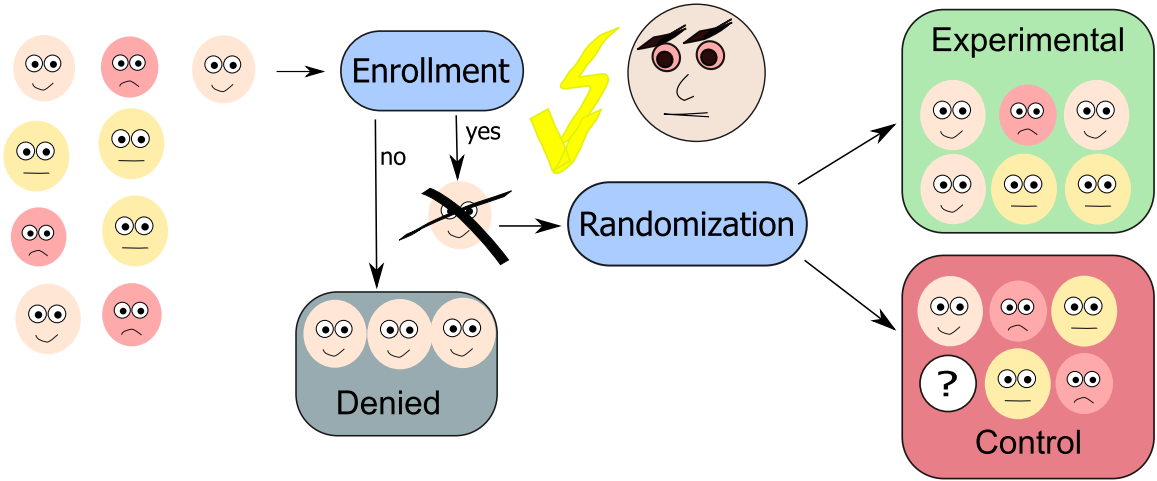
FP7 HEALTH 2013 - 602552

Selection bias



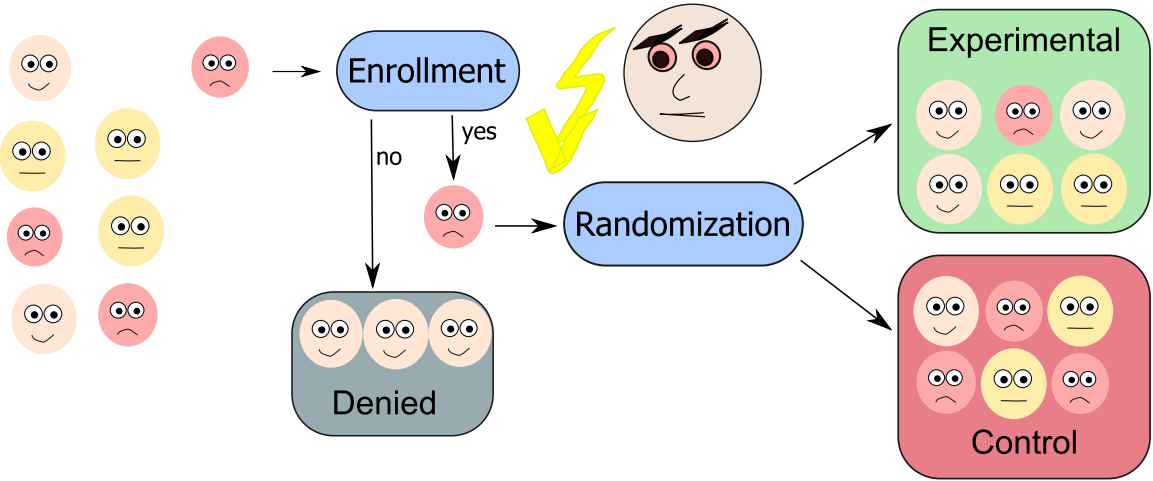
FP7 HEALTH 2013 - 602552

Selection bias



FP7 HEALTH 2013 - 602552

Selection bias

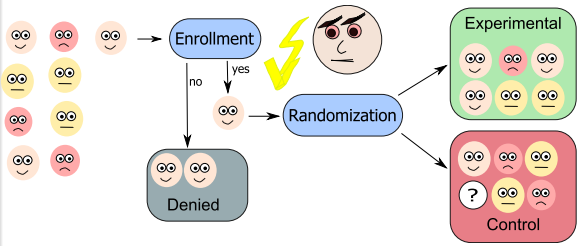


FP7 HEALTH 2013 - 602552

Selection bias

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)

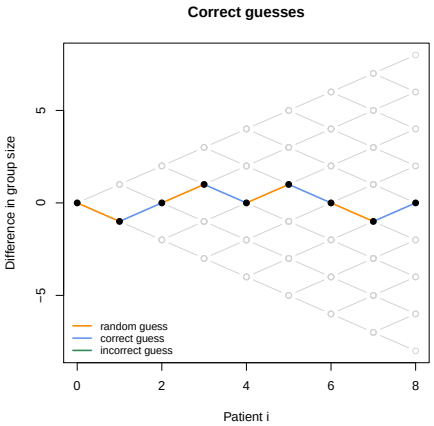
FP7 HEALTH 2013 - 602552

Issue: Correct guesses

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

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

following the so called *convergence strategy* ("CS").



Blackwell and Hodges Jr. (1957)

FP7 HEALTH 2013 - 602552

Expected number of correct guesses

The *expected number of correct guesses* of a randomization procedure $\mathcal{M}$ is given by

$$CG_{\mathcal{M}} := E_{\mathcal{M}}(E_T(\#\{i = 1, \ldots N : g(t_{i+1}) = t_{i+1}\})).$$

The average proportion of correct guesses is

$$avpCG_{\mathcal{M}} := \frac{CG_{\mathcal{M}}}{N}.$$

Blackwell and Hodges Jr. (1957)

Correct guesses - representation in randomizeR

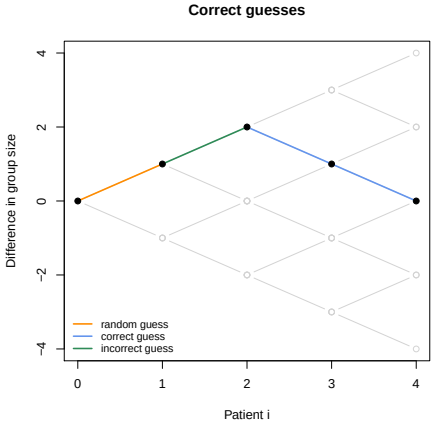
randomizeR formally represents the expected number of correct guesses as follows:

```
> corGuess("CS")
 Object of class "corGuess"

         TYPE = CS
```

Correct guesses - sequence wise assessment

	Sequence	Probability	propCG(CS)
1	CCEE	0.1667	0.625
2	CECE	0.1667	0.750
3	ECCE	0.1667	0.750
4	CEEC	0.1667	0.750
5	ECEC	0.1667	0.750
6	EECC	0.1667	0.625

mean = 0.71

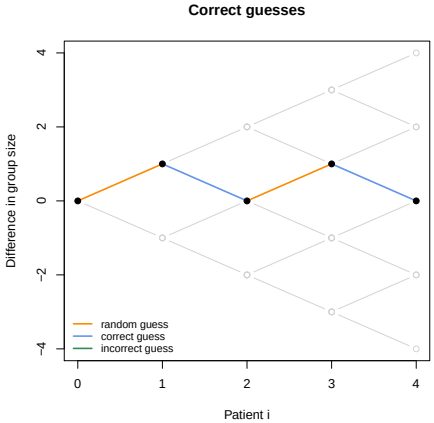


FP7 HEALTH 2013 - 602552

Correct guesses - sequence wise assessment

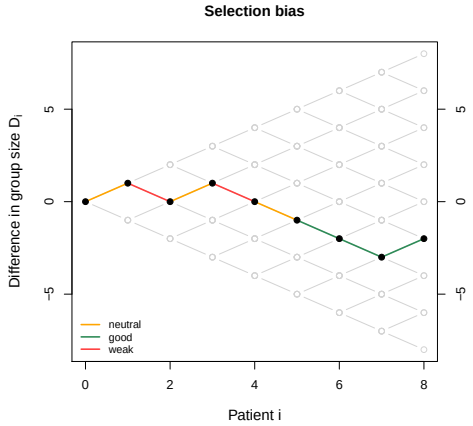
	Sequence	Probability	propCG(CS)
1	CCEE	0.1667	0.625
2	CECE	0.1667	0.750
3	ECCE	0.1667	0.750
4	CEEC	0.1667	0.750
5	ECEC	0.1667	0.750
6	EECC	0.1667	0.625

mean = 0.71



FP7 HEALTH 2013 - 602552

Biasing Policy



Choose patient $i+1$ with expected response

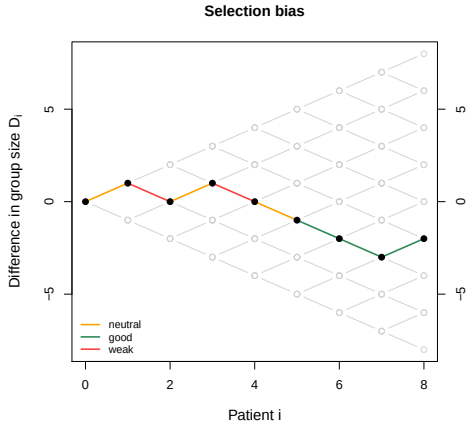
$$E(Y_{i+1}) = \begin{cases} \mu - \eta & N_E(i) > N_C(i) \\ \mu & N_E(i) = N_C(i) \\ \mu + \eta & N_E(i) < N_C(i) \end{cases}$$

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

Proschan (1994)

FP7 HEALTH 2013 - 602552

Selection bias model



The response of patient i is affected by selection bias

$$Y_i \sim \mathcal{N}(\mu + d \cdot T_i - \eta \cdot \text{sign}(D_{i-1}),\ \sigma^2)$$

with *selection effect* $\eta > 0$ and standardized selection effect $\gamma = \eta/\sigma^2$.

Proschan (1994)

FP7 HEALTH 2013 - 602552

Historical: Influence of selection bias on test decision of the z-test

Idea Measure influence of selection bias on the responses in model 1 via the shift of the distribution of the test statistic.

Type-I-error rate

Let $\mathcal{M}$ be such that $D_N(t) = 0 \quad \forall t \in \Gamma_{\mathcal{M}}$, and let $R(t) = \#\{i : D_i(t) = 0\}$ be the number of returns to origin. Let $Z := \bar{Y}_E - \bar{Y}_C$ denote the z-test statistic. Then the asymptotic type-I-error rate of the one-sided z-test is given by

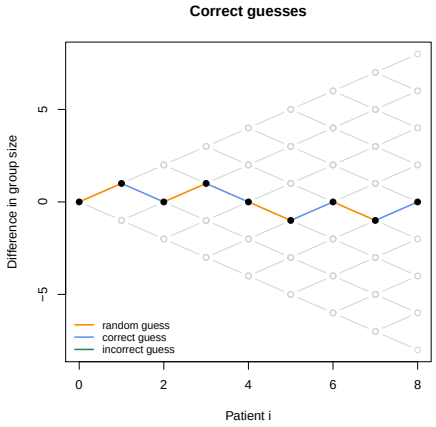
$$P(Z > z_{1-\alpha/2}) = 1 - E\left(\Phi(z_{1-\alpha/2} - \frac{R(T)}{\sqrt{N}} \cdot \gamma)\right)$$

Proschan (1994)

Relation between correct guesses and returns to origin

For a given sequence t, the relation between the correct guesses of t and the number of returns to origin of t are as follows

$$CG(t) = \frac{R(t)}{2} + \frac{N}{2}$$



FP7 HEALTH 2013 - 602552

Issue: Influence of selection bias on test decision of the t-test

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

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

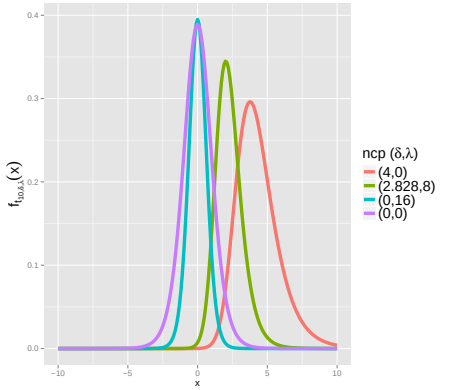
with $\tilde{y}_E = \frac{1}{N_E}\sum_{i=1}^{N} y_i t_i$, $\tilde{y}_C = \frac{1}{N_C}\sum_{i=1}^{N} y_i(1-t_i)$ and $N = N_E + N_C$ is doubly noncentrally t-distributed with parameters δ and λ.

Issue: Influence of selection bias on test decision of the t-test

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

Exact type-I-error selection bias - representation in randomizeR

randomizeR formally represents the exact type-I-error of the t-test in case of selection bias as follows:

```
> selBias("CS", eta=1, method="exact")

 Object of class "selBias"

        ETA = 1
        TYPE = CS
        METHOD = exact
        ALPHA = 0.05
```

Simulated type-I-error selection bias - representation in randomizeR

randomizeR formally represents the simulated type-I-error of the t-test as follows:

```
> selBias("CS", eta=1, method="sim")

 Object of class "selBias"

        ETA = 1
        TYPE = CS
        METHOD = sim
        ALPHA = 0.05
```

Type-I-error selection bias - sequence wise assessment

	Sequence	Probability	rejection prob. exact(CS)	testDec	sim(CS)
1	CCEE	0.1667	0.0490		FALSE
2	CECE	0.1667	0.0952		FALSE
3	ECCE	0.1667	0.0613		TRUE
4	CEEC	0.1667	0.0613		FALSE
5	ECEC	0.1667	0.0952		FALSE
6	EECC	0.1667	0.0490		FALSE

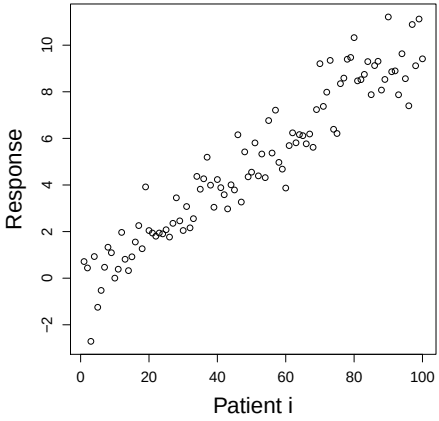
mean = 0.0685

Chronological bias

The expected response of patient i is influenced by a trend $\tau(i, \vartheta)$:

$$Y_i \sim \mathcal{N}(\mu + \tau(i, \vartheta), \sigma^2)$$

$\Rightarrow$ Estimation of treatment effect can be seriously biased.



FP7 HEALTH 2013 - 602552

Shapes of trend

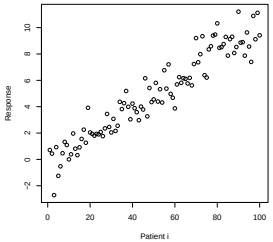


Figure: $\tau(i, \vartheta) = \frac{i}{N}\vartheta$

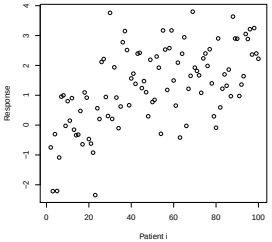


Figure: $\tau(i, \vartheta) = log(\frac{i}{N}\vartheta)$

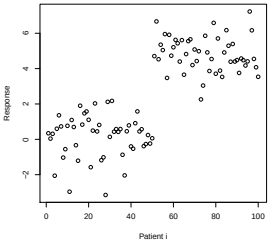


Figure: $\tau(i, \vartheta) = 1_{\{i \geq n_0\}}\vartheta$

Tamm and Hilgers (2014)

FP7 HEALTH 2013 - 602552

Time trend - representation in randomizeR

randomizeR formally represents the type-I-error of the t-test in case of chronological bias as follows:

```
> chronBias("linT", theta=1, method="exact")
```

```
Object of class "chronBias"

        TYPE = linT
        THETA = 1
        METHOD = exact
        ALPHA = 0.05
```

```
Object of class "chronBias"

        TYPE = logT
        THETA = 1
        METHOD = exact
        ALPHA = 0.05
```

$\Rightarrow$ The exact probability of rejecting the null hypothesis can be calculated similarly to selection bias.

Time trend - representation in randomizeR

randomizeR formally represents the type-I-error of the *t*-test in case of chronological bias as follows:

```
> chronBias("stepT", theta=1, method="exact", saltus=50)

 Object of class "chronBiasStepT"

        SALTUS = 50
        TYPE = stepT
        THETA = 1
        METHOD = exact
        ALPHA = 0.05
```

$\Rightarrow$ The exact probability of rejecting the null hypothesis can be calculated similarly to selection bias.

Type-I-error time trend - sequence wise assessment

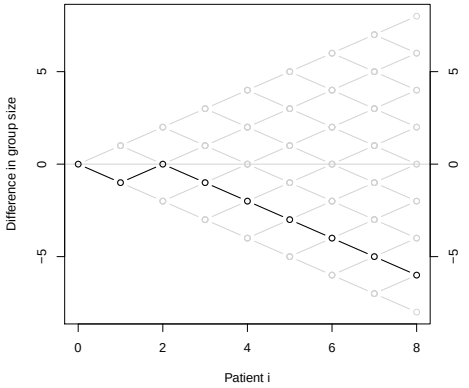
Setting $\sigma^2 = 1$, we get:

	Sequence	Probability	rejection prob. exact(linT)	rejection prob. exact(logT)
1	CCEE	0.1667	0.0525	0.0505
2	CECE	0.1667	0.0493	0.0499
3	ECCE	0.1667	0.0482	0.0497
4	CEEC	0.1667	0.0482	0.0497
5	ECEC	0.1667	0.0493	0.0499
6	EECC	0.1667	0.0525	0.0505

```
meanLinT = 0.05

meanLogT = 0.05
```

Issue: Imbalance in allocation numbers

- Strong imbalances during the trial and at the end of the trial can distort the estimation of the treatment effect.
- Admitting slight imbalances makes the randomization less predictable.
- Types of imbalance:
  - Final imbalance: $D_N(T)$
  - Absolute final imbalance: $|D_N(T)|$
  - Loss: $D_N(T)^2/N$
  - Maximum imbalance: $max_{\{i=1,\ldots,N\}}|D_i(T)|$



Atkinson (2001), Berger (2005), Rosenberger and Lachin (2016)

FP7 HEALTH 2013 - 602552

Imbalance - representation in randomizeR

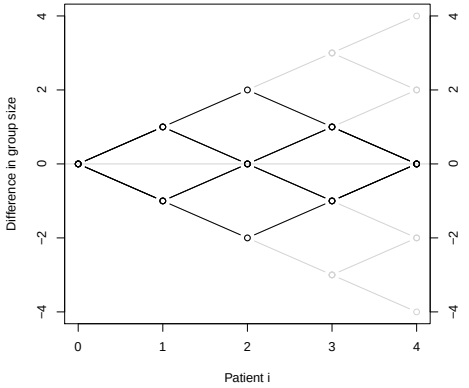
randomizeR formally represents imbalance as follows:

```
> imbal(type="maxImb")
 Object of class "imbal"

         TYPE = maxImb
```

Possible values for type are imb, absImb, loss and maxImb.

Imbalance - sequence wise assessment

	Sequence	Probability	imb	maxImb
1	CCEE	0.1667	0	2
2	CECE	0.1667	0	1
3	ECCE	0.1667	0	1
4	CEEC	0.1667	0	1
5	ECEC	0.1667	0	1
6	EECC	0.1667	0	2



FP7 HEALTH 2013 - 602552

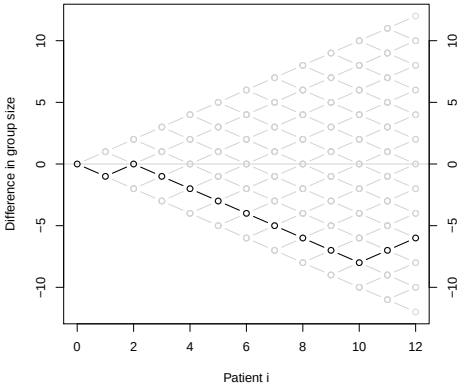
Issue: Power loss due to final imbalance

The power is robust against slight imbalances.

Table: Power loss due to imbalance
$N = 12$, effect size $d = 1.80$, size $\alpha = 0.05$.

$N_E(N,T)$	6	7	8	9	10
$N_C(N,T)$	6	5	4	3	2
1 - $\beta(T)$	0.80	0.77	0.75	0.68	0.55

Lachin (1988)



FP7 HEALTH 2013 - 602552

Power loss - representation in randomizeR

randomizeR formally represents the power loss due to final imbalance as follows:

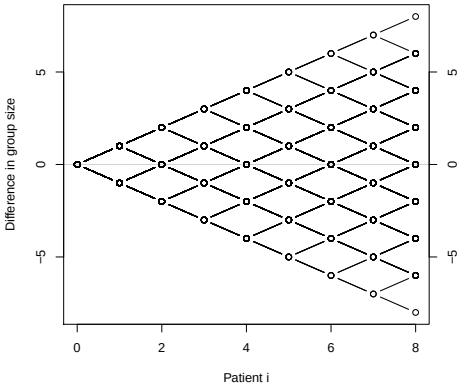
```
> setPower(d = 2, method = "exact", alpha = 0.05)

 Object of class "power"

        D = 2
        METHOD = exact
        ALPHA = 0.05
```

Power loss - sequence wise assessment

Figure: Power loss for total sample size $N = 8$, $\alpha = 0.05$, $\beta = 0.2$ and $d = 2.38$.

	Sequence	Probability	power(exact)
1	EEEEEEEE	0.0039	0.0000
2	CEEEEEEE	0.0039	0.4644
3	ECEEEEEE	0.0039	0.4644
4	CCEEEEEE	0.0039	0.6827
5	EECEEEEE	0.0039	0.4644
6	CECEEEEE	0.0039	0.6827
7	ECCEEEEE	0.0039	0.6827
8	CCCEEEEE	0.0039	0.7749

... mean = 0.53

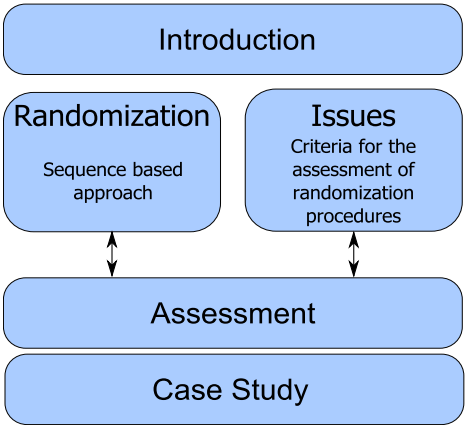


FP7 HEALTH 2013 - 602552

Structure of the workshop

1. Introduction
2. Issues
3. **Randomization procedures**
4. Assessment
5. Case study

At the same time: Get acquainted with the `randomizeR` package $\Rightarrow$ Give feedback!



FP7 HEALTH 2013 - 602552

Randomization procedure

A randomization procedure $\mathcal{M}$ is a probability distribution on Γ. $\mathcal{M}$ *produces* the sequences

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

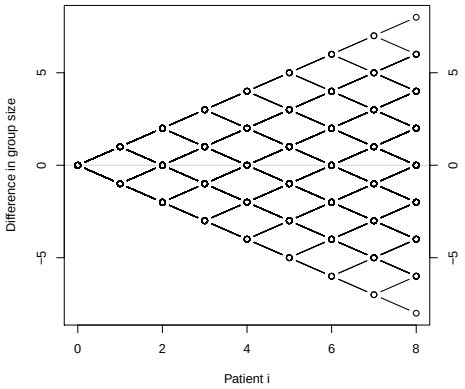


Figure: Visualization of Γ

FP7 HEALTH 2013 - 602552

Random Allocation Rule

RAR Draw without replacement from an urn with $N/2$ blue and red balls each. $\Rightarrow$ Equiprobable sequences with final balance

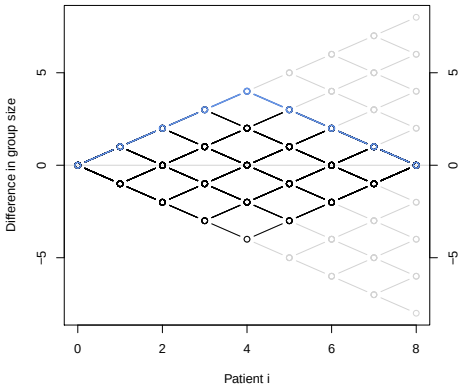


Figure: $N = 8$

FP7 HEALTH 2013 - 602552

Random Allocation Rule

Properties of RAR

- Possible difference in group sizes: $N/2$.
- No difference in group sizes at the end.
- (Small) number of possible sequences:

$$|\Gamma_{\text{RAR}}| = \binom{N}{N/2}$$

- Susceptible to the convergence strategy.

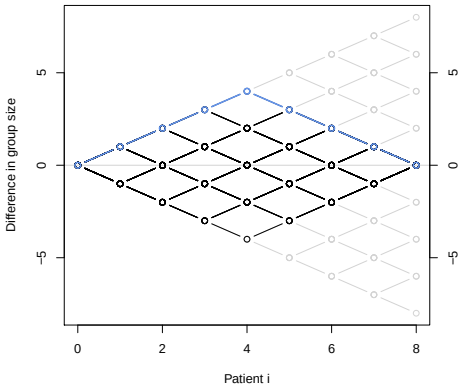


Figure: $N = 8$

FP7 HEALTH 2013 - 602552

Permuted Block Design

$PBR(k)$ Draw without replacement from an urn with $k/2$ blue and red balls each. When the urn is empty, replace all k balls. Repeat until N balls have been drawn.

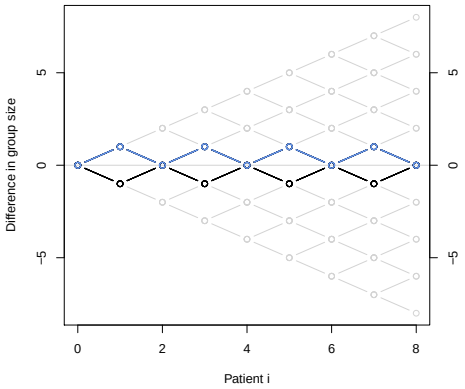


Figure: $N = 8, k = 2$

FP7 HEALTH 2013 - 602552

Permuted Block Design

Properties of PBR

- Possible difference in group sizes: $k/2$.
- No difference in group sizes at the end.
- Small number of possible sequences:

$$|\Gamma_{\text{PBR}}| = \binom{k}{k/2}^{N/k}$$

- Very susceptible to the convergence strategy particularly for small k.

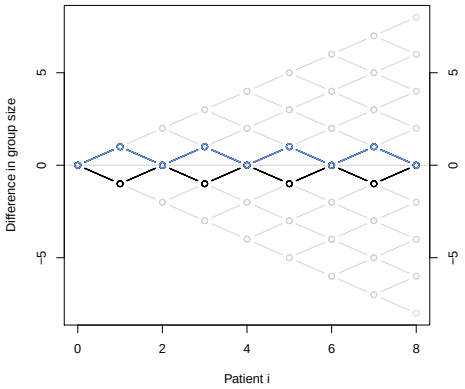


Figure: $N = 8, k = 2$

FP7 HEALTH 2013 - 602552

Big Stick Design

$BSD(b)$ Toss a fair coin N times. If the random walk reaches an imbalance b, make a deterministic allocation in order to reduce the imbalance.

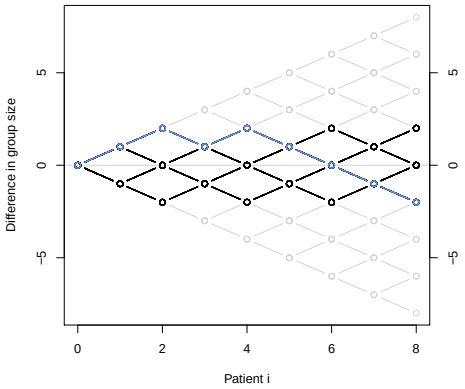


Figure: $b = 2$

FP7 HEALTH 2013 - 602552

Big Stick Design

Properties of BSD

- Possible difference in group sizes: $b/2$.
- (Small) difference in group sizes at the end possible.
- Deterministic allocations when hitting the imbalance boundary.
- Not as susceptible to the convergence strategy as PBR under the assumption of $k/2 = b$.

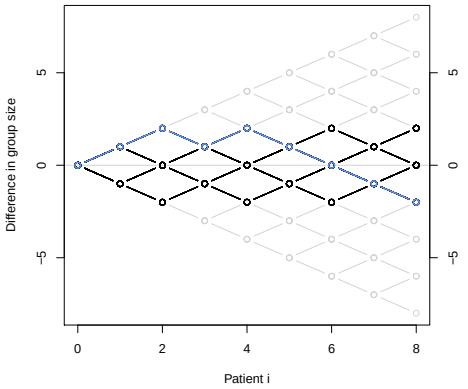
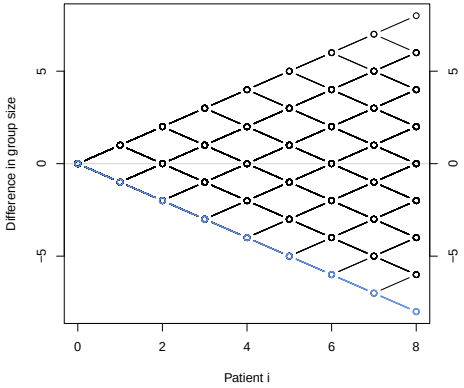


Figure: $b = 2$

FP7 HEALTH 2013 - 602552

Efron's Biased Coin Design

EBC(p) Toss a coin N times: If there is an imbalance, toss let the probability be $0.5 < p < 1$ to reduce the imbalance with the next patient and the probability $1 - p$ to increase the imbalance. If there is no imbalance, toss a fair coin.

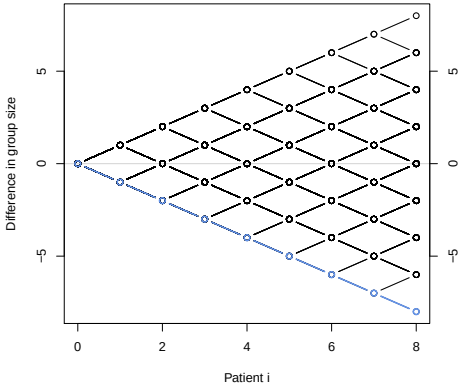


FP7 HEALTH 2013 - 602552

Efron's Biased Coin Design

Properties of EBC

- Possible difference in group sizes: N.
- (High) difference in group sizes at the end possible.
- All sequences are possible, but they are not equiprobable.
- No deterministic allocations during the whole randomization process.

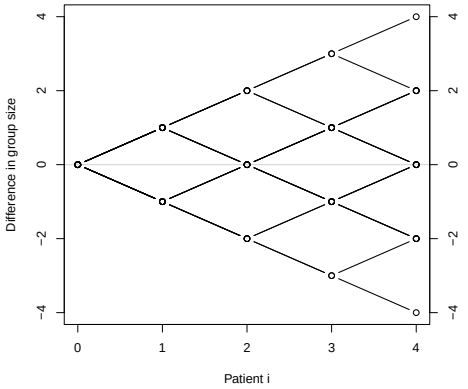


FP7 HEALTH 2013 - 602552

Sequence probabilities ($N = 4$)

	Sequences	RAR	PBR	BSD	EBC
1	E E E E	0.0000	0.00	0.0000	0.0185
2	C E E E	0.0000	0.00	0.0625	0.0555
3	E C E E	0.0000	0.00	0.0625	0.0555
4	C C E E	0.1667	0.00	0.1250	0.0741
5	E E C E	0.0000	0.00	0.1250	0.0370
6	C E C E	0.1667	0.25	0.0625	0.1112
7	E C C E	0.1667	0.25	0.0625	0.1112
8	C C C E	0.0000	0.00	0.0000	0.0370

...

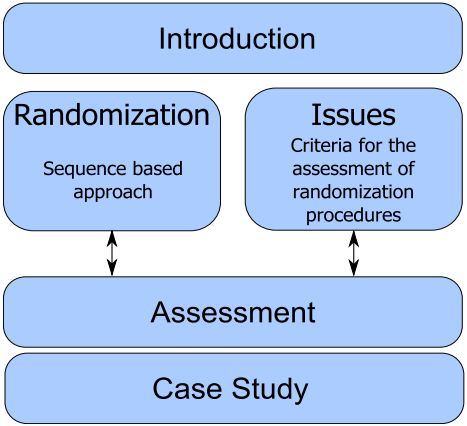


FP7 HEALTH 2013 - 602552

Structure of the workshop

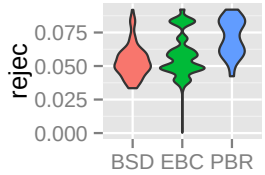
1. Introduction
2. Issues
3. Randomization procedures
4. **Assessment**
5. Case study

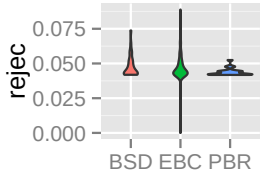
At the same time: Get acquainted with the `randomizeR` package $\Rightarrow$ Give feedback!

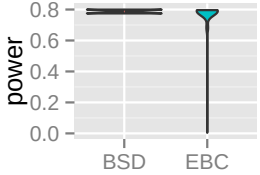


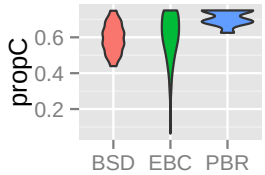
FP7 HEALTH 2013 - 602552

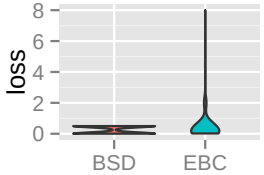
Assessment

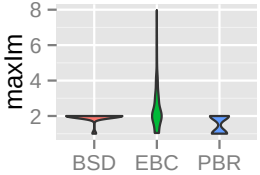






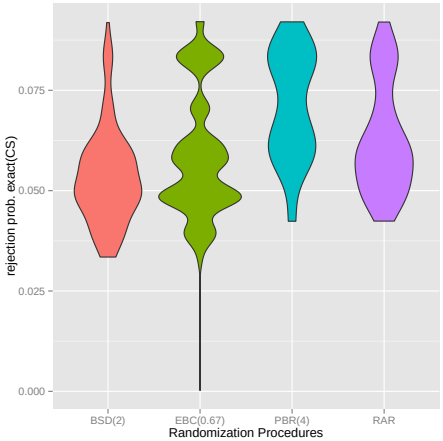






FP7 HEALTH 2013 - 602552

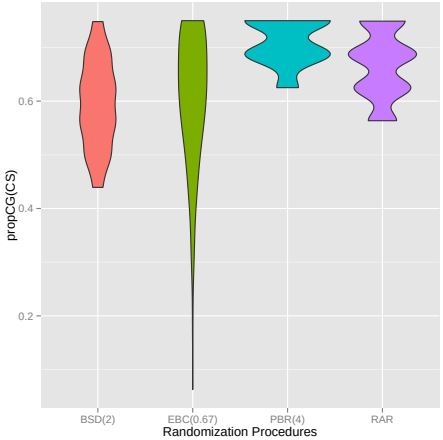
Assessment type-I-error selection bias



- Total sample size $N = 8$.
- Selection bias
  - Selection effect $\eta = d/4$, where $d = 2.38$ denotes the detectable effect of the two sample t-test for $N = 8$.
- Compute the complete set of sequences for each design.
- Calculate the exact type-I-error for each sequence.

FP7 HEALTH 2013 - 602552

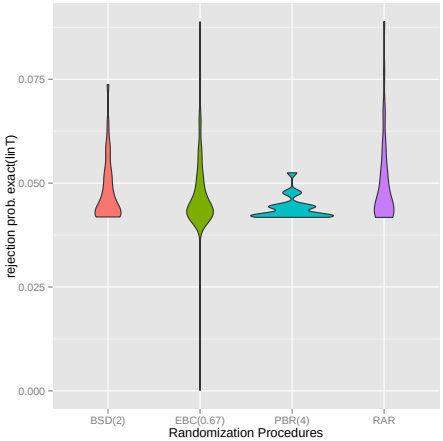
Assessment correct guesses



- Total sample size $N = 8$.
- Compute the complete set of sequences for each design.
- Calculate the proportion of correct guesses for each sequence.

FP7 HEALTH 2013 - 602552

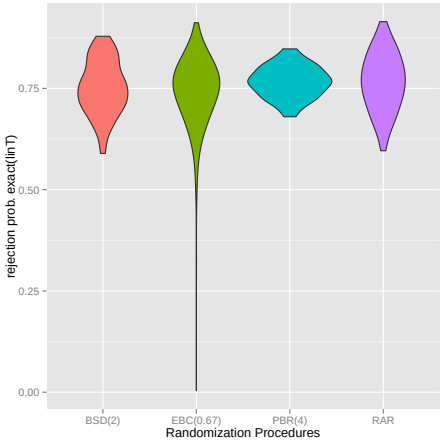
Assessment linear time trend



- Total sample size $N = 8$.
- Linear time trend
  - Strength time trend $\vartheta = 1/N$
- Compute the complete set of sequences for each design.
- Calculate the exact type-I-error for each sequence.

FP7 HEALTH 2013 - 602552

Assessment Power linear time trend



- Total sample size $N = 8$.
- Linear time trend
  - Strength time trend $\vartheta = 1/N$
- Compute the complete set of sequences for each design.
- Calculate the exact power for each sequence.
  - Power for the detection of an effect of size $d = 2.308$ (detectable effect of the t-Test for $N = 8$.

FP7 HEALTH 2013 - 602552

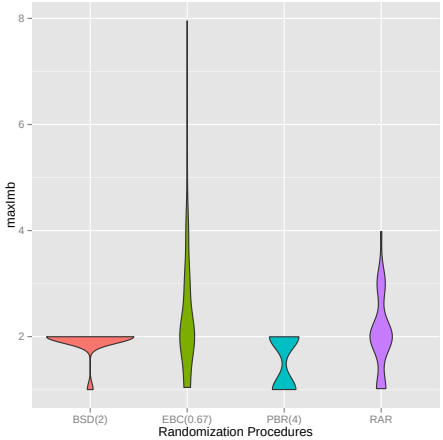
Assessment final imbalance

Comparison for loss

	BSD.2.	EBC.0.67.	PBR.4.	RAR
mean	0.250	0.369	0	0
sd	0.252	0.658	0	0
max	0.500	8.000	0	0
min	0.000	0.000	0	0
x05	0.000	0.000	0	0
x25	0.000	0.000	0	0
x50	0.000	0.000	0	0
x75	0.500	0.500	0	0
x95	0.500	2.000	0	0

- Total sample size $N = 8$.
- Compute the complete set of sequences for each design.
- Calculate the final imbalance of each sequence.

Assessment maximal imbalance



- Total sample size $N = 8$.
- Compute the complete set of sequences for each design.
- Calculate the maximal imbalance of each sequence.

FP7 HEALTH 2013 - 602552

Assessment Power loss due to imbalance

Comparison for power(exact)

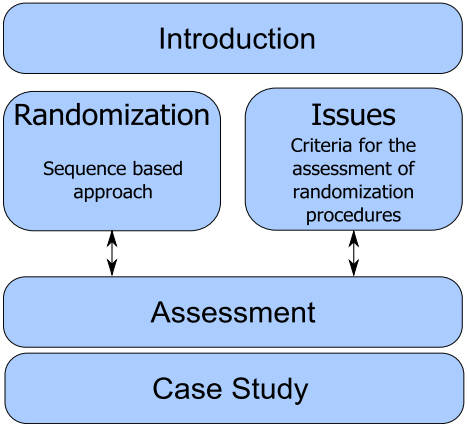
	BSD.2.	EBC.0.67.	PBR.4.	RAR
mean	0.787	0.779	0.8	0.8
sd	0.013	0.045	0.0	0.0
max	0.800	0.800	0.8	0.8
min	0.775	0.000	0.8	0.8
x05	0.775	0.682	0.8	0.8
x25	0.775	0.775	0.8	0.8
x50	0.775	0.800	0.8	0.8
x75	0.800	0.800	0.8	0.8
x95	0.800	0.800	0.8	0.8

- Total sample size $N = 8$.
- Compute the complete set of sequences for each design.
- Calculate the loss of each sequence.

Structure of the workshop

1. Introduction
2. Issues
3. Randomization procedures
4. Assessment
5. **Case study**

At the same time: Get acquainted with the `randomizeR` package $\Rightarrow$ Give feedback!



FP7 HEALTH 2013 - 602552

Template for the selection of a suitable randomization procedure

1. Define the sample size.
2. Identify the issues that might affect the trial.
   - Make assumptions the parameters of the issues.
3. Determine which randomization procedures are available.
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

FP7 HEALTH 2013 - 602552

Template for the selection of a suitable randomization procedure

1. Define the sample size.
2. Identify the issues that might affect the trial.
   - Make assumptions the parameters of the issues.
3. Determine which randomization procedures are available.
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

1. Total sample size $N = 100$
2. Several issues:
   - selection bias of intermediate strength
   - chronological bias due to novel study personnel after $n = 60$
   - power loss due to imbalance
3. *RAR*, *EBC*(2/3), *BSD*(5), *PBR*(10)
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

FP7 HEALTH 2013 - 602552

Template for the selection of a suitable randomization procedure

1. Define the sample size.
2. Identify the issues that might affect the trial.
   - Make assumptions the parameters of the issues.
3. Determine which randomization procedures are available.
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

1. Total sample size $N = 100$
2. Several issues:
   - selection bias of intermediate strength
   - chronological bias due to novel study personnel after $n = 60$
   - power loss due to imbalance
3. *RAR*, *EBC*(2/3), *BSD*(5), *PBR*(10)
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

FP7 HEALTH 2013 - 602552

Template for the selection of a suitable randomization procedure

1. Define the sample size.
2. Identify the issues that might affect the trial.
   - Make assumptions the parameters of the issues.
3. Determine which randomization procedures are available.
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

1. Total sample size $N = 100$
2. Several issues:
   - selection bias of intermediate strength
   - chronological bias due to novel study personnel after $n = 60$
   - power loss due to imbalance
3. *RAR*, *EBC*(2/3), *BSD*(5), *PBR*(10)
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

FP7 HEALTH 2013 - 602552

Template for the selection of a suitable randomization procedure

1. Define the sample size.
2. Identify the issues that might affect the trial.
   - Make assumptions the parameters of the issues.
3. Determine which randomization procedures are available.
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

1. Total sample size $N = 100$
2. Several issues:
   - selection bias of intermediate strength
   - chronological bias due to novel study personnel after $n = 60$
   - power loss due to imbalance
3. RAR, $EBC(2/3)$, $BSD(5)$, $PBR(10)$
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

FP7 HEALTH 2013 - 602552

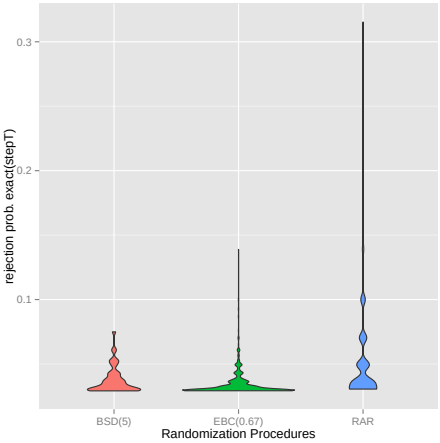
Template for the selection of a suitable randomization procedure

1. Define the sample size.
2. Identify the issues that might affect the trial.
   - Make assumptions the parameters of the issues.
3. Determine which randomization procedures are available.
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

1. Total sample size $N = 100$
2. Several issues:
   - selection bias of intermediate strength
   - chronological bias due to novel study personnel after $n = 60$
   - power loss due to imbalance
3. RAR, $EBC(2/3)$, $BSD(5)$, $PBR(10)$
4. Assess the randomization procedures according to the issues.
5. Choose the most suitable randomization procedure for the conduct of the trial.

FP7 HEALTH 2013 - 602552

Case study - issue chronological bias (type-I-error)

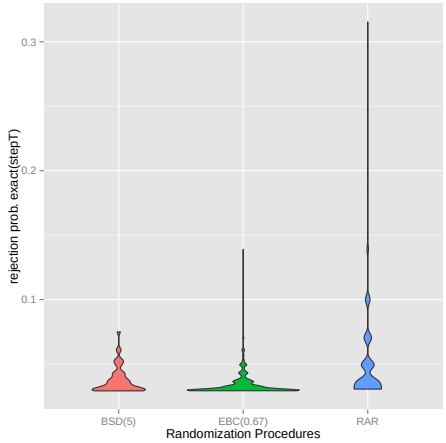


Simulation

- Total sample size $N = 100$.
- Number of repetitions $r = 1,000$
- Chronological bias
  - Step trend after $n = 60$ patients.
  - Strength $\vartheta = 1$.
- Sample r sequences from each randomization procedure and calculate for each sequence `p.value exact (stepT)`.
- `p.value exact (stepT)` is the probability of falsely rejecting the null hypothesis in case of the step trend.

FP7 HEALTH 2013 - 602552

Case study - issue chronological bias (type-I-error)

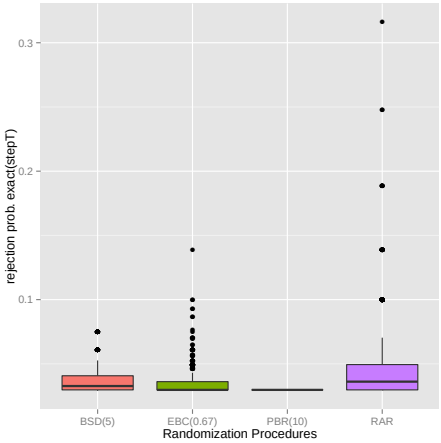


Comparison for rejection prob. exact(stepT)

	RAR	EBC.0.67.	BSD.5.	PBR.10.
mean	0.048	0.034	0.037	0.03
sd	0.029	0.009	0.010	0.00
max	0.316	0.139	0.075	0.03
min	0.030	0.029	0.029	0.03
x05	0.030	0.029	0.029	0.03
x25	0.030	0.030	0.030	0.03
x50	0.036	0.030	0.033	0.03
x75	0.049	0.036	0.041	0.03
x95	0.100	0.049	0.061	0.03

FP7 HEALTH 2013 - 602552

Case study - issue chronological bias (type-I-error)

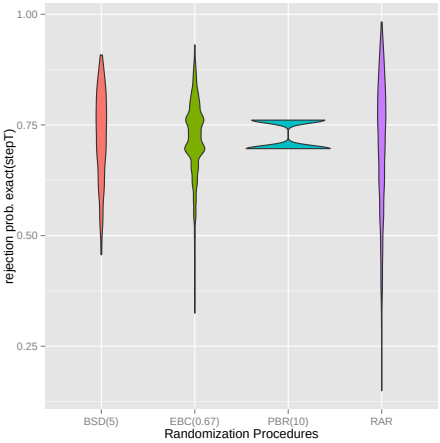


Comparison for rejection prob. exact(stepT)

	RAR	EBC.0.67.	BSD.5.	PBR.10.
mean	0.048	0.034	0.037	0.03
sd	0.029	0.009	0.010	0.00
max	0.316	0.139	0.075	0.03
min	0.030	0.029	0.029	0.03
x05	0.030	0.029	0.029	0.03
x25	0.030	0.030	0.030	0.03
x50	0.036	0.030	0.033	0.03
x75	0.049	0.036	0.041	0.03
x95	0.100	0.049	0.061	0.03

FP7 HEALTH 2013 - 602552

Case study - issue chronological bias (power)

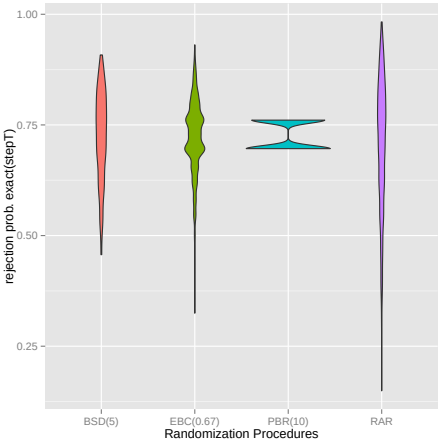


Simulation

- Total sample size $N = 100$.
- Number of repetitions $r = 1,000$
- Chronological bias
  - Step trend after $n = 60$ patients.
  - Strength $\vartheta = 1$.
- Sample r sequences from each randomization procedure and calculate for each sequence `p.value exact (stepT)`.
- `p.value exact (stepT)` is the probability of correctly rejecting the null hypothesis in case of the step trend.

FP7 HEALTH 2013 - 602552

Case study - issue chronological bias (power)

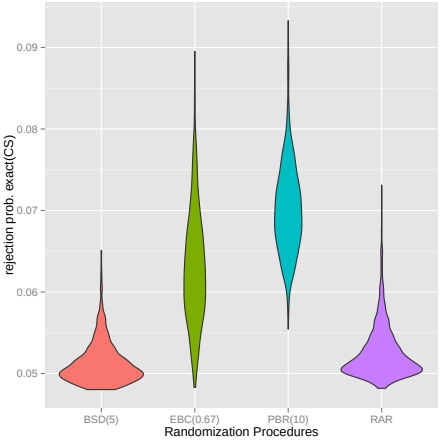


Comparison for rejection prob. exact(stepT)

	RAR	EBC.0.67.	BSD.5.	PBR.10.
mean	0.708	0.723	0.723	0.726
sd	0.147	0.081	0.103	0.033
max	0.983	0.934	0.910	0.761
min	0.149	0.321	0.455	0.696
x05	0.393	0.579	0.533	0.696
x25	0.624	0.696	0.654	0.696
x50	0.761	0.723	0.735	0.696
x75	0.818	0.772	0.808	0.761
x95	0.905	0.847	0.873	0.761

FP7 HEALTH 2013 - 602552

Case study - Issue selection bias (type-I-error)

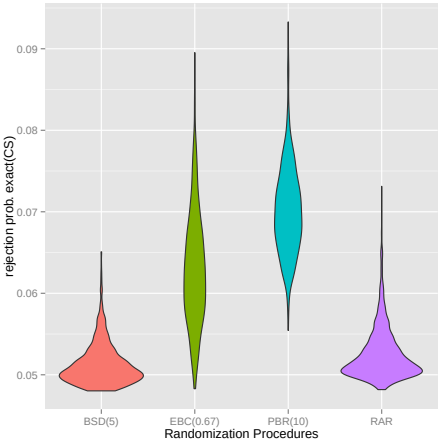


Simulation

- Total sample size $N = 100$.
- Number of repetitions $r = 1,000$
- Selection bias
  - Selection effect $\eta = d/4$, where $d = 0.57$ denotes the detectable effect of the two sample t-test for $N = 100$.
- Sample r sequences from each randomization procedure and calculate for each sequence `p.value exact (CS)`.
- `p.value exact (CS)` is the probability of falsely rejecting the null hypothesis in case of the step trend.

FP7 HEALTH 2013 - 602552

Case study - Issue selection bias (type-I-error)

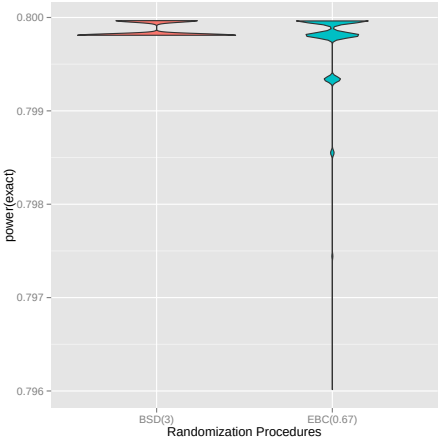


Comparison for rejection prob. exact(CS)

	RAR	EBC.0.67.	BSD.5.	PBR.10.
mean	0.053	0.063	0.051	0.070
sd	0.003	0.007	0.003	0.005
max	0.073	0.090	0.065	0.093
min	0.048	0.048	0.048	0.055
x05	0.049	0.052	0.048	0.062
x25	0.050	0.058	0.049	0.066
x50	0.052	0.062	0.051	0.070
x75	0.054	0.067	0.052	0.073
x95	0.059	0.075	0.056	0.078

FP7 HEALTH 2013 - 602552

Case study - Issue power loss

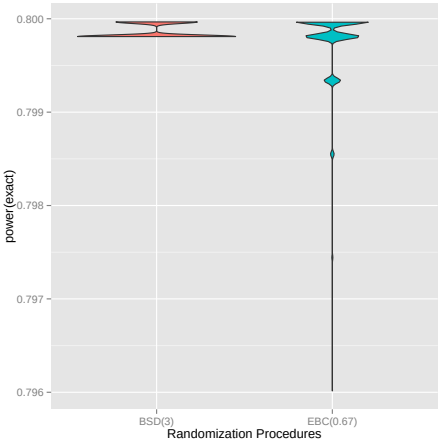


Simulation

- Total sample size $N = 100$.
- Number of repetitions $r = 1,000$
- Power loss
  - How much power do I lose for an effect of $d = 0.58$, the detectable effect for the t-test for $N = 100$
- Sample r sequences from each randomization procedure and calculate for each sequence `power(exact)`.
- `power(exact)` is the probability of correctly rejecting the null hypothesis, taking into account the final imbalance.

FP7 HEALTH 2013 - 602552

Case study - Issue power loss

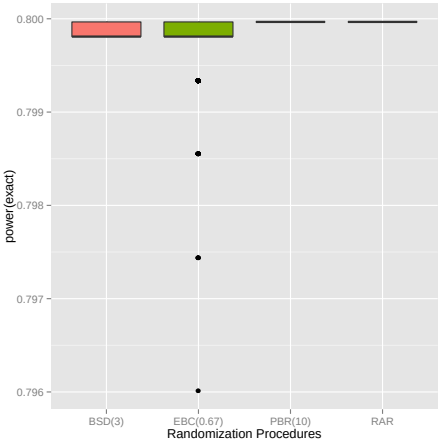


Comparison for power(exact)

	RAR	EBC.0.67.	BSD.3.	PBR.10.
mean	0.8	0.800	0.8	0.8
sd	0.0	0.000	0.0	0.0
max	0.8	0.800	0.8	0.8
min	0.8	0.796	0.8	0.8
x05	0.8	0.799	0.8	0.8
x25	0.8	0.800	0.8	0.8
x50	0.8	0.800	0.8	0.8
x75	0.8	0.800	0.8	0.8
x95	0.8	0.800	0.8	0.8

FP7 HEALTH 2013 - 602552

Case study - Issue power loss

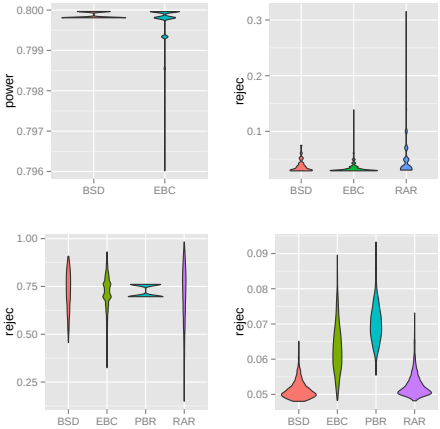


Comparison for power(exact)

	RAR	EBC.0.67.	BSD.3.	PBR.10.
mean	0.8	0.800	0.8	0.8
sd	0.0	0.000	0.0	0.0
max	0.8	0.800	0.8	0.8
min	0.8	0.796	0.8	0.8
x05	0.8	0.799	0.8	0.8
x25	0.8	0.800	0.8	0.8
x50	0.8	0.800	0.8	0.8
x75	0.8	0.800	0.8	0.8
x95	0.8	0.800	0.8	0.8

FP7 HEALTH 2013 - 602552

Conclusion of the case study



- Slight differences between the procedures visible.
- Lack of classification: Is the difference significant?
- Other issues, other recommendations!
  - No general rule detectable.
  - Elaborate simulation study necessary for each scenario.
  - The proposed tool allows high flexibility.
- Include several randomization procedures in the assessment.

FP7 HEALTH 2013 - 602552

Conclusion

The choice of a randomization procedure should follow a sound scientific evaluation of its properties with respect to the constraints in a clinical trial.

FP7 HEALTH 2013 - 602552

Outlook

- Release the `randomizeR` R package on CRAN (end of September 15).
  - Get latest news on `http://www.ideal.rwth-aachen.de/`
  - ..or sign up for the comprehensive IDeAl newsletter!
- Propose unified criterion for the simultaneous assessment of the issues.
- Use randomization tests in order to include the design in the analysis.
- Visit our other talks
  - C02.1: Nicole Heussen: Missing data in randomization tests.
  - I02.1: Ralf-Dieter Hilgers: Integrated Design and AnaLysis in SPG trials (IDeAl).
  - C13.2: Diane Uschner: Selection bias in randomization tests.
  - C13.4: David Schindler: Desirability scores as a linked optimization criterion.
- Please give us your feedback!

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.

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.

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.

FP7 HEALTH 2013 - 602552