# The Risk of Imbalance in Treatment Assignments in Stratified Blocked Randomized Clinical Trials

Bingzhi Zhang* Hui Quan * Lin Wang* Lynn Wei*

**Abstract**

In clinical trials, the stratification and blocking method is frequently used to achieve balance in prognostic factors across treatment groups. However, such randomization strategy has a disadvantage that may lead to overall imbalance in sample size across treatment groups when the number of strata is large. To measure the risk of such imbalance, Hallstrom and Davis (1998) provided the variance of the sample size difference between two treatment groups. However, the distribution assumptions made in their method has limited its validity and applicability. This limitation was confirmed by the simulation studies conducted by Kundt and Glass (2012). Kundt and Glass then proposed utilizing simulation to assess the likelihood of imbalance inclinical trials where Hallstrom and Davis assumptions are no longer valid. In this paper, we proposed a new method deriving the exact probability density function (pdf) of the sample size difference between any two treatment groups in any multiple-arm and stratified blocked randomized clinical trials. It will help the clinical trial designers choose an appropriate stratification strategy at the planning stage to avoid serious imbalanced trials at the end.

**Key Words:** Stratification, blocking, imbalance, randomized clinical trial

## 1. Introduction

The stratification and blocking method is widely used in clinical trial randomization to achieve similar patient characteristics across different treatment groups and near balance in sample sizes within each stratum. However, when the number of strata is large, it is likely to have many incomplete blocks. For example, a dose-ranging study plans to enroll 500 patients from 25 different countries and each patient will be assigned to one of the five treatment groups with equal probability. Country and a two-level prognostic variable are chosen as stratification factors and the block size is five. In an extreme case, we could have 50 incomplete blocks which may jeopardize the overall balance in sample sizes across treatment groups since the target sample size of each group is merely 100. Hence, it is of interest to measure the risk of overall imbalance in sample sizes due to the choices of stratification factor at the time we design a clinical trial.

Let us denote the difference of sample size between two treatment groups as D and the number of patients in the last block of the $i$th stratum by $n_i$. For a study with a balanced design, the expectation of $D$ is zero. Hence, Hallstrom and Davis proposed using the variance of $D$ to measure the risk of imbalance. They derived the formula to calculate the variance of $D$ under two special conditions: one assumes that $n_i$ follows a uniform distribution and the other one assumes that $n_i$ follows a binomial distribution. They argued that the uniform distribution assumption is valid only when the expected sample size in each stratum is large and that the binomial distribution assumption is valid only when the expected sample size in each stratum is small compared to the block size. To investigate the range of validity of Hallstrom and Davis assumptions, Kundt and Glass performed a series of simulation

*Sanofi, 55 Corporate Drive, Bridgewater, New Jersey 08807

studies. In addition, for scenarios not satisfying the above two assumptions, Kundt and Glass suggested evaluating the imbalance using the empirical distribution of $D$.

Both methods have limitations. First, Hallstrom and Davis method is only applicable in the two special cases described above and only provides the variance of $D$ rather than the distribution of $D$. Second, Kundt and Glass' conclusions were based on simulation results from limited scenarios, thus, may be lack of generalizability. In addition, both studies have been conducted in a two-arm trial setting.

To overcome the above limitations, we derived the *pdf* of $D$, which is applicable in any multiple-arm, stratified and blocked randomized clinical trials. This is done by the following two steps. First, we derived the *pdf* of $n_i$. Using this *pdf*, we can easily verify whether Hallstrom and Davis assumptions are satisfied without conducting any simulations. Then, we derived the conditional distribution of $D$ given $n_i$. Combining these two steps, we derived the *pdf* of $D$.

## 2. Methods

To introduce our method, we first need to extend our notations to the multiple treatment group setting. Lets assume $N$ patients are assigned to $A$ treatment groups in $S$ strata with pre-specified probabilities. And let's denote the number of subjects in the $i$th stratum by $N_i$, $i = 1, \cdots, S$. Among the multiple treatment groups, it is usually of interest to perform pairwise comparison between the placebo group and one of the active treatment groups. Let's denote $P_i$ the number of patients on placebo in the last block of the stratum $i$, and $T_i$ the number of patients in the considered active treatment group in the last block of the stratum $i$. Then the sample size difference of the two groups within stratum $i$ is $D_i = T_i - P_i$, and $D = \sum_{i=1}^{S} D_i$ is the total sample size difference of the entire study.

### 2.1 Distribution of $n_i$, number of subjects in the last block of stratum

The size of each stratum $N_i$ largely depends on the randomization strategy, stratification factors, and also the patient population. For example, if we stratify the subjects by country/site and specify a recruitment target for each country/site, the stratum size will be more of a fixed target than a random number which is close to but may not be the same as the target. On the other hand, if we adopt the competitive enrollment strategy to recruit enough patients in time, the sample size in each site will be a random number depending on factors such as disease prevalence in that area. If we stratify the subjects by certain prognostic factors, such as age, gender or biomarkers, the sample size in each stratum will also be a random number if we recruit patients freely. In this paper, we only consider the latter cases where the stratum size is random. We also assume that $(N_1, N_2, \cdots, N_S)$ follows multinomial distribution with total sample size $N$ and probabilities $(p_1, \cdots, p_s)$. The total sample size $N$ usually equals to or has a target of $K \cdot B$ for some positive integer K. Here we only consider the ideal case where $N = K \cdot B$. This assumption is slightly in favor of the balance across treatments but it still has its generalizability. In the case that $N$ is not a multiple of $B$, it can be written as $N = K \cdot B + b$ for some positive integer $b$ less than $B$. The additional sample size difference between two treatment groups brought in by the residual $b$ will be very limited. Following the above assumptions, the joint probability function of $\mathbf{n} = (\mathbf{n}_1, \cdots, \mathbf{n}_S)$ can be

written as

$$\begin{aligned} & P(\mathbf{n}_1 = n_1, \cdots, \mathbf{n}_S = n_s) \\ = & \sum_{(k_1,k_2,\cdots,k_s)} P(N_1 = k_1 B + n_1, \cdots, N_s = k_s B + n_s). \end{aligned}$$

$n_i$, $i = 1, \cdots, S$, takes integer values from 0 to $B-1$. With this notation, we say that the number of patients in the last block is 0 if there is no incomplete block in that stratum. The summation of $n_i's$ is a multiple of the block size $B$. And $\mathbf{k} = (k_1, \cdots, k_S)$ is a vector of integer elements indicating the number of full blocks in each stratum, which satisfies the following condition

$$\sum_{i=1}^{S} k_i = K - \left( \sum_{i=1}^{S} n_i \right) / B,$$

so that the total sample size is $N = K \cdot B$.

It follows that $N_i$ has marginal binomial distribution $\mathrm{B}(N, p_i)$. And the *pdf* of $n_i$ is

$$P(\mathbf{n_i} = n_i) = \begin{cases} \sum_{k_i=0}^{K-1} P(N_i = k_i B + n_i) & \text{if } n_i = 1, \cdots, B-1, \\ \sum_{k_i=0}^{K} P(N_i = k_i B) & \text{if } n_i = 0, \end{cases}$$

where

$$P(N_i = x) = \binom{N}{x} p_i^x (1-p_i)^{N-x}, \quad i = 1, \cdots, S,$$

for any $x = 0, \cdots, N$. Given $p_i$, the distribution of $n_i$ is independent of $p_j$, $j \neq i$.

For readers to get an idea of the distribution of $n_i$ in different scenarios, the *pdf* plots of $n_i$ are provided in Figure 1. Here we assume that $N$ subjects are assigned to 5 treatment groups with block size 5 and $N$ varying from 5 to 600. The probability $p_i$ of a subject belonging to the $i$th stratum is chosen from $0.01, 0.05, 0.1, 0.2$. In each plot, the probabilities of $n_i = 0, 1, 2, 3, 4$ were given with varying total sample size $N$. In reality, for cases where the probability of belonging to a stratum is 0.01 and the total sample size is 5 or 50, the stratum will not be considered. But here we still consider these scenarios because they may still happen in reality due to unpredictable reasons and we want to evaluate Hallstrom and Davis assumption.

To compare the true *pdf* with Hallstrom and Davis' binomial assumptions, we also provide the *pdf* plots of $n_i$ (Figure 2) over values from 0 to 4, assuming it follows binomial distribution with parameter $N$ and $p_i$, as references. With binomial assumption and fixed $p_i$, the probability of $n_i \leq 4$ is close to 1 only when $N_i$ is small, and it decreases as $N_i$ increases. From the *pdf* plots, it is obvious that the binomial assumption does not hold when $N_i$ is big, as only a small amount of probability is allocated on the possible values of $n_i : 0, 1, 2, 3, 4$.

The difference between the real *pdf* of $n_i$ and the uniform distribution can be easily seen by looking at Figure 1 alone. It shows that when the expected number of subjects $N \cdot p_i$ in a stratum is greater than the block size 5, the distribution of subjects in the last block is very close to uniform.

And by comparing Figure 1 and 2, we can see that for the cases where $N \cdot p_i$ is relatively small compared to the block size, for example when $N \cdot p_i = 0.05, 0.5, 2.5,$

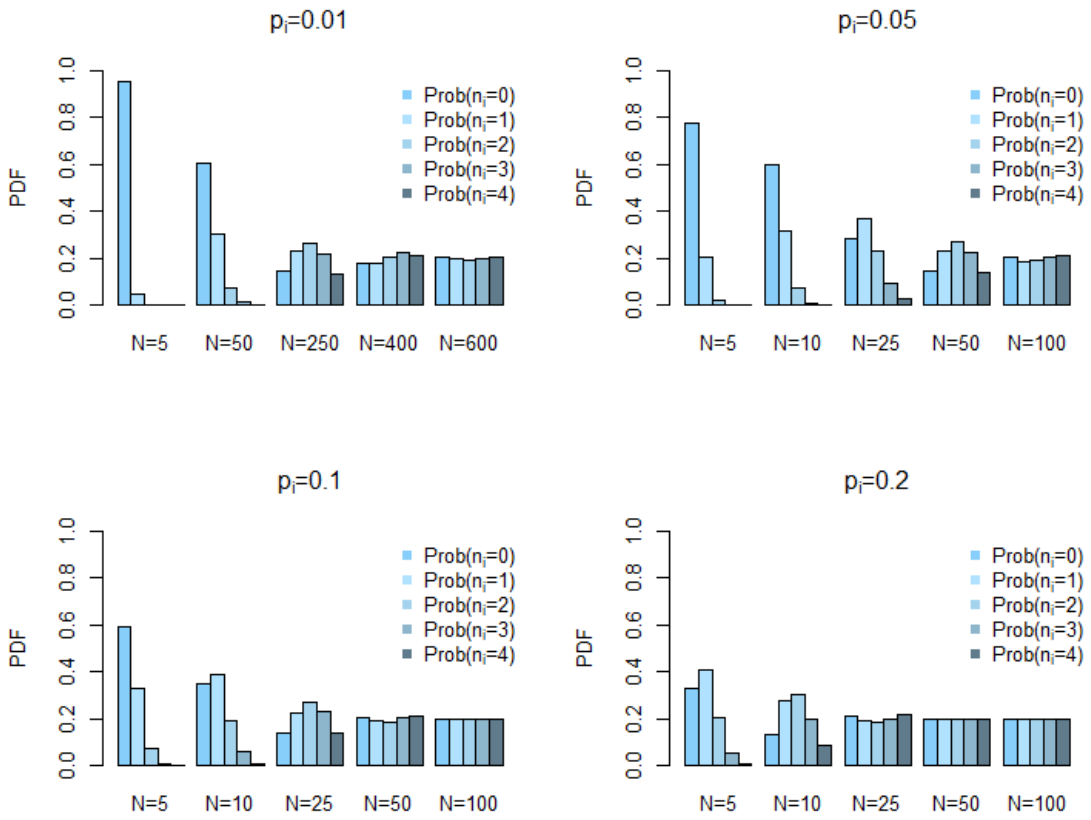


**Figure 1**: True *pdf* of $n_i$

the distribution of $n_i$ is almost exactly same as the binomial distribution with parameters $N$ and $p_i$. In the cases where $N \cdot p_i$ is smaller than but close to the block size, the distribution of $n_i$ is neither binomial nor uniform. For example, in the scenario where the total sample size is 400 and the probability of subjects belonging to the stratum is 0.01, the expected number of subject in this stratum is 4 and the distribution of $n_i$ in this stratum is neither close to binomial nor to uniform. For such cases, the formula of $Var(D)$ provided by Hallstrom and Davis is not applicable. Instead, we could calculate the variance and *pdf* of $D$ as shown in the next section.

### 2.2 Distribution of $D$, difference in sample sizes between two groups

Lets first consider trials where the treatment assignments within each stratum are balanced within each block. In such case, the block size $B$ is a multiple of $A$ and it can be written as $B = A \cdot m$, where $m$ is a positive integer. Given $n_i$, we have the conditional probability of $D_i = d_i$, $d_i >= 0$, as

$$
\begin{aligned}
&P(D_i = -d_i|n_i) = P(D_i = d_i|n_i) \\
&= \begin{cases} \left[\sum_{k=0}^{m-d_i} C_{(A-2)m}^{n_i-(d_i+2k)} C_m^{d_i+k} C_m^k\right] / C_B^{n_i}, & \text{if } d_i \le min(m, n_i) \\ 0, & \text{other.} \end{cases}
\end{aligned}
$$

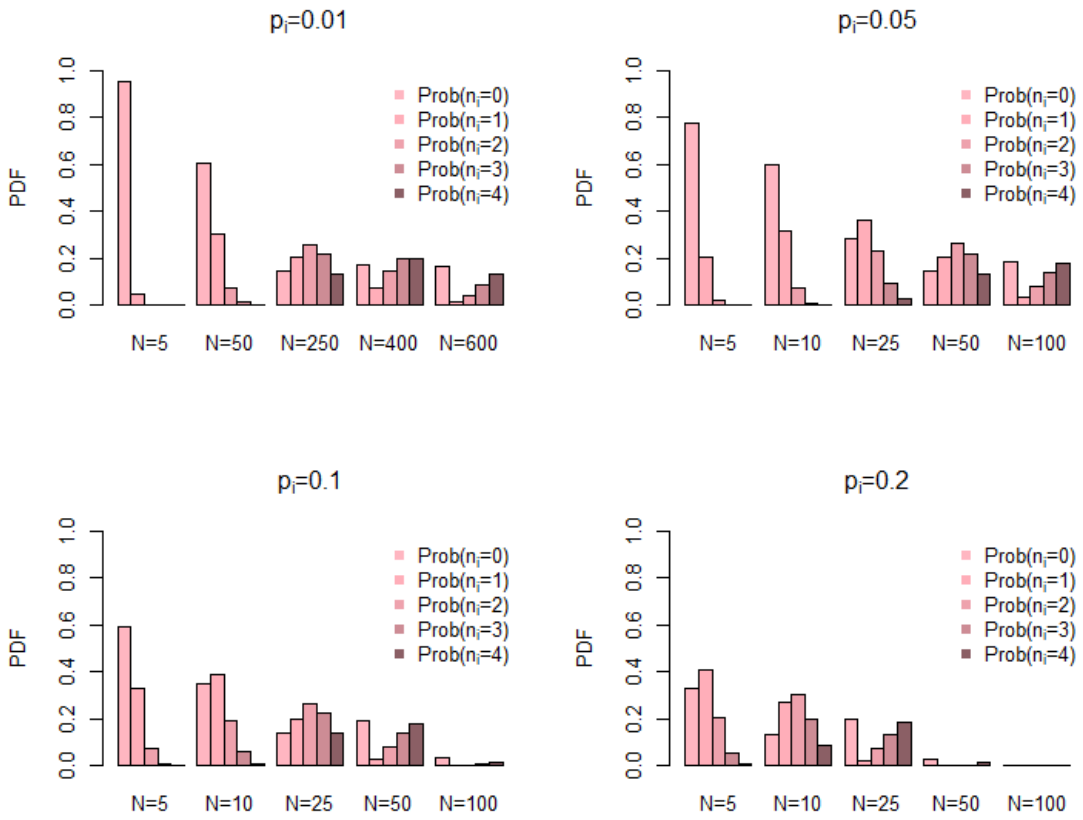


**Figure 2**: *pdf* of $n_i = 0, \cdots, 4$ assuming $n_i$ follows $Bin(N_i, p_i)$

The expected value of $D_i$ is $E(D_i) = E(D_i|n_i) = 0$ and $E(D) = \sum_{i=1}^{S} E(D_i) = 0$. So the variance of $D$ is

$$\begin{aligned} Var(D) &= E(\mathrm{Var}(\sum_{i=1}^{S} D_i|(n_1, n_2, \cdots, n_S))) \\ &= E(\sum_{i=1}^{S} \mathrm{Var}(D_i|n_i)) = \sum_{i=1}^{S} \mathrm{Var}(D_i) \\ &= \sum_{i=1}^{S} \sum_{j=1}^{B-1} \left[ P(n_i = j) \sum_{d_i=1}^{m} d_i^2 \cdot 2P(D_i = d_i|n_i = j) \right]. \end{aligned}$$

Further we could also write the *pdf* of $D$ as

$$P(D = d) = \sum_{\mathbf{n}:\mathrm{S}(\mathbf{n}) \geq d} \left[ P(\mathbf{n}) \cdot \sum_{\mathbf{d}:\mathrm{S}(\mathbf{d})=d} \prod_{i=1}^{S} P(d_i|n_i) \right], \tag{1}$$

where $\mathbf{n} = (n_1, n_2, \cdots, n_S)$, $\mathbf{d} = (d_1, d_2, \cdots, d_S)$ and $\mathrm{S}(\mathbf{n})$, $\mathrm{S}(\mathbf{d})$ are the summations of $\mathbf{n}$ and $\mathbf{d}$ correspondingly.

The above formula can be easily generalized to accommodate the clinical trials with imbalanced design. For example, let us assume the ratio of probabilities of assigning patients to arm P and arm T is 1: 2 and there are $m$ and $2m$ patients in

**Table 1**: *pdf* and *cdf* of $|D|$

| $\mid D \mid$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| pdf | 0.20877 | 0.36346 | 0.24413 | 0.12256 | 0.04621 | 0.01234 | 0.00227 | 0.00025 | 0.00001 |
| cdf | 0.20877 | 0.57223 | 0.81636 | 0.93892 | 0.98513 | 0.99747 | 0.99974 | 0.99999 | 1.00000 |

arm P and T in each block correspondingly. In this case, given $n_i$, the conditional probability of $D_i = d_i$, $d_i >= 0$, is

$$P(D_i = d_i|n_i) = \begin{cases} \left[\sum_{k=0}^{min(2m-d_i,m)} C_{B-3m}^{n_i-(d_i+2k)} C_{2m}^{d_i+k} C_m^k\right] / C_B^{n_i}, & \text{if } d_i \leq min(n_i, 2m) \\ 0, & \text{other.} \end{cases}$$

and

$$P(D_i = -d_i|n_i) = \begin{cases} \left[\sum_{k=0}^{m-d_i} C_{B-3m}^{n_i-(d_i+2k)} C_{2m}^{k} C_m^{d_i+k}\right] / C_B^{n_i}, & \text{if } d_i \leq min(n_i, m) \\ 0, & \text{other.} \end{cases}$$

And the *pdf* of D is in the same form as (1), except that the conditional distribution is replaced with the ones derived above.

### 2.3 Example

In this section, we provide an example to examine the risk of imbalance by implementing our method. Let us assume 24 patients are assigned with equal probability to 3 treatment groups, $P$, $T_1$, and $T_2$, in 8 strata with block size 3. We also assume that patients have equal probability to be from each strata. We are interested in the potential imbalance between the placebo group $P$ and treatment group $T_1$. Following formula (1), the *pdf* and cumulative density function (*cdf*) of the absolute difference $|D|$ between number of patients in P and $T_1$ are summarized in Table 1. The probability of having $|D| \geq 3$ is about 18.4% when the expectation of the sample size in each treatment group is 8. Compared to the sample size, the difference of 3 is not negligible. With the knowledge about the risk of imbalance before the trial, trial planner could modify the randomization strategy to find the balance point between the balance in sample size across treatment groups and the balance of prognostic factors.

## 3. Results and Discussion

The *pdf* of $D$ provides us a complete view on the risk of imbalance in treatment assignment for any multiple-arm stratified and blocked randomized trials. It can deal with cases that are beyond the valid range of Hallstrom and Daviss assumptions. We note that this formula can be applied to all cases, however, the computation time will increase significantly for trials with big sample size and large number of strata as the computation will have to go through all combinations of $N_i's$ and $n_i's$.

On the other hand, the variance of $D$ can be calculated in much less time as the formula is largely simplified by taking advantage of conditional independence of $D_i's$ given $n_i's$. So it can serve as a simple and convenient alternative measure of the risk of imbalance.

## REFERENCES

Hallstrom, Al and Davis, K. (1988), "Imbalance in Treatment Assignments in Stratified Blocked Randomization," *Controlled Clinical Trials* 9, 375-382.

Kundt, G. and Glass, A. (2012), "Evaluation of Imbalance in Stratified Blocked Randomization - Some Remarks on the Range of Validity of the Model by Hallstrom and Davis," Methods Inf Med 51, 55-62.