Assessment of randomization procedures based on single sequences under selection and chronological bias

Ralf-Dieter Hilgers, Simon Langer

Department of Medical Statistics, RWTH-Aachen University

Friday 10th July 2015
Table of contents

1 Introduction

2 Setting

3 Assessment

4 Conclusion

5 References
Introduction

- 21 out of 63 Orphan drug legislations involve open label studies (Joppi, 2013)

- *treatment comparisons should involve consideration of the potential contribution of bias to the p-value* (ICH E9)

- no recommendation to give scientific arguments for selection of randomization procedure

- no uniform performance of randomization procedures

- Clinical Scenario Evaluation (Benda 2010)
Propose a tool for assessing the impact of selection bias as well as chronological bias on the type one error probability for a given randomization sequence (procedure) and thus enabling a scientific discussion of the appropriate choice of the randomization procedure.
Setting - Model

Two arm parallel group design

\[ y_i = \mu_A Z_i + \mu_B (1 - Z_i) + \tau_i + \epsilon_i, \quad 1 \leq i \leq n_A + n_B \]  

- allocation

\[ Z_i = \begin{cases} 
1 & \text{if patient } i \text{ is allocated to group } A \\
0 & \text{if patient } i \text{ is allocated to group } B 
\end{cases} \]

- \( \mu_j \) expected response under treatment \( j = A, B \)

- \( \tau_i \) denotes the fixed unobserved "bias" effect acting on the response of patient \( i \)

- errors \( \epsilon_i \) iid \( N(0, \sigma^2) \)
Setting - Test Statistic

Aim: test the hypotheses $H_0 : \mu_A = \mu_B$ vs. $H_1 : \mu_A \neq \mu_B$

use t-Test (under misspecification)

$$T = \sqrt{\frac{n_An_B}{n_A+n_B}} \left( \tilde{y}_A - \tilde{y}_B \right)$$

$$\frac{1}{n_A+n_B-2} \left( \sum_{i=1}^{n} Z_i(y_i - \tilde{y}_A)^2 + \sum_{i=1}^{n} (1 - Z_i)(y_i - \tilde{y}_B)^2 \right)$$

where $\tilde{y}_A = \frac{1}{n_A} \sum_{i=1}^{n} y_i Z_i$; $\tilde{y}_B = \frac{1}{n_B} \sum_{i=1}^{n} y_i (1 - Z_i)$; $n = n_A + n_B$
Theorem: Under $H_0 : \mu_A = \mu_B$ the type 1 error probability in (1) (under misspecification) for the allocation sequence $Z = (Z_1, \ldots, Z_n)$ is

$$P \left( \left| T \right| > t_{n_A+n_B-2}(1 - \alpha/2) \right| Z \right) = F_{n_A+n_B-2,\theta_1,\theta_2} \left( t_{n_A+n_B-2}(\alpha/2) \right) + F_{n_A+n_B-2,-\theta_1,\theta_2} \left( t_{n_A+n_B-2}(\alpha/2) \right).$$

$F_{n_A+n_B-2,\theta_1,\theta_2}$ denotes the distribution function of the doubly non-central t-distribution with $n_A + n_B - 2$ degrees of freedom and parameters

$$\theta_1 = \frac{1}{\sigma} \sqrt{\frac{n_An_B}{n_A+n_B}} \left( \tilde{\tau}_A - \tilde{\tau}_B \right), \quad \theta_2 = \frac{1}{\sigma^2} \left[ \sum_{i=1}^{n} \tau_i^2 - n_A \tilde{\tau}_A^2 - n_B \tilde{\tau}_B^2 \right]$$

where \( \tilde{\tau}_A = \frac{1}{n_A} \sum_{i=1}^{n} \tau_i Z_i \); \( \tilde{\tau}_B = \frac{1}{n_B} \sum_{i=1}^{n} \tau_i (1 - Z_i) \)
Sketch of the proof:

for the given allocation vector \( \mathbf{Z} = (Z_1, \ldots, Z_n) \)

\[
\tilde{y}_A - \tilde{y}_B \sim N \left( \mu_A - \mu_B + \tilde{\tau}_A - \tilde{\tau}_B, \sigma^2 \frac{n_A + n_B}{n_A n_B} \right)
\]

where \( \tilde{\tau}_A = \frac{1}{n_A} \sum_{i=1}^{n} \tau_i Z_i \) ; \( \tilde{\tau}_B = \frac{1}{n_B} \sum_{i=1}^{n} \tau_i (1 - Z_i) \)
for the allocation vector \( \mathbf{Z} = (Z_1, \ldots, Z_n) \)

\[
\sum_{i=1}^{n} Z_i(y_i - \tilde{y}_A)^2 + \sum_{i=1}^{n} (1 - Z_i)(y_i - \tilde{y}_B)^2 \sim \chi_{n_A + n_B - 2}(\theta_2)
\]

with non-centrality parameter

\[
\theta_2 = \frac{1}{\sigma} \left( \sum_{i=1}^{n} Z_i(\tau_i - \tilde{\tau}_A)^2 + \sum_{i=1}^{n} (1 - Z_i)(\tau_i - \tilde{\tau}_B)^2 \right)
\]

\[
= \frac{1}{\sigma} \left[ \sum_{i=1}^{n} \tau_i^2 - n_A\tilde{\tau}_A^2 - n_B\tilde{\tau}_B^2 \right]
\]
Thus $T$ follows a doubly non-central $t$ distribution with $n_A + n_B - 2$ degrees of freedom and non-centrality parameters (Johnson, Kotz, Balakrishnan, 1995, Robins, 1948)

$$\theta_1 = \frac{1}{\sigma} \sqrt{\frac{n_An_B}{n_A + n_B}} (\mu_A - \mu_B + \tilde{\tau}_A - \tilde{\tau}_B) = \frac{1}{\sigma} \sqrt{\frac{n_An_B}{n_A + n_B}} (\tilde{\tau}_A - \tilde{\tau}_B)$$

$$\theta_2 = \frac{1}{\sigma^2} \left[ \sum_{i=1}^{n} \tau_i^2 - n_A\tilde{\tau}_A^2 - n_B\tilde{\tau}_B^2 \right]$$

using the properties of the distribution (Kocherlakota, 1991)

$$F_{\nu,\theta_1,\theta_2}(t) = 1 - F_{\nu,-\theta_1,\theta_2}(-t)$$
Assessment - 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 $k$ Blocks of size $m_s$, $1 \leq s \leq k$

**BSD($a$)**  (Big Stick design) CR allow for imbalance within a limit $a$
Assessment - Selection Bias Model

With $N_j(i - 1)$ is the number of treatment $j$ assignments after $(i - 1)$ assignments and

$$p_A(i - 1) = \frac{n_A - N_A(i - 1)}{(n_A + n_B) - (N_A(i - 1) + N_B(i - 1))}, \quad n_A = n_B$$

the selection biasing policy (Tamm, 2012), according to convergence strategy (Blackwell, 1957), $q \in \left[\frac{1}{2}, 1\right]$

$$E(y_i) = \begin{cases} \mu_A Z_i + \mu_B (1 - Z_i) + \eta & \text{if } p_A(i - 1) > q \\ \mu_A Z_i + \mu_B (1 - Z_i) & \text{if } 1 - q \leq p_A(i - 1) \leq q \\ \mu_A Z_i + \mu_B (1 - Z_i) - \eta & \text{if } p_A(i - 1) < 1 - q \end{cases}$$

Then

$$\tau_i = \eta \left[1_{(q, 1]}(p_A(i - 1)) - 1_{[0, 1-q]}(p_A(i - 1))\right]$$
Table: Empirical type 1 error probability by different cut-off values for a two sided t-test with $2n = 40; \alpha = 0.05, \beta = 0.2$ and selection effect $\eta = \frac{\delta}{2} = 0.45$ under PBR

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>Type 1 Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PBR(4)</td>
</tr>
<tr>
<td>1/2</td>
<td>0.191</td>
</tr>
<tr>
<td>2/3</td>
<td>0.145</td>
</tr>
<tr>
<td>1</td>
<td>0.050</td>
</tr>
</tbody>
</table>

using SAS with 10 000 replications
**Table:** Empirical type 1 error probability of a two sided t-test with $2n$; $\alpha = 0.05$, $\beta = 0.2$ and selection bias effect $\eta = \frac{\delta(2n)}{2}$

<table>
<thead>
<tr>
<th>$2n$</th>
<th>$\delta(2n)$</th>
<th>CR</th>
<th>RAR</th>
<th>PBR(4)</th>
<th>BSD (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2.381</td>
<td>0.058</td>
<td>0.102</td>
<td>0.141</td>
<td>0.064</td>
</tr>
<tr>
<td>20</td>
<td>1.325</td>
<td>0.054</td>
<td>0.082</td>
<td>0.177</td>
<td>0.075</td>
</tr>
<tr>
<td>32</td>
<td>1.024</td>
<td>0.055</td>
<td>0.072</td>
<td>0.188</td>
<td>0.083</td>
</tr>
<tr>
<td>40</td>
<td>0.909</td>
<td>0.053</td>
<td>0.071</td>
<td>0.195</td>
<td>0.088</td>
</tr>
</tbody>
</table>

using R with 100 000 replications
Tamm (2012) found:

- empirical type I error increases with smaller blocksize under PBR
- empirical type I error decreases with higher $q$
- elevation are substantial even for $q = 2/3$
Assessment - Chronological Bias Tamm (2014)

- 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

\[
\tau_i = \lambda \begin{cases} 
\frac{i}{n_A+n_B} & \text{linear time trend} \\
1_{i \geq c}(i) & \text{stepwise trend} \\
\log\left(\frac{i}{n_A+n_B}\right) & \text{log trend}
\end{cases}
\]
Tamm (2014) found, that

- using t-test in presence of time trends results in conservative test decisions under permuted block randomizations even for small time trends; Blocked ANOVA should be used
- medium block sizes already achieve a good reduction of the inflated type I error rate in worst-case scenarios
Assessment - Selection and chronological bias for special sequences

all Sequences assume $n_A = n_B = n/2$

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Bias</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>type 1 error probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternating</td>
<td>Selection</td>
<td>$-\frac{\eta}{\sigma}\sqrt{n}/2$</td>
<td>0</td>
<td>0.074</td>
</tr>
<tr>
<td>10101...</td>
<td>Time Trend</td>
<td>$\frac{\lambda}{\sigma}\sqrt{n}/2$</td>
<td>$\frac{\lambda^2}{\sigma^2}n(n^2 - 4)/12$</td>
<td>0.000</td>
</tr>
<tr>
<td>Separating</td>
<td>Selection</td>
<td>$-\frac{\eta}{\sigma}\sqrt{1/n}$</td>
<td>$\frac{\eta^2}{\sigma^2}(n - 2)/n$</td>
<td>0.050</td>
</tr>
<tr>
<td>111...000...</td>
<td>Time Trend</td>
<td>$\frac{\lambda}{\sigma}\sqrt{n^3/4}$</td>
<td>$\frac{\lambda^2}{\sigma^2}n(n^2 - 4)/48$</td>
<td>0.999</td>
</tr>
<tr>
<td>avoid TT</td>
<td>Selection</td>
<td>$-\frac{\eta}{\sigma}\sqrt{n}/2$</td>
<td>$\frac{\eta^2}{\sigma^2}$</td>
<td>0.050</td>
</tr>
<tr>
<td>1001 1001...</td>
<td>Time Trend</td>
<td>0</td>
<td>$\frac{\lambda^2}{\sigma^2}n(-1 + n^2)/12$</td>
<td>0.000</td>
</tr>
</tbody>
</table>

type 1 error probability setting: $n_A = n_B = 6, \eta = \frac{\delta(12)}{4} = 0.45, \lambda = \sigma = 1$
Assessment - Joint additive Bias

weighted additive (selection and chronological) bias model

\[ \tau_i = \lambda \frac{i}{n_A + n_B} + \eta \text{sign}(N_A(i - 1) - N_B(i - 1)) \]

- weights via definition of \( \lambda \) and \( \eta \)
- different shape of time trend can be incorporated
- relaxed version of selection bias possible
- multiplicative could also be done
Assessment - Assessment of Bias for RAR (N=12)

setting: \( n_A = n_B = 6, \eta = \frac{\delta_{12}}{4} = 0.45, \lambda = \sigma = 1 \)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Selection Bias</th>
<th>Linear Time Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>0.0397</td>
<td>0.0415</td>
</tr>
<tr>
<td>x05</td>
<td>0.0481</td>
<td>0.0414</td>
</tr>
<tr>
<td>x25</td>
<td>0.0518</td>
<td>0.0423</td>
</tr>
<tr>
<td>x50</td>
<td>0.0611</td>
<td>0.0449</td>
</tr>
<tr>
<td>mean</td>
<td>0.0635</td>
<td>0.0500</td>
</tr>
<tr>
<td>x75</td>
<td>0.0724</td>
<td>0.0520</td>
</tr>
<tr>
<td>x95</td>
<td>0.0991</td>
<td>0.0736</td>
</tr>
<tr>
<td>max</td>
<td>0.1086</td>
<td>0.1188</td>
</tr>
<tr>
<td>sd</td>
<td>0.0151</td>
<td>0.0115</td>
</tr>
</tbody>
</table>
Assessment - Assessment of Bias for BSD(2) (N=12)

setting: $n_A = n_B = 6$, $\eta = \frac{\delta(12)}{4} = 0.45$, $\lambda = \sigma = 1$

<table>
<thead>
<tr>
<th>Measure</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selection Bias</td>
<td>Linear Time Trend</td>
</tr>
<tr>
<td>min</td>
<td>0.0342</td>
</tr>
<tr>
<td>x05</td>
<td>0.0370</td>
</tr>
<tr>
<td>x25</td>
<td>0.0483</td>
</tr>
<tr>
<td>x50</td>
<td>0.0537</td>
</tr>
<tr>
<td>mean</td>
<td>0.0557</td>
</tr>
<tr>
<td>x75</td>
<td>0.0619</td>
</tr>
<tr>
<td>x95</td>
<td>0.0792</td>
</tr>
<tr>
<td>max</td>
<td>0.1086</td>
</tr>
<tr>
<td>sd</td>
<td>0.0129</td>
</tr>
</tbody>
</table>
Assessment - Comparison RAR and BSD(2) (N=12)

**Figure: RAR**

**Figure: BSD(2)**

(setting: $n_A = n_B = 6$, $\eta = \frac{\delta(12)}{4} = 0.45$, $\lambda = \sigma = 1$)
Conclusion

- presented a framework for scientific evaluation of randomization procedures on bias
- other randomization procedures are easy to implement
- this evaluation should be part of the TSAP
- dichotomous endpoint / survival endpoint could be evaluated by simulation (*work under progress by Marcia Rückbeil*)
- Cave: this is the wrong test (!) but mirrors the practical situation, use randomization tests to reflect the randomization argument
- R package (*randomizeR*) coming soon, see website (http://www.ideal.rwth-aachen.de)
References


Tamm M, Hilgers RD. *Methods of Information in Medicine* 2012; **51**:138-143.