Influence of Selection Bias on the Test Decision

A Simulation Study

M. Tamm1; E. Cramer2; L. N. Kennes1; N. Heussen1
1Department of Medical Statistics, RWTH Aachen University, Aachen, Germany;
2Department of Medical Statistics, RWTH Aachen University, Aachen, Germany;
M. Tamm1; E. Cramer2; L. N. Kennes1; N. Heussen1

Keywords
Selection bias, randomized clinical trials, permuted block randomization, allocation concealment, masking

Summary
Background: Selection bias arises in clinical trials by reason of selective assignment of patients to treatment groups. Even in randomized clinical trials with allocation concealment this phenomenon can occur if future assignments can be predicted due to knowledge of former allocations.

Objectives: Considering unmasked randomized clinical trials with allocation concealment the impact of selection bias on type I error rate under permuted block randomization is investigated. We aimed to extend the existing research into this topic by including practical assumptions concerning misclassification of patient characteristics to get an estimate of type I error close to clinical routine. To establish an upper bound for the type I error rate different biasing strategies of the investigator are compared first. In addition, the aspect of patient availability is considered.

Methods: To evaluate the influence of selection bias on type I error rate under several practical situations, different block sizes, selection effects, biasing strategies and success rates of patient classification were simulated using SAS.

Results: Type I error rate exceeds 5 percent significance level; it reaches values up to 21 percent. More cautious biasing strategies and misclassification of patient characteristics may diminish but cannot eliminate selection bias. The number of screened patients is about three times larger than the needed number for the trial.

Conclusions: Even in unmasked randomized clinical trials using permuted block randomization with allocation concealment the influence of selection bias must not be disregarded evaluating the test decision. It should be incorporated when designing and reporting a clinical trial.

1. Introduction

When conducting a clinical trial to compare two different treatments, it is an important aspect to ensure that the two generated treatment groups are in fact comparable. Selective allocation to the treatment groups based on different patient characteristics can bias the test decision. This effect is called selection bias. For example, healthier patients may be elected for one treatment whereas sicker patients are allocated to the other therapy. Random allocation of treatments to patients is known to reduce this effect of selection bias. In many trials [1] the random allocation is based on permuted block randomization (PBR) ([2] pp 41–42). The block sizes could be fixed or a design with random choice of block sizes could be applied. An advantage of PBR is that the groups are balanced over the time and chronological bias can be controlled ([3] pp 65). The allocations must be concealed to prevent enrollment of participants according to the next assignment specified by the randomization list. However, allocation concealment cannot totally eliminate selection bias ([4] pp 42–80). If the next allocation is indeed concealed but predictable in some way, patients could be enrolled according to the treatment expected to be assigned next. Thus, missing masking of previous allocations causes possible prediction of the next assignment. Due to the fact that the treatments are balanced within one block, information about the former allocations might be abused to predict the following assignment. If the block size is known because of details in the protocol or predictable by the information about previous allocations, probabilities of the next assignment are determinable. Some assignments could even be determined with certainty. Bias resulting by reason of unmasked previous allocations is called third-order selection bias [5]. We investigate the influence of third-order selection bias in this scenario of an unmasked randomized clinical trial with allocation concealment. In fact, there are some trials for which masking is hard to conduct, e.g. in case of surgery treatment where the physician has to know the allocated treatment ([2] p 19).

Proschan [6] and Kennes et al. [7] have already presented theoretical results investigating the impact of third-order selection bias on type I error. Their derivations assume, among others, unlimited availability...
of patients with desired expected response to therapy and a correct discrimination between these patients. Most likely, in practice these assumptions will not be fulfilled. Therefore we extended their theoretical results carrying out a simulation study to investigate these aspects. In the simulation we chose small block sizes as well as larger block sizes to meet the wide range of suggestions in the literature regarding the choice of block size. For example, Altman and Bland suggest the choice of small block sizes to achieve balanced groups [8]. Pocock ([9] p 77) notes that “[…] in general a trial without stratification should have a reasonably large block size so as to reduce predictability […]."

Moreover, different strategies of patient selection — named biasing strategies — are investigated first. Berger [5] and Follmann and Proschan [10] quantified the impact of selection bias under different biasing strategies with the expected covariate imbalance and the so called “probability of successful bias minus the probability of unsuccessful bias”. We aimed to investigate the influence of different biasing strategies on type I error rate to establish an upper bound for the increase of type I error rate.

2. Model

We consider an unmasked randomized clinical trial with allocation concealment comparing two treatments, experimental (E) and control (C). The randomization is executed in permuted blocks of fixed length. The future allocations are concealed, whereas the past assignments to group E or C are known to the investigator, who is involved in the recruiting process of patient enrollment. We assume that the investigator prefers the experimental treatment of this index is incorporated in the appraisal of patient. However, misjudgment of this index is incorporated in the simulation. We investigate the influence of selection bias on type I error under the null hypothesis of general we set \( \alpha_i = 0, \delta_i = 1, 2 \). We assume \( \pi_{(i,j)} \) to have values \(-\eta, 0, \eta \) with \( \eta > 0 \) only representing weak, neutral, and strong response to therapy, respectively. The index \( (i, j) \) characterizes different groups of patients with different characteristics resulting in distinct expected response to therapy. To characterize the impact of selection bias the shift \( \eta \) in patient effect caused by the different patient characteristics has to be related to the overall standard deviation \( \sigma \). The ratio \( \gamma = \eta/\sigma \) is named selection effect [6].

Let 2k be the fixed block size of the permuted block design and 2n be the total sample size. \( N_E \) and \( N_C \) denote the actual numbers of assignments to treatment group E or C within one block. We assume that the investigator knows the block size. Then, the (conditional) probability of the next patient included in the study being assigned to the experimental group can be determined as \( P_E = (k - N_E)/(2k - (N_E + N_C)) \). Note that the last allocation within each block is always predictable with certainty. In preference of the experimental group the investigator may enroll a patient with strong expected response to therapy if the probability of next allocation to the experimental treatment is high enough. ‘High enough’ could be expressed in terms of exceeding a fixed cut-off value \( q \) with \( q \geq 1/2 \). Similarly a patient with weak expected response will be included if the probability of allocation to experimental is below \( 1 - q \). For intermediate probabilities a neutral patient may be included. Following this biasing policy the investigator chooses a patient with expected response to therapy:

\[
\begin{cases}
\mu + \eta, & \text{if } P_E > q, \\
\mu, & \text{if } 1-q \leq P_E \leq q, \\
\mu - \eta, & \text{if } P_E < 1-q,
\end{cases}
\]

with \( q \geq 1/2 \). In Ivanova et al. [11] a similar biasing policy assuming a binary outcome is described. It should be noted that for \( q = 1/2 \) Formula 2 reduces to the decision \( N_E > N_C, N_E = N_C, \) or \( N_E < N_C \) for choosing the corresponding patient. In this case the block size does not need to be known.

The setting above presumes that enough patients with corresponding characteristics are available and identifiable. In the following simulations we suppose that the investigator is able to reject all patients based on external (eligibility assessment) or internal (interpretation of inclusion or exclusion criteria) reason until a patient with required characteristics appears and is included ([4], p 11). The quantity of patients screened but not included due to this procedure is additionally counted in the simulations. Especially we concentrate on the question of the identifiability of patient characteristics. The occurrence of misclassifications and therefore the choice of a ‘wrong’ patient are incorporated. Moreover, the effect of different cut-offs on the deviation of type I error from nominal significance level is quantified. A cut-off point \( q = 1/2 \) means that the investigator biases as soon as the assignment to one of the groups is more likely. Higher cut-off values imply that patients with strong or weak expected response to therapy are assigned only if there is a greater imbalance in the group sizes between the experimental and control group. In this situation the next allocation can be predicted correctly with higher probability.

3. Simulation Study

Considering different scenarios the impact of selection bias on the test decision is in-
vestigated via simulation. In particular, we focus on the effect of different cut-off values and mainly on the effect of errors in judging the different patient characteristics. Independent data sets were simulated and analyzed using SAS (SAS 9.1.3 & 9.2, SAS Institute Inc., Cary, NC, USA) under Windows XP (Service Pack 3). The generation of required pseudo random numbers was achieved using the SAS Call routine RANDGEN of the IML procedure. The initialization of the random number stream was done with the RANDSEED call to get independent and reproducible samples.

### 3.1 Data Generation

To investigate the specified scenario in a first step a randomization list based on the permuted block design was generated. The patients' data and the allocation of the patients to the experimental or control group by the investigator were simulated based on the generated randomization list and the preferences of the investigator. As described above, the investigator rejects all possible participants until a patient with desired characteristics arrives and is included in the study. Equal probabilities for the three possible patient characteristics were assumed.

The sample size was related to the difference of the expected response of patients to therapy. A sample size of 200 is necessary to detect a treatment effect size of 0.4 with 80 percent power using the two-tailed t-test on the 5 percent significance level. Adapted from this calculation a sample size of 200 with selection effect (η/σ) of half, quarter and eighth of the assumed effect size was chosen for the simulation.

At first, different cut-off values for the biasing policy were analyzed given a selection effect (η/σ) of 1/5. We assumed a correction for the expected response between the two groups E and C. The performance of the test decision was evaluated with the empirical type I error rate based on the null hypothesis of no treatment differences [12]. The empirical type I error rate was calculated as the proportion of simulated data sets with p-values less than 5%. Deviation of type I error rate from the 5% significance level indicates biasing. Furthermore the number of patients screened but not included in the randomized trial for reason of selection effect was counted for each simulated trial and the average was calculated for each of the one million runs.

### 3.2 Analysis of Generated Data

For each combination of relevant parameters the entire process of data generation was repeated one million times. Each generated data set was analyzed using the two-tailed t-test to investigate differences in the expected response between the two groups E and C.

#### Table 1
Correct classification by investigator

<table>
<thead>
<tr>
<th>Classification by investigator</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>1</td>
</tr>
<tr>
<td>Neutral</td>
<td>0</td>
</tr>
<tr>
<td>Weak</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Table 2
Misclassification by investigator (Level 1)

<table>
<thead>
<tr>
<th>Classification by investigator</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>0.7</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.2</td>
</tr>
<tr>
<td>Weak</td>
<td>0.1</td>
</tr>
</tbody>
</table>

#### Table 3
Misclassification by investigator (Level 2)

<table>
<thead>
<tr>
<th>Classification by investigator</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>0.5</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.3</td>
</tr>
<tr>
<td>Weak</td>
<td>0.2</td>
</tr>
</tbody>
</table>
4. Simulation Results

4.1 Different Cut-offs

The biasing policy with cut-off $q = 0.5$ results in the strongest deviation from the nominal 5% significance level with a type I error rate up to 21% for block size 4 (Table 4). Larger block sizes or higher cut-off values reduce the deviation of type I error rate from the nominal 5% significance level (Tables 4–6). These results coincide with effects observed by Berger [5] and Follmann and Proschan [10]. Although the deviation of type I error rate from the nominal 5% significance level is reduced in the case of biasing policies with high cut-off values, it is still essential. In particular, the choice of ‘strong’ and ‘weak’ patients just when allocation can be predicted with certainty still results in strong deviation from nominal 5% significance level as it can be seen in Tables 4–6. E.g. for block size 4 and $q = 2/3$ the type I error rate is 18% (Table 4).

We already assumed that the investigator waits for inclusion of a patient according to the randomization list until a candidate with supposed response appears. In the meantime patients will not be included in the clinical trial. As expected by theoretical means, we found that the number of screened patients is three times larger than the needed number for the trial. In all scenarios about 400 possible participants were rejected due to the reason that the investigator does not include any patient not matching the expected allocation to treatment following the biasing strategy.

4.2 Misclassification of Patient Characteristics by the Investigator

In Section 4.1 it was observed that the biasing strategy with $q = 0.5$ results in the strongest deviation of 5% significance level. For this reason only this worst case is assumed and investigated in this section. Table 7 presents the type I error rate in the scenario that the investigator always classifies the patients correctly (Table 1). We found that especially for small block sizes and a selection effect of 1/5 the type I error exceeds the 5 percent significance level noticeably. In the worst case (block size 4, selection effect 1/5), the type I error rate increases up to 21%. This result coincides with the result of Section 4.1. Tables 8 and 9 show the results for the weaker and probably more realistic assumption of misclassification of patients by the investigator. As before, selection bias should not be disregarded in these scenarios. Assuming the amount of misclassification specified in Table 2, the type I error rate is more than twice of the nominal significance level considering a block size of 4 and a selection effect of 1/5. However using very large block sizes along with minor selection effects the type I error rate is close to the nominal 5% significance level. The influence of selection bias is negligible. The type I error rates given in Table 9 are based on the setting described in Table 3. In this setting only a slight tendency of correct appraisal of the expected response to therapy of patients by the investigator is supposed. Hence, one might expect that bias of the test decision is negligible. Nevertheless the type I error rate increases up to 6.8% for a block size of 4 and a selection effect of 1/5.

As already mentioned in Section 4.1, the investigator – following the biasing strategy of Formula 2 – rejects a lot of patients which would have probably been included in the clinical trial. The average numbers of prospects are given in the last row of Tables 7–9 for the different scenarios of this section.

5. Discussion and Conclusion

The aim of the simulation study was to investigate how selection bias will inflate the type I error rate in unmasked randomized clinical trials (permuted block design) with allocation concealment assuming realistic conditions close to clinical routine. At first, different cut-offs for the biasing policy were considered. If the conditional probability of being assigned to the experimental group E or to the control group C exceeds the fixed cut-off, either a patient with strong expected response (experimental E) or a patient with weak expected response (control C) was chosen. Otherwise a ‘neutral’ patient was enrolled. We have seen that the biasing policy with a cut-off of 0.5 results in the strongest increase of the type I error rate.

© Schattauer 2012

Methods Inf Med 2/2012
If the number of patients is limited or it is not enough time to wait for the patient desired by the investigator, biasing behavior based on higher cut-offs and hence better prediction of allocation could result in the strongest deviation. This aspect remains to be analyzed in more detail in future research.

Based on the worst case (q = 0.5) we have analyzed the effect of selection bias under the realistic assumption that the investigator is only able to predict the expected response to treatment for a particular patient with a certain probability. Even though the effect of selection bias is reduced in this scenario, the test decision is still considerably biased especially in the case of strong selection effect and small block size. The size of selection effect is hard to determine in advance and it has to be estimated on the basis of present data. Designing a clinical trial one should assume the worst case of strong effects of different patient characteristics and therefore a possibly remarkable bias. Small block sizes are commonly used to achieve balanced groups throughout the trial [1]. Kundt [13] and Berger et al. [14] proposed less restrictive randomization procedures reducing predictability by keeping a maximum tolerable imbalance between groups. In particular, Kundt provides indications of choosing the optimal block size given a tolerable imbalance only achieved or exceed with low probability. It turns out that quite large block sizes can satisfy the balancing condition. We have seen that also under practical assumptions the influence of selection bias on type I error rate is remarkable for small block sizes. Therefore we advise again choosing small block sizes due to balancing concerns. The choice of parameter values for other randomization procedures is discussed by Kundt in [15]. Another important aspect is that for q = 0.5 the block size has not to be known. This implies that any strategy to protect against guessing the block size, for example the choice of random blocks, achieves no reduction of type I error assuming the setting of subsection 3.1. Furthermore it should be mentioned that the choice of larger sample sizes exacerbate the elevation of type I error rate. Theoretical results can be found in [7]. Moreover, we have found that if the investigator follows the biasing policy of Formula 2 strictly, the number of patients screened for the trial heavily exceeds the number of patients included in the trial. Such a proceeding certainly extends the recruitment phase of a clinical trial. Besides the biasing of the test decisions this is a further negative aspect, because almost every trial is faced with the challenge to recruit the planned number of patients in the specified duration of study.

Overall even in randomized clinical trials selection bias is an important aspect to consider when designing the study and interpreting the study results. Berger and Christophi [16] indicated how to report a clinical trial regarding to randomization technique and selection bias. Furthermore Berger and Exner [17] and Ivanova et al. [11] present methods to detect selection bias and adjust for it under certain conditions. Concealment of future allocations but also masking of past assignments to treatment groups are main factors which should be realized if possible.

Acknowledgment
This research was funded by the Excellence Initiative of the German federal and state governments.

References

### Table 7 Type I error rate assuming correct classification by investigator

<table>
<thead>
<tr>
<th>Selection effect</th>
<th>Block size</th>
<th>200</th>
<th>100</th>
<th>10</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/5</td>
<td>0.05271</td>
<td>0.060376</td>
<td>0.133857</td>
<td>0.211986</td>
<td></td>
</tr>
<tr>
<td>1/10</td>
<td>0.051371</td>
<td>0.052499</td>
<td>0.070852</td>
<td>0.089427</td>
<td></td>
</tr>
<tr>
<td>1/20</td>
<td>0.050533</td>
<td>0.050569</td>
<td>0.055155</td>
<td>0.059678</td>
<td></td>
</tr>
<tr>
<td>Average number of prospects</td>
<td>598</td>
<td>598</td>
<td>600</td>
<td>601</td>
<td></td>
</tr>
</tbody>
</table>

### Table 8 Type I error rate assuming misclassification by investigator (Level 1)

<table>
<thead>
<tr>
<th>Selection effect</th>
<th>Block size</th>
<th>200</th>
<th>100</th>
<th>10</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/5</td>
<td>0.052380</td>
<td>0.054584</td>
<td>0.086820</td>
<td>0.119636</td>
<td></td>
</tr>
<tr>
<td>1/10</td>
<td>0.050862</td>
<td>0.051398</td>
<td>0.059143</td>
<td>0.067910</td>
<td></td>
</tr>
<tr>
<td>1/20</td>
<td>0.050323</td>
<td>0.050710</td>
<td>0.052412</td>
<td>0.054779</td>
<td></td>
</tr>
<tr>
<td>Average number of prospects</td>
<td>651</td>
<td>646</td>
<td>615</td>
<td>597</td>
<td></td>
</tr>
</tbody>
</table>

### Table 9 Type I error rate assuming misclassification by investigator (Level 2)

<table>
<thead>
<tr>
<th>Selection effect</th>
<th>Block size</th>
<th>200</th>
<th>100</th>
<th>10</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/5</td>
<td>0.050349</td>
<td>0.050705</td>
<td>0.059192</td>
<td>0.067933</td>
<td></td>
</tr>
<tr>
<td>1/10</td>
<td>0.050343</td>
<td>0.050455</td>
<td>0.052626</td>
<td>0.054361</td>
<td></td>
</tr>
<tr>
<td>1/20</td>
<td>0.049892</td>
<td>0.050100</td>
<td>0.050526</td>
<td>0.051262</td>
<td></td>
</tr>
<tr>
<td>Average number of prospects</td>
<td>652</td>
<td>647</td>
<td>615</td>
<td>597</td>
<td></td>
</tr>
</tbody>
</table>