Assessment of Randomization Procedures with Respect to the Influence of Bias on Type 1 error Elevation

Ralf-Dieter Hilgers

Department of Medical Statistics, RWTH Aachen University

MODA 11, 2016, June 12-17th

FP7 HEALTH 2013 - 602552

Table of contents

FP7 HEALTH 2013 - 602552

Study Design in Practice

(ICH E9): 2.3.3 Randomization: In combination with blinding, randomisation helps to avoid possible bias in the selection and allocation of subjects arising from the predictability of treatment assignments.

- no randomization procedure performs best with all criteria
  - Rosenberger (2016), Atkinson (2014), etc.
- no recommendation to give scientific arguments for the choice of randomization procedure
  - ICH Guidelines
  - CONSORT Statement
- 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-I-error probability for a given randomization procedure
2. understanding the properties of randomization procedures in practical settings
3. stimulate a discussion of the selection of an appropriate randomization procedure based on scientific arguments

Clinical Scenario Evaluation (CSE)

evaluate various designs with respect to the clinical situation

1. **Introduction**
2. **Objective** - select the best practice RP to improve the level of evidence
3. **CSE framework**
   1. **Assumptions**
   2. **Options**
   3. **Metrics**
4. **Evaluation Methods**
5. **Software**
6. **Result**
7. **Discussion**
   1. **Evaluation concept** select the best practice (RP)
   2. **Clinical implication**
8. **Conclusion** choice of randomization design

(Benda, 2011)

FP7 HEALTH 2013 - 602552

3.2 CSE - Options: Randomization Procedures

Fixed sample procedures, no adaptive randomization procedures.

- CR Complete randomization is accomplished by tossing a fair coin, so the probability that patient i will receive treatment 1 is always $\frac{1}{2}$
- RAR Random Allocation rule, fix total sample size N. Randomize so that half the patients receive treatment 1
- PBR (Permuted Block Randomization) Implementation of RAR within B Blocks of size $b_s, 1 \leq s \leq B$
- BSD(a) (Big Stick design) CR allow for imbalance within a limit a
- EBC(p) (Efrons Biased Coin) flip a biased coin (p) in favour of the treatment which is allocated less frequently

...etc.

3.3 CSE - Metric: type-I-error probability

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.

- per sequence (conditional) approach
- averaged (unconditional) approach

Metric of CSE randomization

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

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 (Kennes, 2011), (Langer, 2014)
  - multiarm parallel group with continuous endpoint (Tasche, 2016)
  - two arm parallel group with time to event endpoint (Rückbeil, 2015)
- bias specification
  - selection bias (Kennes, 2011), (Tamm, 2011), (Rückbeil, 2015), (Tasche, 2016)
  - chronological bias (Tamm, 2014)

FP7 HEALTH 2013 - 602552

4. CSE - Evaluation: Statistical Model

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

(Proschan 1994)
(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-I-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

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

where $\tilde{\tau}_E = \frac{1}{N_E}\sum_{i=1}^{N}\tau_i T_i$; $\tilde{\tau}_C = \frac{1}{N_C}\sum_{i=1}^{N}\tau_i(1-T_i)$ *(Langer, 2014)*

6. CSE - Result: Selection Bias

Empirical type-I-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

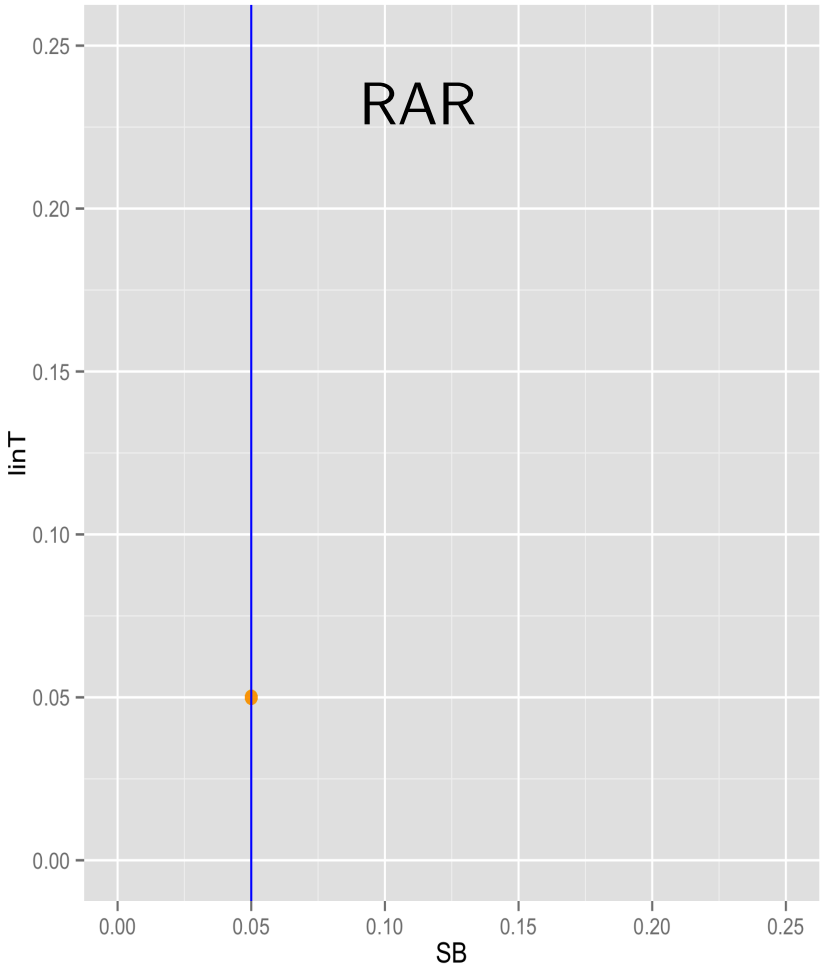
- $N_E = N_C, N_E + N_C = N$
- $\delta(N) : \alpha = 0.05, 1 - \beta = 0.8$
- selection bias effect $\eta = \frac{\delta(N)}{2}$

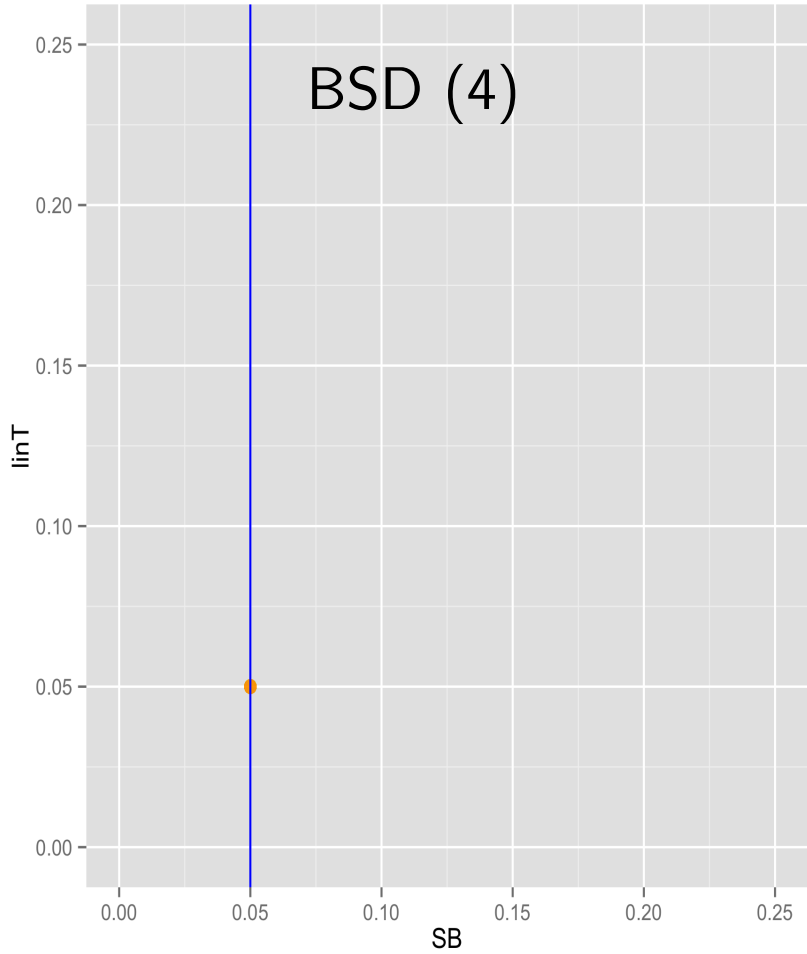
using R with 100 000 replications

(calculation by Tamm, 2015)

6. CSE - Result: Selection Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0.0 \times \delta;\ \theta = 0$

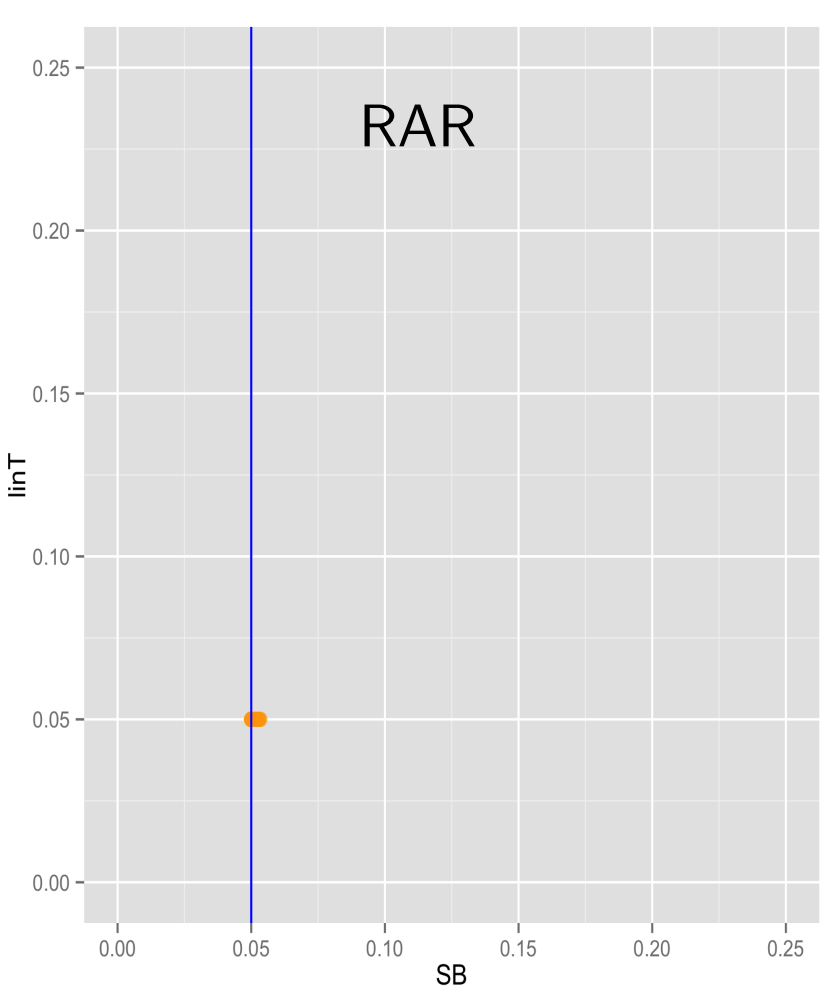


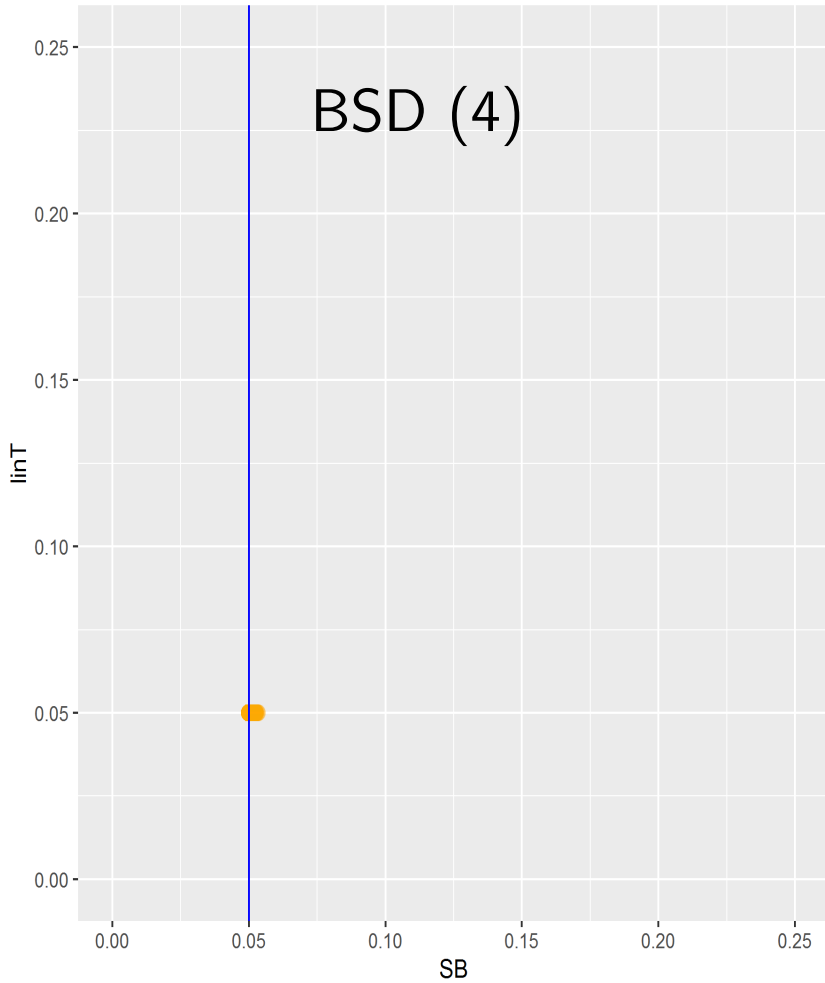


FP7 HEALTH 2013 - 602552

6. CSE - Result: Selection Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0.1 \times \delta;\ \theta = 0$

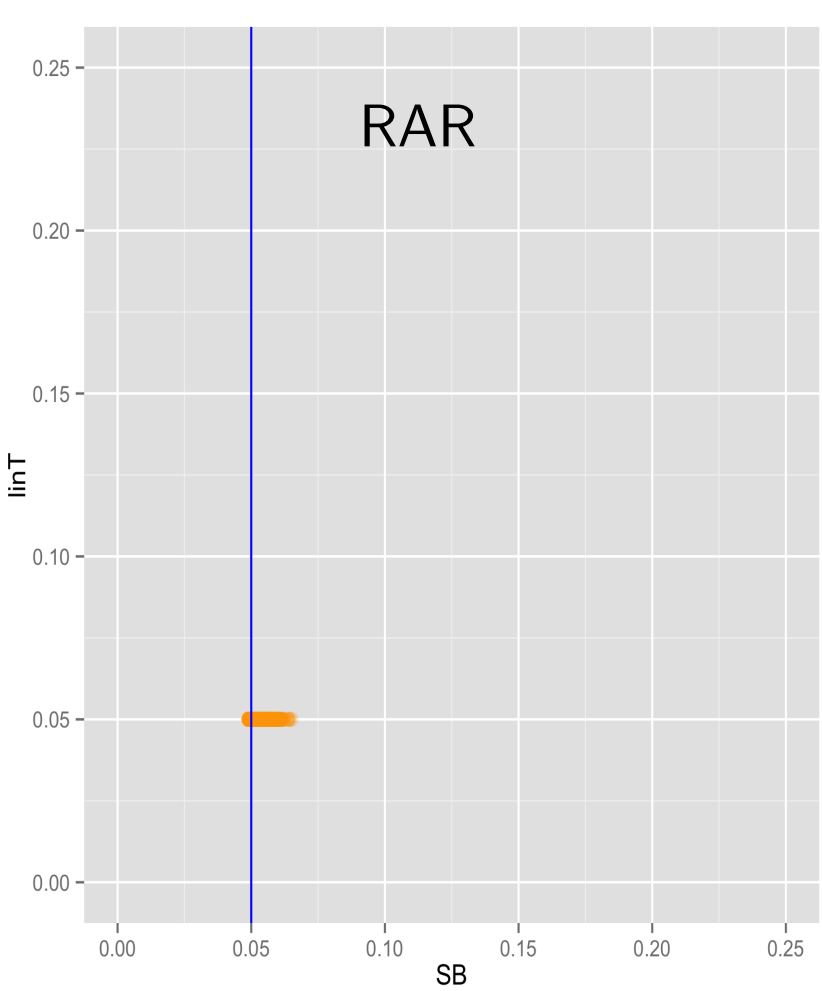


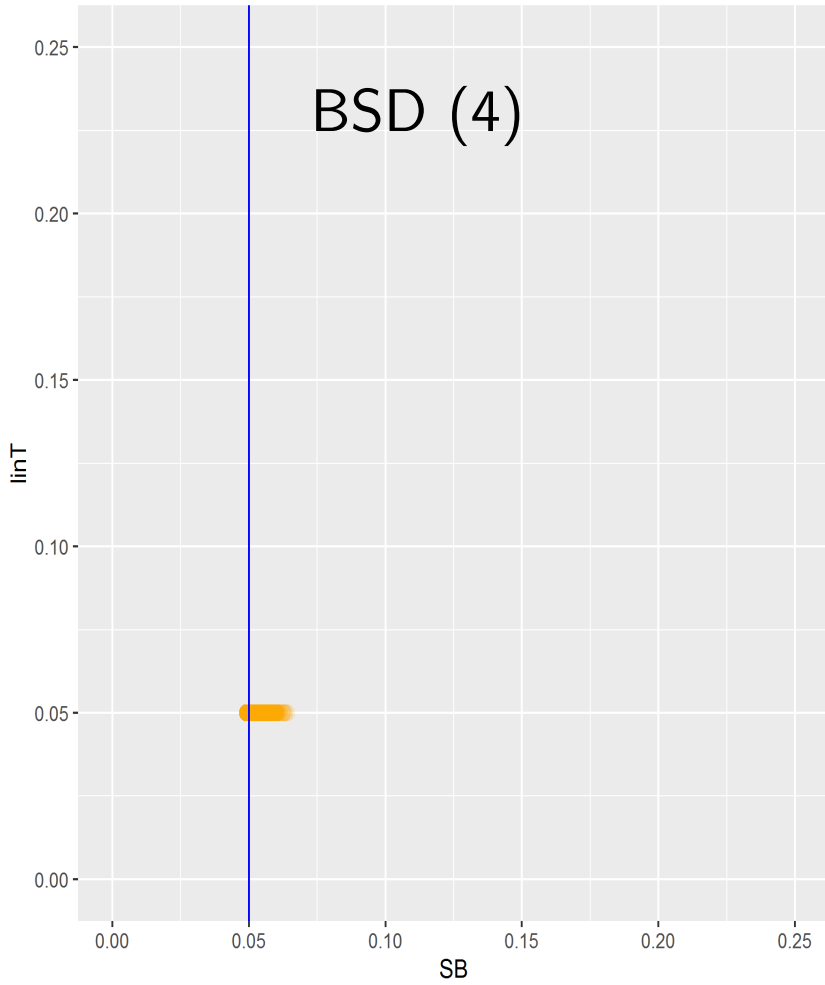


FP7 HEALTH 2013 - 602552

6. CSE - Result: Selection Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0.2 \times \delta;\ \theta = 0$

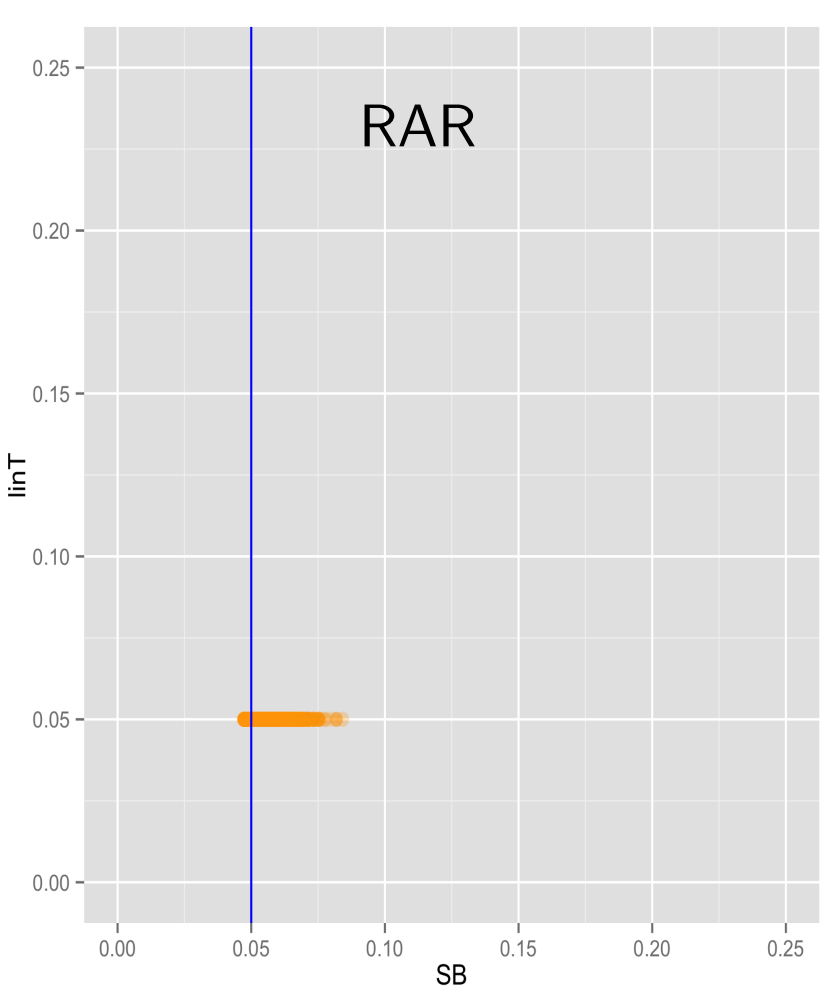


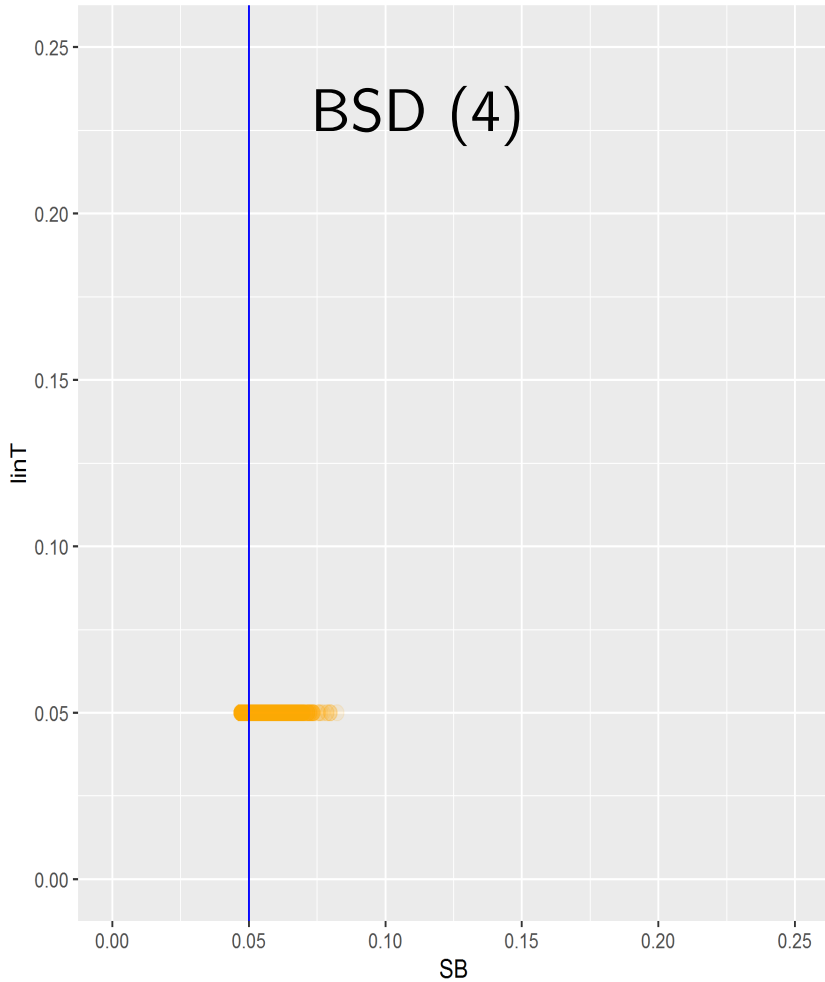


FP7 HEALTH 2013 - 602552

6. CSE - Result: Selection Bias (N=96)

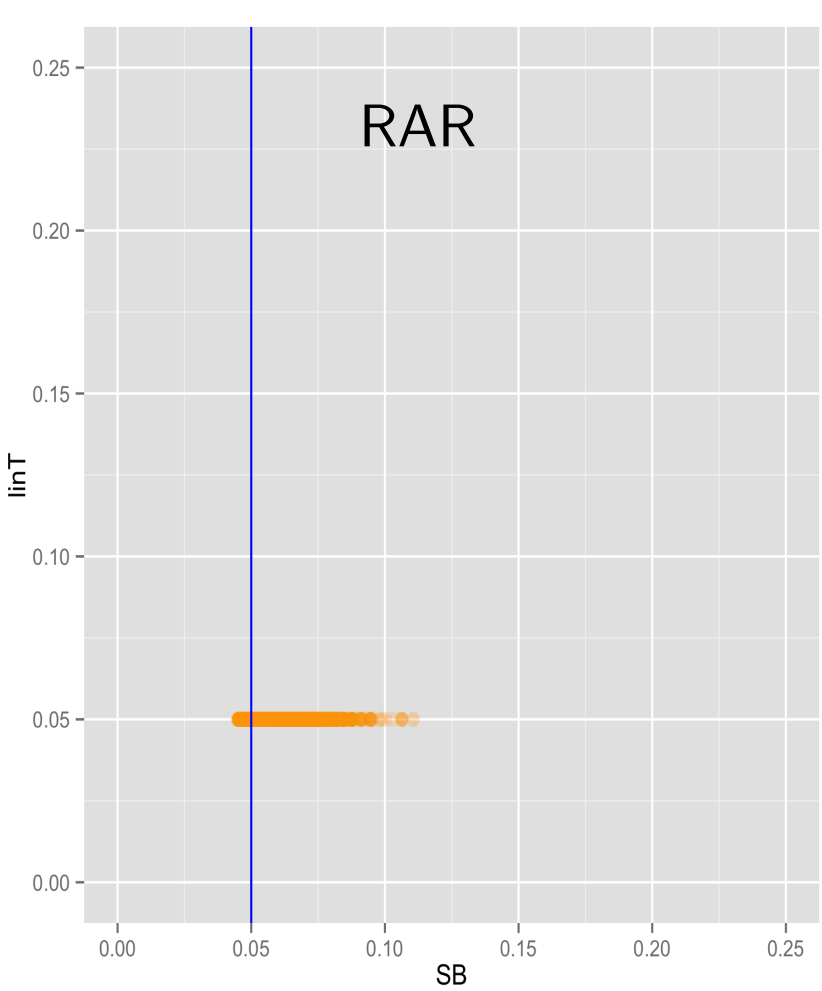
setting: $N_E = N_C = 48, \eta = 0.3 \times \delta;\ \theta = 0$

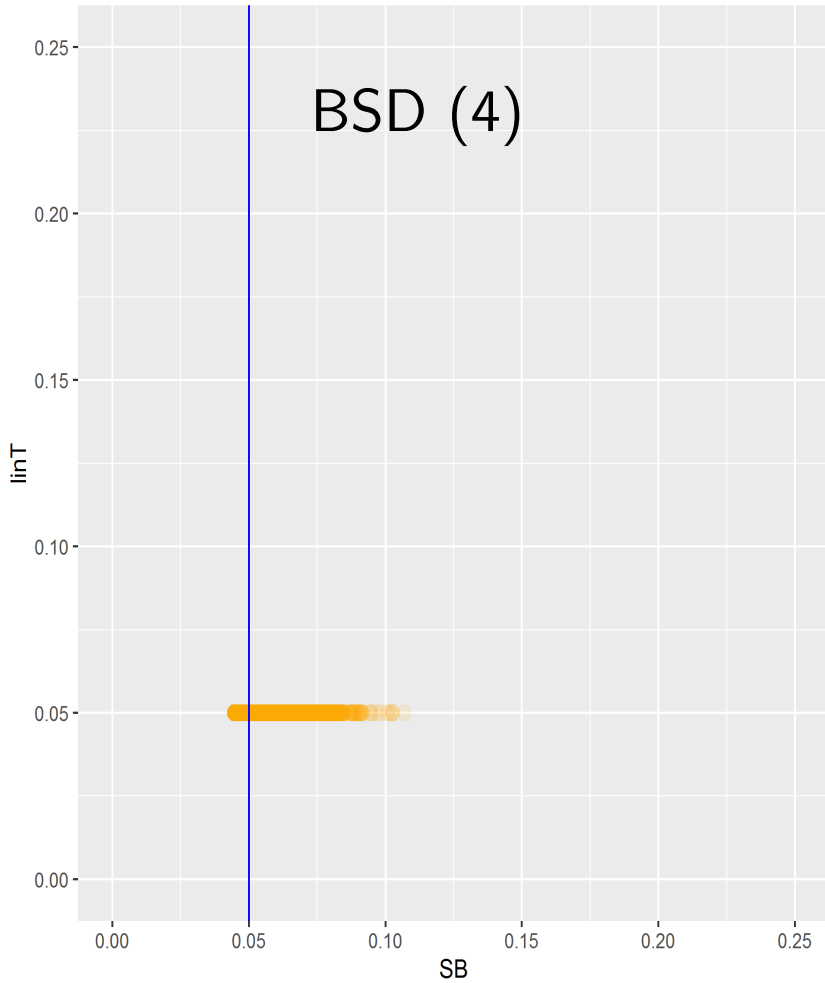




6. CSE - Result: Selection Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0.4 \times \delta; \theta = 0.8$

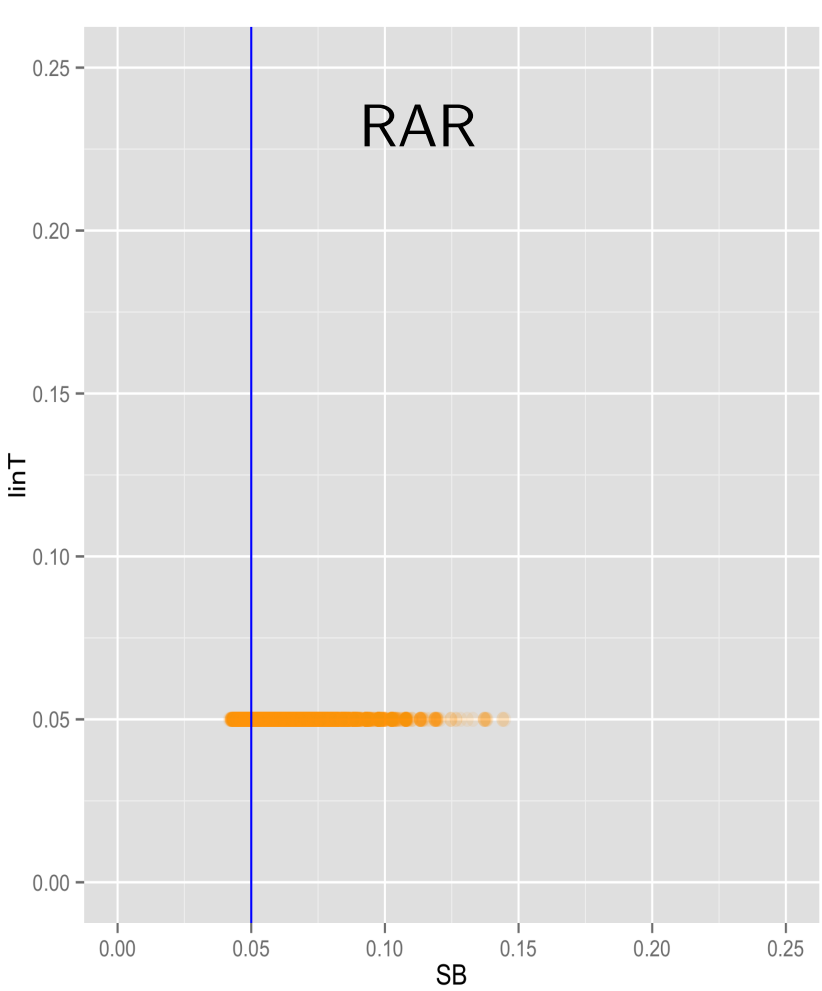


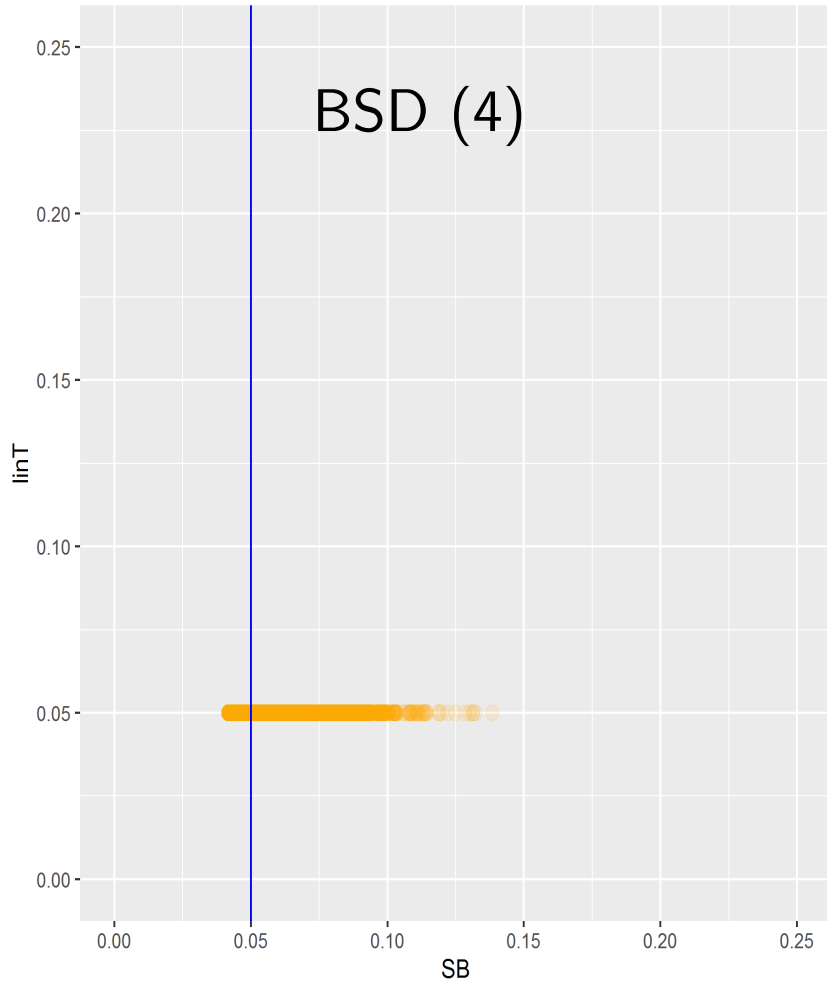


FP7 HEALTH 2013 - 602552

6. CSE - Result: Selection Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0.5 \times \delta;\ \theta = 0$





FP7 HEALTH 2013 - 602552

4. CSE - Evaluation: Time Trend Bias Model

two arm parallel group trial continuous endpoint

Biasing policy according to convergence strategy

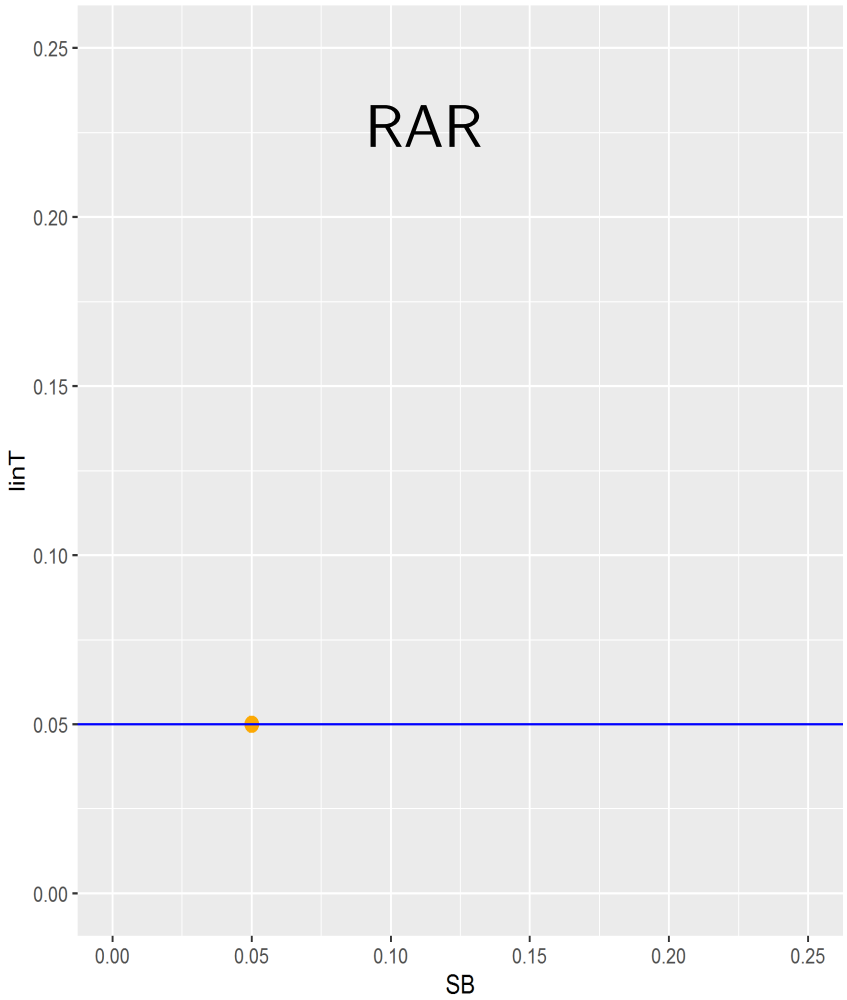
$$\tau_i = \theta \times \begin{cases} \frac{i}{N_E+N_C} & \text{linear time trend} \\ \mathbb{1}_{i \geq S}(i) & \text{stepwise trend} \\ \log(\frac{i}{N_E+N_C}) & \text{log trend} \end{cases}$$

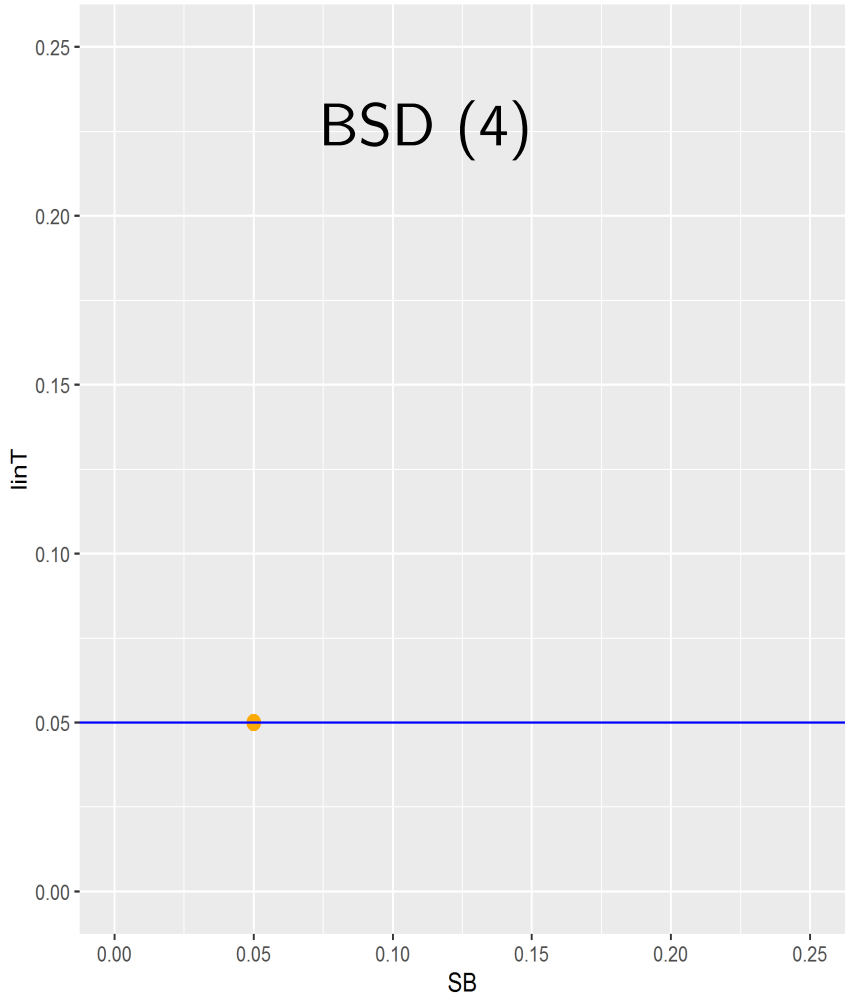
- θ proportional to variance
- other functions are possible
- long recruitment time in rare diseases, (EMA, 2006)
  - changes in population characteristics
  - learning effect in therapy / surgical experience (Hopper, 2007)
  - change in diagnosis (FDA, 2011), etc.
- special form of accidental bias,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

FP7 HEALTH 2013 - 602552

6. CSE - Result: Linear Time Trend Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0; \theta = 0.0 \times \sigma$

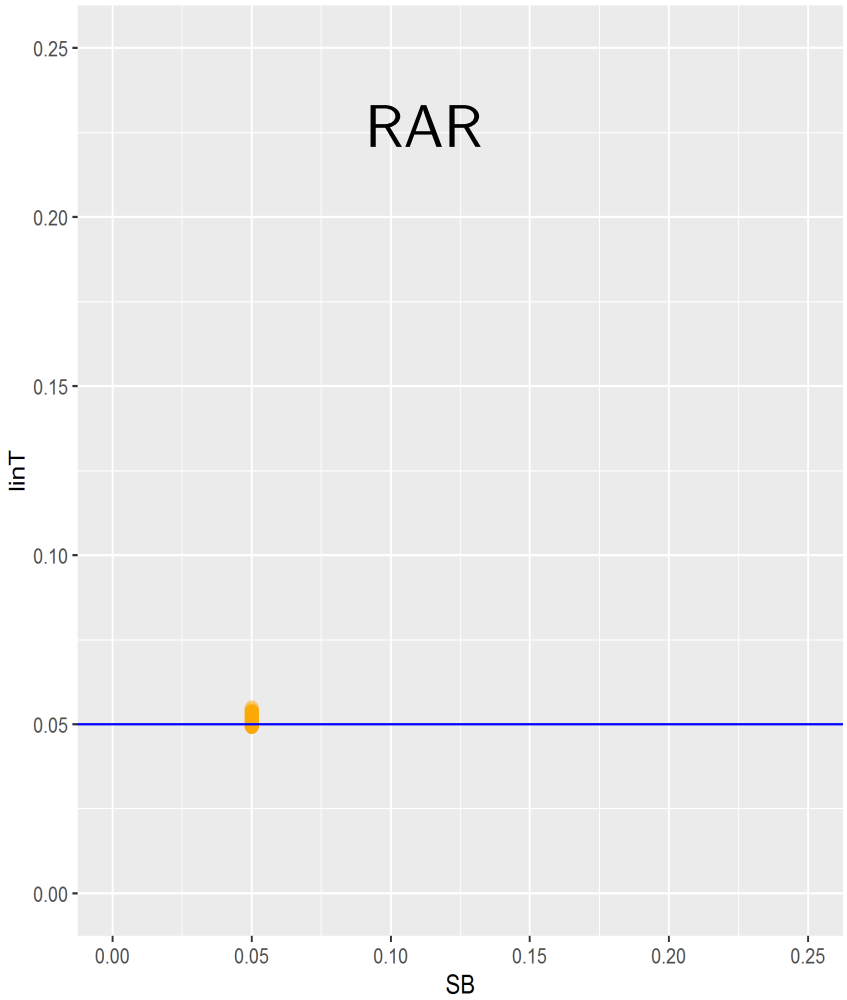


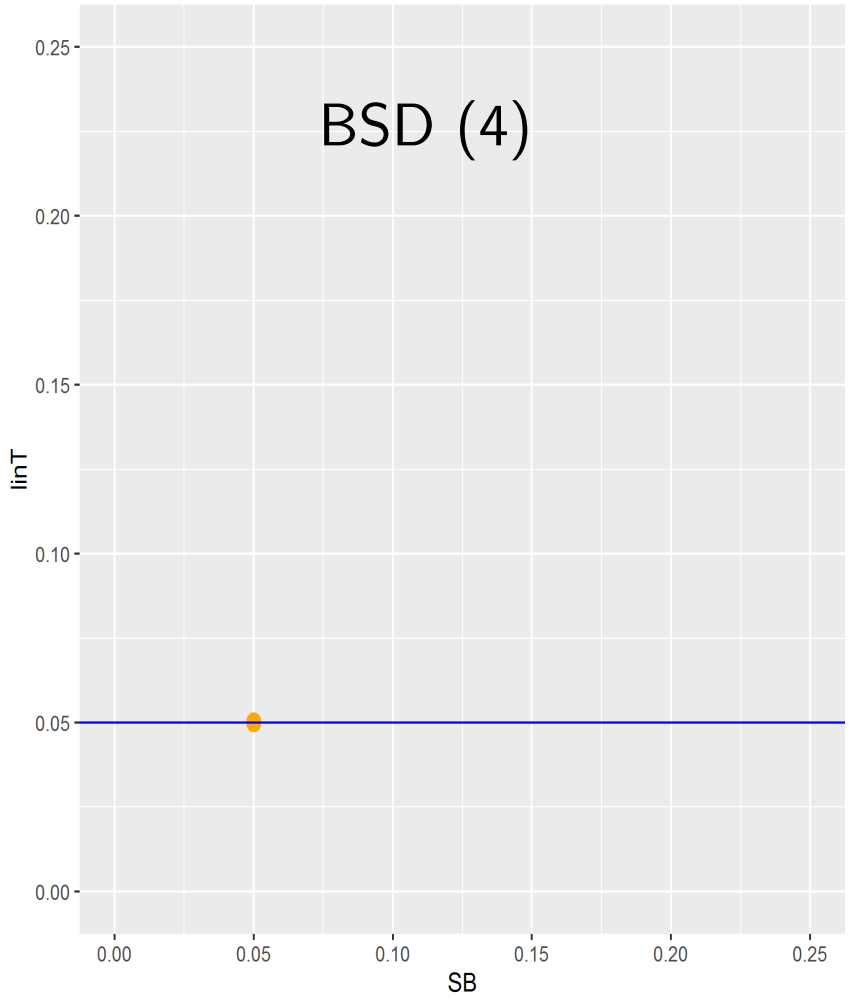


FP7 HEALTH 2013 - 602552

6. CSE - Result: Linear Time Trend Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0; \theta = 0.2 \times \sigma$

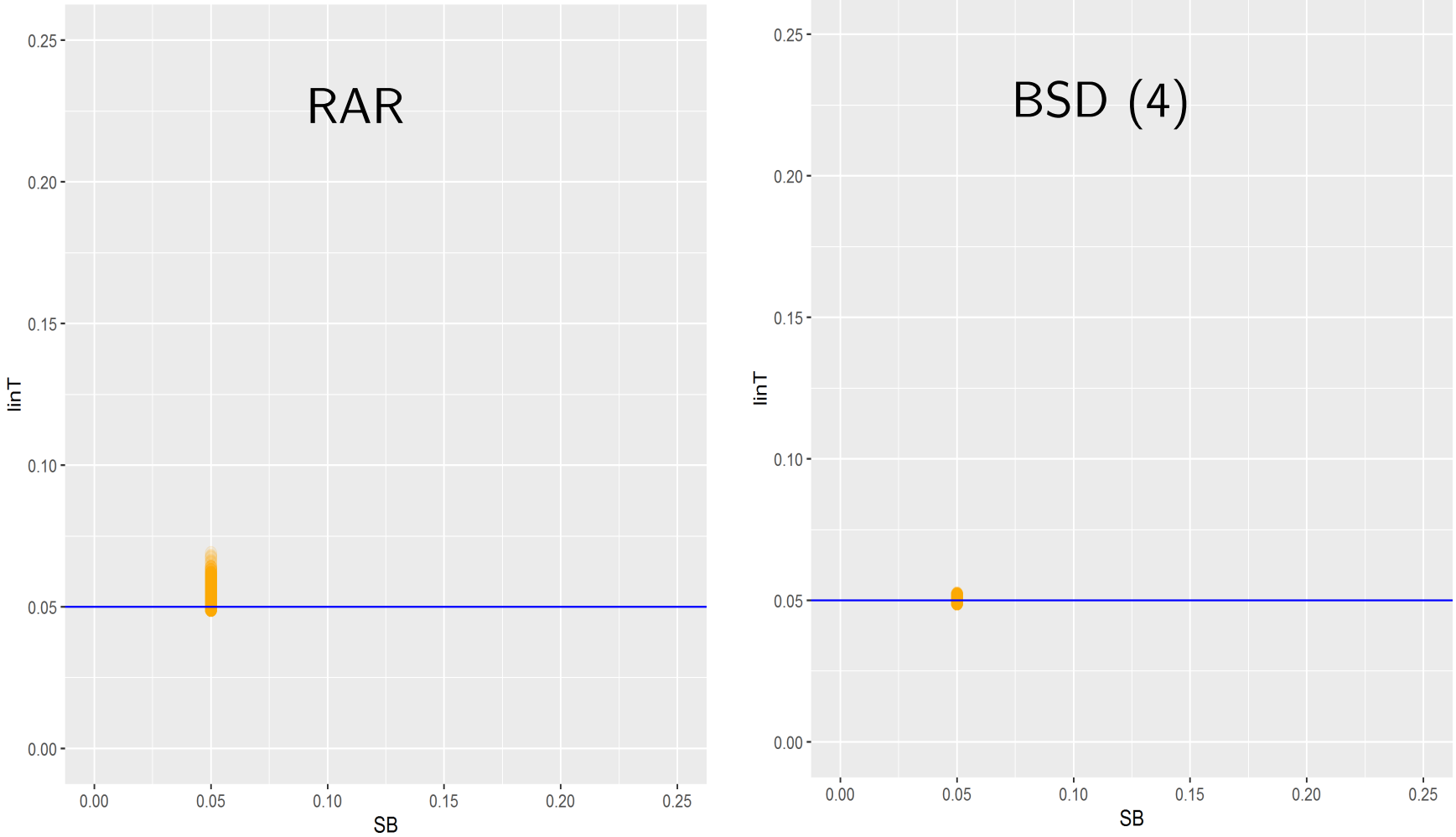




FP7 HEALTH 2013 - 602552

6. CSE - Result: Linear Time Trend Bias (N=96)

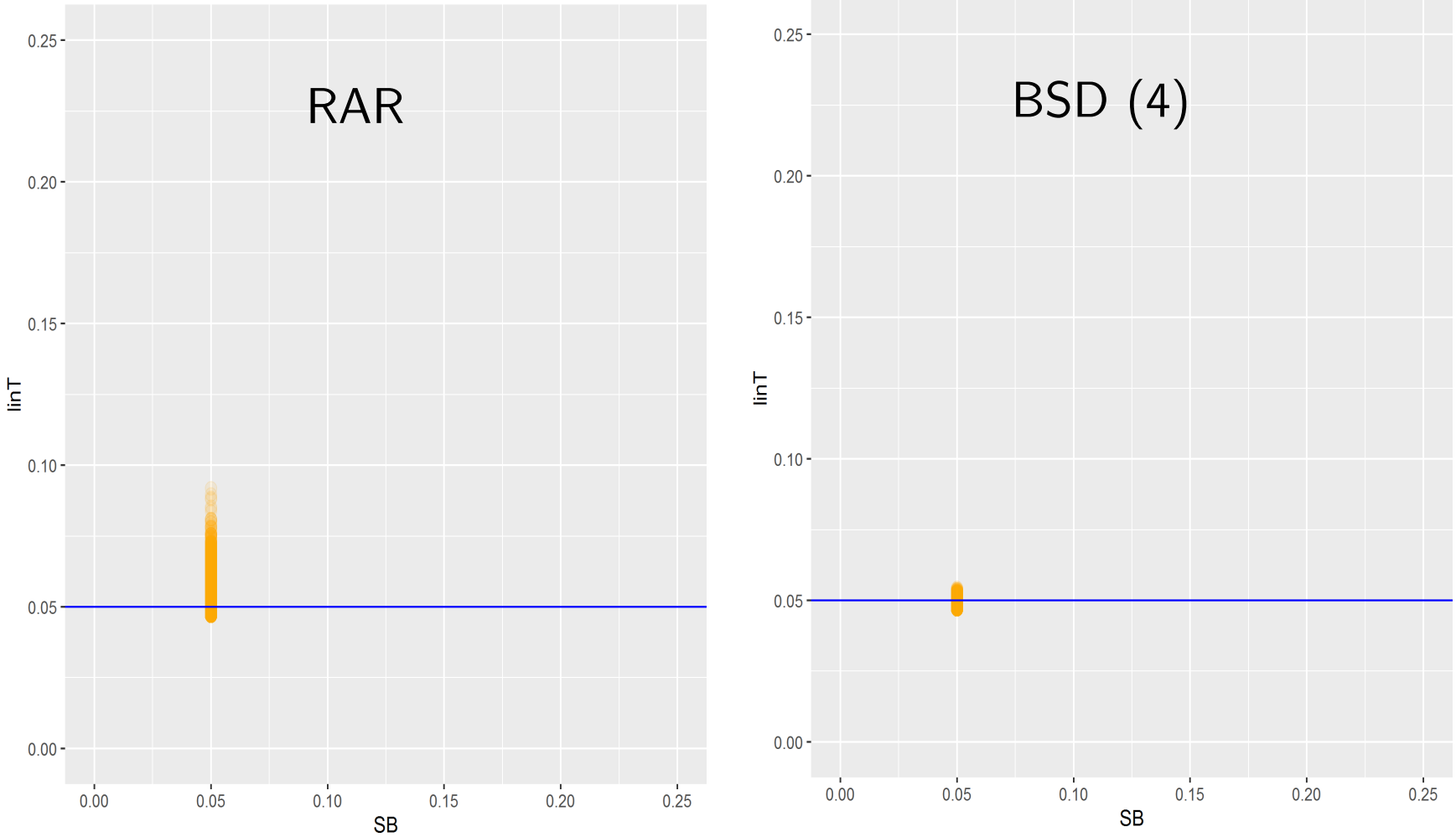
setting: $N_E = N_C = 48, \eta = 0; \theta = 0.4 \times \sigma$



FP7 HEALTH 2013 - 602552

6. CSE - Result: Linear Time Trend Bias (N=96)

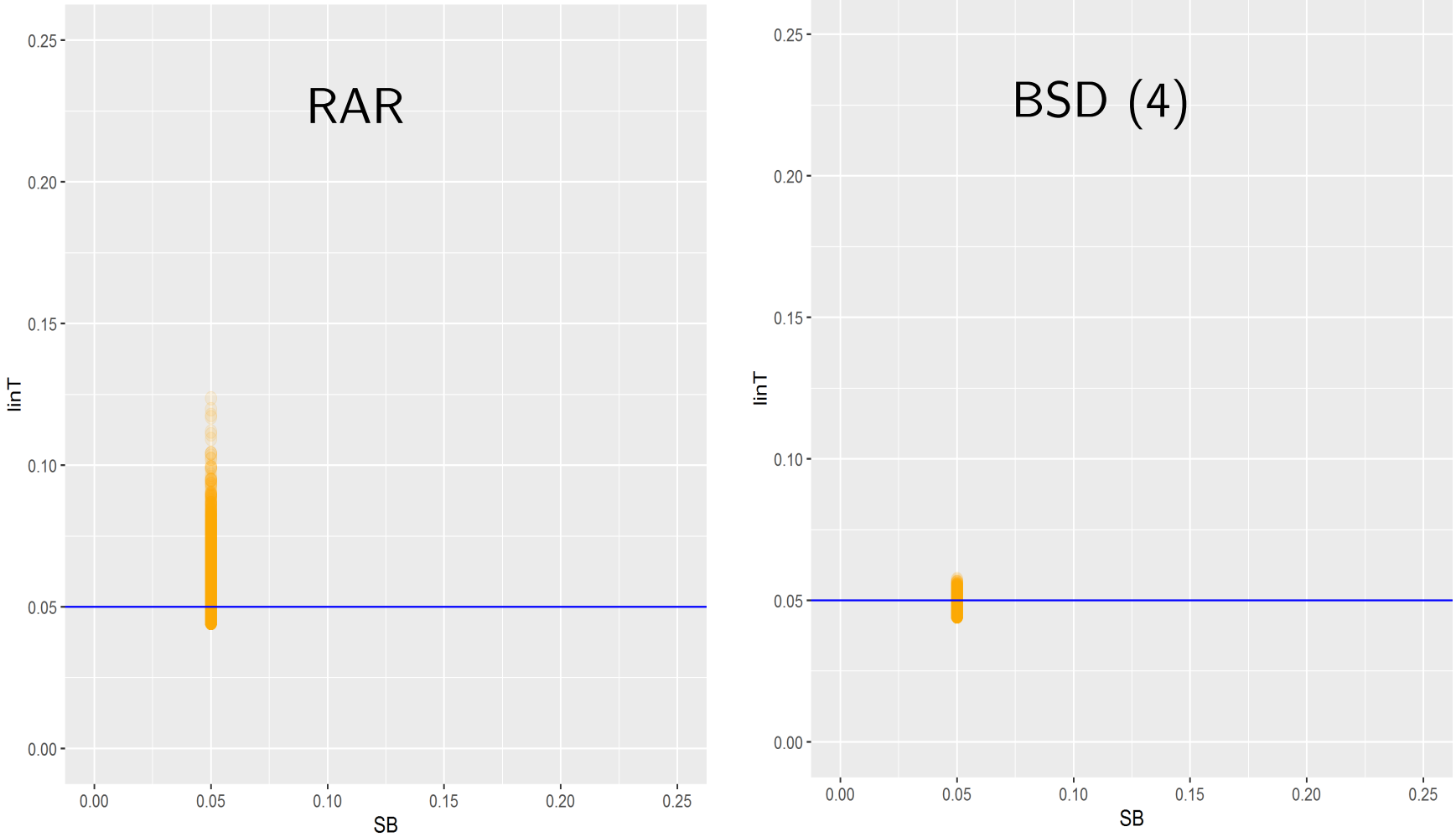
setting: $N_E = N_C = 48, \eta = 0; \theta = 0.6 \times \sigma$



FP7 HEALTH 2013 - 602552

6. CSE - Result: Linear Time Trend Bias (N=96)

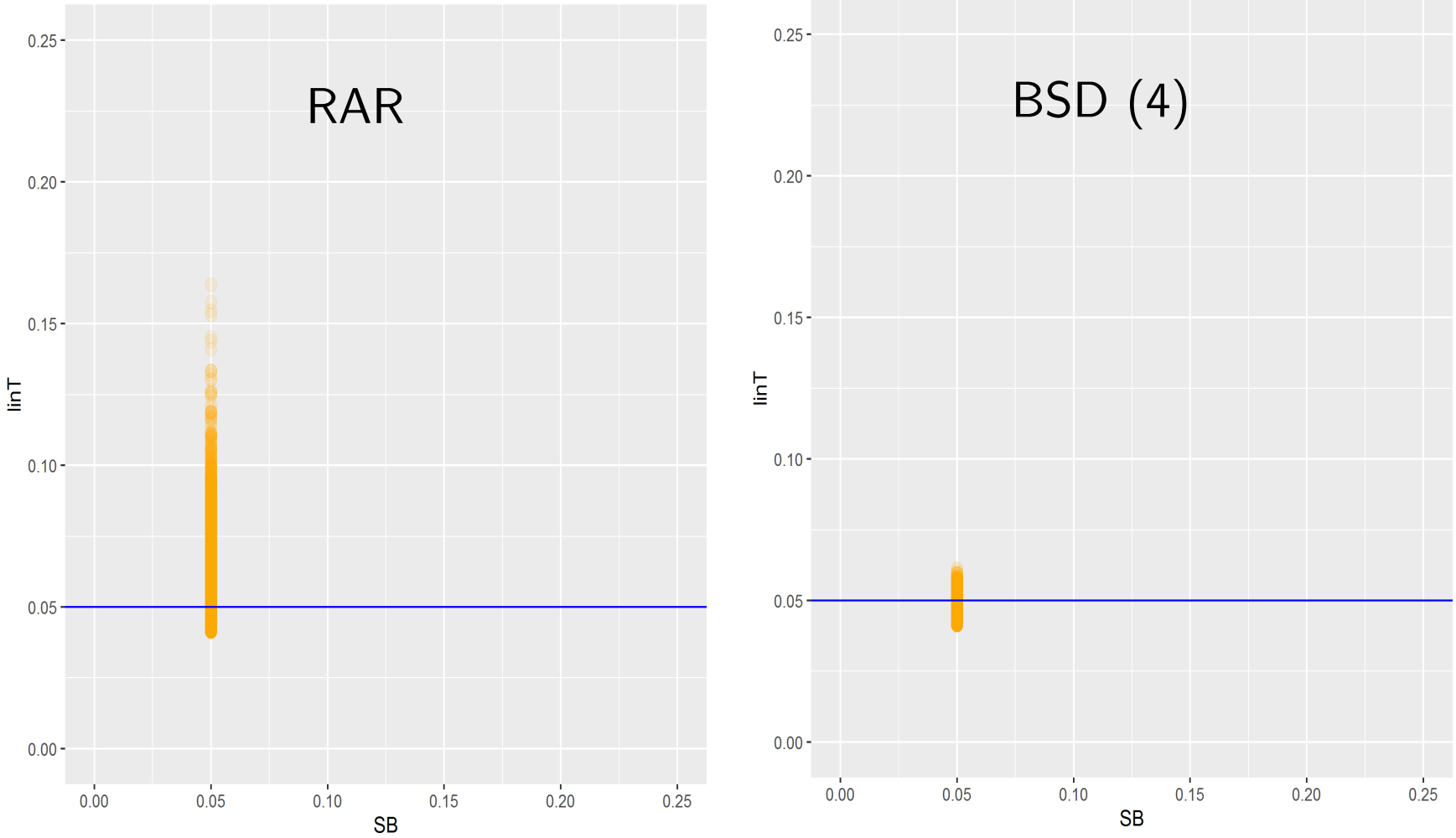
setting: $N_E = N_C = 48, \eta = 0; \theta = 0.8 \times \sigma$



FP7 HEALTH 2013 - 602552

6. CSE - Result: Linear Time Trend Bias (N=96)

setting: $N_E = N_C = 48, \eta = 0; \theta = 1.0 \times \sigma$



FP7 HEALTH 2013 - 602552

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

two arm parallel group trial continuous endpoint

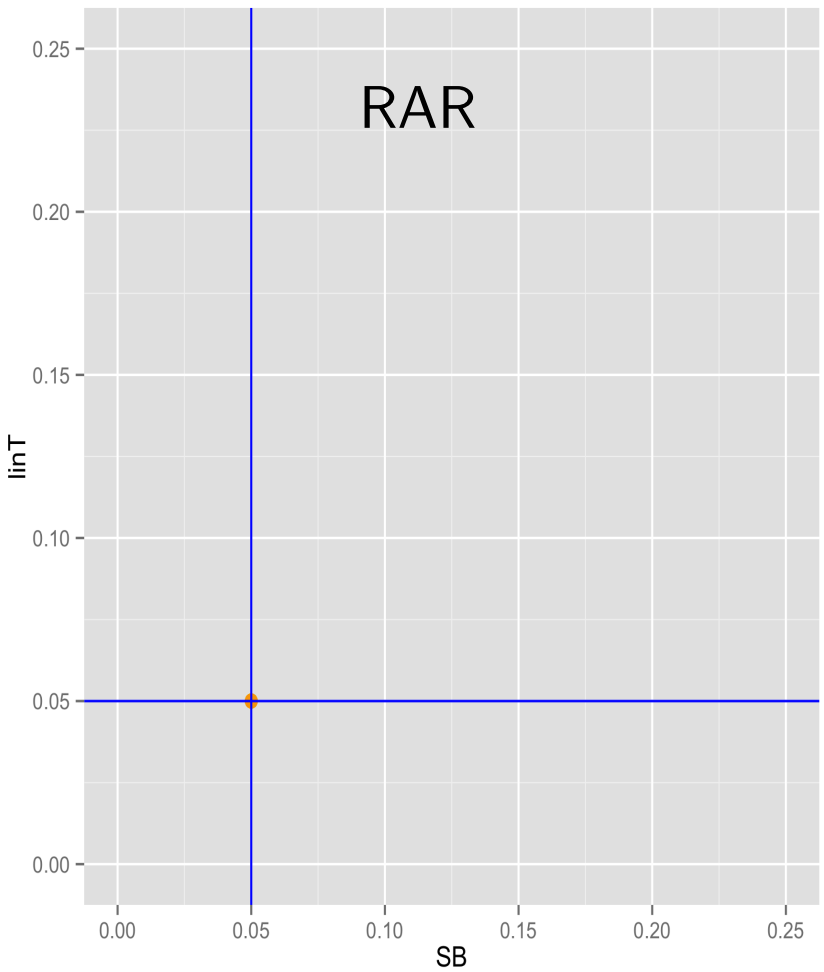
Joint Additive Bias

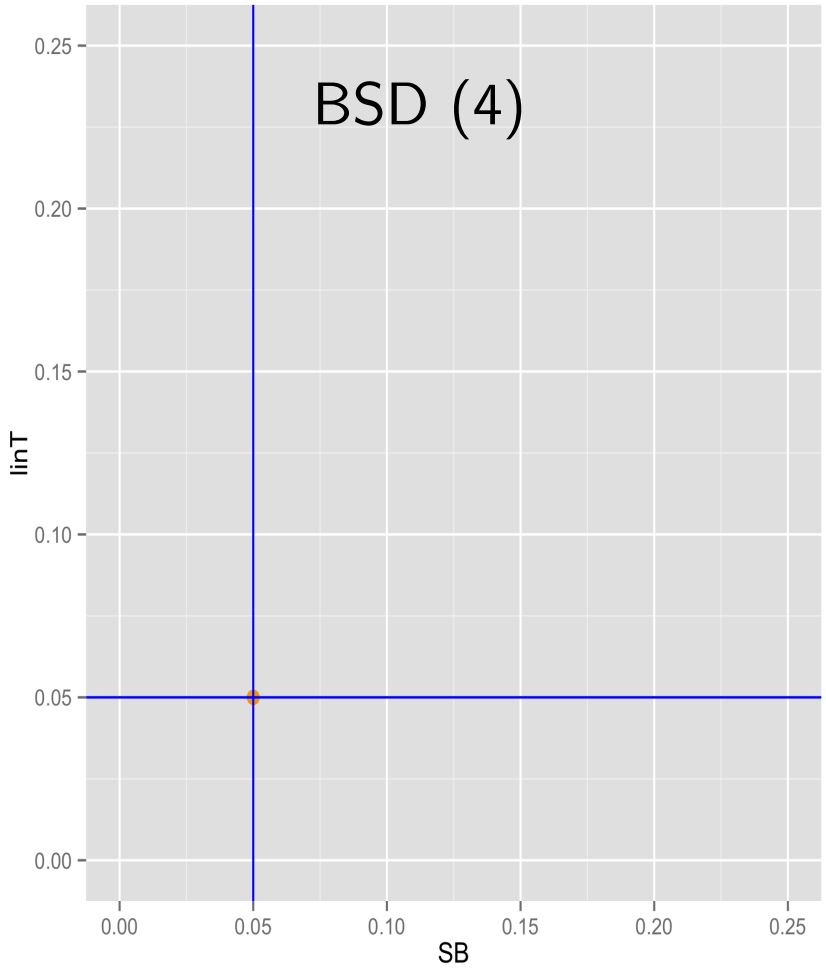
$$\tau_i = \underbrace{\theta \frac{i}{N_E + N_C}}_{\text{time trend}} + \underbrace{\eta \ [\text{sign}(\ n_E(i-1) - n_C(i-1)\ )]}_{\text{selection bias}}$$

- 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, 2014)
- relaxed version of bias policy (non strict decision, random η)

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

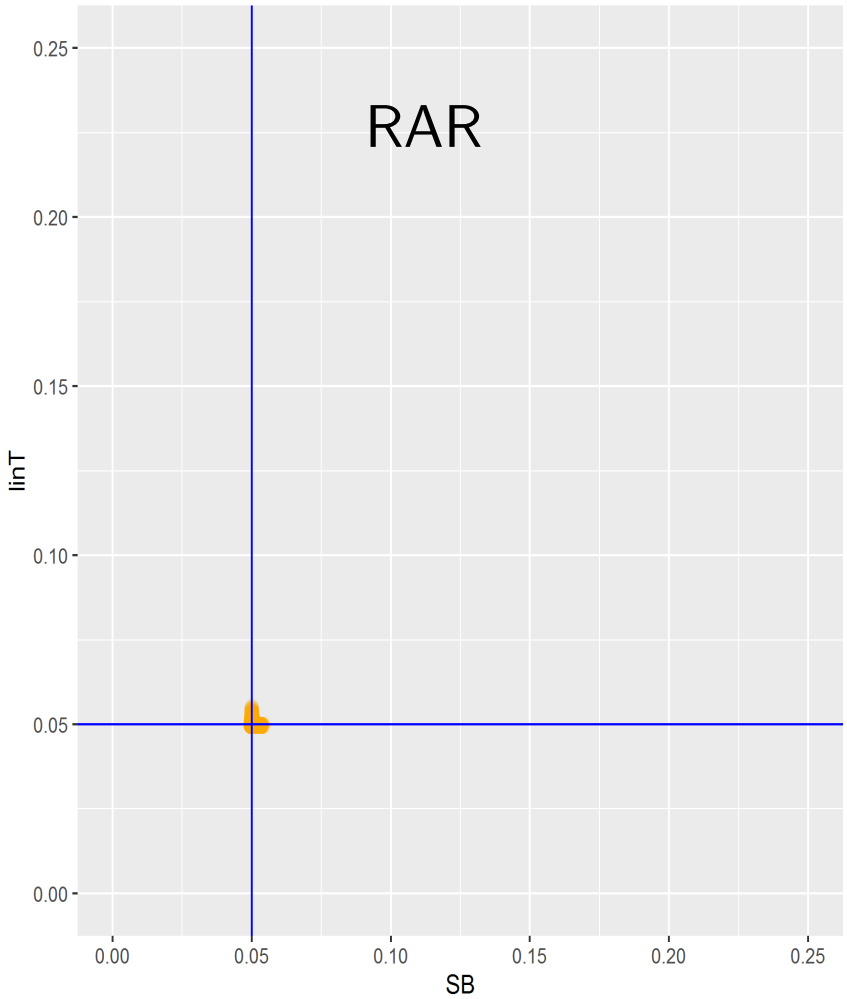
setting: $N_E = N_C = 48, \eta = 0.0 \times$effectsize (δ), $\theta = 0.0 \times \sigma$

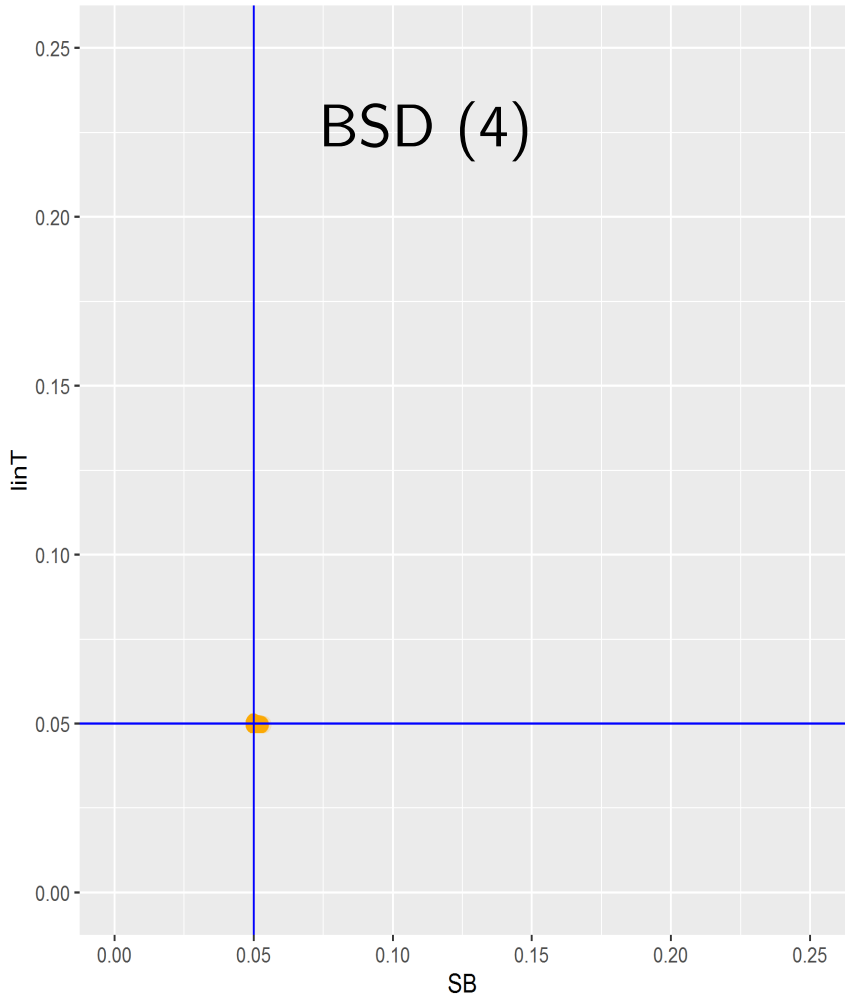




6. CSE - Result: both Biases for (N=96)

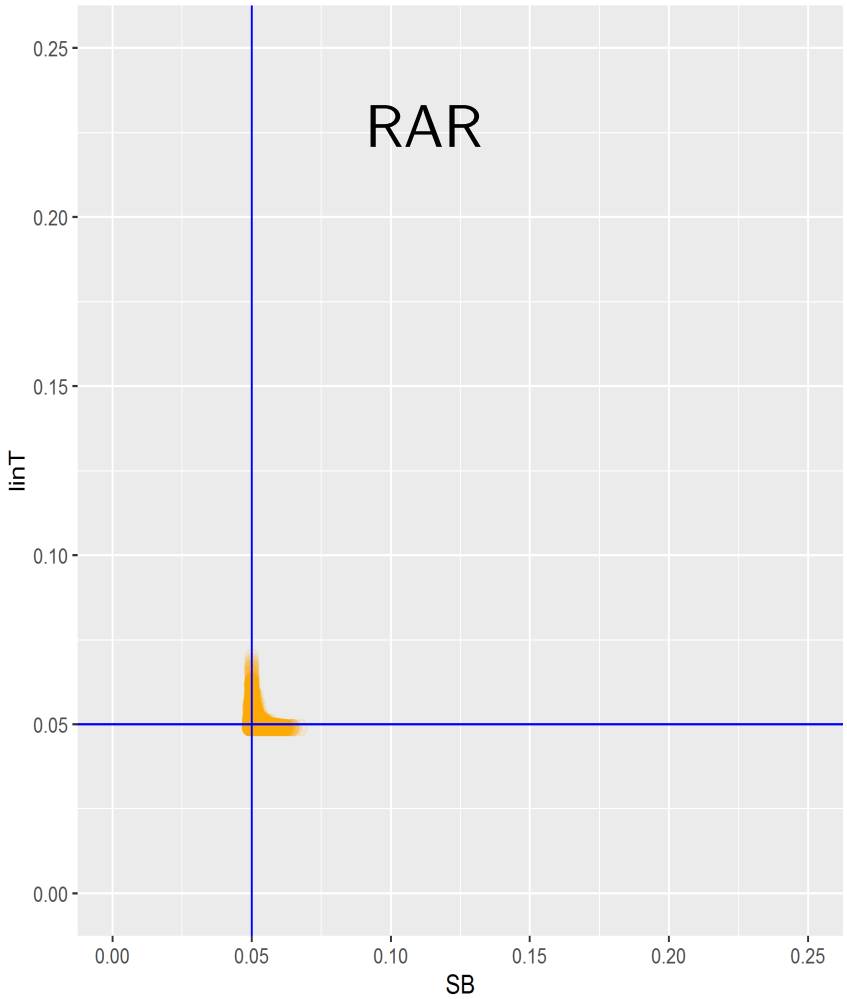
setting: $N_E = N_C = 48, \eta = 0.1 \times \delta, \theta = 0.2 \times \sigma$

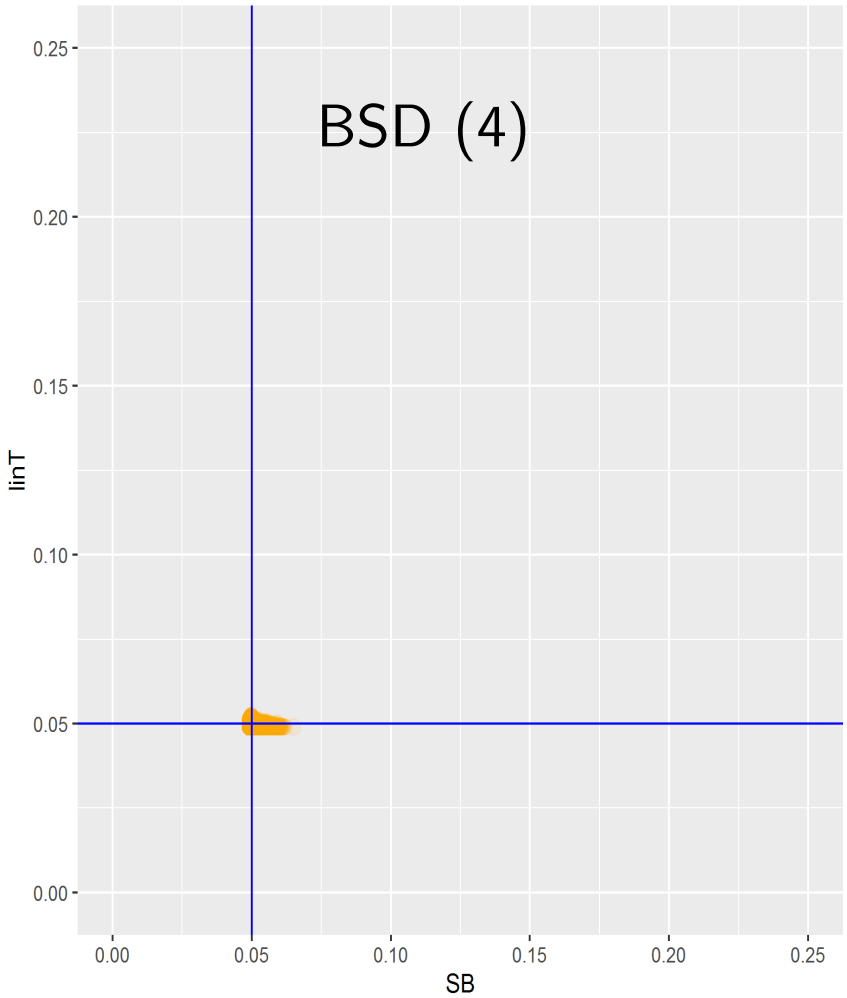




6. CSE - Result: both Biases for (N=96)

setting: $N_E = N_C = 48, \eta = 0.2 \times \delta, \theta = 0.4 \times \sigma$

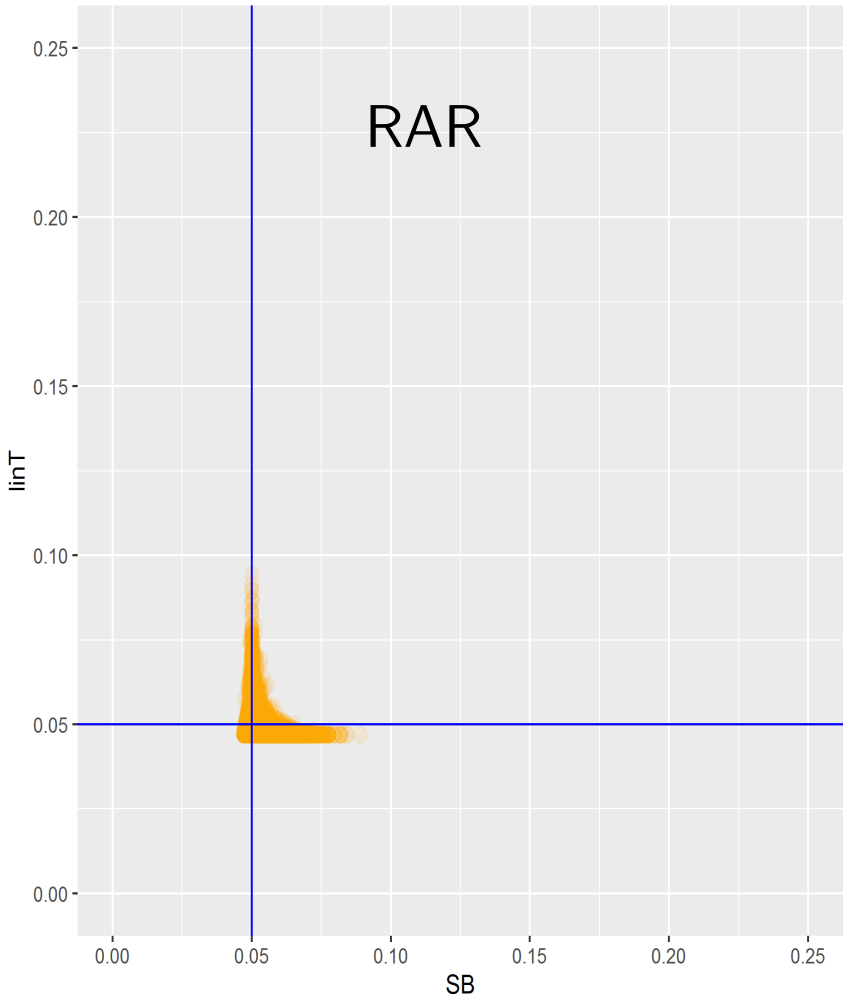


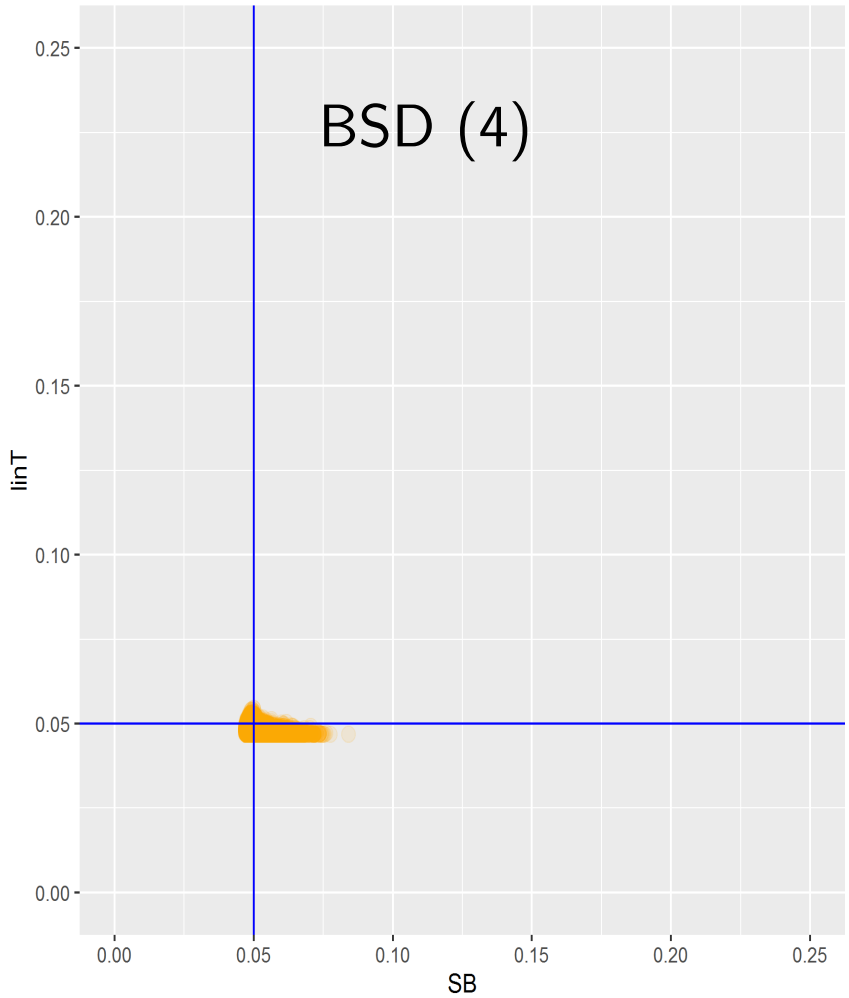


FP7 HEALTH 2013 - 602552

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

setting: $N_E = N_C = 48, \eta = 0.3 \times \delta, \theta = 0.6 \times \sigma$

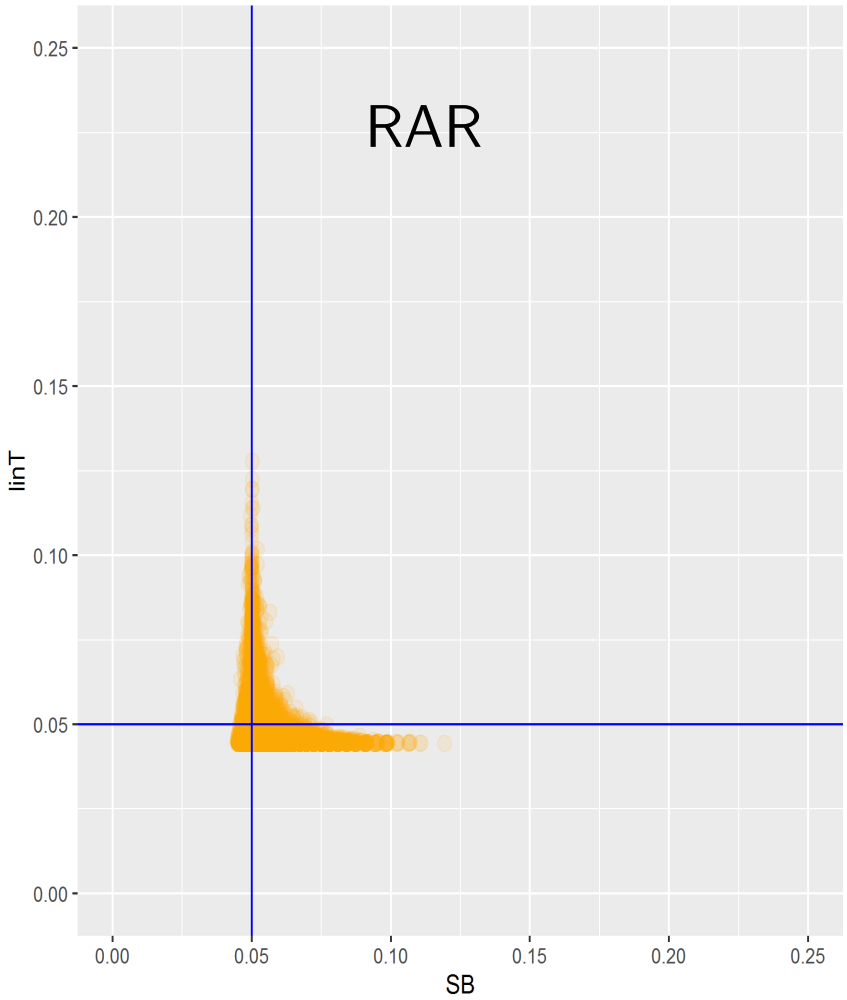


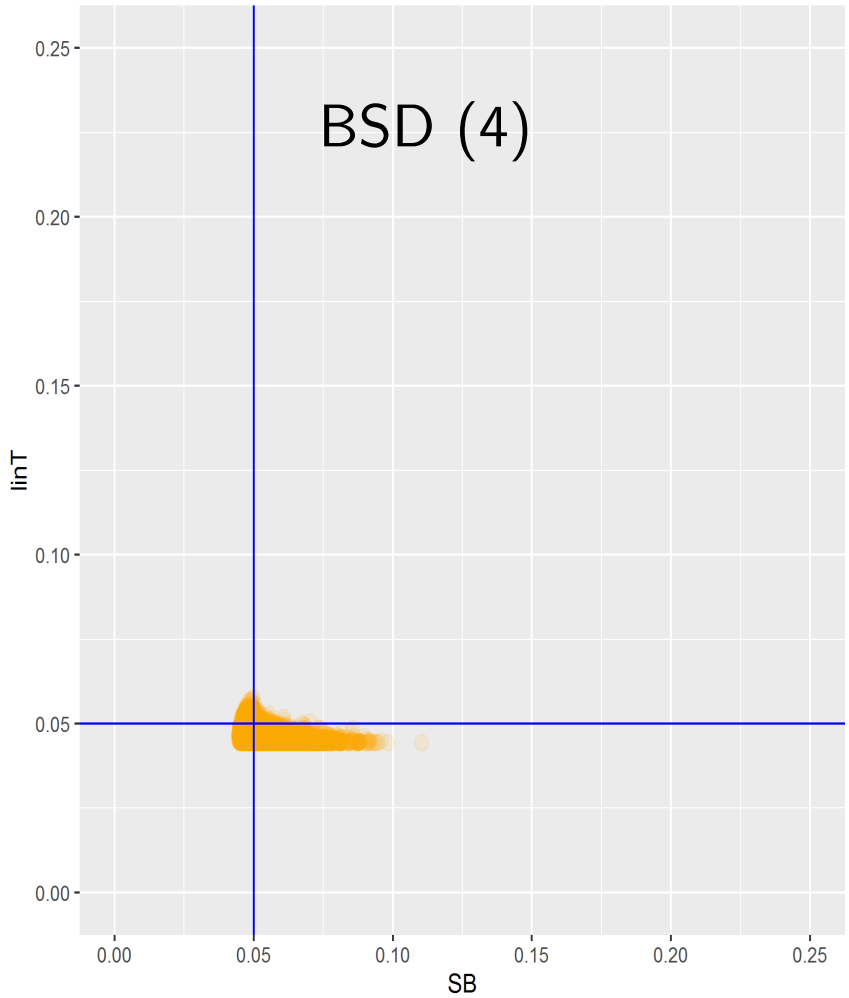


FP7 HEALTH 2013 - 602552

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

setting: $N_E = N_C = 48, \eta = 0.4 \times \delta, \theta = 0.8 \times \sigma$

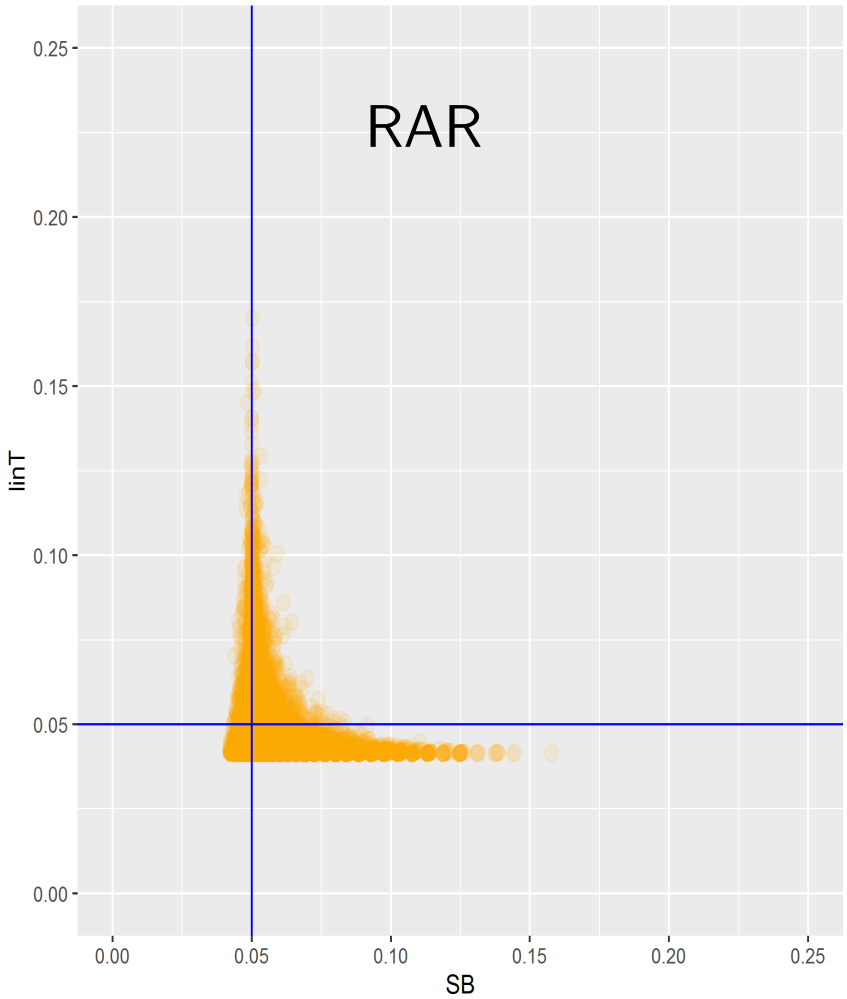


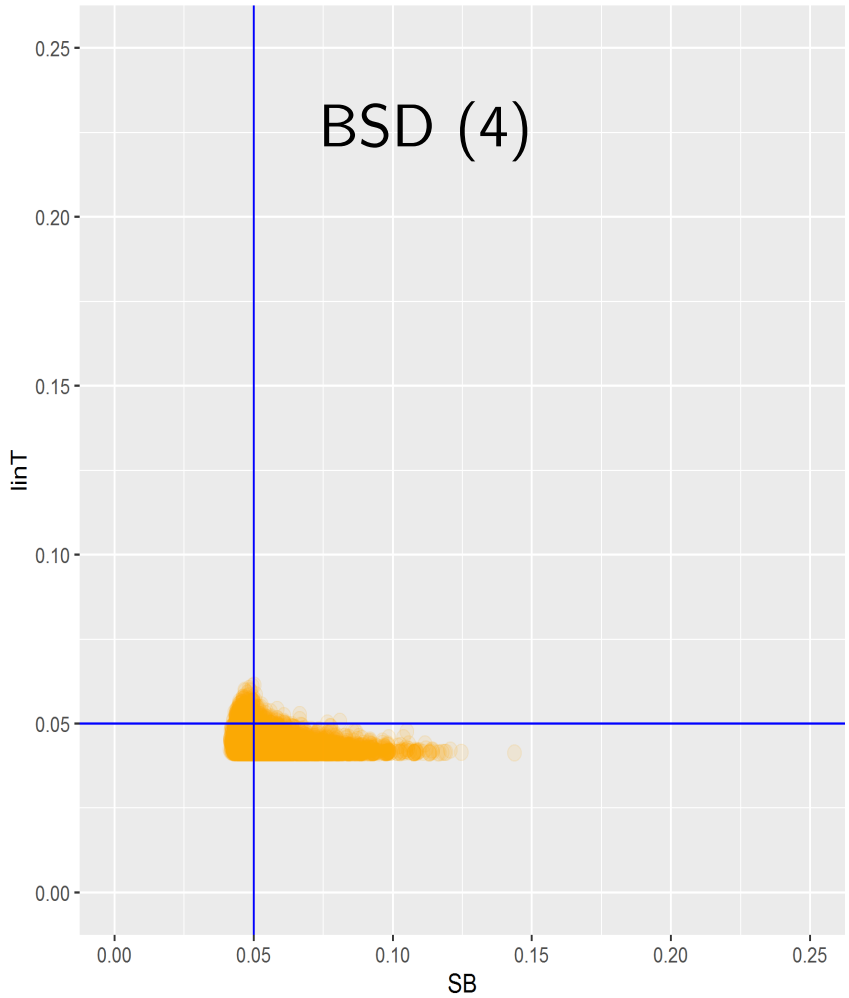


FP7 HEALTH 2013 - 602552

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

setting: $N_E = N_C = 48, \eta = 0.5 \times \delta, \theta = 1.0 \times \sigma$





FP7 HEALTH 2013 - 602552

5. CSE - Software: randomizeR

. . . will use randomizeR, to conduct the evaluation 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 *(Uschner, 2016)*
- development of a shiny app

7.2 CSE - Practical Implications

- among other it is shown, that none of the randomization procedures performs uniformly best.
- ignoring the influence of selection bias may affect the test decision, by means of type-I-error rate probability
- the effect may be, that conservative or anticonservative test decisions occure
- practical settings may affect the choice of a randomization procedure, e.g. the choice the magnitude of η and θ have to be discussed within the practical context
- at least a minimum effect of bias (related to the clinical important effect size) should be assumed
- discussion of theses topics may help to understand the selection a randomization procedure within the particular/practical study settings

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 may 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 Team in Aachen

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

FP7 HEALTH 2013 - 602552

References

- Kennes, L. N., Cramer, E., Hilgers, R. E., and Heussen, N. (2011). The impact of selection bias on test decisions in randomized clinical trials *Statistics in Medicine* 2011; **30**:2573-2581.
- Kennes, L. N. (2012). The impact of selection bias on test decisions in randomized clinical trials *Master Thesis Mathematics RWTH Aachen*
- Kennes, L. N., Rosenberger William F., Hilgers, R.-D., (2015). Inference for blocked randomization under a selection bias model *Biometrics* 2015; **71**:y 979?984. doi.org/10.1111/biom.12334.
- Langer S. The modified distribution of the t-test statistic under the influence of selection bias based on random allocation rule *Master Thesis, RWTH Aachen University, Germany*, 2014
- Rückbeil M. The impact of selection bias on test decisions in survival analysis *Master Thesis, RWTH Aachen University, Germany*, 2015
- Tasche A. Selection Bias bei mehr als zwei Behandlungsgruppen *Studienarbeit, RWTH Aachen University, Germany*, 2016
- Tamm M, Cramer E, Kennes LN, Heussen N Influence of Selection Bias on the Test Decision - A Simulation Study *Methods of Information in Medicine* 2012; **51**:138-143. DOI: 10.3414/ME11-01-0043.
- Tamm M, Hilgers RD. Chronological Bias in Randomized Clinical Trials Arising from Different Types of Unobserved Time Trends *Methods of Information in Medicine* 2014; **53**:501-510. DOI: 10.3414/ME14-01-0048.

FP7 HEALTH 2013 - 602552