Impact of Stratified Randomization in Clinical Trials

Vladimir V. Anisimov

Abstract This paper deals with the analysis of randomization effects in clinical trials. The two randomization schemes most often used are considered: unstratified and stratified block-permuted randomization. A new analytic approach using a Poisson-gamma patient recruitment model and its further extensions is proposed. The prediction of the number of patients randomized in different strata to different treatment arms is considered. In the case of two treatments, the properties of the total imbalance in the number of patients on treatment arms caused by using stratified randomization are investigated and for a large number of strata a normal approximation of imbalance is proved. The impact of imbalance on the power of the trial is considered. It is shown that the loss of statistical power is practically negligible and can be compensated by a minor increase in sample size. The influence of patient dropout is also investigated.

1 Introduction

The properties of various types of randomization schemes are studied in the papers Hallstrom and Davis (1988), Lachin (1988), Matts and Lachin (1988), and books by Pocock (1983), Rosenberger and Lachin (2002). However, the impact of randomness in patient recruitment and the prediction of the number of randomized patients in the case of multiple centres have not been fully investigated.

To investigate these phenomena, a new analytic approach using a Poisson-gamma patient recruitment model developed in Anisimov and Fedorov (2006, 2007) is proposed. The model accounts for the variation in recruitment over time and in recruitment rates between strata. The prediction of the number of patients randomized in different strata to different treatment arms is considered. In the case of two treat-

---
Prof. Vladimir V. Anisimov
Research Statistics Unit, GlaxoSmithKline, New Frontiers Science Park (South), Third Avenue, Harlow, Essex, CM19 5AW, United Kingdom, e-mail: Vladimir.V.Anisimov@gsk.com

ments, the properties of the total imbalance in the number of patients randomized to different treatment arms caused by using stratified randomization are investigated as well. For a large number of strata a normal approximation of imbalance is proved. These results are used for investigating the impact of randomization on the power and sample size of the trial. Note that in a special case of a centre-stratified randomization some results in these directions are obtained in Anisimov (2007). The effect of patient dropout is also considered. These results form the basis for comparing randomization schemes using combined criteria including statistical power, study costs, drug supply costs, etc.

2 Recruitment in Different Strata

Consider a multicentre clinical trial carried out with the aim to recruit in total n patients. Suppose that the patient population is divided into S strata. Strata can stand for different countries, centres or regions, groups of population specified by some covariates, etc. Upon registration, patients are randomized to one of the treatment arms according to some randomization scheme. The recruitment is stopped when the total number of recruited patients reaches n. Assume that the patients in different strata are recruited independently. Accounting for a natural variation in recruitment between strata, we can consider the following model: the recruitment in s-th stratum is described by a Poisson process with rate μ_s, where μ_s is viewed as a realization of a gamma distributed variable with parameters $(\alpha N_s, \beta)$ (shape and rate parameters), and the values N_s reflect the sizes of strata. Denote $N = \sum_s N_s$.

As a natural illustration of this model, assume that there are N clinical centres divided among S regions, where a region s has N_s centres. Let us associate the region s with s-th stratum. Suppose that the recruitment in centres is described by a Poisson-gamma model (Anisimov and Fedorov, 2006,2007): in centre i the patients are recruited according to a Poisson process with rate λ_i, where $\{\lambda_i\}$ are viewed as a sample from a gamma distributed population with parameters (α, β). Then the recruitment in s-th region is described by a Poisson process with rate μ_s which is gamma distributed with parameters $(\alpha N_s, \beta)$. For this case, in Anisimov and Fedorov (2007) a ML-procedure for estimating parameters is proposed.

Consider now the prediction of the total number of patients n_s recruited in a particular strata s. The variable n_s has a mixed binomial distribution with parameters (n, g_s) where $g_s = \mu_s/\mu$, $\mu = \sum_{s=1}^{S} \mu_s$. Thus, μ has a gamma distribution with parameters $(\alpha N, \beta)$ and g_s has a beta distribution with parameters $(\alpha N_s, \alpha(N - N_s))$. Denote by $\mathscr{B}(a,b)$ a beta function. Then n_s has a beta-binomial distribution and $\mathbf{P}(n_s = k) = P(n, N, N_s, \alpha, k)$, where

$$P(n, N, N_s, \alpha, k) = \binom{n}{k} \frac{\mathscr{B}\Big(\alpha N_s + k, \alpha(N - N_s) + n - k\Big)}{\mathscr{B}\Big(\alpha N_s, \alpha(N - N_s)\Big)}, \; k = 0, .., n. \quad (1)$$

3 Randomization Effects

Description of randomization schemes can be found in the books by Pocock (1983), Rosenberger and Lachin (2002). Consider the two often used in clinical trials randomization schemes: unstratified and stratified block-permuted randomization. Unstratified randomization means that the patients registered for the study are randomized to treatment arms according to the independent randomly permuted blocks of a fixed size without regard to stratum. Stratified randomization means that the patients are randomized according to randomly permuted blocks separately in each stratum. Clearly, unstratified randomization minimizes the imbalance in the number of patients on treatment arms for the whole study, but in general is likely to increase the imbalance within each stratum compared to stratified randomization.

Assume that there are K treatments with the allocations $(k_1,..,k_K)$ within a randomly permuted block of a size $K_1 = \sum_{j=1}^{K} k_j$. Denote by $n_s(j)$ the number of patients randomized to treatment j in s-th stratum.

Consider first an unstratified randomization. Assume that the value $M = n/K_1$ is integer. Then there are Mk_j patients on treatment j and all patients can be divided into K groups with Mk_j patients in group j, $j = 1,..K$. Within each group the patients are distributed among strata independently of other groups according to a beta-binomial distribution as described in section 2. Thus, for any stratum s,

$$\mathbf{P}(n_s(j) = i_j, \ j = 1,..,K) = \prod_{j=1}^{K} P(Mk_j, N, N_s, \alpha, i_j). \tag{2}$$

Consider now a stratified randomization. In this case in each stratum randomization is carried out independently of other strata according to block-permuted randomization. If in some stratum s, n_s is not a multiple of K_1, then the last block is incomplete. The incomplete block may contain an unequal number of patients on treatment arms and cause an imbalance in this stratum. Many incomplete blocks in different strata may cause an imbalance between the total number of patients on treatment arms and this may lead to power loss in the study.

Assume that s-th stratum contains an incomplete block of size $m, m = 1,..,K_1 - 1$, and denote by $\xi_j(m)$ the number of instances of treatment j in this block. Then $\xi_j(m)$ has a hypergeometric distribution and $\mathbf{P}(\xi_j(m) = l) = \binom{k_j}{l}\binom{K_1-k_j}{m-l}\binom{K_1}{m}^{-1}$, $l = 0, 1,..,min(k_j,m)$. Therefore, $\mathbf{E}[\xi_j(m)] = k_j m/K_1$, $\mathbf{V}ar[\xi_j(m)] = k_j m(K_1 - k_j) \times (K_1 - m)/(K_1^2(K_1 - 1))$. Let $int(a)$ be the integer part of a, and $mod(a,k) = a - int(a/k)k$. Then

$$n_s(j) = int(n_s/K_1)k_j + \xi_j(mod(n_s, K_1)). \tag{3}$$

As the distribution of n_s is given by (1), the characteristics of $n_s(j)$ can be numerically calculated. Closed-form expressions for the mean and the variance of $n_s(j)$ also can be derived. In the case when strata are associated with different geographical regions, these results allow prediction of supply needed to cover patient demand in regions, number of places in hospitals, etc.

3.1 Impact of Randomization on the Power and Sample Size

Let us consider the impact of randomization scheme on the sample size and the power of a statistical test. If one might expect a statistically significant stratum-by-treatment interaction, then stratified randomization should be preferable from a statistical point of view as it provides better balance within each stratum. Therefore, let us assume that there is no stratum-by-treatment interaction. As stratified randomization in general causes the random imbalance between treatment arms, one would expect that unstratified randomization should be preferable. However, we prove that in general the size of imbalance is rather small compared to the total sample size and its impact on the power and sample size is practically negligible.

3.1.1 Properties of Imbalance in Stratified Randomization

Assume for simplicity that there are only two treatments, *a* and *b* with equal treatment allocations. Denote by $\eta_s = n_s(a) - n_s(b)$ an imbalance in stratum *s*. Let n_j^* be the total number of patients on treatment *j*, $j = a, b$, and $\Delta = n_a^* - n_b^*$ be the total imbalance in the number of patients on both treatments. Then $\Delta = \sum_{s=1}^{S} \eta_s$.

Theorem 1. *For large enough n and S such that $n \min(N_s)/N \geq K_1$, the imbalance Δ is well approximated by a normal distribution with mean zero and variance $s_0^2 S$, where $s_0^2 = (K_1 + 1)/6$.*

Proof. For equal treatment proportions $k_j = K_1/2$ and $\mathbf{E}[\xi_j(m)] = m/2$, $\mathbf{V}ar[\xi_j(m)] = m(K_1 - m)/(4(K_1 - 1))$, $j = 1, 2$. Thus, if in *s*-th stratum the incomplete block has a size *m*, then the imbalance in this stratum is $\eta_s(m) = \xi_1(m) - (m - \xi_1(m)) = 2\xi_1(m) - m$, and $\mathbf{E}[\eta_s(m)] = 0$, $\mathbf{V}ar[\eta_s(m)] = 4\mathbf{V}ar[\xi_1(m)] = m(K_1 - m)/(K_1 - 1)$. In general, in stratum *s* the imbalance η_s is a random variable: $\eta_s = \eta_s(m)$ with probability $q_m(n, N_s, K_1)$, $m = 0, .., K_1 - 1$, where $\eta_s(0) = 0$, and $q_m(n, N_s, K_1) = \mathbf{P}(mod(n_s, K_1) = m)$. Thus, $\mathbf{E}[\eta_s] = 0$ and from (1) it follows

$$q_m(n, N_s, K_1) = \sum_{l=0}^{n/K_1 - 1} P(n, N, N_s, \alpha, m + lK_1), \ m = 0, 1, .., K_1 - 1. \quad (4)$$

Furthermore, if on average the number of patients in a stratum is not less than $2K_1$, one can use the approximation $q_m(\cdot) \approx 1/K_1$ (compare with Hallstrom and Davis (1988)). This is also supported by numerical calculations and Monte Carlo simulations (Anisimov 2007). For example, for $n = 60, S = 6, N_s = 1$ (on average 10 patients in a stratum), $K_1 = 4$ and $\alpha = 1.2$, numerical calculations give $(q_0, q_1, q_2, q_3) = (0.269, 0.259, 0.244, 0.228)$ and simulated values for 10^6 runs coincide with these values up to 3 digits.

Thus, using the approximation $q_m(n, N_s, K_1) = 1/K_1, m = 0, .., K_1 - 1$, we have $\mathbf{V}ar[\eta_s] \approx s_0^2 = (K_1 + 1)/6$. The variables η_s and η_p are not correlated as $s \neq p$ and conditionally independent. Thus, $\mathbf{E}[\eta_s \eta_p] = 0$, $\mathbf{V}ar[\Delta] \approx s_0^2 S$, and at large *S*, Δ is

approximated by a normal distribution with parameters $(0, s_0^2 S)$. This is supported by Monte Carlo simulations (Anisimov 2007). □

Remark 1. As shown above, for large enough numbers of patients the imbalance η_s in each stratum can be approximated by a mixed hypergeometric distribution $\widetilde{\eta}_s = 2\xi(U) - U$, where $\mathbf{P}(U = m) = 1/K_1, m = 0, .., K_1 - 1$, $\mathbf{E}\widetilde{\eta}_s = 0$, $\mathbf{Var}\widetilde{\eta}_s = s_0^2$, and the variables $\widetilde{\eta}_s$ are independent. Thus, for a few strata ($S < 10$), the imbalance Δ can be approximated by the variable $\widetilde{\Delta} = \sum_{s=1}^{S} \widetilde{\eta}_s$, where $\mathbf{E}\widetilde{\Delta} = 0$, $\mathbf{Var}\widetilde{\Delta} = s_0^2 S$.

3.1.2 Impact of Imbalance on the Power and Sample Size

In general imbalance is rather small compared to the sample size. Theorem 1 implies that with probability $1 - \varepsilon$, for large S ($S \geq 10$), $|\Delta| \leq s_0 \sqrt{S} z_{1-\varepsilon/2}$. If $S < 10$, then $|\Delta| \leq s_0 \sqrt{S/\varepsilon}$ (basing on Remark 1 and Chebyshev inequality). In particular, for $n \geq 100$, $K_1 \leq 4$ with probability 0.95, $|\Delta| \leq 8$ as $S = 20$, and $|\Delta| \leq 6$ as $S = 6$.

Let us evaluate the increase in sample size required to maintain the same power as for the balanced study accounting for possible imbalance. Consider as an example a standard test that compares means in two patient populations.

Assume that n patients are randomized to two treatments, a and b, in S strata. If one can expect a stratum-by-treatment interaction, then the stratified randomization should be more preferable from a statistical point of view. Consider the case where there is no stratum-by-treatment interaction. Then general guidelines indicate that unstratified randomization should be more preferable from a statistical point of view. However, we prove that stratified randomization leads practically to the same results.

Consider a stratified randomization by blocks of size K_1 and equal treatment allocations. Let n_j^* be the total number of patients randomized to treatment j, $j = a, b$, and $\{x_1, x_2, .., x_{n_a^*}\}$ and $\{y_1, y_2, .., y_{n_b^*}\}$ be the patient responses on each treatment. Suppose that the observations are independent with unknown means m_a and m_b and the known variance σ^2. It is known that for testing the hypothesis: $H_0 : m_a - m_b = 0$ against $H_1 : m_a - m_b \geq h$ with probabilities γ and δ of type I and type II errors, the values n_a^* and n_b^* should satisfy the relation

$$h\left(\sigma\sqrt{1/n_a^* + 1/n_b^*}\right)^{-1} = z_{1-\gamma/2} + z_{1-\delta}. \tag{5}$$

For a balanced study $n_a^* = n_b^* = n/2$ (assuming that n is even). Thus, in the balanced case a sample size is $n_{bal} = 4\sigma^2(z_{1-\gamma/2} + z_{1-\delta})^2/h^2$. Denote by $\Delta = n_b^* - n_a^*$ the imbalance between treatment arms. Let us evaluate a sample size increase $n_+ = n - n_{bal}$ required to achieve the same power as for a balanced trial.

Theorem 2. *At small S/n_{bal}^2, $n_+ \approx s_0^2 S(1 + \sqrt{2} z_{1-\delta})(1 + \zeta)/n_{bal}$, where ζ is the error term of approximation, $\zeta = O(s_0^2 S/n_{bal}^2)$.*

Proof. Consider a standard test statistic

$$T^* = \frac{\bar{x}_a - \bar{y}_b}{\sigma\sqrt{1/n_a^* + 1/n_b^*}}, \tag{6}$$

where $\bar{x}_a$ and $\bar{y}_b$ are sample means. Under the hypothesis H_0 for large enough n_a^* and n_b^*, $T^* \approx \mathcal{N}(0.1)$, where $\mathcal{N}(0,1)$ has a standard normal distribution. Thus, for testing H_0 with error probabilities γ and δ, the acceptance region is the interval $(-z_{1-\gamma/2}, z_{1-\gamma/2})$, and under the hypothesis H_1 it should be

$$\mathbf{P}_{H_1}(T^* \le z_{1-\gamma/2}) = \delta. \tag{7}$$

Accounting for random imbalance, let us find n satisfying (7). Let ζ_i be the values of the magnitude $O(s_0^2 S/n_{bal}^2)$. Then, under the hypothesis H_1, given the imbalance Δ and assuming that $m_a - m_b = h$ and Δ/n is small, one can use the approximation: $T^* \approx \frac{h}{2\sigma}\sqrt{n}(1 - \Delta^2(1 + \zeta_1)/(2n^2)) + \mathcal{N}(0,1)$. As $z_{1-\gamma/2} + z_{1-\delta} = \sqrt{n_{bal}}\frac{h}{2\sigma}$, relation (7) is asymptotically equivalent to a quadratic equation $n_+^2 + n_{bal}n_+ - Q(1 + \zeta_2) = 0$, where $Q = s_0^2 S(1 + \sqrt{2}z_{1-\delta})$. Thus, $n_+ = \frac{n_{bal}}{2}(\sqrt{1 + 4Q(1 + \zeta_2)/n_{bal}^2} - 1) = Q(1 + \zeta_3)/n_{bal}$. Results of Monte Carlo simulation support this statement for rather wide range of parameters and even for not so large n, e.g. $n = 30$.   □

As usually $S < n_{bal}/2$ and for two treatments $K_1 = 4$, this implies that in general $n_+ \le 2$. Thus, both randomization schemes lead practically to the same sample size.

Note that the impact of imbalance is concentrated in the term $\Delta^2/2n^2 = O(S/n^2)$ and is negligible at large n. This is in agreement with Lachin (1988).

3.1.3 Impact of patient dropout

Consider the impact of a random patient dropout on a sample size for both randomization schemes on the example of the test that compares means (see Section 3.1.2). Assume that each patient randomized to treatment j will stay till the end of the trial with probability p_j, $j = a, b$. Only these patients will be included into the analysis. The values $q_j = 1 - p_j$, $j = a, b$, define the probabilities of dropout. Let v_j be the number of patients initially randomized to treatment j. Assume that $v_a - v_b = G$, where G is a random variable with mean zero and variance D^2. As $v_a + v_b = n$, then $v_a = n/2 + G/2$, $v_b = n/2 - G/2$. In this general setting we can combine together the cases of unstratified and stratified randomization, as in the first case $G = 0$, and in the second case $G = \Delta$ and according to Theorem 1, $D^2 \approx s_0^2 S$.

Let n_j^* be the remaining number of patients on treatment j after dropout. Then $n_a^* = \text{Bin}(n/2 + G/2, p_a)$, $n_b^* = \text{Bin}(n/2 - G/2, p_b)$, where $\text{Bin}(k, p)$ is a binomial variable with parameters (k, p). If G is random, n_a^* and n_b^* are dependent and $\mathbf{E}[n_j^*] = np_j/2$, $\mathbf{Var}[n_j^*] = np_j q_j/2 + D^2 p_j^2/4$, $\mathbf{E}[n_a^* n_b^*] = p_a p_b(n^2 - D^2)/4$. Thus, at large n

$$(n_a^*, n_b^*) \approx \left((n/2)p_a(1 + \psi_a \xi_a/\sqrt{n}), (n/2)p_b(1 + \psi_b \xi_b/\sqrt{n}) \right), \tag{8}$$

where $\psi_j = \sqrt{2q_j/p_j + D^2/n}$, $j = a, b$, and vector (ξ_a, ξ_b) has a bivariate normal distribution, $\mathbf{E}\xi_j = 0$, $\mathbf{Var}\xi_j = 1$, $\mathbf{E}[\xi_a \xi_b] = -D^2/(n\psi_1 \psi_2)$. Denote

$$M = \frac{h}{\sigma^2}\sqrt{\frac{p_a p_b}{2(p_a + p_b)}}, \quad R = \frac{D^2 p_a p_b (p_a - p_b)^2}{2n}, \quad B^2 = \frac{h^2}{\sigma^2}\frac{q_a p_b^3 + q_b p_a^3 + R}{4(p_a + p_b)^3}.$$

Under the hypothesis H_1, after some algebra one can get an approximation for statistic (6) in the form $T^* \approx \sqrt{n}M + \sqrt{1 + B^2}\mathcal{N}(0,1)$. This relation together with (7) implies the relation for the required sample size:

$$n \approx \frac{2\sigma^2(p_a + p_b)}{h^2 p_a p_b}\left(z_{1-\gamma/2} + \sqrt{1 + B^2}z_{1-\delta}\right)^2. \qquad (9)$$

Consider now the averaged design (the number of patients on treatments a and b are fixed and equal to $(n/2)p_a$ and $(n/2)p_b$, respectively). Using (5) one can easily establish that the sample size for the averaged design is

$$n_{aver} \approx \frac{2\sigma^2(p_a + p_b)}{h^2 p_a p_b}\left(z_{1-\gamma/2} + z_{1-\delta}\right)^2.$$

Thus, the sample size increase compared to the averaged design is concentrated in the term B^2 and is practically negligible. For example, if B^2 is rather small,

$$n - n_{aver} \approx \frac{q_a p_b^3 + q_b p_a^3 + R}{2p_a p_b (p_a + p_b)^2}z_{1-\delta}(z_{1-\gamma/2} + z_{1-\delta}). \qquad (10)$$

In particular, for $\gamma = \delta = 0.05$ and $p_a = p_b = p$, in the region $p \geq 0.4$ (dropout less than 60%), $n - n_{aver} \leq 2$ (sample size increases by no more than two patients).

The impact of the randomization scheme is concentrated in the term R. For unstratified randomization $R = 0$, while in the case of stratified randomization $R = s_0^2 S p_a p_b (p_a - p_b)^2/(2n)$ and is also rather small. Calculations show that using stratified randomization practically does not lead to sample size increase.

Table 1: Sample size calculations.

h	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
Averaged design	409	284	209	160	127	103	85	71	61	53	46
Unstratified	411	286	211	162	129	105	87	73	63	55	48
Stratified	411	286	211	162	129	105	87	73	63	55	48

Table 1 shows the calculated values of sample sizes for a particular scenario. Consider a study with $S = 10$ strata of equal sizes ($N_s = 1$). Let $\gamma = 0.05, \delta = 0.05, p_a = 0.4, p_b = 0.7, K = 2$, block size $K_1 = 4$. Consider three cases: averaged design (randomness in dropout is not accounted for), unstratified randomization and stratified randomization. We set $\sigma^2 = 1$. The sample size is calculated for different values of h in interval $[0.5, 1.5]$. As one can see, a sample size increase accounting for random patient dropout is only two patients, and using stratified randomization does not

lead to an additional sample size increase compared to unstratified randomization. Similar results are true for other scenarios and large number of strata.

4 Conclusions

Using the advanced patient recruitment model allows prediction at the design stage of the number of patients randomized to different treatment arms in different strata and investigation of the properties of imbalance caused by stratified randomization and its impact on the power and sample size of the trial. For two treatment arms with interest in a statistical test that compares means, it is shown, that the sample size increase required to compensate for random imbalance is practically negligible. Randomness in patient dropout also leads to a negligible sample size increase compared to averaged design (fixed number of randomized patients). These results show that stratified randomization even for a large number of strata does not lead to a visible sample size increase compared to unstratified randomization.

The type of randomization may affect other characteristics of the trial, e.g. centre-stratified randomization in general requires less drug supply compared to unstratified randomization. Thus, in the cases when the choice of randomization is not dictated by the type of data, it is beneficial to use various criteria accounting for sample size, recruitment and supply costs, etc., when choosing a randomization scheme.

References

Anisimov, V. V. (2007). Effect of imbalance in using stratified block randomization in clinical trials. *Bulletin of the International Statistical Institute - LXII, Proc. of the 56 Annual Session, Lisbon*, 5938–5941.

Anisimov, V. V. and V. V. Fedorov (2006). Design of multicentre clinical trials with random enrolment. In *Advances in Statistical Methods for the Health Sciences* (N. Balakrishnan, J. L. Auget, M. Mesbah, and G. Molenberghs eds). Berlin: Birkhäuser, 387–400.

Anisimov, V. V. and V. V. Fedorov (2007). Modeling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in Medicine 26*, 4958–4975.

Hallstrom, A. and K. Davis (1988). Imbalance in treatment assignments in stratified blocked randomization. *Controlled Clinical Trials 9*, 375–382.

Lachin, J. M. (1988). Statistical properties of randomization in clinical trials. *Controlled Clinical Trials 9*, 289–311.

Matts, J. P. and J. M. Lachin (1988). Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials 9*, 327–345.

Pocock, S. J. (1983). *Clinical Trials. A Practical Approach*. New York: Wiley.

Rosenberger, W. F. and J. M. Lachin (2002). *Randomization in Clinical Trials*. New York: Wiley.