Aspects for the scientific evaluation of randomization procedures in small clinical trials

Ralf-Dieter Hilgers
Nicole Heussen

Department of Medical Statistics, RWTH Aachen University

R. Clifton Bailey Seminar Series Spring 2016
Department of Statistics George Mason University, April 22

FP7 HEALTH 2013 - 602552

Table of contents

FP7 HEALTH 2013 - 602552

Study Design in Practice

- no randomization procedure performs best with all criteria
- no recommendation to give scientific arguments for selection of randomization procedure
- 21 out of 63 Orphan drug legislations involve open label studies (Joppi, 2013)

FP7 HEALTH 2013 - 602552

Objective

1. Present a framework for assessment of the impact of bias (both, selection and chronological) on the type one error probability for a given randomization procedure
2. Stimulate a scientific discussion of the appropriate choice of the randomization procedure
3. Understanding the properties of randomization tests in practical settings

FP7 HEALTH 2013 - 602552

Clinical Scenario Evaluation (CSE)

1. **Introduction** - state the general intend of the CSE
2. **Objective** - state clearly the general objectives of the CSE
3. **CSE framework**
   - **Assumptions** - describe the range of different assumptions
   - **Options** - state the different options, e.g. competing designs
   - **Metrics** - describe the measures to evaluate different options
4. **Evaluation Methods** - describe statistical models
5. **Software** - describe software used
6. **Result** - report the results
7. **Discussion** -discuss wether CSE met the planned objectives, assumptions and options
   - **Evaluation concept** explain the concept of evaluation
   - **Clinical implication** how can results be applied
8. **Conclusion** draw conclusions

(Benda, 2011)

FP7 HEALTH 2013 - 602552

Assumptions of CSE-Randomization

focus on the magnitude of the selection bias effect η and the time trend θ based on reasonable assumptions

- time trend as a synonym for chronological bias
- practical experience
- reporting standard is weak, no recommendation to report about the randomization list or randomization procedure

3.2 CSE - Options

Options of CSE-Randomization

select a set of randomization procedures with respect to the study settings by showing the influence of bias on the study results (take into account various parameters of the RP's)

- **Randomization Procedures**: Rosenberger Lachin (2016) Randomization in Clinical Trials. Wiley, New Jersey.

3.3 CSE - Metrics

Choose a measure which reflects the impact of bias on the results of the trial with respect to the randomization procedure:

ICH E9: The interpretation of statistical measures of uncertainty of the treatment effect and treatment comparisons should involve consideration of the potential contribution of bias to the p-value, confidence interval, or inference.

Metric of CSE randomization

- $\rightarrow$ **empirical type-I-error rate**
- $\rightarrow$ **confidence interval**

Evaluation Methods of CSE - Randomization

use a specific design, e.g. two arm parallel group with continuous endpoint, to analyse the impact of various randomization procedures with respect to the study settings (bias specifications) on the study results e.g. type I error probability

- model
  - two arm parallel group with continuous endpoint
  - multiarm parallel group with continuous endpoint
  - two arm parallel group with time to event endpoint
- bias specification
  - selection bias
  - chronological bias

4. CSE - Evaluation: Statistical Model (1)

two arm parallel group design, continuous endpoint

Aim: test the hypotheses $H_0 : \mu_E = \mu_C$ vs. $H_1 : \mu_E \neq \mu_C$

Model for two arm parallel group design with continuous endpoint

$$Y_i = \mu_E T_i + \mu_C(1 - T_i) + \tau_i + \epsilon_i, \quad 1 \leq i \leq N_E + N_C$$

- allocation
$$T_i = \begin{cases} 1 & \text{if patient } i \text{ is allocated to group } E \\ 0 & \text{if patient } i \text{ is allocated to group } C \end{cases}$$
- μ_j expected response under treatment $j = C, E$
- τ_i denotes the fixed unobserved "bias" effect acting on the response of patient i
- errors ϵ_i iid $\mathcal{N}(0, \sigma^2)$

4. CSE - Evaluation: (general) Selection Bias Model

two arm parallel group trial continuous endpoint

Biasing policy according to convergence strategy

$$\tau_i = \begin{cases} \eta & \text{if } p_E(i-1) > q \\ 0 & \text{if } 1-q \leq p_E(i-1) \leq q\,, \quad q \in [\frac{1}{2}, 1] \\ -\eta & \text{if } p_E(i-1) < 1-q \end{cases}$$

with $N_E = N_C$ and the number of treatment j assignments $n_j(i)$ after i assignments:

$$p_E(i-1) = \frac{N_E - n_E(i-1)}{(N_E+N_C)-(n_E(i-1)+n_C(i-1))}$$

---

with $\mathbb{1}_{(q,1]}(p_E(k)) = 1$, if $q < p_E(k) \leq 1$ write briefly:

$$\tau_i = \eta \left[\mathbb{1}_{(q,1]}(p_E(i-1)) - \mathbb{1}_{[0,1-q)}(p_E(i-1))\right]$$

(Tamm, 2012)

4. CSE - Evaluation: (general) Selection Bias Model

- two sided t-test
- PBR
- $delta = 0.909 : N_E = N_C = 20; \alpha = 0.05, 1 - \beta = 0.8$
- selection effect $\eta = \frac{\delta}{2} = 0.45$

cutoff q	empirical type 1 error rate PBR(4)	PBR(8)	PBR(10)
1/2	0.191	0.134	0.118
2/3	0.145	0.134	0.086
1	0.050	0.050	0.048

using SAS with 10 000 replications

4. CSE - Evaluation: Selection Bias Model (1)

two arm parallel group trial continuous endpoint

Biasing policy according to convergence strategy

$$\tau_i = \begin{cases} \eta & \text{if } n_E(i-1) < n_C(i-1) \\ 0 & \text{if } n_E(i-1) = n_C(i-1) \\ -\eta & \text{if } n_E(i-1) > n_C(i-1) \end{cases}$$

- η proportional to effect size δ
- $\tau_i = \eta \ [\text{sign}(\ n_E(i-1) - n_C(i-1)\ )]$
- $n_j(k)$: assignments to treatment j after k allocations

(Proschan 1994)
(Kennes 2011)

4. CSE - Evaluation: Statistical Model (2)

multiarm parallel group design, continuous endpoint, here $k = 3$

Aim: test the hypotheses $H_0 : \mu_E = \mu_{C_1} = \mu_{C_2}$ vs. $H_1 : \mu_E \neq \mu_{C_1} \neq \mu_{C_2}$

Model for multiarm parallel group design with continuous endpoint

$$Y_i = \sum_{j=0}^{k} \mu_j \, \mathbb{1}_{\{j\}}(T_i) + \tau_i + \epsilon_i, \quad 1 \leq i \leq \sum_{j=0}^{k} N_j$$

- allocation $T_j \in \{0, \ldots, k\}$ indicate the $(k+1)$ treatments, here k=2, treatments $\{E, C_1, C_2\}$
- $\mathbb{1}_{\{j\}}(T_i) = 1$ if $T_i = j$ else $\mathbb{1}_{\{j\}}(T_i) = 0$
- μ_j expected response under treatment j
- τ_i denotes the fixed unobserved "bias" effect acting on the response of patient i
- errors ϵ_i iid $\mathcal{N}(0, \sigma^2)$

FP7 HEALTH 2013 - 602552

4. CSE - Evaluation: Selection Bias Model (2)

three arm trial with treatments E, C_1 and C_2

Biasing policy according to convergence strategy

$$\tau_i = \begin{cases} \eta & \text{if } n_E(i-1) < \min\{n_{C_1}(i-1), n_{C_2}(i-1)\}, \\ -\eta & \text{if } n_E(i-1) > \max\{n_{C_1}(i-1), n_{C_2}(i-1)\} \\ 0 & \text{otherwise} \end{cases}$$

- η proportional to effect size δ
- $n_j(k)$: assignments to treatment j after k allocations, $j \in \{E, C_1, C_2\}$
- $\rightarrow$ ANOVA model

(Tasche, 2016)

4. CSE - Evaluation: Statistical Model (3)

wo arm parallel group design, time to event data

Aim: test the hypotheses $H_0 : \lambda_E = \lambda_C$ vs. $H_1 : \lambda_E \neq \lambda_C$

Model for 2-arm parallel group design with time to event endpoint and additive proportional hazard

$$h(y_i) = \lambda_C \exp(\beta T_i + \tau_i) \quad 1 \leq i \leq N_E + N_C$$

- allocation

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

- $h(y_i)$ hazard rate of ith patient
- λ_j hazard rate under treatment $j = C, E$ with $\lambda_E = \lambda_C \exp(\beta)$
- τ_i fixed unobserved "bias" effect acting on the response of patient i

4. CSE - Evaluation: Selection Bias Model (3)

two arm trial with time to event data

Biasing policy according to convergence strategy

$$\tau_i = \begin{cases} \delta & \text{if } n_E(i-1) > n_C(i-1) \\ 0 & \text{if } n_E(i-1) = n_C(i-1) \\ -\delta & \text{if } n_E(i-1) < n_C(i-1) \end{cases}$$

- $\delta \in (0, \infty)$
- $n_j(k)$: assignments to treatment j after k allocations,
- $\rightarrow$ F test model

(Rückbeil, 2015)

4. CSE - Evaluation: Time Trend Bias Model

two arm parallel group trial continuous endpoint

Biasing policy according to convergence strategy

$$\tau_i = \theta \times \begin{cases} \frac{i}{N_E+N_C} & \text{linear time trend} \\ \mathbb{1}_{i \geq S}(i) & \text{stepwise trend} \\ \log(\frac{i}{N_E+N_C}) & \text{log trend} \end{cases}$$

- θ proportional to variance
- other functions are possible
- long recruitment time in Rare Diseases, (EMA, 2006)
  - changes in population characteristics
  - learning effect in therapy / surgical experience (Hopper, 2007)
  - change in diagnosis (FDA, 2011), etc.
- special form of accidental bias, when considering a time-heterogeneous covariate *(Tamm, 2014)*

4. CSE - Evaluation: Joint Additive Bias Model (2)

two arm parallel group trial continuous endpoint

Joint Additive Bias

$$\tau_i = \underbrace{\theta \frac{i}{N_E + N_C}}_{\text{time trend}} + \underbrace{\eta \ [\text{sign}(\ n_E(i-1) - n_C(i-1)\ )]}_{\text{selection bias}}$$

- weighted additive (selection and chronological) bias model
- weights via definition of θ and η
- multiplicative could also be done
- different shape of time trend can be incorporated (Tamm et al., 2014)
- relaxed version of bias policy (non strict decision, random η)

4. CSE - Evaluation: Statistical Test for Model (1)

two arm parallel group trial continuous endpoint

Aim: test the hypotheses $H_0 : \mu_E = \mu_C$ vs. $H_1 : \mu_E \neq \mu_C$
use Z-Test (Gaussian test) (under misspecification)

$$S = \frac{(\tilde{y}_E - \tilde{y}_C)}{\sqrt{2n\sigma^2}}$$

where $\tilde{y}_E = \frac{1}{N_E} \sum_{i=1}^{N} y_i T_i$; $\tilde{y}_C = \frac{1}{N_C} \sum_{i=1}^{N} y_i (1 - T_i)$; $N = N_E + N_C$

4. CSE - Evaluation: Statistical Test for Model (1)

Under $H_0 : \mu_E = \mu_C$ the type 1 error probability for the two arm parallel group normal model (under misspecification) for

$$P(|Z| > z_{1-\alpha/2}) = 1 - \sum_{i_1=1}^{k_1} \cdots \sum_{i_M=1}^{k_M} \left\{ \Phi\left(z_{1-\frac{\alpha}{2}} - \frac{i_1 + \ldots + i_M}{\sqrt{2n}} \cdot \gamma_c \right) - \Phi\left(-z_{1-\frac{\alpha}{2}} - \frac{i_1 + \ldots + i_M}{\sqrt{2n}} \cdot \gamma_c \right) \right\} \prod_{j=1}^{M} P(N'_{2k_j} = i_j)$$

- PBR$(k_1, \ldots, k_M)$, M blocks of length
- N'_{2k_j}: number of returns to the origin in $j-$th block, $j \in \{1, \ldots, M\}$
- $\gamma_c := \frac{\eta_c}{\sigma}$ selection bias effect

(Kennes, 2011)

4. CSE - Evaluation: Statistical Test for Model (1)

two arm parallel group trial continuous endpoint

Aim: test the hypotheses $H_0 : \mu_E = \mu_C$ vs. $H_1 : \mu_E \neq \mu_C$

use t-Test (under misspecification)

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

where $\tilde{y}_E = \frac{1}{N_E}\sum_{i=1}^{N} y_i T_i$; $\tilde{y}_C = \frac{1}{N_C}\sum_{i=1}^{N} y_i(1-T_i)$; $N = N_E + N_C$

4. CSE - Evaluation: Statistical Test for Model (1)

Theorem: Under $H_0 : \mu_E = \mu_C$ the type 1 error probability for the two arm parallel group normal model (under misspecification) for the allocation sequence $\mathbf{T} = (T_1, \ldots, T_{N_E+N_C})$ is

$$P\left(|S| > t_{N_E+N_C-2}(1-\alpha/2)\middle|\mathbf{T}\right)$$
$$= F_{N-2,\vartheta,\lambda}\left(t_{N_E+N_C-2}(\alpha/2)\right) + 1 - F_{N_E+N_C-2,\vartheta,\lambda}\left(t_{N_E+N_C-2}(1-\alpha/2)\right).$$

$F_{N_E+N_C-2,\vartheta,\lambda}$ denotes the distribution function of the doubly non-central t-distribution with $N_E + N_C - 2$ degrees of freedom and parameters

$$\vartheta = \frac{1}{\sigma}\sqrt{\frac{N_E N_C}{N_E + N_C}}\,(\tilde{\tau}_E - \tilde{\tau}_C) \quad \lambda = \frac{1}{\sigma^2}\left[\sum_{i=1}^{N}\tau_i^2 - N_E\tilde{\tau}_E^2 - N_C\tilde{\tau}_C^2\right]$$

where $\tilde{\tau}_E = \frac{1}{N_E}\sum_{i=1}^{N}\tau_i T_i$; $\quad \tilde{\tau}_C = \frac{1}{N_C}\sum_{i=1}^{N}\tau_i(1-T_i)$

4. CSE - Evaluation: Statistical Test for Model (1)

Sketch of the proof:

for the given allocation vector $\mathbf{T} = (T_1, \ldots, T_{N_E+N_C})$

$$\tilde{y}_E - \tilde{y}_C \sim \mathcal{N}\left(\mu_E - \mu_C + \tilde{\tau}_E - \tilde{\tau}_C, \sigma^2 \frac{N_E + N_C}{N_E N_C}\right)$$

where $\tilde{\tau}_E = \frac{1}{N_E} \sum_{i=1}^{N} \tau_i T_i$; $\quad \tilde{\tau}_C = \frac{1}{N_C} \sum_{i=1}^{N} \tau_i (1 - T_i)$

4. CSE - Evaluation: Statistical Test for Model (1)

for the allocation vector $\mathbf{T} = (T_1, \ldots, T_{N_E+N_C})$

$$\sum_{i=1}^{N} T_i(y_i - \tilde{y}_E)^2 + \sum_{i=1}^{N}(1 - T_i)(y_i - \tilde{y}_C)^2 \sim \chi_{N_E+N_C-2}(\lambda)$$

with non-centrality parameter

$$\begin{aligned}\lambda &= \frac{1}{\sigma}\left(\sum_{i=1}^{N} T_i(\tau_i - \tilde{\tau}_E)^2 + \sum_{i=1}^{N}(1 - T_i)(\tau_i - \tilde{\tau}_C)^2\right) \\ &= \frac{1}{\sigma}\left[\sum_{i=1}^{N} \tau_i^2 - N_E\tilde{\tau}_E^2 - N_C\tilde{\tau}_C^2\right]\end{aligned}$$

4. CSE - Evaluation: Statistical Test for Model (1)

Thus T follows a doubly non-central t distribution with $N_E + N_C - 2$ degrees of freedom and non-centrality parameters (Johnson, Kotz, Balakrishnan, 1995, Robins, 1948)

$$\vartheta = \frac{1}{\sigma}\sqrt{\frac{N_E N_C}{N_E + N_C}}\,(\mu_E - \mu_C + \tilde{\tau}_E - \tilde{\tau}_C) = \frac{1}{\sigma}\sqrt{\frac{N_E N_C}{N_E + N_C}}\,(\tilde{\tau}_E - \tilde{\tau}_C)$$

$$\lambda = \frac{1}{\sigma^2}\left[\sum_{i=1}^{N}\tau_i^2 - N_E\tilde{\tau}_E^2 - N_C\tilde{\tau}_C^2\right]$$

using the properties of the distribution (Kocherlakota, 1991)

$$F_{\nu,\vartheta,\lambda}(t) = 1 - F_{\nu,-\vartheta,\lambda}(-t)$$

5. CSE - Evaluation Method: randomizeR

. . . will use randomizeR, to conduct the evalaution and report the findings

current status of randomizeR

- implemented randomization procedures: `CR`, `RAR`, `PBR`, `RPBR`, `HADA`, `MP`, `BSD`, `UD`, `TBD`, `EBC`, `GBC`, `CD`, `BBC`

$\Rightarrow$ generating / saving a randomization sequence as `.csv` file

- implemented assessment criteria: `selBias`, `chronBias`, `corGuess`, `imbal`, `setPower`, `combineBias`

$\Rightarrow$ assessment and comparison of randomization procedures possible

in progress\next steps

- assessment of linked criteria, randomization tests, time to event model, multiarm model
- bias corrected test
- development of a shiny app

6. CSE - (general) Result: Selection Bias

with PBR, the empirical type I error is elevated (higher than 5%)

- ... substantial even up to $q = 2/3$ (Tamm, 2012)
- ... with smaller blocksize (Tamm, 2012; Kennes, 2011)
- ... with smaller blocksize and misclassification of patients performance (Tamm, 2012)
- ... with smaller blocksize in multiarm trials (Tasche, 2016)
- ... with smaller blocksize in time to event trials (Rückbeil, 2015)

empirical type I error elevation is reduced

- ... with a randomization list, which is too long
- ... with multicenter trial, where biasing policies in centers are in opposite directions
- ... with number of arms in multiarm trials (Tasche, 2016)

6. CSE - Result: Selection Bias

Empirical type 1 error probability of a two sided t-test

N	$\delta(N)$	BSD (2)	CR	EBCD ($\frac{2}{3}$)	MP(2)	PBR(4)	RAR
8	2.381	0.064	0.058	0.089	0.118	0.141	0.102
20	1.325	0.075	0.054	0.093	0.129	0.177	0.082
32	1.024	0.083	0.055	0.097	0.137	0.188	0.072
40	0.909	0.088	0.053	0.100	0.140	0.195	0.071

- $N_E = N_C, N_E + N_C = N$
- $\delta(N) : \alpha = 0.05, 1 - \beta = 0.8$
- selection bias effect $\eta = \frac{\delta(N)}{2}$

using R with 100 000 replications

6. CSE - (general) Result: Chronological Bias

Rosenkranz (2011) investigated CR, RAR, BCD, ABCD, TBD and found

- anticonservative behaviour of BCD and ABCD
- CR and RAR maintain the level
- TBD marked type I error probability elevation

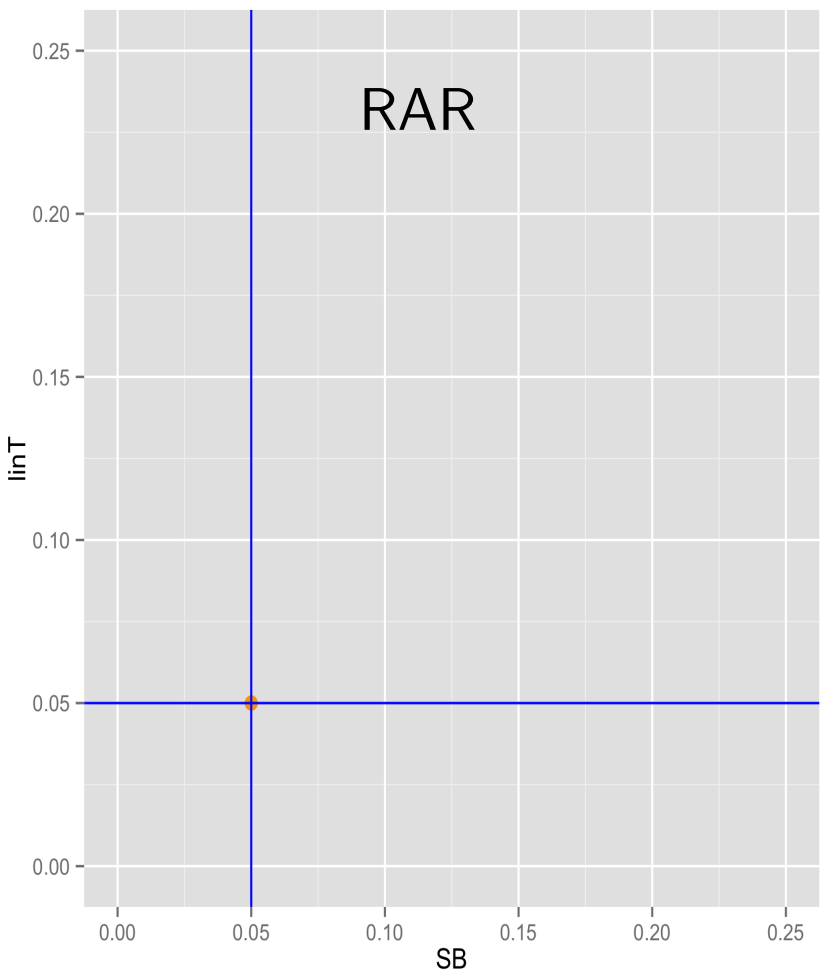
Tamm (2014) investigated PBR and found

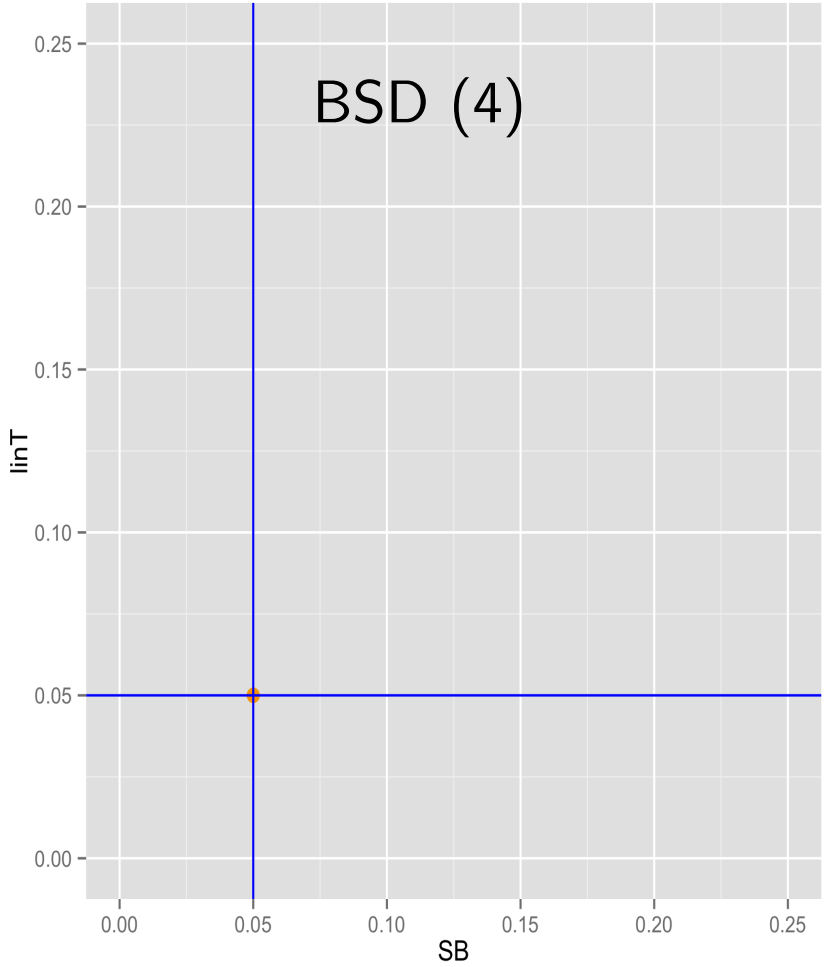
- large blocksizes linear time trend affect the empirical type I error rate toward conservative test decisions.
- medium block sizes because they already restrict chronological bias to an acceptable extent
- include blocklength in statistical analysis
- checked for possible time trends by using the graphical methods suggested by Altman and Royston

FP7 HEALTH 2013 - 602552

6. CSE - Result: both Biases for (N=96)

setting: $\eta = 0 \times \delta$; $\theta = 0 \times \sigma$

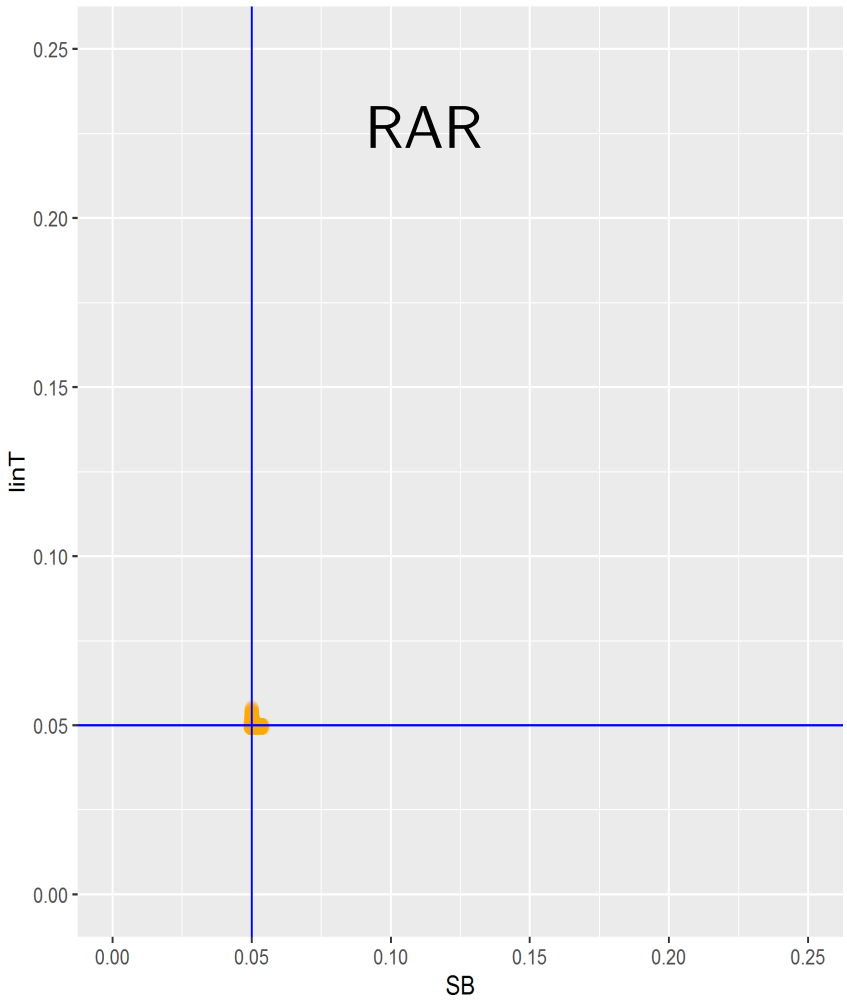


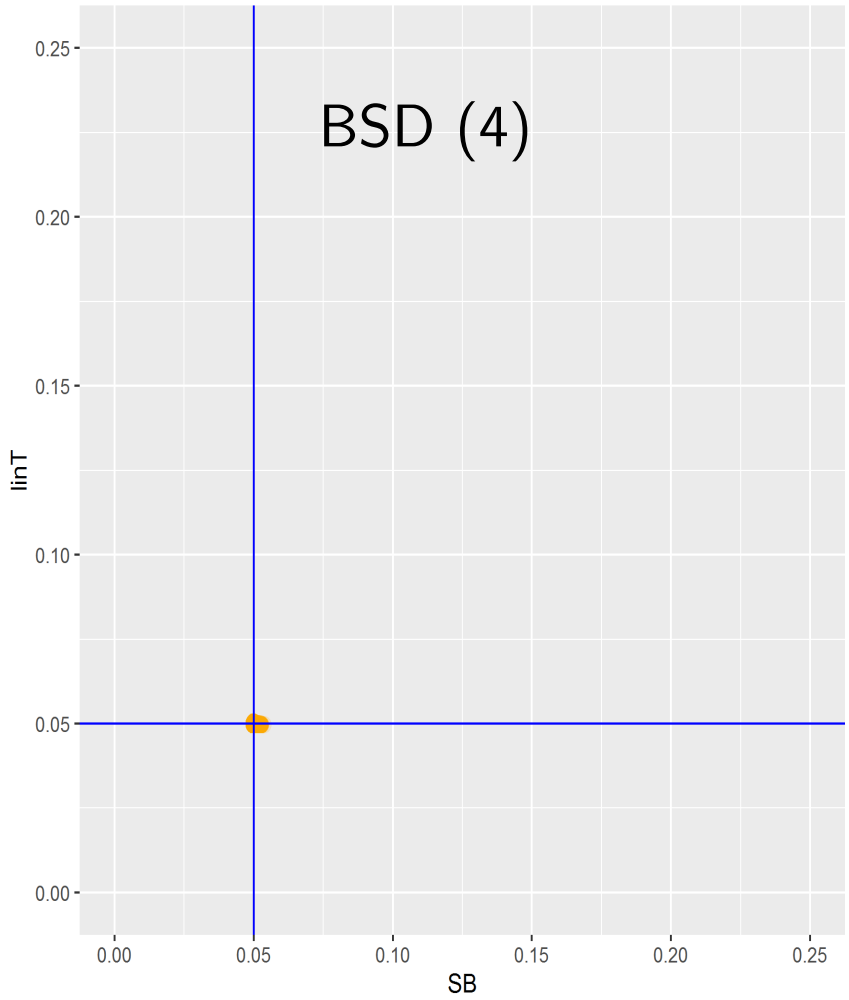


FP7 HEALTH 2013 - 602552

6. CSE - Result: both Biases for (N=96)

setting: $\eta = 0.1 \times \delta$; $\theta = 0.2 \times \sigma$

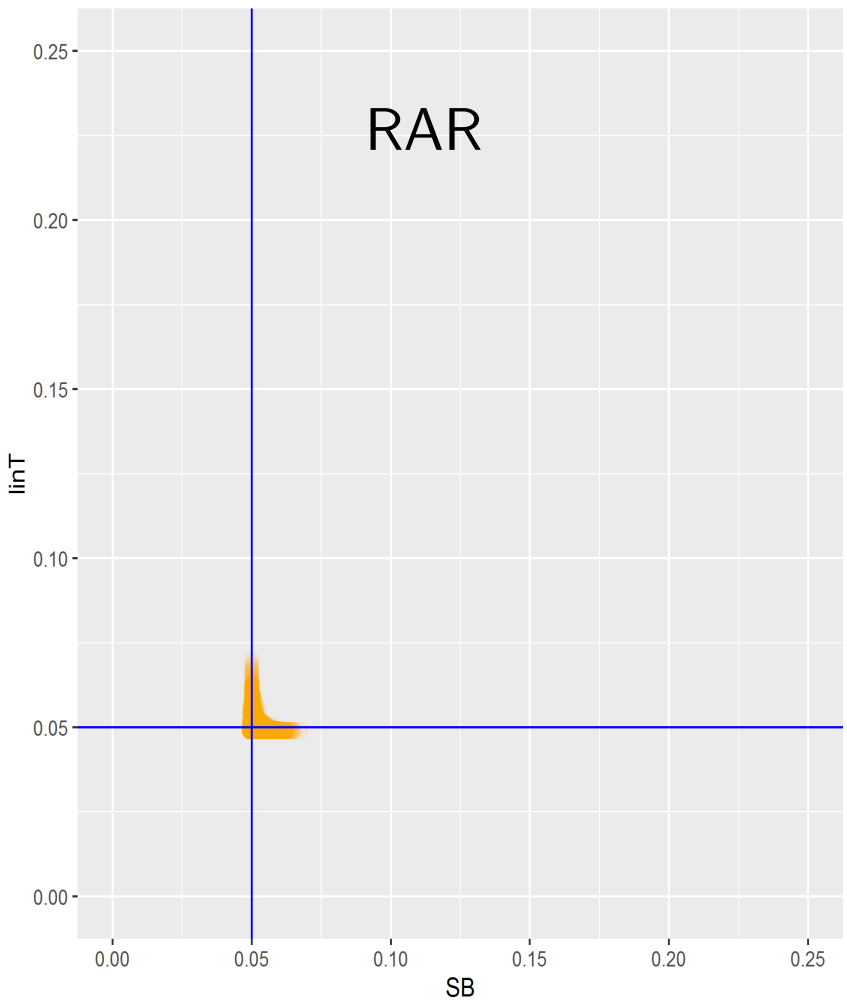


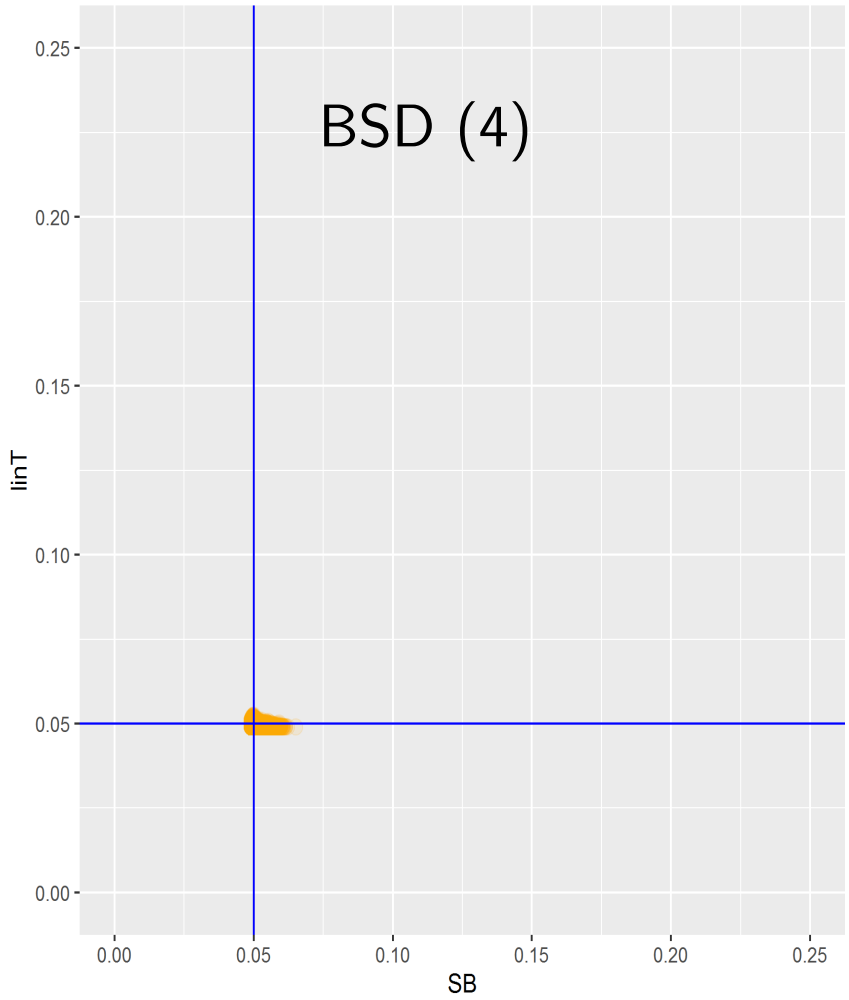


FP7 HEALTH 2013 - 602552

6. CSE - Result: both Biases for (N=96)

setting: $\eta = 0.2 \times \delta$; $\theta = 0.4 \times \sigma$

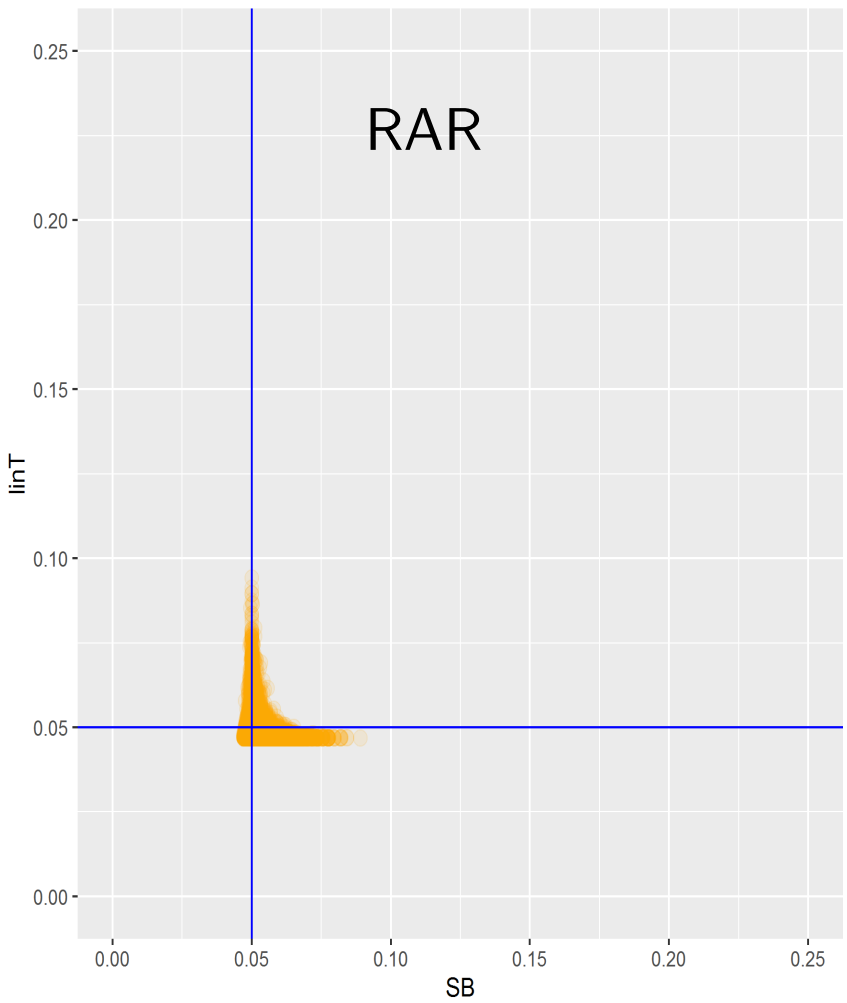


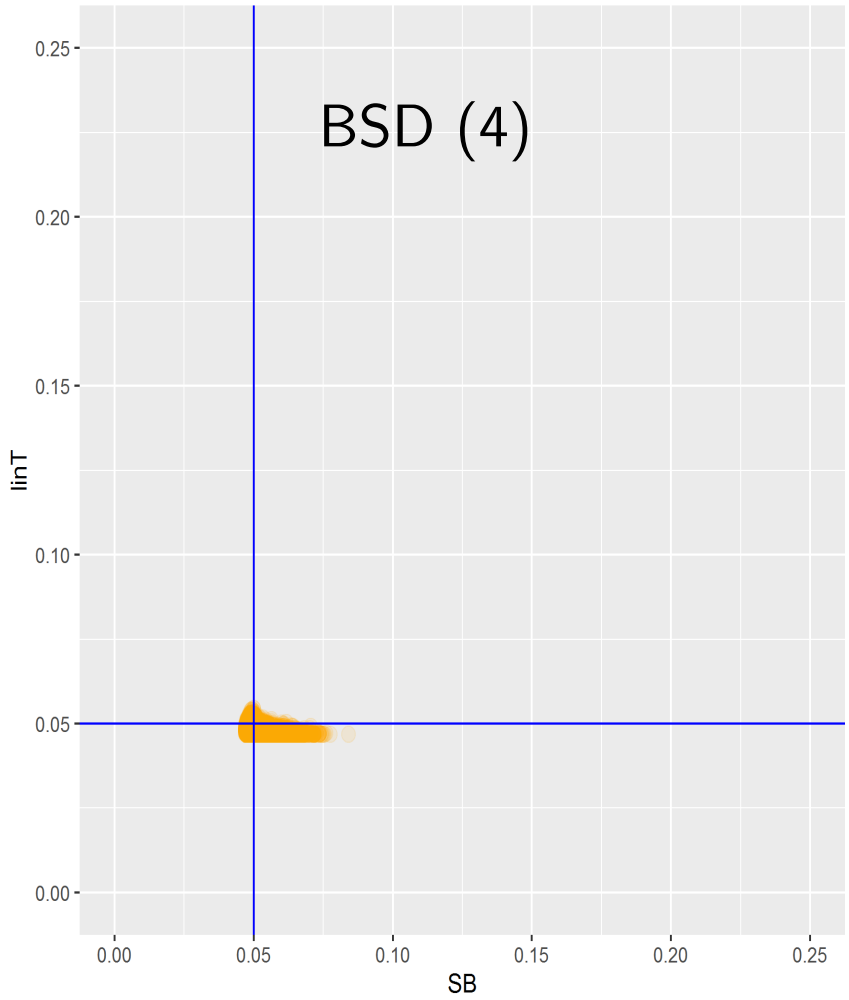


FP7 HEALTH 2013 - 602552

6. CSE - Result: both Biases for (N=96)

setting: $\eta = 0.3 \times \delta$; $\theta = 0.6 \times \sigma$

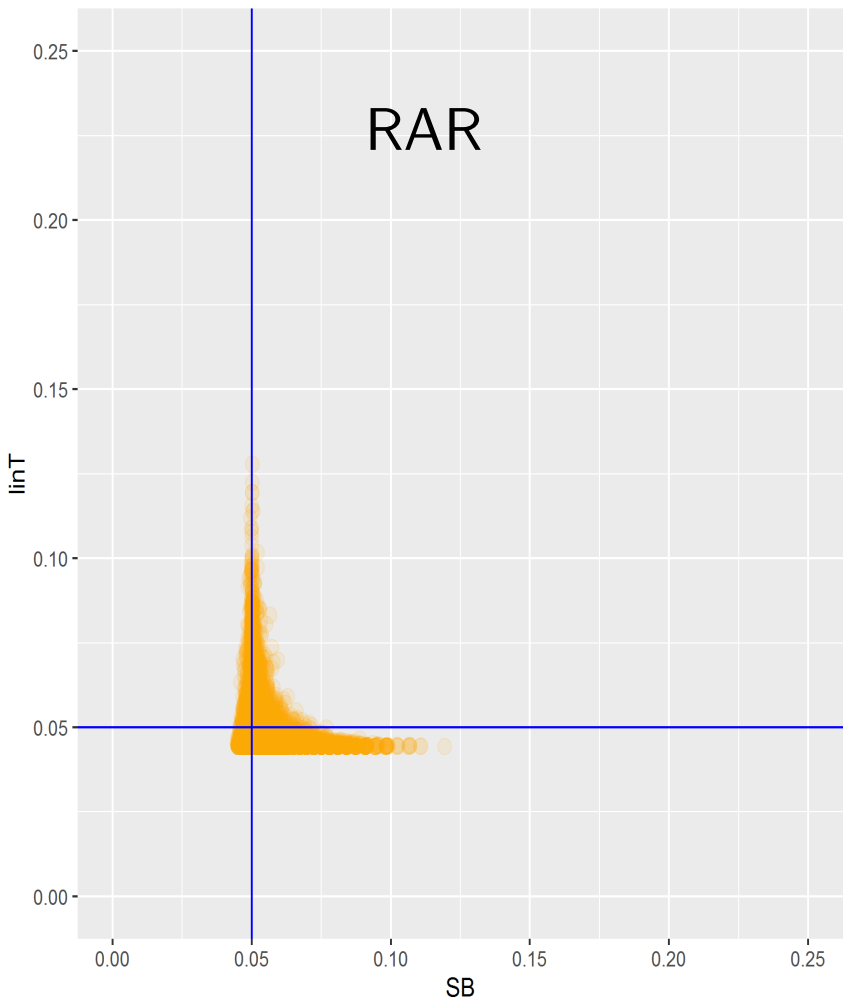


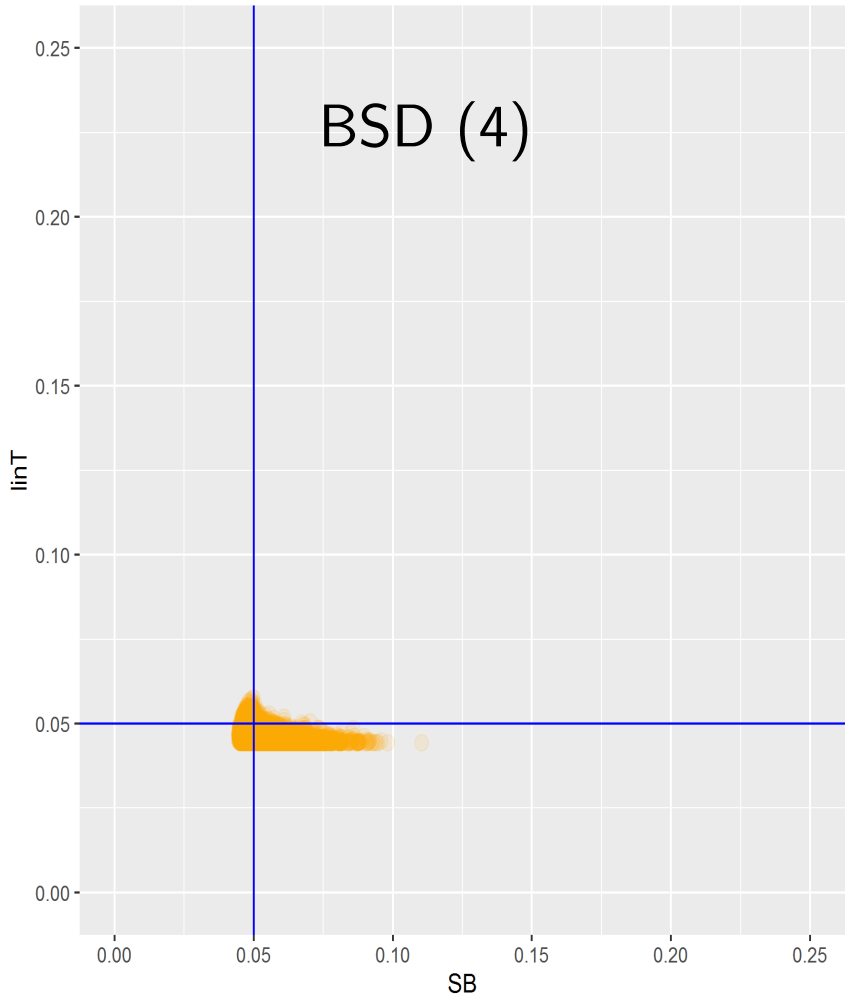


FP7 HEALTH 2013 - 602552

6. CSE - Result: both Biases for (N=96)

setting: $\eta = 0.4 \times \delta$; $\theta = 0.8 \times \sigma$

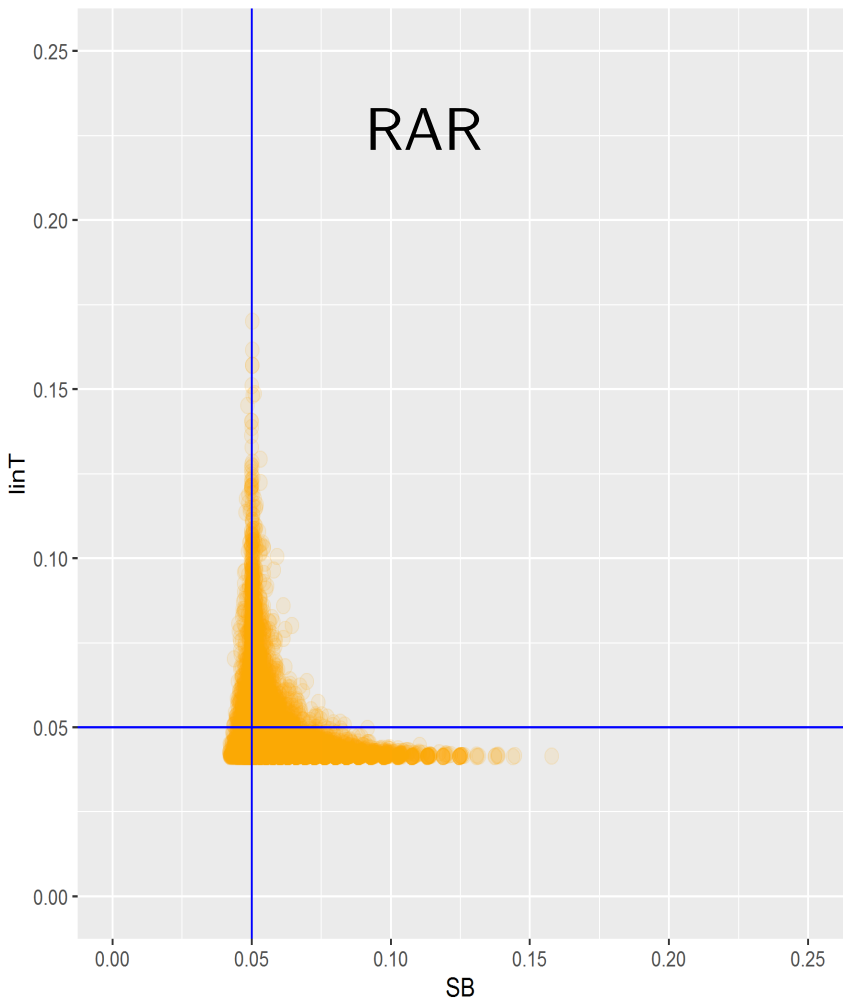


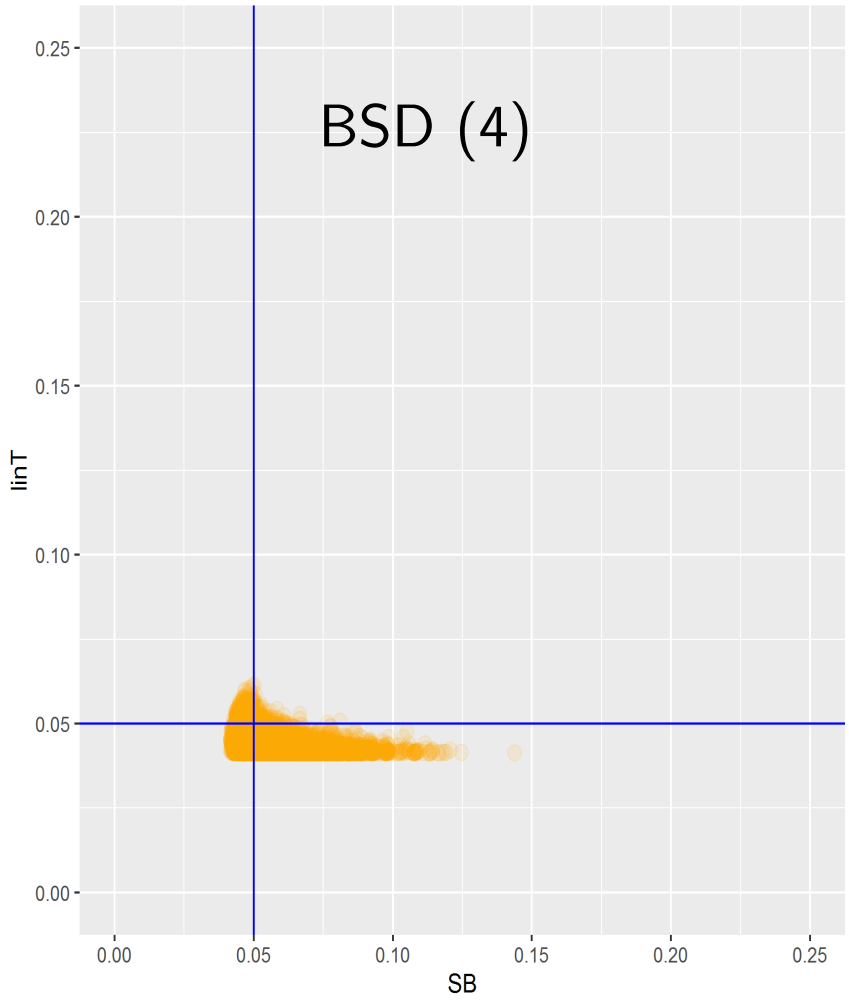


FP7 HEALTH 2013 - 602552

6. CSE - Result: both Biases for (N=96)

setting: $\eta = 0.5 \times \delta$; $\theta = 1 \times \sigma$





FP7 HEALTH 2013 - 602552

Summary: CSE - Randomization

1. **Introduction** - intend select the best practice randomization procedure (RP) to improve the level of evidence
2. **Objective** - state the RP with respect to impact on α
3. **CSE framework**
   - **Assumptions** - selection and time trend bias
   - **Options** - set of RP's
   - **Metrics** - (empirical) type I error rate
4. **Evaluation Methods** - e.g. parallel group, continuous endpoint,
5. **Software** - randomizeR
6. **Result** - report
7. **Discussion**
   - **Evaluation concept**
   - **Clinical implication** select the best practice (RP)
8. **Conclusion** choice of randomization design

FP7 HEALTH 2013 - 602552

Conclusion

- presented a framework for scientific evaluation of randomization procedures in the presence of bias, to be included in trial documents
- understand that the treatment effect could be hidden by bias, which max result from a randomization sequence
- Software to do assessment is available, R package *(randomizeR)*

- start understanding effects with time to event data *(Rückbeil, 2015)*
- start understanding effects with multifactorial designs *(Tasche, 2016)*
- start understanding the effect of missing values on the test decision based on randomization test

- no yet completely developed a bias corrected test *(Kennes, 2015)*

FP7 HEALTH 2013 - 602552

Research Group

- Ralf-Dieter Hilgers
- Nicole Heussen
- Lieven Kennes
- Miriam Tamm
- David Schindler
- Diane Uschner
- Marcia Rückbeil
- Simon Langer
- Antje Tasche
- Mui Pham
- Martin Manolov
- Christina Fitzner

FP7 HEALTH 2013 - 602552

Developments - A brief History

- **Randomization Procedures**: Rosenberger Lachin (2016) Randomization in Clinical Trials. Wiley, New Jersey.

Selection Bias	**Chronlogical Bias**
Blackwell & Hodges (1957)	Altman & Royston (1998)
Proschan (1994)	Rosenkranz (2011)
Kennes, Cramer, Hilgers &Heussen (2011)	Tamm & Hilgers (2011)
Tamm, Cramer, Kennes & Heussen (2011)	
Langer (2014)	
Rückbeil (2015)	
Kennes, Rosenberger & Hilgers (2015)	
Tasche (2016)	