Evaluation of Imbalance in Stratified Blocked Randomization

Some Remarks on the Range of Validity of the Model

by Hallstrom and Davis

G. Kundt; A. Glass
University of Rostock, School of Medicine, Institute of Biostatistics and Informatics in Medicine and Ageing Research, Rostock, Germany

Keywords
Stratified permuted-block randomization, balancing properties

Summary

Objectives: If in a clinical trial prognostic factors are known in advance, it is often recommended that randomization of patients should be stratified. The best-known method is permuted-block randomization within strata. But it suffers from the disadvantage that imbalance still occurs in the trial as a whole if there are a large number of strata, or/and the block sizes are too large for the number of patients. The results of Hallstrom and Davis are appropriate for evaluating the risk of such a troubled situation by using two special cases of their general variance formula. But it is merely generally argued for whichever practical situations these special cases are valid. Consequently, additional investigations are required to reveal the conditions for correct application.

Methods: We investigated the range of validity of special cases by performing computer simulations, varying a number of trial characteristics, and discuss the application of results for practical situations.

Results: The validity of special cases is not given in each situation. Depending on block size, a binomial distribution model is valid for a permitted average maximum number of patients per stratum between 36% and 57% of considered block sizes, whereas the uniform distribution model works adequately from at least 70%. In an intermediate range of invalidity, implementation of a simulation study is necessary to compute the probability distribution of differences.

Conclusions: Our results are important if choosing the stratified permuted-block randomization to estimate the risk for an intolerable overall imbalance when planning a trial.

Correspondence to:
Guenther Kundt
University of Rostock
School of Medicine
Institute of Biostatistics and Informatics in Medicine and Ageing Research
Ernst-Heydemann-Str. 8
18057 Rostock
Germany
E-mail: guenther.kundt@uni-rostock.de

Methods Inf Med 2012; 51: 55–62
doi: 10.3414/ME10-01-0078
received: October 28, 2010
accepted: March 27, 2011
prepublished: July 26, 2011

1. Introduction

The randomized allocation of treatment is a fundamental principle of experimental design which has been incorporated into clinical trials. There are a variety of different methods which have been proposed and used for allocating treatments to patients in sequential clinical trials. The most basic procedure is the well-known complete randomization (analogous to repeated coin-tossing in the case of two therapies), whereby each allocation is made completely at random. However, complete randomization has the disadvantage that repeated use will occasionally result in trials in which treatment totals are highly imbalanced (especially in small trials, or in the early stages of a trial) [1], and/or in which prognostic profiles across the treatment groups are insufficiently similar. These risks were the motivation in the development of several restricted randomization procedures that control the probability of obtaining a sequence of random assignments with an undesirable imbalance in the numbers assigned to each treatment, e.g. permuted-block randomization [2] or the biased coin design [3] with a suitable choice of values of parameters [4] and different balance effectiveness [5].

Stratification of randomization process is a way to prevent the second failure possibility of imbalance of prognostic factors. Very importantly, it produces a good protection against the effect of recruitment center dropout because it assures that patients in each center are balanced for treatment assignment. Therefore, withdrawal of a center will not result in an imbalance among remaining patients in a stratified randomization trial. With good reason, the ICH guidelines of the EMEA emphasize that, in multicenter trials, clinics should be used for stratification. As clinical populations can be strikingly different “it is advisable to have a separate random scheme for each center, i.e. to stratify by center or to allocate several whole blocks to each center” [6].

In practice, the procedures of blocking and stratifying are often used in combination, e.g. by using the well-known permuted-block randomization within strata. Generally, not only complete randomization but also all restricted randomization methods have both desirable and undesirable properties. A disadvantage of the permuted-block randomization within strata is the risk of considerable imbalance in the trial as a whole, especially if there is a large number of strata, or/and the block sizes are
too large in relation to the number of patients enrolled. Hallstrom and Davis [7] derived a formula for calculating the variance of difference in the number of patients assigned to each treatment, depending on the block size and the number of strata, and hence, for assessing the overall imbalance of a trial. To our knowledge, this has until now constituted the most important tool for users to assess the overall imbalance and to decide whether and with what probability the aforementioned disadvantage might or might not be present in a concrete practical situation.

They considered two special cases that specify the general variance formula to become feasible for application. It has to be proved for whichever practical situations (described by block size, number of randomized patients and number of strata) of both special cases the formula is valid. If it is valid, imbalance can be estimated easily; if not, simulations are required. In this paper we focus on this topic.

In Section 2, we briefly explain the method of stratified permuted-block randomization, and describe the rationale behind the investigations of Hallstrom and Davis. Later on, in Section 3, we present the methods applied and in Section 4 the main results of our own simulation study regarding the range of validity of the two special cases of the model of Hallstrom and Davis, followed by evaluation of practical situations and a detailed example. In Section 5 we discuss the results.

As the most widely employed design for clinical trials is the two-group comparison design, we consider exclusively the case of two therapies, denoted by the letters A and B, where approximately equal sample sizes \( N_A \) and \( N_B \) are desired and \( AD_s = |N_A - N_B| \) represents the absolute value of the difference between the numbers of patients in two treatment groups after \( n \) assignments.

### 2. Permutated-block Randomization within Strata and the Hallstrom-Davis Model

Permuted-block design was the earliest attempt to protect against chance imbalances. It is probably the most commonly used randomization procedure in practice. The rationale is that each block has an equal number of assignments for each treatment, and the assignments are listed in random order. As each block is filled, the numbers of treatments assigned to each group are brought back into perfect balance. The method was generalized to incorporate strata. If stratification is used, blocking may be done within strata to ensure equal treatment groups within each stratum. However, the effectiveness of stratified blocked randomization is reduced, especially when the number of strata is large, or/and block sizes are too large in relation to the number of patients enrolled [2, 8–10]. As this will result with high probability in empty strata or poorly filled initial blocks in each stratum, the whole procedure of stratification may fail to achieve its basic aim and considerable imbalance across treatments for any factor and overall could still exist. Such imbalances appear awkward and may lead to some loss of credibility for the trial, especially for persons not oriented to statistics.

In planning a trial, the risk of the aforementioned disadvantage must be determined to decide whether “trouble” will be present with high probability at the end of the study. The results of Hallstrom and Davis are very important for this to be achieved and should be used. They derive the variance for the difference in the number of patients assigned to each group as a function of the block size and the number of strata. If patients are to be assigned to treatment groups in \( k \) strata, \( b_i \) is the block size in the \( i \)th stratum and \( N_i \) is the number of assignments issued in the last block of stratum \( i \), then the overall variance of the difference is

\[
\text{var}D = \frac{\sum_{i=1}^{k} (b_i + 1)/6}{\sum_{i=1}^{k} b_i^2 - 1}, \tag{1}
\]

where \( E(N_i) \) is the expected value of the number of patients in the \( i \)th stratum in the last block ([7], page 376).

Of course, Formula 1 can be used if and only if something is known about the distribution of the \( N_i \), more precisely, the expected value of \( N_i \). For this, Hallstrom and Davis made two assumptions and defined two special cases. First, if the expected number of patients in each stratum is large relative to the block size, it is reasonable to assume that the distribution of \( N_i \) is uniform over the integers 1 to \( b_i \). In this special case the overall variance is

\[
\text{var}D = \sum_{i=1}^{k} (b_i + 1)/6 \tag{Appendix}. \tag{2}
\]

Or secondly, if the expected number of patients in each stratum is small relative to the block size, the distribution of \( N_i \) is essentially binomial with parameters \( n \) and \( p_i \). Then the overall variance is

\[
\text{var}D = n \left( 1 - \frac{n - 1}{\sum_{i=1}^{k} b_i^2 - 1} \right) \tag{Appendix}. \tag{3}
\]

In Equation 3, \( n \) is the total number of patients, and \( p_i \) is the probability that a patient is enrolled in \( i \)th stratum of the trial.

We investigated the range of validity of both binomial and uniform distribution model.

### 3. Methods

#### 3.1 Simulation

For computation of probability distribution we used the formulae of special cases (2) and (3) introduced by Hallstrom and Davis. In a simulation study, the distribution of the absolute difference after \( n \) assignments \( AD_s = |N_A - N_B| \) ascertained by the theoretical model was compared to the distribution received after 5,000 simulation runs. The following conditions were used for the simulation:

- sample sizes were \( n = 100, 200, 300, 400 \) and 500,
- block sizes were \( b = 4, 6, 8, \) and 10,
- generally, a patient is associated to one of the strata with the same probability \( p = 1/\text{number of strata} \).

In a process of a stepwise increasing the number of strata, (i) for the uniform distribution of \( N_i \) the minimum number of patients within a block and (ii) for the binomial distribution of \( N_i \) the maximum number of patients within the block,
necessary for the validity of the model is determined.

3.2 Evaluation

The distribution of the absolute difference after \( n \) assignments \( AD_n \) ascertained by the theoretical model was compared to the distribution received after 5,000 simulation runs by using the Kolmogorov-Smirnov test. The validity of the model was rejected for a two-sided \( p \)-value less than 0.05.

4. Results

4.1 Range of Validity of Two Considered Distribution Models

For the five sample sizes \( n = 100, 200, 300, 400 \) and 500 and for different block sizes \( b = 4, 6, 8 \) and 10, results of the simulations are averaged and are presented in Figure 1. We found that for \( k > 35 \) strata (\( n = 100 \)), \( k > 72 \) strata (\( n = 200 \)), \( k > 105 \) strata (\( n = 300 \)), \( k > 145 \) strata (\( n = 400 \)) and \( k > 185 \) strata (\( n = 500 \)) likewise, the discrepancy between the uniform distribution model and the result of simulation study reached significance (\( b = 4 \)). Thus, at least 100/35 = 2.86 patients (\( n = 100 \)), 200/72 = 2.78 patients (\( n = 200 \)), 300/105 = 2.86 patients (\( n = 300 \)), 400/145 = 2.76 patients (\( n = 400 \)) and 500/180 = 2.78 patients (\( n = 500 \)), that means 2.81 patients on average have to be recruited per stratum to apply the uniform distribution model.

For binomial distributed \( N_i \) the validity of the model was lost for \( k < 72 \) strata (\( n = 100 \)), \( k < 140 \) strata (\( n = 200 \)), \( k < 215 \) strata (\( n = 300 \)) and \( k < 275 \) strata (\( n = 400 \)) and \( k < 350 \) strata (\( n = 500 \)) and therefore, at most 100/72 = 1.39 patients (\( n = 100 \)), 200/140 = 1.43 patients (\( n = 200 \)), 300/215 = 1.40 patients (\( n = 300 \)), 400/275 = 1.45 patients (\( n = 400 \)), 500/350 = 1.43 patients (\( n = 500 \)), that means 1.42 patients on average can be accrued per stratum.

To assess the probability of an overall imbalance outside the validity ranges of the investigated uniform and binomial distributions, it is necessary to refer to the result of a simulation study. The concerned ranges of blocks are presented in Figure 1 by the white boxes.

From Figure 1, we recognize that the ranges of validity of both binomial and uniform distribution models will be increased absolutely with the increasing size of the last block within a stratum. More feasibly, we want to express the range of validity in relation to block size (% patients per block size), and here validity is increased for binomial (36–57%), but decreased for uniform distribution (30–23%) for increasing block size.

In detail, for block size \( b = 4 \) validity is given if the permitted average maximum percentage of patients per stratum is at most 1.42 \( \times \) 100/4 = 35.5% of the block, whereas for \( b = 10 \) the percentage can be up to 57%. The model of uniform distribution works acceptably if the permitted average maximum percentage of patients per stratum is at least 70% (\( b = 4 \)), which means it works only for a remaining range of 30%. And for \( b = 10 \), the model is valid only for

Fig. 1 Validity of probability models as a function of block size and number of patients. Range of validity of binomial distribution model (check-box) and uniform distribution model (grey box) for block sizes of \( b = 4, 6, 8 \) and \( b = 10 \). Range of validity will be limited by the permitted average maximum number of patients per stratum (binomial distribution model) and by the permitted average minimum number of patients per stratum (uniform distribution model).
23% of block size, namely, if the permitted average maximum percentage of patients per stratum is at least 77%.

For determining the expected probability distributions of absolute difference of \( NA \) to \( NB \), the application of these results is advantageous.

### 4.2 Evaluation of Several Practical Situations

To illustrate the scope for application, we would like to consider three different constellations of an assumed trial, with respective probabilities for reaching or exceeding values of the absolute difference \( AD_n \geq 8 \) or \( AD_n \geq 10 \).

We consider a clinical trial where the total number of patients is \( n = 100 \). The three prognostic factors site (5 levels), severity (3 levels) and gender (2 levels) should be accounted for, resulting in \( k = 5 \times 3 \times 2 = 30 \) strata. Then for \( b = 6 \), neither of the two models is valid (100/30 = 3.3 patients are expected in average per stratum, \( \triangleright \) Fig. 1) and we have \( \text{Pr}(AD_n \geq 8) = 0.209 \) and \( \text{Pr}(AD_n \geq 10) = 0.111 \) from simulation (\( \triangleright \) Fig. 2, dashed line), which is obviously too high a risk. This could be decreased by modifying the trial characteristics for stratification, either by limiting the number of strata, or by decreasing block size, or both. Let’s consider both strategies in detail.

By limiting to the two most important prognostic factors (site and severity) with \( k = 5 \times 3 = 15 \) strata, uniform distribution model is valid and, hence, we will get smaller probability values \( \text{Pr}(AD_n \geq 8) = 0.056 \) and \( \text{Pr}(AD_n \geq 10) = 0.017 \). Further, by using only the most important factor (site), resulting in \( k = 5 \) strata, we will get more reduced probabilities \( \text{Pr}(AD_n \geq 8) < 0.001 \) and \( \text{Pr}(AD_n \geq 10) < 0.001 \). But in practice, a reduction of prognostic factors, as considered here hypothetically, is not trivial. So, if all three prognostic factors are essen-

---

**Fig. 2**
Probabilities \( \text{Pr}(AD_n \geq d) \) of \( AD_n = |NA - NB| \) for stratified blocked randomization with block size \( b = 6 \) in every stratum and several numbers of strata \( k = 30 \), \( k = 15 \), or \( k = 5 \) (\( n = 100 \) patients).

**Fig. 3**
Probabilities \( \text{Pr}(AD_n \geq d) \) of \( AD_n = |NA - NB| \) for stratified blocked randomization with \( k = 30 \) strata and different block sizes \( b = 4 \) and \( b = 6 \) (\( n = 100 \) patients).
tial and from the medical point of view
there is no way to decrease the number of
$k = 30$, the block size could be decreased to
$b = 4$ to reduce the probability of imbal-
ances. Then, the uniform model is valid and
we obtain for $b = 4$ \( Pr(AD_n \geq 8) = 0.11 \) and
\( Pr(AD_n \geq 10) = 0.046 \) (Fig. 3).

It is possible to refine these investiga-
tions. Let’s go back to the first example
\((n = 100 \) patients, three prognostic factors,
\( k = 30 \) strata). Assuming that we expect
most of the patients (about 80%) in only
20% of the strata, this means that 80% of
strata will incorporate only about 20% of
patients. By choosing $b = 6$ for the few
“crowded” strata, which contain 80% of pa-
ients, we can apply the uniform distribu-
tion model. In the remaining “sparse” stra-
ta we choose a smaller block size $b = 4$ and
can apply the binomial distribution model
(Fig. 1, 20/24 = 0.83). Then, for a com-
bined model with $b = 4$ and $b = 6$, respec-
tively, we obtain \( Pr(AD_n \geq 8) = 0.030 \) and
\( Pr(AD_n \geq 10) = 0.007 \) (Fig. 4, line with triangles),
instead of \( Pr(AD_n \geq 8) = 0.209 \) and
\( Pr(AD_n \geq 10) = 0.111 \) as in the case of
generally $b = 6$ (Fig. 4, dashed line).

Furthermore, if we include only two
prognostic factors with three and five lev-
els, respectively, resulting in $k = 3 \times 5 = 15$
strata, we would obtain clearly reduced
probabilities for the combined model by
\( Pr(AD_n \geq 8) = 0.016 \) and \( Pr(AD_n \geq 10) = 0.002 \), compared to the case of $b = 6$
(Fig. 4, line with dots). Remarkably,
notably reduced probabilities for any ab-
solute difference will be obtained by the
combined model of $k = 30$ strata com-
pared to $k = 15$ strata for $b = 6$, indicated by
the curve of the combined model of $k = 30$
strata being always below the curve of
$k = 15$ strata and block size $b = 6$.

Altogether, the probability distribution
of absolute difference of sample sizes $N_A$
and $N_B$ is determinable in advance of a trial
by the theoretical results of Hallstrom and
Davis [7], by our results of validity of special
cases, or by a simulation study. From this eval-
uation it is possible to assess for a specific practical situation if the dis-
advantage of considerable imbalance for
the trial as a whole will be present in reality.

4.3 Example

Case 1: Uniform distribution model used
in planning a trial

The Intramyocardial TransPlantation of
BonE MaRow stem Cells For ImprovE-
ment of Post-Infarct MyoCardial Regen-
eraTion in Addition to CABG Surgery
(PERFECT) study [11] was started in 2009
in Rostock, Germany. This study aims to
provide proof of the quality, efficacy, and
safety of cardiac stem cell therapy. For ran-
domization, the permuted-block design
within strata should be used [12]. The ran-
domization procedure should be stratified
by study site. 142 participating patients are
due to be enrolled from three study centers.
Thus, 142/3 = 47.3 patients in average per
stratum are expected. For all block sizes $b = 4, 6, 8$ and 10 in every stratum the uniform
distribution model is valid. The variance of
an overall imbalance for, say, $b = 6$, is by

\[
\text{Formula 2 } \sigma^2 = \frac{D(n)}{N} + \frac{W(n)}{N^2}
\]

where $D(n)$ is the absolute difference at
the trial as a whole will be present in reality.

4.3 Example

Case 1: Uniform distribution model used
in planning a trial

The Intramyocardial TransPlantation of
BonE MaRow stem Cells For ImprovE-
ment of Post-Infarct MyoCardial Regen-
eraTion in Addition to CABG Surgery
(PERFECT) study [11] was started in 2009
in Rostock, Germany. This study aims to
provide proof of the quality, efficacy, and
safety of cardiac stem cell therapy. For ran-
domization, the permuted-block design
within strata should be used [12]. The ran-
domization procedure should be stratified
by study site. 142 participating patients are
due to be enrolled from three study centers.
Thus, 142/3 = 47.3 patients in average per
stratum are expected. For all block sizes $b = 4, 6, 8$ and 10 in every stratum the uniform
distribution model is valid. The variance of
an overall imbalance for, say, $b = 6$, is by

\[
\text{Formula 2 } \sigma^2 = \frac{D(n)}{N} + \frac{W(n)}{N^2}
\]

where $D(n)$ is the absolute difference at
Case 2: Binomial distribution model used in planning a trial
Referring to the PERFECT study again, but assuming now seven instead of three study centers to shorten the recruitment phase, an additional prognostic factor of importance, the "ejection fraction", with levels "low" and "high", should be included. This results in \( k = 7 \times 2 = 14 \) strata, leading to an expectation of about 10 patients in average per stratum, and to the application of the uniform distribution model for estimating the probability of an overall imbalance. (The estimated probabilities are \( \text{Pr}(\text{AD}_d \geq 6) = 0.191 \) and \( \text{Pr}(\text{AD}_d \geq 9) = 0.05 \), considering \( b = 8 \). We will come back to these probabilities in the following case 3.

Let's assume, that in our example trial, an interim analysis is planned after a quarter of patients (\( n = 35 \)) have been recruited, for estimating the degree of efficacy of stem cell therapy compared to control. Assuming at this time point \( 35/14 = 2.5 \) patients on average per stratum, the binomial distribution model could be used for \( b = 6, 8, 10 \), or a simulation study for \( b = 4 \), to estimate the probability of expected imbalance at this time point. Here, we will get for \( b = 6, p_1 = 1/14, k = 14 \) by
\[
\text{Formula 3 } \text{var } D = 35 \times (1 - (34 \times 14/ (5 \times 142))) = 18, \text{ a standard deviation of } 4.24 \text{ and } \text{Pr}(\text{AD}_d \geq 9) = 0.034.
\]

Besides a planned interim analysis, binomial distributed differences could also occur in trials, if in the first instance – during a pre-study phase – responder patients are identified, to reassign therapies later on during the subsequent trial, as happened in the CAPS study [13]. In this case of a two-study-phases design, the number of patients per stratum is hardly, first, to be estimated in advance and, second, to be increased just before starting the subsequent trial. But nevertheless, a randomization process has to be planned, and Formula 3 could be used in planning to evaluate the anticipated effect of different choices for \( k \) and \( b \) on the overall balance in assignments.

Case 3: Combination of uniform and binomial models used in planning a trial
In a trial like the PERFECT study the situation could arise that patients are expected from seven sites, which differ considerably in their recruitment rates. Let’s assume four sites which are expected to contribute only about 15% of all patients (\( n = 21 \)), and three sites which are thus able to provide the remaining 85% (\( n = 121 \)). As two levels of ejection fraction are considered for stratification here too, we will receive \( k = 14 \) strata in total as in case 2, but this time dividing into \( k = 2 \times 4 = 8 \) “sparse” strata (57%) comprising few patients, and \( k = 2 \times 3 = 6 \) “crowded” strata (43%), respectively. For a blocking size of \( b = 8 \), we can estimate the probability of the total imbalance of “sparse” strata by using the binomial model (\( 21/8=2.6 \) patients per stratum, \( \text{var } D = 13.5, \text{SD} = 3.67, \text{Pr}(\text{AD}_d \geq 26) = 0.102 \)), and of “crowded” strata by using the uniform model (\( 121/6 = 20.2 \) patients per stratum, \( \text{var } D = 9, \text{SD} = 3, \text{Pr}(\text{AD}_d \geq 6) = 0.046 \)). Weighting both probabilities by their fraction of strata, for the total imbalance of the combined model consisting of a binomial and a uniform part this results in \( \text{Pr}(\text{AD} \geq 26) = 0.078 \), which is just about a third of \( \text{Pr}(\text{AD} \geq 6) = 0.191 \) from case 2 (uniform model for the whole trial).

If, in addition, the blocking factor were reduced from 8 to 6 for “sparse” strata, we could still apply the binomial model (\( \text{var } D = 10.5, \text{SD} = 3.24, \text{Pr}(\text{AD} \geq 6) = 0.064 \)), and further reduce the probability for the total imbalance of the combined model by about 2%, from 0.078 to \( \text{Pr}(\text{AD} \geq 6) = 0.056 \).

The application of a binomial-uniform combination of probability models in a trial as described here enables us to model the differences and, respectively, to estimate the probability of the total imbalance in a more adequate manner, compared to a less differentiated consideration using only one probability model for all strata. By modifying the choice of \( b \) (and/or \( k \), if possible) referring to this estimation, we can impact the probability of risk where necessary.

5. Discussion and Conclusions
Randomized controlled trials set the methodological standard of excellence in medical research. The key word is “randomized”, and this must be done properly.

If in a clinical trial prognostic factors are known in advance to be associated with the outcome of a patient, it is often worthwhile that the randomization for a clinical trial should be stratified on these factors, particularly in a multi-center trial. The most well-known and practical stratified randomization method used is permuted-block randomization within strata. The main disadvantage of this method is that, in certain situations, there is a significant risk that imbalance will still occur to a considerable extent for the trial as a whole.

For a specific practical situation it needs to be decided in advance whether or not the aforementioned “trouble” will with high probability be present at the end of the study. To be able to make a competent decision here, the results of Hallstrom and Davis [7] are of relevance. By using a particular part of their results, the probability of observing a difference in absolute value of unacceptable magnitude can be calculated. If this probability is small enough not to be of concern, permuted-block randomization within strata could be carried out as planned. Calculation can be done only by using two special cases, the range of validity of which has been unknown until now.

We analyzed the validity of two special cases by computer simulations, varying a number of trial characteristics to assess their effects on balance. It is hardly surprising that the validity of special cases was not given for any constellation of a trial characterized by the total number of patients \( n \), the block size \( b \), and the number \( k \) of strata. Figure 1 reveals a non-marginal range of about 20–34% of considered block sizes, for which none of the two available distribution models is valid. Here, the implementation of a simulation study is necessary for calculating probability distribution.

On the other hand, a sizeable range of block size, representing the range of validity of available distribution models, is given. Depending on block size, the model of binomial distribution is valid for a permitted average maximum number of patients per stratum between at most 36–57% of considered block size, whereas the model of uniform distribution works adequately from at least 70% of patients per last block.

For trials with an expected fraction of “crowded” and a fraction of “sparse” filled strata we are able to account for such special circumstances by using a combined...
model consisting of different parts (binomial, uniform or simulation) for the calculation of the probability of an overall imbalance of the trial. We demonstrated that the application of a combined model, if reflecting circumstances of a trial more appropriately, gives more unbiased estimates for the risk of imbalance than less differentiated modelling. So, if details are known about the expected number of patients per stratum, it is recommended that this information be realized by applying a combined model when planning a trial.

In summary, regarding the risk of the appearance of an intolerable overall imbalance, our results are important for the evaluation of permuted-block randomization within strata in a more accurate way.

References

Appendix
1. Distribution of $N_i$ is uniform over the integers 1 to $b_i$. It is

$$E(N_i) = \frac{b_i}{2} \sum_{r=1}^{b_i} p(N_i = r) = \frac{1}{b_i} \times \frac{b_i}{2} \sum_{r=1}^{b_i} r = \frac{1}{b_i} \times \frac{(b_i + 1) \times b_i}{2} = \frac{b_i + 1}{2}$$

and

$$E(N_i^2) = \frac{b_i}{2} \sum_{r=1}^{b_i} r^2 = \frac{1}{b_i} \times \frac{b_i}{2} \times (b_i + 1) \times (2 \times b_i + 1) = \frac{(b_i + 1) \times (2 \times b_i + 1)}{6}$$

Then for Equation 1 we obtain

$$\text{var } D = \sum_{i=1}^{k} \left[ b_i \cdot E(N_i^2) - E(N_i^2) \right] / (b_i - 1) = \sum_{i=1}^{k} \left[ b_i \times \frac{b_i + 1}{2} - \frac{(b_i + 1) \times (2 \times b_i + 1)}{6} \right] / (b_i - 1)$$

$$= \sum_{i=1}^{k} \left[ \left( \frac{b_i + 1}{2} \right) \times \left( b_i - \frac{2 \times b_i + 1}{3} \right) \right] / (b_i - 1) = \sum_{i=1}^{k} \left( \frac{b_i + 1}{2} \right) \times \left( \frac{b_i - 1}{3} \right) / (b_i - 1) = \sum_{i=1}^{k} \frac{b_i + 1}{6}$$

2. Distribution of $N_i$ is binomial with parameters $n$ and $p_i$. Then it is valid

$$E(N_i^2) = E(N_i^2) - E(N_i) + E(N_i) = E(N_i^2 - N_i) + E(N_i) = E(N_i \times (N_i - 1)) + E(N_i)$$

Furthermore, we have

$$E(N_i \times (N_i - 1)) = \sum_{r=0}^{\infty} r \times (r-1) \times \binom{n}{r} \times p_i^r \times (1 - p_i) = \sum_{r=0}^{n} n \times (n-1) \times \binom{n-2}{r-2} \times p_i^r \times (1 - p_i)$$

$$= n \times (n-1) \times p_i^2 \times \sum_{r=2}^{n} \binom{n-2}{r-2} \times p_i^{r-2} \times (1 - p_i) = n \times (n-1) \times p_i^2 \times (p_i + (1 - p_i))^{r-2}$$

$$= n \times (n-1) \times p_i^2$$

© Schattauer 2012
Then it is

$$E(N_i^2) = n \times x(n-1) \times p_i^2 + n \times p_i = n^2 \times p_i^2 - n \times p_i^2 + n \times p_i = n \times p_i \times (1-p_i) + (n \times p_i)^2$$

And we obtain for Formula 1

$$\text{var } D = \sum_{i=1}^{k} \left[ b_i \times E(N_i) - E(N_i^2) \right] / (b_i - 1) = \sum_{i=1}^{k} \left[ b_i \times n \times p_i \times (n \times p_i \times (1-p_i) + (n \times p_i)^2) \right] / (b_i - 1)$$

$$= \sum_{i=1}^{k} \left[ n \times p_i \times (b_i - (1-p_i) - n \times p_i) / ((b_i - 1)) \right] = \sum_{i=1}^{k} \left[ n \times p_i \times (b_i - 1) + n \times p_i \times (p_i - n \times p_i) \right] / b_i - 1$$

$$= n + \sum_{i=1}^{k} \frac{n \times p_i \times (1-n)}{b_i - 1} = n - \sum_{i=1}^{k} \frac{n \times p_i^2 \times (n-1)}{b_i - 1} = n \times \left( 1 - (n-1) \times \frac{\sum_{i=1}^{k} p_i^2}{\sum_{i=1}^{k} b_i - 1} \right)$$