The selection of a randomization procedure to avoid the impact of bias on the test result in clinical trials? A case study!

Ralf-Dieter Hilgers

Department of Medical Statistics, RWTH Aachen University

WBC, 2016, November 9th

This research is part of WP 1 of the IDeAl project funded from the European Union Seventh Framework Programme [FP7 2007-2013] under grant agreement No. 60255.

Study Design in Practice

- What the theory tells us:
  - no randomization procedure performs best with all criteria, Rosenberger (2016), Atkinson (2014)
- What applied scientist mostly feel:
  - scepticism to randomization
  - do not well understood randomization principle
  - is just allocation and think unequal group size is a major problem
  - think that randomization is for balancing covariates but does mostly not work
  - select a procedure by opinion or software availability
- What the literature mirrors:
  - no training in randomization
  - no recommendation to give scientific arguments for the choice of randomization procedure, neither ICH Guidelines nor CONSORT Statement

FP7 HEALTH 2013 - 602552

ERDO

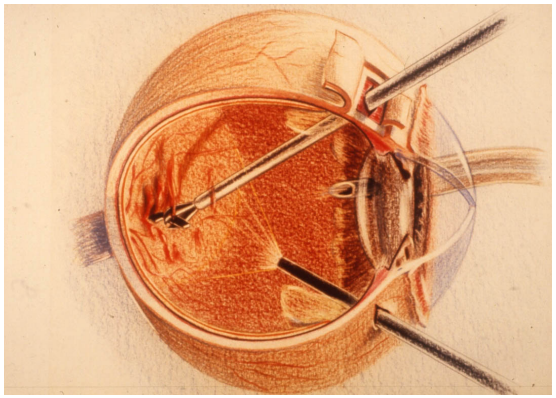
Evaluation of Randomization Procedures for Trial Design Optimization

1. **Introduction** - intend select the best practice randomization procedure (RP) to improve the level of evidence
2. **Objective** - select a best practice RP
3. **ERDO framework**
   - **Assumptions** - incl. design, clinical setting
   - **Options** - suitable set of RP's
   - **Metrics** - evaluation criterion e.g. averaged (empirical) type I error rate
4. **Evaluation Methods** - incl. statistical model, software, presentation of results, decision rule
5. **Result and Decison**
6. **Discussion and Clinical implication** - select the best practice (RP)
7. **Conclusion** choice of randomization design

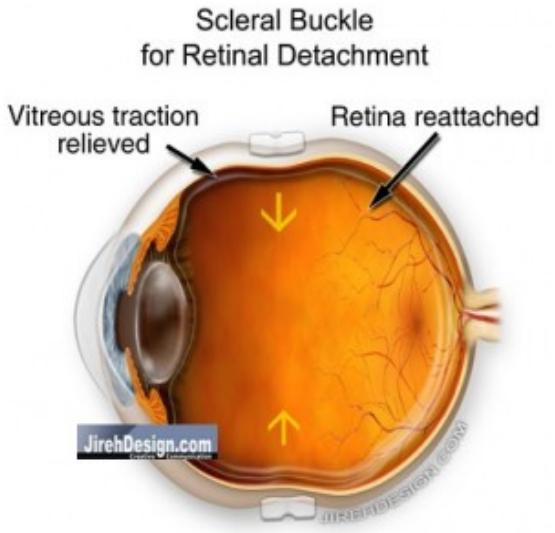
FP7 HEALTH 2013 - 602552

1. Introduction - Case Study

- scleral buckling (SB) with primary pars plana vitrectomy (PPV) in rhegmatogenous retinal detachment (SPR-Study, Heimann 2007)



- Additional encircling band might improve one year best corrected visual acuity results in the scleral buckling group.



http://www.retinaeyedoctor.com/tag/eye/

2. Objective - Case Study

- to design a new clinical trial (EnBand-Study) with respect to the selection of a (best practice) randomization procedure
- Tasks:
  - investigate the potential impact of selection and chronological bias on the test decision

3. Assumptions Design of the Case Study

Primary Endpoint

- Change in Best corrected Visual Acuity one year after surgery to baseline

Clinical Trial Layout

- parallel group design
- targeted allocation ratio 1:1, with a fixed sample design

Sample Size Calculation from SPR study data

- Sample size: 65 patients per group (VA: 0.52 (SD 0.77) with 0.90 (SD 0.73) without encircling band, t-test, two-sided 5% significance level, power of 80%, pooled standard deviation 0.765)

3. Assumptions Clinical Settings of the Case Study

- in almost all surgical trials learning effects may be present

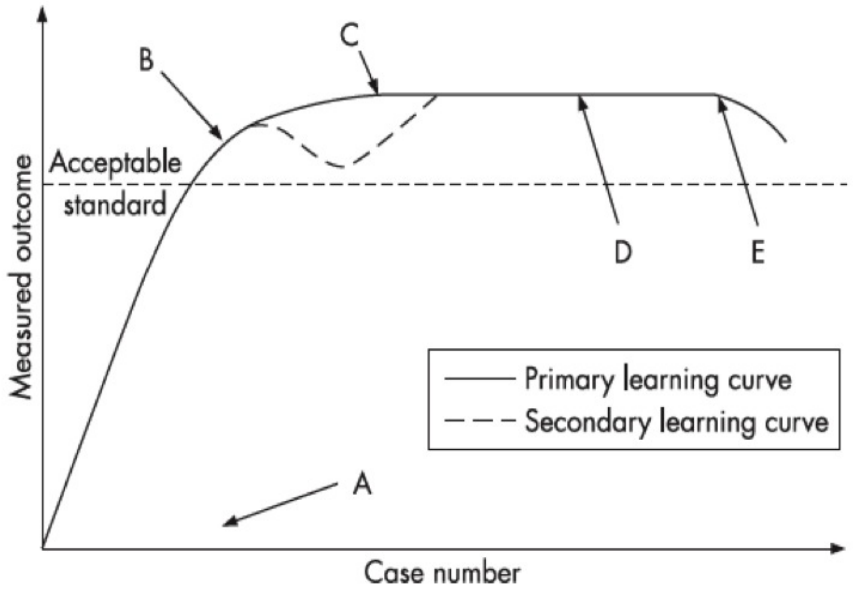


Figure 1 Idealised surgical learning curve.

Hopper AN, Jamison MH, Le WG. Learning curves in surgical practice. Postgrad Med J 2007; **83***: 777-779*

Linear time trend from SPR study data

- linear time trend of $0.14 - 0.26\ i/n$.

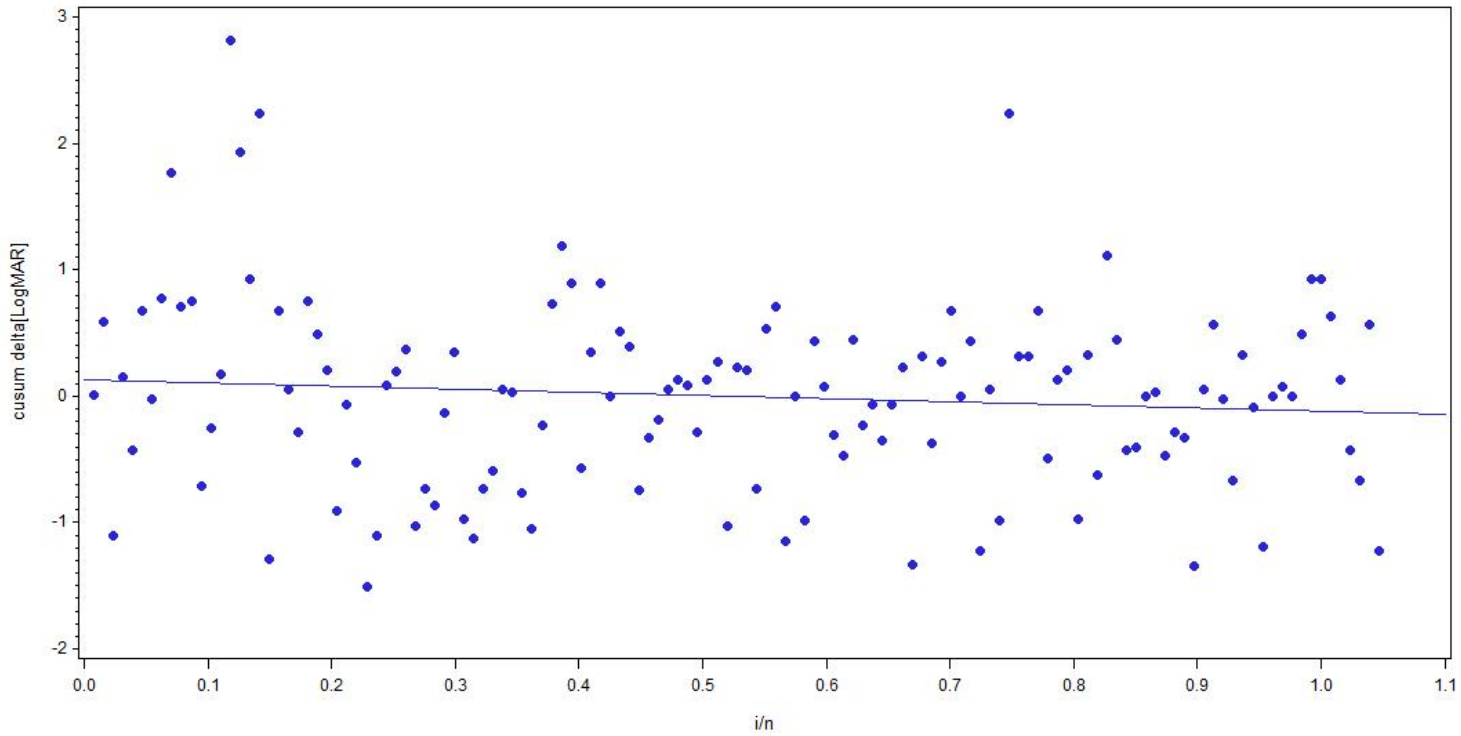


Figure: CUSUM - plot of change in best corrected visual acuity

FP7 HEALTH 2013 - 602552

3. Assumptions - Clinical Settings of the Case Study

Selection effect can be estimated from the data of a past study or may be estimated as portion of the effect size.

Selection effect from SPR study data

- Selection bias effect of a reasonable 30% of the treatment difference is assumed (i.e. 0.114 or standardized in terms of the pooled variance $\eta = 0.15$)

3. Options of the Case Study

Options of ERDO-Various Randomization Procedures

most procedure may be embedded in the "Biased Coin" Terminology

- pure ones
- with maximum tolerated imbalance
- with maximum tolerated and final imbalance
- with adaption
- with adaption and terminal balance
- with adaption and with maximum tolerated imbalance

3. Options of the Case Study

Options of ERDO-Various Randomization Procedures

CR complete randomization, tossing a fair coin, so the probability that patient i will receive treatment E is always $\frac{1}{2}$

RAR random allocation rule, fix total sample size N and randomize so that half the patients receive treatment E

PBR(b) permuted block randomization with block size $b \in \{2, 10\}$, implement RAR within each block

BSD(a) big stick design, use CR allow for a maximum tolerated imbalance $a \in \{3, 4, 5\}$

EBC(p) Efron' s biased coin, flip a biased coin $p = 2/3$ in favor of the less frequently allocated treatment

Chen(p,a) (Chen's Design) use EBC(p) allow for a maximum tolerated imbalance of $a \in \{2, 4\}$.

3. Metric of the Case Study

several evaluation metrics are possible, averaged number of best guesses, balancing behavior, loss in estimation, etc.

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.

Assess the various randomization procedures with respect to

Two metrics

- averaged type 1 error probability over all sequences
- proportion of sequences which maintain the 5% significance level (Level of Evidence)

4. Evaluation Method - Joint Bias Model

Model for two arm parallel group design with continuous endpoint

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

- test the hypotheses $H_0 : \mu_E = \mu_C$ vs. $H_1 : \mu_E \neq \mu_C$
- $T_i = 0$ or $T_i = 1$ if patient i is allocated to group C (without EB) or E (with additional EB)
- μ_j expected response under treatment $j = C, E$
- errors ϵ_i iid $\mathcal{N}(0, \sigma^2)$

Joint additive bias model to design the EnBand-Study

$$\tau_i = 0.26 \frac{i}{65 + 65} + 0.15 \operatorname{sign}(N_E(i-1) - N_C(i-1))$$

denotes the fixed unobserved "bias" effect

4. Evaluation Method of the Case Study - Procedure

- Identify the best practice randomization procedure for the EnBand-Study by a comprehensive simulation study.
- Conduct a sensitivity study use η between 0.1, 0.15 and 0.2 and θ between 0.2, 0.25 and 0.3 as suitable values.

Decision

- Select the design with the proportion of sequences ≤ 0.05 as close as possible to CR

4. Software used in the Case Study

. . . 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`, `RTBD`, `EBC`, `ABCD`, `GBC`, `Chen`, `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

(Uschner, 2016)

5. Results of the Case Study

Table: Impact of selection bias and time trend on probability of type I error for different randomization procedures

Randomization Procedure	Selection Bias	Linear-Time Trend Bias	Type I Error Probability [mean]	Type I Error Probability ≤ 0.05
CR	0.150	0.260	0.051	0.522
RAR	0.150	0.260	0.053	0.272
PBR(2)	0.150	0.260	0.134	0.000
PBR(10)	0.150	0.260	0.079	0.000
BSD(3)	0.150	0.260	0.057	0.034
BSD(4)	0.150	0.260	0.053	0.220
BSD(5)	0.150	0.260	0.052	0.351
EBC(2/3)	0.150	0.260	0.069	0.007
Chen(2)	0.150	0.260	0.085	0.000
Chen(4)	0.150	0.260	0.071	0.000

5. Results of the Case Study

Table: Impact of selection bias on probability of type I error for different randomization procedures

Randomization Procedure	Selection Bias	Linear-Time Trend Bias	Type I Error Probability [mean]	Type I Error Probability ≤ 0.05
CR	0.150		0.051	0.484
RAR	0.150		0.053	0.134
PBR(2)	0.150		0.135	0.000
PBR(10)	0.150		0.080	0.000
BSD(3)	0.150		0.057	0.008
BSD(4)	0.150		0.054	0.131
BSD(5)	0.150		0.052	0.267
EBC(2/3)	0.150		0.070	0.001
Chen(2)	0.150		0.085	0.000
Chen(4)	0.150		0.072	0.000

Discussion of the Case Study Results

Decision

- Because of the results presented in the table, BSD (5) should be used for the EnBand Study. This is rather robust if only time trend or selection bias are present in the data.

- large differences between the performance of the randomization procedures
- even the complete randomization procedure does not prevent against selection bias overall
- misleading "expected type one error probability" which shows comparable results for CR, RAR and BSD, but for CR, almost 50% of trials exhibit a type one error probability elevation; in case of BSD(5) this amounts to nearly 2/3.

Conclusion from the Case Study Results

In the planned "encircling band" study, assuming moderate amount of selection and time trend bias effects the BSD (5) procedure was recommended for designing the clinical trial.

FP7 HEALTH 2013 - 602552

Conclusion for the Evaluation Process

- 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) (Uschner, 2016)*

- start understanding effects with time to event data *(Rückbeil, 2016)*
- start understanding effects with multifactorial designs *(Tasche, 2016)*
- start understanding the effect of missing values on the test decision based on randomization test *(Heussen, 201X)*

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

FP7 HEALTH 2013 - 602552

References

- Kennes LN, Cramer E, Hilgers RD and Heussen N. (2011). The impact of selection bias on test decisions in randomized clinical trials *Statistics in Medicine* 2011; **30**:2573-2581.
- Kennes LN. (2012). The impact of selection bias on test decisions in randomized clinical trials *Master Thesis Mathematics RWTH Aachen*
- Kennes LN, Rosenberger WF and Hilgers RD. (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, Hilgers RD and Heussen N. Impact of third order selection bias on test decisions in survival analysis 2016 submitted
- Tasche A. Selection Bias bei mehr als zwei Behandlungsgruppen *Studienarbeit, RWTH Aachen University, Germany*, 2016
- Tamm M, Cramer E, Kennes LN and 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 and 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.
- Uschner D, Schindler D, Heussen N and Hilgers RD. randomizeR: An R Package for the Assessment and Implementation of Randomization in Clinical Trials 2016 submitted

FP7 HEALTH 2013 - 602552

5. Results of the Case Study

Table: Impact of time trend on probability of type I error for different randomization procedures

Randomization Procedure	Selection Bias	Linear-Time Trend Bias	Type I Error Probability [mean]	Type I Error Probability ≤ 0.05
CR		0.260	0.050	0.667
RAR		0.260	0.050	0.671
PBR(2)		0.260	0.049	1.000
PBR(10)		0.260	0.049	1.000
BSD(3)		0.260	0.049	1.000
BSD(4)		0.260	0.050	0.985
BSD(5)		0.260	0.050	0.944
EBC(2/3)		0.260	0.049	0.987
Chen(2)		0.260	0.049	1.000
Chen(4)		0.260	0.049	1.000