



Research Article

Copy Right@ Weijia Mai

# On Random Center Grouping in Multicenter Clinical Trials

Weijia Mai\* and Shein-Chung Chow

Department of Biostatistics and Bioinformatics, Duke University School of Medicine, North Carolina

\*Corresponding author: Weijia Mai, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, 2424 Erwin Road, Hock Suite 1102, Durham, North Carolina.

To Cite This Article: Weijia Mai\* and Shein-Chung Chow. On Random Center Grouping in Multicenter Clinical Trials. Am J Biomed Sci & Res. 2021 - 11(5). AJBSR.MS.ID.001667. DOI: 10.34297/AJBSR.2021.11.001667.

Received: 📅 January 12, 2021; Published: 📅 January 26, 2021

## Abstract

With the attention of achieving desired power within a pre-specified timeframe in clinical development, a multicenter trial is often conducted to expedite the patient recruitment. However, a multicenter trial with plethoric number of centers is very likely to result in numerous small centers, which will cause the problem of treatment imbalance (unequal number of patients per arm within a center) and/or center imbalance (a few large centers with a number of small centers). In practice, treatment imbalance within and between centers inevitably occurs regardless the randomization models and/or methods used. Treatment imbalance within center and/or center imbalance will not only (i) increase the probability of observing treatment-by-center interaction, but also (ii) decrease the power for detecting clinically meaningful difference (or treatment effect) of the test treatment under investigation. In this paper, we propose a method of determining whether an observed significant interaction is a false alarm and a method of random grouping for an unbiased and reliable assessment of treatment effect when significant treatment and center imbalance are observed. The impact of treatment and center imbalance will also be examined in terms of the potential loss of power.

**Keywords:** Treatment Imbalance; Center Imbalance; Treatment-by-Center Interaction; Simple Sampling without replacement

## Introduction

As indicated in the 1988 guidance entitled *Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications* published by the United States Food and Drug Administration (FDA), adequate clinical trials need to be conducted in good control in order to provide ample evidence of the safety and effectiveness for endorsement of a new medication. In practice, a multicentered, randomized, well-controlled clinical trial is often considered for this purpose. When multiple clinical trials are concurrently conducted at various centers but follow the same protocol, these single trials can be viewed together as a single study called a multicenter trial. With the intention of achieving desired power within a pre-specified timeframe in clinical development, a multicenter trial is often performed to speed up the patient recruitment. There are two main purposes of using a multicenter clinical trial in clinical development: Firstly, it is to show that the clinical results are duplicable and not happening by chance from

center to center. Secondly, it is to illustrate that the clinical results can be generalized to populations from different geographical locations. However, the occurrence of redundant small centers may result in treatment imbalance and center imbalance, which consequently cause treatment-by-center interaction and make it difficult for us to combine the centers for an overall assessment. In practice, it is suggested that appropriate statistical tests for treatment-by-center interaction should be performed.

As indicated in Chow and Liu [1], a multicenter trial is not equivalent to separate single-site trials, since the purpose is to gather data from different centers together and make an overall analysis. It is suggested that proper statistical tests for homogeneity across centers should be conducted to check possible quantitative or qualitative treatment-by-center interactions [2,3] before grouping data for an overall assessment of safety and effectiveness of the study medication. When significant treatment effect differences



occur in different directions across centers, a qualitative interaction is said to occur. Whereas when the treatment differences are with different magnitude but still in the same direction across centers, a quantitative interaction, which is relatively less important comparing to qualitative interaction [4] is said to occur. Also, as pointed out by Guohua and Douglas (1997), an overall statistical inference regarding the treatment effect is not valid to be made if there is a substantial qualitative interaction between treatment and center, in which case that centers cannot be grouped, and the treatment effect must be assessed by study center.

For a multicenter trial, [5] posted the several questions, which are helpful for the design and analysis of multicenter trials. The first one is whether the existence of extremely small centers will affect the reliability of separate interpretation of results. Another helpful one is in contrast to the first question, which is whether large centers with large number of data will dominate the results of analysis. Moreover, whether there exist outliers in results is also an important question. Finally, is the question that we care most about, which is whether treatment-by-center interaction will cause the trial to be invalid. In this paper, however, our focus will be placed on statistical evaluation of random center grouping when a multicenter trial results in a number of small centers. In this paper, we propose a method of determining whether the observed significant treatment-by-center is a false alarm (which may be due to too many small centers) through center grouping based on the optimal selection of number proposed by [8]. Following Shao and Chow [8] idea, we also propose a method of random center grouping to handle the issue of treatment and center imbalance.

In the next section, following the idea of two-stage sampling strategy proposed by Shao and Chow (1993) [8] a rule of thumb for optimal selection of the number of centers is proposed. Section 3 discusses the impact of treatment imbalance on statistical power for testing treatment effect. Section 4 illustrates the use of center grouping for determining whether the observed significant treatment-by-center interaction (as a result of too many small centers) is a false alarm. In Section 5, we propose a method for center grouping when the multicenter trial results in a number of small centers followed by a simulation study of power change. The method utilizing random center grouping is proposed in Section 6. Section 7 compares the two methods of center grouping. An example concerning a multicenter trial for treating patients with breast cancer is given in the Section 8 to illustrate the application of the proposed methods. Discussions and concluding remarks are given in the last section.

### The Selection of The Number of Centers

As mentioned in the previous section, one purpose of multicenter trial is to reduce the time consumption of patient recruitment process. The more the centers, the sooner the recruitment process

and hence the quicker the study would be completed. Nevertheless, more centers would certainly result in small centers (i.e., fewer patients in each center).

Since for comparative clinical trials, comparisons are made between patients within the same center to test the treatment effects, if the number of patients in each center is too small, the assessments and inferences made may be statistically invalid and biased. We thereby intend to combine the data of small centers to form a larger dummy center. Nevertheless, it is required by both FDA and ICH guidelines that statistical tests for homogeneity across centers must be conducted in order to check for the existence of significant treatment-by-center interaction before pooling data. Therefore, under the situations where a qualitative treatment-by-center interaction is observed and tested to be significant, the sponsor may be required to examine the treatment effect of each center independently instead of pooling together due to statistical invalidation.

The increase in number of centers will certainly increase the chance for significant qualitative treatment-by-center interaction to occur, especially when both large and small centers exist.

The observed statistical significance may be due to two following reasons [9]:

**Heterogeneity among centers:** Some centers exhibit relatively large variability.

**Heterogeneity across centers:** Some centers do not constitute a representative sample of the target patient population.

Accordingly, the way to determine the optimal number of centers is noteworthy for investigators. However, for simplicity and convenience, centers are often grouped without considering from statistical perspective in many multicenter trials. The method of random grouping in this paper will be based on the idea regarding two-stage sampling strategy in clinical trials proposed by Shao and Chow [8], we propose the following rule of thumb for selection of the number of centers given that the total sample size of the intended trial is  $N$ . We propose the rule of thumb that the number of patients in each center should be at least equal to the number of centers for achieving optimal statistical properties for treatment assessment. Thus, if the intended clinical trial calls for  $N$  patients in total, it is suggested that around  $\sqrt{N}$  study centers with around  $\sqrt{N}$  patients in each be selected for achieving optimal statistical properties for treatment assessment.

### The Impact of Treatment Imbalance on Power

Without loss of generality, consider a multicenter clinical trial with two treatments. In practice, sample size is often chosen for the purpose of detecting clinically meaningful significance at pre-specified significance level with a desired power [10].

Under the assumption of normality, sample size of a balanced trial (i.e., same number of patients in each arm) is often calculated by:

$$n = \frac{2\sigma^2(z_{\frac{\alpha}{2}} + z_{\beta})^2}{\Delta^2}$$

with a power of

$$1 - (\Phi[z_{\frac{\alpha}{2}} - \frac{\Delta}{\sigma\sqrt{\frac{2}{n}}}] + \Phi[z_{\frac{\alpha}{2}} + \frac{\Delta}{\sigma\sqrt{\frac{2}{n}}}]$$

where the standard deviation of the random error is written as  $\sigma$ ,  $z_{\alpha}$  represents  $\alpha^{th}$  quantile of a standard normal distribution, and  $\Delta$  means the difference of clinical importance.

Since under normal distribution, the part  $(\Phi[z_{\frac{\alpha}{2}} + \frac{\Delta}{\sigma\sqrt{\frac{2}{n}}}]$  is very small and negligible, for simplicity, the power can be rewritten as follows

$$1 - (\Phi[z_{\frac{\alpha}{2}} - \frac{\Delta}{\sigma\sqrt{\frac{2}{n}}}] \tag{1}$$

This expression of power is obtained by neglecting both center effect and the effect due to treatment and center interaction, hence is the optimal power that we can possibly get.

Treatment imbalance (i.e., different number of patients in each arm) is sometimes inevitable in spite of the plans of having treatment groups of same size, which may therefore cause differences among centers. Under these circumstances, the expression of power becomes

$$1 - \Phi[z_{\frac{\alpha}{2}} - \frac{\Delta}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}] \tag{2}$$

which is obviously less than the power of balanced trial. Hence in order to achieve the optimal power as shown in (1), we set (1) and (2) to be equal, which leads to

$$\frac{\Delta}{\sigma\sqrt{\frac{2}{n}}} = \frac{\Delta}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{3}$$

It is clearly that even if the sample size  $N$  is fixed,  $n_i, i = 1, 2$ , are not fixed, since we are not able to predict how many patients will be in each center after the completion of the trial. Thus, from (3) we can know that increasing total sample size  $N$  is the only way to make the equation true (i.e., to achieve optimal power) when variance  $\sigma$  is assumed to be unchanged. It is noteworthy that the variance of the test statistic calculated with equal number of patients in each center will be the same as if it is a single center trial. Hence, the minimum variance of test statistics is achieved since  $\frac{2}{n} \leq \frac{1}{n_1} + \frac{1}{n_2}$ .

### Interaction Detection Using Center Grouping

Statistical test for homogeneity across centers is the most common way to detect potential interaction between treatment and center, however, center grouping provides another way of doing this.

Suppose there are 5 centres: 3 small ones with 4, 4 and 2 patients respectively, 2 large ones with 12 and 10 patients respectively. Following are two different scenarios:

**Scenario 1:** As we can see from figure 1, the result of one small center is different from others, which possibly cause significant interaction. One way to check the significance of interaction is to do the F test, and we can get the following table 1, from which we know that the interaction is significant as p value is less than 0.05. Now we combine the first 3 small centers and remark them as center 1, center 4 and 5 will remain the same and be remarked as center 2 and 3. The effects after center grouping is as following figure 2 shows, which matches with the conclusion that we drew from the F test result since interaction still exists after center grouping.

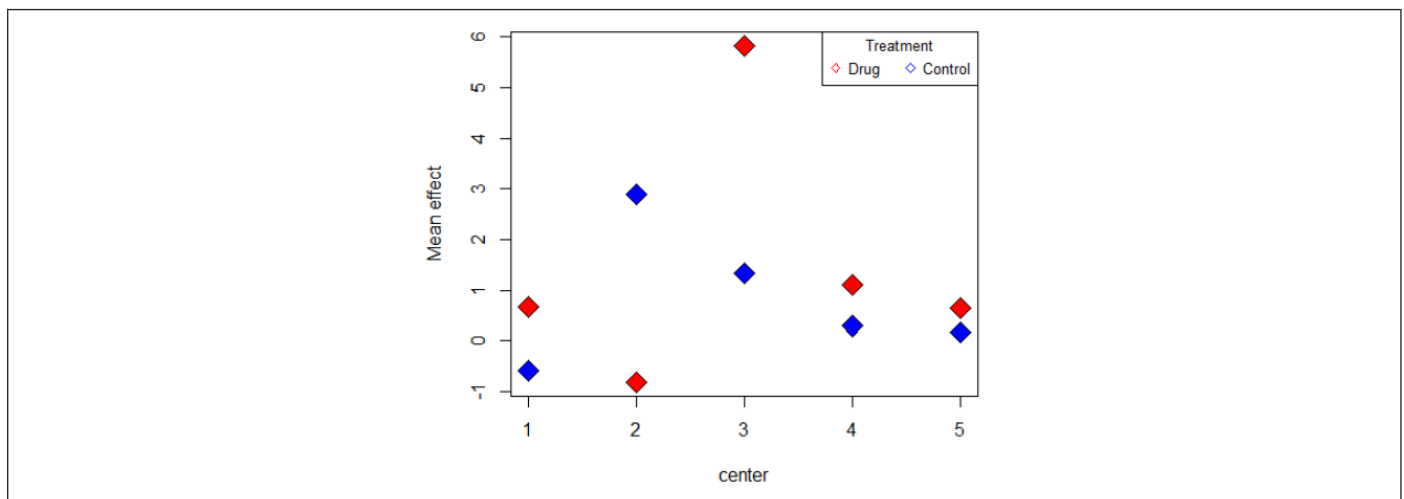


Figure 1: Plot of treatment effects before grouping.

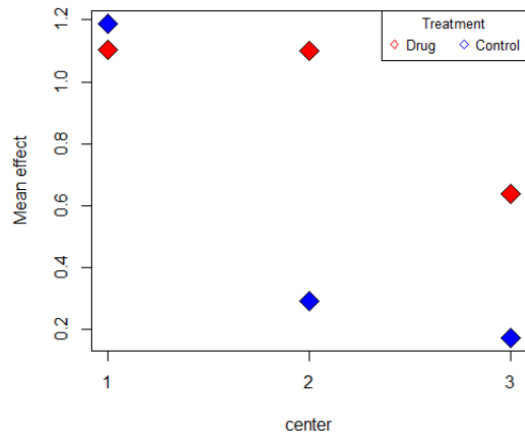


Figure 2: Plot of treatment effect after grouping.

Interaction Test			
F	Interaction DoF	Error DoF	p value
7.300906	4	22	0.0006700413

Table 1: Results of interaction test.

**Scenario 2:** In this scenario, there is again one small center provides opposite trend compared to other centers, and cause interaction to occur, which can be seen in figure 3. Then by conducting F test, we get the following results which suggest that there is no significant interaction (table 2). Again, we combine the

three small centers and do the analysis, figure 4 illustrates that the interaction disappears after center grouping, which matches with the outcome of F test. Thereby we conclude that the method of detecting interaction using center grouping is valid, and if the interaction disappears after grouping, we are then able to make a general assessment of the drug across centers.

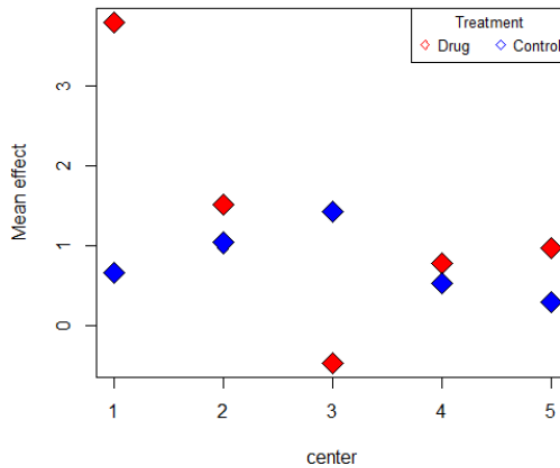


Figure 3: Plot of treatment effect before grouping.

Interaction Test			
F	Interaction DoF	Error DoF	p value
2.131331	4	22	0.1109714

Table 2: Results of interaction test.

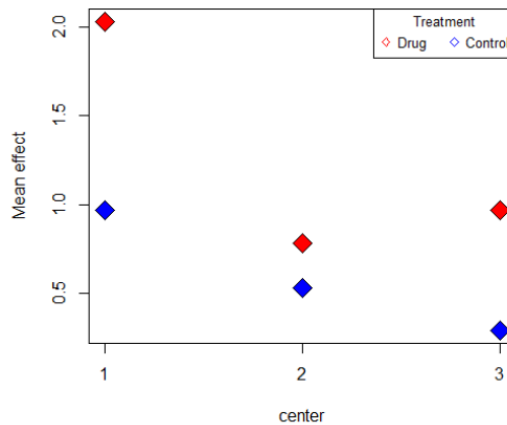


Figure 3: Plot of treatment effect before grouping.

### Center Grouping

Without loss of generality, consider this two-way factor random effects model [6]:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

$$i = 1, 2, \dots, I; j = 1, 2, \dots, J; k = 1, 2, \dots, K,$$

where  $\mu$  is the general mean,  $\alpha_i = A_i - \mu$  is the main effect of the  $i^{th}$  level of A with  $A_i$  to be the fixed effect due to the  $i^{th}$  treatment (factor A),  $\beta_j = B_j - \mu$  is the main effect of the  $j^{th}$  level of B with  $B_j$  to be the random effect due to the  $j^{th}$  center (factor B). We then define  $\mu_{ij}$  to be the cell mean and thereby define random effect due to interaction to be  $(\alpha\beta)_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$ . Following are some constraints on this model:

$$\alpha_{\cdot} = \sum_i \alpha_i = 0$$

$$\beta_{\cdot} = \sum_j \beta_j = 0$$

$$(\alpha\beta)_{i\cdot} = \sum_j (\alpha\beta)_{ij} = 0 \quad \forall i$$

$$(\alpha\beta)_{\cdot j} = \sum_i (\alpha\beta)_{ij} = 0 \quad \forall j$$

Both  $\alpha_i$ ,  $\beta_j$  are independently identically normally distributed with variance  $\sigma_A^2$  and  $\sigma_B^2$  and  $\varepsilon_{ijk}$  is the random error which is also independently identically normally distributed with mean 0 and variance  $\sigma^2$  [7]. Then as indicated in Scheffé (1959), the sum of squares is calculated as follows:

$$SS_{\alpha} = JK \sum_{i=1}^I (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$SS_{\beta} = IK \sum_{j=1}^J (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2,$$

$$SS_{\alpha\beta} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2,$$

$$SS_{\varepsilon} = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - \bar{y}_{ij\cdot})^2$$

where

$$\bar{y}_{i\cdot} = \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K y_{ijk}$$

$$\bar{y}_{\cdot j} = \frac{1}{IK} \sum_{i=1}^I \sum_{k=1}^K y_{ijk}$$

$$\bar{y}_{ij\cdot} = \frac{1}{K} \sum_{k=1}^K y_{ijk},$$

$$\bar{y}_{\cdot\cdot} = \frac{1}{JK} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K y_{ijk}.$$

If centers (factor B) are combined, then it is easy to show that the new  $SS_{\varepsilon}$  becomes

$$New SS_{\varepsilon} = SS_{\beta} + SS_{\alpha\beta} + SS_{\varepsilon}.$$

Therefore, we can get

$$E(New SS_{\varepsilon}) = [(J-1) + (I-1)(J-1) + IJ(K-1)]\sigma^2 + (J-1)IK\sigma_B^2$$

$$E(New SS_{\varepsilon}) = I(JK-1)\sigma^2 + IK(J-1)\sigma_B^2$$

which implies that

$$E(New MS_E) = \sigma^2 \left( 1 + \frac{K(J-1)\sigma_B^2}{(JK-1)\sigma^2} \right)$$

From these expressions, we make following conclusions. First,

$$\frac{K(J-1)\sigma_B^2}{(JK-1)\sigma^2} > 0 \text{ if } \sigma_B^2 \text{ is not 0. Meanwhile, } \frac{K(J-1)\sigma_B^2}{(JK-1)\sigma^2} \text{ increases as}$$

number of centers ( $J$ ) being combined increases, and  $\frac{K(J-1)\sigma_B^2}{(JK-1)\sigma^2}$  also depends upon the similarity of centers ( $\sigma_B^2$ ) being combined. Based on this expression, it can be deduced that the change in  $MS_E$  might depend on not only  $I, J, K$ , but also on the ratio of  $\frac{\sigma_B^2}{\sigma^2}$ . Before combining the  $J$  centers, the treatment effect can be tested by

$$\frac{MS_A}{MS_E} = \frac{\frac{SS_\alpha}{I-1}}{\frac{SS_\epsilon}{IJ(K-1)}} \sim \frac{IJ(K-1)}{I-1} \times \frac{\chi^2(I-1, \delta)}{\chi^2(IJ(K-1))} \sim F_{I-1, IJ(K-1)}(\delta),$$

which follows a non-central F distribution under alternative hypothesis with non-centrality parameter  $\delta = \frac{JK}{\sigma^2} \sum \alpha_i^2$  and degrees of freedom  $I - 1$  and  $IJ(K - 1)$ .

After grouping, we can test the treatment effect by using the new  $SS_\epsilon$ :

$$\frac{MS_A}{New MS_E} = \frac{\frac{SS_\alpha}{I-1}}{\frac{New SS_\epsilon}{I(JK-1)}}$$

We used simulation based to 20,000 iterations to generate

tables illustrating the power before and after center grouping. As discussed above, since  $I, J, K$ , the ratio  $\frac{\sigma_B^2}{\sigma^2}$  and  $\delta$  might affect the test results, we change these values and summarized the results of power analysis in (Table 3-5). When conducting simulations,  $\sigma^2$  is set to be 1,  $\sigma_A^2$  and  $\sigma_B^2$  are then set according to the ratios. In the tables,  $p_1$  denotes the power before grouping,  $p_2$

denotes the power after grouping, which is the power within each dummy center. The relative improvement is demonstrated by

$$\Delta \times 100 = \frac{p_2 - p_1}{p_1} \times 100\%$$

Since even combining some small centers into a dummy center can have power increased, then it is certain that by grouping all dummy centers together, we can get a power which is significantly higher than before. We can thereby use the relative improvement in power as an indicator of the center grouping effect, and we aim to combine centers in the manner of reaching maximum  $\Delta$ .

From Tables 3-5, following statements can be concluded:

- A. Power can increase significantly if the small centers are properly grouped.
- B. Maximum power increase in percentage ( $\Delta$ ) can be reached by making appropriate choice of the number of centers in each dummy center ( $J$ ) if sample size per arm ( $K$ ) of small centers is fixed.
- C. Under some circumstances with certain simulated data, however, power may also decrease after center grouping.

Power Comparison for $\sigma_B^2/\sigma^2 = 0.01, I = 2$										
$r\_alpha^a$	J = 2			J = 3			J = 4			
	$p1^b$	$p2^c$	$\Delta^d$	$p1$	$p2$	$\Delta$	$p1$	$p2$	$\Delta$	
K=2	0.0	0.707	0.756	6.882	0.750	0.785	4.738	0.358	0.451	25.894
	0.1	0.584	0.636	8.895	0.651	0.692	6.362	0.516	0.585	13.377
	0.2	0.545	0.598	9.863	0.626	0.671	7.165	0.557	0.626	12.488
	0.3	0.528	0.584	10.504	0.628	0.675	7.476	0.587	0.658	12.155
	0.4	0.522	0.578	10.861	0.628	0.678	7.889	0.602	0.676	12.304
	0.5	0.518	0.576	11.262	0.638	0.689	8.005	0.619	0.695	12.355
	0.6	0.522	0.581	11.280	0.639	0.692	8.247	0.631	0.709	12.336
	0.7	0.524	0.584	11.460	0.647	0.701	8.380	0.644	0.723	12.184
	0.8	0.526	0.586	11.546	0.654	0.710	8.458	0.649	0.731	12.502
	0.9	0.529	0.590	11.599	0.657	0.713	8.567	0.661	0.742	12.248
1.0	0.530	0.592	11.684	0.656	0.714	8.745	0.664	0.746	12.325	
K=3	0.0	0.481	0.570	18.369	0.060	0.196	227.463	0.059	0.201	243.036
	0.1	0.579	0.647	11.760	0.237	0.402	69.713	0.305	0.443	44.981
	0.2	0.611	0.677	10.916	0.344	0.512	48.766	0.355	0.502	41.213
	0.3	0.631	0.699	10.792	0.415	0.585	41.031	0.379	0.532	40.227
	0.4	0.647	0.717	10.796	0.456	0.629	38.029	0.398	0.555	39.420
	0.5	0.658	0.731	10.943	0.491	0.686	35.597	0.404	0.568	40.690
	0.6	0.676	0.749	10.765	0.512	0.691	34.910	0.410	0.575	40.370
	0.7	0.687	0.761	10.799	0.538	0.714	32.831	0.415	0.586	41.169
	0.8	0.690	0.767	11.052	0.554	0.732	32.284	0.422	0.592	40.361
	0.9	0.700	0.777	10.927	0.569	0.746	31.285	0.426	0.599	40.582
1.0	0.704	0.783	11.106	0.582	0.759	30.512	0.428	0.601	40.214	

<sup>a</sup> $r\_alpha = (1/n) * \sum \alpha^2$   
<sup>b</sup> $p1$ : power before combine  
<sup>c</sup> $p2$ : power after combine  
<sup>d</sup> $\Delta$ : power increase percentage

Table 3: Power Comparison.

Power Comparison for  $\sigma^2 B^2/\sigma^2 = 0.1, I = 2$

r_alpha <sup>a</sup>	J = 2			J = 3			J = 4			
	p1 <sup>b</sup>	p2 <sup>c</sup>	Δ <sup>d</sup>	p1	p2	Δ	p1	p2	Δ	
K=2	0.0	0.208	0.363	74.417	0.596	0.618	3.694	0.822	0.891	8.483
	0.1	0.340	0.476	39.992	0.663	0.680	2.523	0.603	0.739	22.601
	0.2	0.404	0.532	31.809	0.675	0.692	2.559	0.577	0.729	26.309
	0.3	0.439	0.567	29.169	0.691	0.709	2.607	0.579	0.741	27.893
	0.4	0.466	0.594	27.586	0.711	0.729	2.610	0.585	0.752	28.694
	0.5	0.481	0.611	27.148	0.725	0.744	2.658	0.590	0.765	29.505
	0.6	0.499	0.631	26.426	0.738	0.757	2.685	0.595	0.774	30.261
	0.7	0.514	0.648	26.005	0.747	0.767	2.741	0.602	0.784	30.296
	0.8	0.522	0.657	25.967	0.753	0.774	2.820	0.598	0.788	31.652
	0.9	0.531	0.668	25.819	0.763	0.784	2.809	0.603	0.794	31.645
1.0	0.534	0.674	26.268	0.769	0.791	2.822	0.605	0.798	31.871	
K=3	0.0	0.370	0.482	30.056	0.758	0.815	7.559	0.000	0.000	-14.099
	0.1	0.494	0.576	16.530	0.616	0.695	12.885	0.012	0.014	13.343
	0.2	0.507	0.590	16.248	0.606	0.693	14.344	0.052	0.057	10.540
	0.3	0.522	0.608	16.324	0.611	0.705	15.300	0.104	0.113	8.679
	0.4	0.534	0.621	16.474	0.618	0.717	15.935	0.160	0.172	7.488
	0.5	0.539	0.629	16.812	0.630	0.731	16.102	0.209	0.223	6.607
	0.6	0.552	0.645	16.778	0.630	0.735	16.690	0.252	0.267	6.124
	0.7	0.556	0.651	17.017	0.635	0.742	16.782	0.286	0.303	5.780
	0.8	0.560	0.657	17.166	0.639	0.748	17.119	0.317	0.334	5.485
	0.9	0.569	0.668	17.257	0.643	0.753	17.091	0.347	0.365	5.204
1.0	0.576	0.674	17.077	0.649	0.761	17.256	0.370	0.388	5.123	

<sup>a</sup>r\_alpha = (1/n) \* Σσ<sup>2</sup>  
<sup>b</sup>p1: power before combine  
<sup>c</sup>p2: power after combine  
<sup>d</sup>Δ: power increase percentage

Table 4: Power Comparison.

Power Comparison for  $\sigma^2 B^2/\sigma^2 = 1, I = 2$

r_alpha <sup>a</sup>	J = 2			J = 3			J = 4			
	p1 <sup>b</sup>	p2 <sup>c</sup>	Δ <sup>d</sup>	p1	p2	Δ	p1	p2	Δ	
K=2	0.0	0.095	0.166	75.194	0.218	0.241	10.549	0.548	0.726	32.347
	0.1	0.177	0.268	51.172	0.388	0.409	5.562	0.628	0.770	22.553
	0.2	0.248	0.345	38.938	0.453	0.474	4.803	0.645	0.791	22.619
	0.3	0.307	0.407	32.642	0.488	0.511	4.611	0.665	0.815	22.566
	0.4	0.351	0.453	29.013	0.514	0.537	4.490	0.683	0.836	22.470
	0.5	0.386	0.490	26.827	0.533	0.557	4.456	0.698	0.854	22.430
	0.6	0.418	0.523	25.095	0.547	0.572	4.441	0.708	0.868	22.595
	0.7	0.441	0.547	24.131	0.560	0.585	4.455	0.719	0.880	22.435
	0.8	0.461	0.569	23.436	0.570	0.596	4.453	0.721	0.888	23.064
	0.9	0.484	0.592	22.261	0.581	0.606	4.394	0.730	0.896	22.813
1.0	0.497	0.606	21.834	0.586	0.612	4.462	0.733	0.901	22.937	
K=3	0.0	0.925	0.950	2.704	0.196	0.465	136.850	0.015	0.066	350.193
	0.1	0.636	0.743	16.889	0.412	0.624	51.193	0.146	0.268	83.140
	0.2	0.599	0.717	19.746	0.484	0.691	42.822	0.274	0.413	50.684
	0.3	0.593	0.719	21.163	0.520	0.731	40.532	0.347	0.493	42.131
	0.4	0.594	0.726	22.195	0.548	0.762	39.162	0.408	0.558	36.725
	0.5	0.593	0.731	23.225	0.568	0.787	38.572	0.450	0.603	33.976
	0.6	0.602	0.742	23.383	0.578	0.801	38.468	0.486	0.639	31.469
	0.7	0.604	0.749	24.035	0.593	0.817	37.905	0.510	0.663	30.001
	0.8	0.605	0.753	24.439	0.601	0.828	37.786	0.528	0.682	29.180
	0.9	0.611	0.761	24.493	0.607	0.835	37.678	0.551	0.704	27.854
1.0	0.614	0.766	24.654	0.613	0.842	37.350	0.564	0.718	27.331	

<sup>a</sup>r\_alpha = (1/n) \* Σσ<sup>2</sup>  
<sup>b</sup>p1: power before combine  
<sup>c</sup>p2: power after combine  
<sup>d</sup>Δ: power increase percentage

Table 5: Power Comparison.

### Random Center Grouping

As mentioned above, the power of dummy centers could be higher than that of smaller centers, hence it is suggested to group these smaller centers together with the goal of increasing power. Since we would like to keep the generalizability of multicenter trial, it is not a good choice to group by geographical location although it is the easiest way to think of. The reason for this is that efficacy varies significantly between population groups with different demographic factors such as culture, environment, and income, therefore it is a better idea to group patients with different demographic factors together to remove the heterogeneity in results caused by these differences. It is recommended to form dummy centers by grouping small centers randomly, since the outcomes may be invalid with non-random center grouping [9].

We know that generally, the between-center variability can be evaluated by considering the following equation

$$E\left(\frac{SS_B}{J-1}\right) = \sigma^2 + IK\sigma_B^2$$

therefore, we argue that achieving an unbiased estimate for  $\sigma^2 + IK\sigma_B^2$  means that the randomization is valid. Assume we would like to group  $J$  small centers into some dummy centers each consists of  $n$  small centers, we propose the following randomization method and believe it to be valid. First of all, randomly permutate the index  $j = 1, \dots, J$  into  $j_1, j_2, \dots, j_J$ . Then we assign first  $n$  indices in  $j_1, j_2, \dots, j_J$  to the first dummy center (Chow, 2011) and similar procedure for the rest dummy centers.

It is obvious that the above method for randomization is simple random sampling without replacement, hence according to [9], the

mean of each center ( $\bar{y}_{.jt}, t = 1, \dots, n$ ) within the dummy center can be regarded as a simple random sampling without replacement from a population of  $J$  small centers each with mean  $\bar{y}_{.j}, j = 1, \dots, J$ . Thus, the sum of squares within each dummy center can be written as

$$SS_{\beta}^* = IK \sum_{t=1}^n (\bar{y}_{.jt} - \frac{1}{n} \sum \bar{y}_{.j})^2$$

Therefore  $\frac{SS_{\beta}^*}{n-1}$  can be used to estimate  $\sigma^2 + IK\sigma_B^2$ . Lohr (1999) proved that given  $\bar{y}_{.j}, j = 1, \dots, J, \frac{SS_{\beta}^*}{n-1}$  is an unbiased estimator for

$$\frac{SS_{\beta}}{J-1} = \frac{IK}{J-1} \sum_{j=1}^J (\bar{y}_{.j} - \bar{y}_{..})^2$$

Thus,

$$E\left(\frac{SS_{\beta}^*}{n-1}\right) = E\left[E\left(\frac{SS_{\beta}^*}{n-1} \mid y_{.j}, j = 1, \dots, J\right)\right]$$

$$= E\left(\frac{SS_{\beta}}{J-1}\right)$$

$$= \sigma^2 + IK\sigma_B^2$$

Thus, the randomization procedure that we proposed for center grouping is valid.

### Procedure for Random Center Grouping

In this section, we will introduce two methods for random center grouping based on the idea of selecting an optimal number of centers for achieving optimal statistical properties for treatment assessment in multicenter trials proposed by Shao and Chow

(1993). The proposed method under two scenarios is summarized below.

#### Scenario 1

**Step 1:** From the rule of thumb proposed by Shao and Chow (1993), we start with the optimal selection of approximately  $\sqrt{N}$  study centers with approximately  $\sqrt{N}$  patients in each center if there are  $N$  patients in total. Hence the first step is to calculate the value of  $\sqrt{N}$  to see around how many patients is optimal for a center to have and how many centers should we have.

**Step 2:** Let  $n \approx \sqrt{N}$  be optimal number of patients per center,  $K \approx \sqrt{N}$  be the optimal number of centers such that  $N = nK$ . Then the study sites with number of patients greater than or equal to  $n$ , or not much less than  $n$  will retain and suppose the number of retained centers is  $K_1$ . The remaining  $K_{prenew} = K_{pre} - K_1$  centers should be randomly grouped into  $K_2 = K - K_1$  dummy centers if  $K_{pre}$  is the number of centers before grouping.

If  $N_1 = \sum_{i=1}^{K_1} n_i = n_1 + n_2 + \dots + n_{K_1}$  is the total number of patients in the retained centers, then  $N_2 = N - N_1$  would be the total number of patients in the centers which need to be grouped. Hence,  $N_2 = \sum_{j=1}^{K_2} n_j$  and it is optimal that we can have  $N_2 = K_2 \times n_{K_2}$ , which means that all dummy centers have the same number of patients  $n_{K_2}$ . However, the exact value of  $K_2$  should be determined using the tables like Table 1-3, since we would like to maximize the power increase.

#### Scenario 2

**Step 1:** Step one here is the same as step one of Scenario 1.

	Scenario 1	Scenario 2
<b>Advantages</b>	<ul style="list-style-type: none"> <li>✓ Decrease center imbalance and increase power</li> <li>✓ More likely to achieve optimal statistical properties for treatment</li> </ul>	<ul style="list-style-type: none"> <li>✓ Decrease center imbalance and increase power</li> <li>✓ Less likely to have very different center size</li> </ul>
<b>Disadvantages</b>	<ul style="list-style-type: none"> <li>✓ Center imbalance may still exist after grouping</li> </ul>	<ul style="list-style-type: none"> <li>✓ Not achieving the possible optimal statistical properties since number of centers is not chosen to be <math>\sqrt{N}</math> (as proposed by Shao and Chow 1993)</li> </ul>

**Table 6:** Comparison between two scenarios.



**Step 2:** Let  $n \approx \sqrt{N}$  be optimal number of patients per center,  $K \approx \sqrt{N}$  be the optimal number of centers such that  $N=nK$ . Then if some study sites have number of patients much greater than  $n$ , then these centers will retain and suppose the number of retained centers is  $K_1$ . The remaining  $K_{prenew} = K_{pre} - K_1$  centers should be randomly grouped into dummy centers such that the size of study site after grouping is approximately equal to the size of retained centers so that the overall power can be increased as mentioned in the previous section. Similarly, the exact number of dummy centers should be determined using tables like (Table 3-5) in order to achieve a maximum increase in power. Following is a table comparing the advantages and disadvantages of these two methods (Table 6).

### An Example

Here is a demonstration of the proposed method for doing center grouping in multicenter trials with the existence of small centers. The example used here is a real clinical trial testing a drug against placebo for treating patients with metastatic breast cancer, where the trial is comparative, parallel-group, randomized, and double-blinded. In the study protocol, 288 patients should be recruited in approximately 43 centers so that the desired statistical power for evaluation can be achieved, and therefore each center should have 6 to 7 patients. Although the 43 centers speeded up the patient recruitment process, 7 centers had more than 10 patients in each while the other 36 centers enrolled less than 10 patients each, which cause a significant variation among centers. In consequence,

it is essential for these small centers to be grouped into dummy centers with similar size not only to address Lewis' questions [5] but also to make an unbiased and reliable assessment on safety and efficacy of the test drug.

Since 288 patients in total suggests that  $\sqrt{288} \approx 17$  would be the optimal number of patients in each center for the trial to have. As mentioned in the Section 6, the centers with more than 10 patients will remain as single centers. Secondly, grouping of the 36 centers with patients less than 10 must be put in consideration. Since among these 36 centers, 24 have patients in both treatment groups and 12 have patients in only one treatment group, we first consider grouping the 24 centers. Suppose

$\frac{\sigma_B^2}{\sigma^2} = 0.01, \frac{\sum \alpha_i^2}{\sigma^2} = 0.5$  has around 2 patients per treatment group, then as it can be seen from table 6, if the size of the dummy center is selected to be 2, i.e.,  $J = 2$ , the power will improve 11.262 %, i.e.,  $\Delta = 11.262$  %. Whereas if  $J = 3$  or  $J = 4$  is selected, the value of  $\Delta$  would be 8.005 % or 12.355 % respectively. As we aim to maximize the improvement in power, according to the table, the size of dummy center should be chosen to be 4. It is thereby suggested to group these 24 centers into 6 dummy centers randomly using the random method mentioned in Section 5 and each dummy center will then contain 4 centers selected at random from the 24 small centers. The total number of centers will be 13, and then randomly assign the patients from the 12 centers with patients in only one treatment group to the 13 centers. This example is summarized in Table 7.

**Table 7:** Example of Centre Grouping.

Centre Characteristic	Before centre grouping	After centre grouping
total # of centres	43	13
# of centres with patients in only one group	12	0
# of centres with more than 10 patients in each	7	7
# of centres with less than 10 patients in each	36	0
# of dummy centres	0	6

### Concluding Remarks

As we discussed before, the observed significant treatment-by-center interaction may just be by chance alone or may be due to the existence of large number of small centers. We thereby introduce random grouping to provide a way to check false-positive of the observed treatment-by-center interaction. Also, by random grouping one can remove heterogeneity within/across centers and therefore achieve better statistical properties (e.g. power) to make statistical inference. The two scenarios discussed in Section 7 provides the optimal ways for grouping under two different circumstances, and the selection of method should base on the actual situation of centers after the clinical trials.

There are certain limitations to the methods that we proposed. Although we aim to increase the power through combining small centers, under some circumstances, however, center grouping can also decrease the power during simulation. Also, when doing the simulation, we assume the ideal case, which is to have same number of patients in each treatment group per center before and after grouping. Nevertheless, this is usually not the case in practice, hence it is possible in some real cases that the combined centers have number of patients still smaller than the optimal number of patients that we want to have in each center, but the power after combining centers is believed to increase in most cases.

## References

1. Chow SC, Liu JP (1998) Design and Analysis of Clinical Trials. (3<sup>rd</sup> edn) John Wiley & Sons, New York, pp: 892.
2. FDA (1988) Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications. US Food and Drug Administration, Rockville, Maryland, USA.
3. ICH (1995) Guideline for Structure and Content of Clinical Study Report. International Conference on Harmonization, Yokohama, Japan.
4. Bayman EÖ, Chaloner K, Cowles MK (2010) Detecting qualitative interaction: A Bayesian approach. *Statistics in Medicine* 29(4): 455-463.
5. Lewis JA (1995) Statistical issues in regulation of medicine. *Statistics in Medicine* 14(2):127-136.
6. Siddhartha Nandy (2018) Chapter 19: Two-way ANOVA [lecture notes]. *Applied Regression Analysis STAT* 512.
7. Scheffé H (1959) *The Analysis of Variance*. Wiley, New York.
8. Shao J, Chow SC (1993) Two-stage sampling with pharmaceutical applications. *Statistics in Medicine* 12(21): 1999-2008.
9. Chow SC (2011) *Controversial Statistical Issues in Clinical Trials*. Chapman and Hall/CRC Press, Taylor & Francis, New York.
10. Chow SC, Shao J, Wang H, Lokhnygina Y (2018) Sample Size Calculations in Clinical Research. 3<sup>rd</sup> Edition, Chapman and Hall/CRC Press, Taylor & Francis, New York, 107(31-32): 552-556.
11. Sharon L, Lohr (1999) *Sampling Design and Analysis*. Duxbury Press, Pacific Grove, CA 93950.
12. Pan G, Wolfe DA (1997) Test for qualitative interaction of clinical significance. *Statistics in Medicine* 16(14): 1645-1652.