Selection of a Randomisation Procedure
Does it matter? How it works!

Ralf-Dieter Hilgers
with the help of
Nicole Heussenn, Marcia Rückbeil, David Schindler, Diane Uschner

Department of Medical Statistics, RWTH Aachen University

IDeAI Webinar Series (II), 2016, October 6th

This research is part of WP 2 of the IDeAI project funded from the European Union Seventh Framework Programme [FP7 2007-2013] under grant agreement No. 60255.
Where we are?

- **Assessment of Randomization**
  - Ralf-Dieter Hilgers, RWTH Aachen

- **Simulation of Clinical Trials**
  - Mats Karlsson, Upsalla Univ.

- **Extrapolating Dose-Response Information**
  - Holger Dette, RUB Bochum

- **Selection of Genetic Factors**
  - Malgorzata Bogdan, Warschau Univ.

- **Design of Pharmacogenetic Trials**
  - Stephen Senn, LIH Luxemburg

- **Surrogate Endpoints**
  - Geert Molenberghs, Hasselt Univ.

- **Adaptive Design Studies**
  - Franz König, MUW Wien

- **Optimal Design in Mixed Models**
  - France Mentré, INSERM Paris

- **Decision Analysis**
  - Carl-Fredrik Burman, Chalmers Univ.

- **Dissemination**
  - Christoph Male, MUW Wien
What is this Webinar about?

- Randomisation is a key feature to protect against bias in randomised clinical trials.
- Does randomisation protect against bias?
- Study the impact of bias on the study results depending on the randomisation procedure.
- Show the instruments.
- Show related aspects.
ICH E9 guideline recommends to study the potential contribution of bias to the $p$–value.

Neither the ICH guidelines nor the CONSORT statement refer to a scientific guided decision about a particular randomisation procedure and reporting standard is weak.

The specific randomisation procedure is seldom or only partly considered in the statistical analysis.

It seems that the selection is left to the scientists opinion.

So why do make things complicated?
natural choice of treatment assignments before Bradford Hill introduces the randomisation in clinical trials (*D’Arcy Hart P. 1940*)
still sometime favoured nowadays (*Mathe, 2005*)
What do you think about this?
### Can the effect of selection bias be measured?

<table>
<thead>
<tr>
<th>total sample size N</th>
<th>12</th>
<th>24</th>
<th>48</th>
<th>96</th>
<th>192</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect size (efs)</td>
<td>1.796</td>
<td>1.197</td>
<td>0.826</td>
<td>0.578</td>
<td>0.408</td>
</tr>
<tr>
<td>$\frac{efs}{4}$</td>
<td>0.45</td>
<td>0.30</td>
<td>0.21</td>
<td>0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.10867</td>
<td>0.10799</td>
<td>0.10783</td>
<td>0.10787</td>
<td>0.10827</td>
</tr>
</tbody>
</table>
Ok, “tossing coin” randomisation or more precisely use complete randomisation.

**Balancing of sample size?**
Ok, keep the number of allocations to the treatment groups fixed or more precisely use random allocation rule.

Ok, let’s do block randomisation, or more precisely use permuted block randomisation.

Balancing of sample size important?

Table: Balancing behaviour and Power, t-test at twosided 5% significance level

<table>
<thead>
<tr>
<th>total sample size N</th>
<th>12</th>
<th>24</th>
<th>48</th>
<th>96</th>
<th>192</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio ( \frac{n_1}{n_2} )</td>
<td>2:1</td>
<td>2:1</td>
<td>2:1</td>
<td>2:1</td>
<td>2:1</td>
</tr>
<tr>
<td>Power</td>
<td>75%</td>
<td>75%</td>
<td>75%</td>
<td>75%</td>
<td>75%</td>
</tr>
</tbody>
</table>
Does Randomisation Do a Good Job?

**SPR-Study**: Multicenter, randomised (open label), prospective clinical trial comparing scleral buckling (SB) versus primary vitrectomy (PPV) in rhegmatogenous retinal detachment of medium complexity, change in BCVA

**Surgeon no 12** (n = 36)
- SB $-1.046(0.711)$, PPV $-0.390(0.662)$
  - Welch t-test: $p = 0.0071$
- LR test for treatment: 5.59 ($p = 0.0179$).
- LR test for selection bias: 0.007 ($p = 0.9338$).

**Surgeon no 13** (n = 34)
- SB $-0.707(0.622)$, PPV $-0.680(1.048)$
  - Welch t-test: $p = 0.9278$
- LR test for treatment: 1.27 ($p = 0.2593$).
- LR test for selection bias: 3.577 ($p = 0.0586$).

(Heimann, 2007)
(Kennes, 2015)
there seem to be an impact of selection bias irrespective of randomisation

no randomisation procedure performs best with all criteria

no recommendation to give scientific arguments for the choice of randomisation procedure
  ▶ ICH Guidelines
  ▶ CONSORT Statement
Work program of WP2: Randomisation

- Assessment of randomisation procedures with respect to impact of bias on test results
- Development / Recommendation of adequate randomisation procedures for small population groups
- Development of randomisation test for small population groups
Summary: CSE - Randomisation

1. **Introduction** - intend select the best practice randomisation procedure (RP) to improve the level of evidence

2. **Objective** - select a RP with respect to impact on $\alpha$

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 randomisation design
### 3.1 CSE - Assumptions: Statistical 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 \( C \) or \( E \)
- \( \mu_j \) expected response under treatment \( j = C, E \)
- \( \tau_i \) denotes the fixed unobserved "bias" effect
- errors \( \epsilon_i \) iid \( \mathcal{N}(0, \sigma^2) \)
two arm parallel group trial with continuous endpoint

Joint Additive Bias

\[ \tau_i = \theta \frac{i}{N_E + N_C} + \eta \left[ \text{sign} \left( n_E(i - 1) - n_C(i - 1) \right) \right] \]

- \( n_j(i) \): assignments to treatment \( j \) after \( i \) allocations
- different shape of time trend or selection bias can be incorporated
- weighted additive or multiplicative bias model
3.2 CSE - Options: Randomisation Procedures

Options of CSE-randomisation

- **various randomisation procedures and their parameter settings**

**RAR** Random Allocation rule, fix total sample size $N$. randomise so that half the patients receive treatment 1

**BSD($a$)** (Big Stick design) CR allow for imbalance within a limit $a$ . . . etc.
ICH E9: The interpretation of statistical measures of uncertainty of the treatment effect and treatment comparisons should involve considerations of the potential contribution of bias to the p-value, confidence interval, or inference.

**Metric of CSE randomisation**

- → *empirical type-I-error rate*

study the empirical type-I-error rate or empirical test size via simulation
5. CSE - Software: randomizeR

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

**current status of randomizeR**

- implemented randomisation procedures: CR, RAR, PBR, RPBR, HADA, MP, BSD, UD, TBD, EBC, GBC, CD, BBC
- generating / saving a randomisation sequence as .csv file
- implemented assessment criteria: selBias, chronBias, corGuess, imbal, setPower, combineBias
- assessment and comparison of randomisation procedures possible

**in progress\next steps**

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

(Uschner, 2016)
setting: $N_E = N_C = 48, \eta = 0.0 \times \delta; \theta = 0$
setting: \( N_E = N_C = 48, \eta = 0.1 \times \delta; \theta = 0 \)
setting: \( N_E = N_C = 48, \eta = 0.2 \times \delta; \theta = 0 \)
setting: \( N_E = N_C = 48, \eta = 0.3 \times \delta; \theta = 0 \)

### RAR

- \( P: 0.146 \)

### BSD (4)

- \( P: 0.228 \)
setting: $N_E = N_C = 48, \eta = 0.4 \times \delta; \theta = 0$
setting: $N_E = N_C = 48$, $\eta = 0.5 \times \delta$; $\theta = 0$
setting: $N_E = N_C = 48$, $\eta = 0$; $\theta = 0.0 \times \sigma$
setting: $N_E = N_C = 48, \eta = 0; \theta = 0.2 \times \sigma$
setting: $N_E = N_C = 48$, $\eta = 0$; $\theta = 0.4 \times \sigma$
setting: $N_E = N_C = 48, \eta = 0; \theta = 0.6 \times \sigma$
setting: $N_E = N_C = 48, \eta = 0; \theta = 0.8 \times \sigma$
6. CSE - Result: Linear Time Trend Bias (N=96)

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

RAR

BSD (4)
6. CSE - Result: Both Biases for \((N=96)\)

setting: \(N_E = N_C = 48, \eta = 0.0 \times \delta, \theta = 0.0 \times \sigma\)
6. CSE - Result: Both Biases for \((N=96)\)

setting: \(N_E = N_C = 48, \eta = 0.1 \times \delta, \theta = 0.2 \times \sigma\)
setting: $N_E = N_C = 48, \eta = 0.2 \times \delta, \theta = 0.4 \times \sigma$
6. CSE - Result: Both Biases for \((N=96)\)

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

![Graph showing RAR and BSD (4)]
6. CSE - Result: Both Biases for (N=96)

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

RAR

BSD (4)
6. CSE - Result: Both Biases for (N=96)

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

RAR

BSD (4)
Clinical implication

- 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
- in the context or rare diseases, this implies the risk to overlook an effective treatment
Conclusion

- presented a framework for scientific evaluation of randomisation 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 randomisation sequence
- software to do assessment is available, R package *(randomizeR)*
Conclusion

What is about other metrics / criteria?

- developed a uniform assessment criteria (*Schindler, 2016*)
Conclusion

What is about multiarm clinical trials?

- start understanding effects with multifactorial designs (Tasche, 2015, Uschner, 2016)
What is about time to event endpoints?

- start understanding effects with time to event data (Rückbeil, 2015)
Open problems??

- start understanding the effect of missing values on the test decision based on randomisation test (*Heussen, 2016*)

- no yet completely developed a bias corrected test (*Kennes, 2015*)
The IDeAl Series of Webinars

- **VISIT THE IDeAl WEBPAGE**
  - http://www.ideal.rwth-aachen.de

- Get **LinkedIn IDEAL ? FP7 Project**
  - http://www.linkedin.com/groups/IDEAL-FP7-Project-6556030

- **Twitter @ideal_fp7**
  - https://twitter.com/ideal_fp7

- **Get your own experience with WP 2 findings:**
  - Software Download: https://cran.r-project.org/web/packages/randomizr/index.html
  - Online Tutorials: https://youtu.be/o1xBft_VoBA

Kennes LN. (2012). The impact of selection bias on test decisions in randomised clinical trials *Master Thesis Mathematics RWTH Aachen*


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


Uschner D, Schindler D, Heussen N and Hilgers RD. randomiseR: An R Package for the Assessment and Implementation of randomisation in Clinical Trials 2016 submitted