

# UK-Cubic Complete Randomized Design of Experiment: A Complete Three Way Layout for Large Multi-Center Clinical Trials

N. Uthayakumaran

NIE, ICMR, Chennai, Tamilnadu, India

**Abstract:** *In clinical trial research, if health researcher wants to know whether particular drug which has been invented is beneficial or not in the treatment of particular disease, there are many experimental designs to approach, in which most important one is Latin Square Design (LSD). The LSD has its own advantages and disadvantages. LSD is assumed to be a three factor experiment but actually it has only two-way stratification. Analysis is done with the missing blocks since it is an incomplete three way layout. In LSD, the experimental material is divided into rows and columns, each having the same number of experimental units which is equal to the number of treatments. A new design of experiment described in this paper is an attempt to extend LSD and produce a complete three way layout so that analysis is made without missing blocks and number of experimental units in blocks is equal. The number of experimental units is not equal to the number of treatments, in turn; it is possible to handle large sample size. It also controls more variation and results in a smaller mean square error. In this design, it is possible to include larger number of participants, different geographic locations, wider range of population groups and the ability to compare results among centers, all of which increase the generalizability of the multi-center clinical trial study. This design will be useful for large multi-center clinical trials. To reveal the efficiency of this design and its good randomization, it reflects the results as per the data simulated through linear model. In many cases, efficacy will vary significantly between population groups with different backgrounds, like, genetic, environmental, ethnic, etc. This design can properly evaluate efficacy by adopting these groups in dispersed trial approach.*

**Key words:** Design of experiment with a complete three way layout; Include larger number of participants; Different geographic locations; Wider range of population groups; Compare results among centers; Simulated linear model; Increase the generalizability of the large multi-center clinical trial study; Dispersed trial approach.

## 1. Introduction

A statistical design is a plan for the collection and analysis of data. Design of experiment means how to design an experiment in the sense that how the observations or measurements should be obtained to answer a query in a valid, efficient and economical way. It mainly deals with the parameters like, replication, randomization and local control. However, the selection of an efficient design requires careful planning in advance of data collection and analysis. Proceeding along these lines, the usefulness of the statistical and experimental designs for collection, analysis and inference has been demonstrated by many authors, for example see [1], [2] and [3]. The experimental designs are the way to carefully plan the experiments in advance so that results are both objective and valid. Ideally, the experimental designs should describe how participants are allocated to experimental groups [4]. There are many types of experimental designs like, Complete Randomized Design (CRD), Randomized Complete Block Design (RCBD) and Latin Square Design (LSD) etc. These experimental designs are widely used in the agricultural research. These designs are also extensively used in the fields of biology, medical, social science and business research [10]. Among these designs, LSD plays an important role and also different from CRD and RCBD which are one-way and two-way layouts respectively. LSD is a three-way layout and also controls more variation than CRD and RCBD by the simple analysis. The analysis of data in LSD is conditional in the sense it depends on which Latin square is used for allocating the treatments. LSD is an incomplete three-way layout in which each of the three factors, viz,

rows, columns and treatments, is at  $n$  levels each and observations only on  $2n$  of the  $3n$  possible treatment combinations are taken. Each treatment combination contains one level of each factor [4]. It may be noted that UK-cubic complete randomized design (UKCD) is an attempt to extend LSD and produces a complete three-way layout in which each of the factors is at  $n$  levels each and observations on all the  $3n$  levels. Each treatment contains one level of each factor.

## 2. Methods

### 2.1 UKCD- Design of Experiment

UKCD of order  $n$  is an arrangement [5], [6], [7], [8] and [9] of  $n$  symbols in  $n^3$  blocks arranged in  $n$  rows,  $n$  columns and  $n$  altitudes such that each symbol occurs only once in each row, each column and each altitude. That is, it forms the homogeneous blocks naturally. For example, to frame UKCD of order 4, choose four symbols A, B, C and D. These letters are used as symbols for treatments. Form them in an initial design layout in such a way that each of the letters out of A, B, C and D occurs only once in each row, each column and each altitude in the natural and cyclic order.

The process of randomization is done with the help of random number table. For this UKCD, select first set of random numbers randomly. Use the ranks of these random numbers to select the existing row numbers of each altitude-level from the initial design layout and arrange them in the first modified design layout accordingly.

Column numbers of each altitude-level should be randomized in a similar way, using the same procedure for the rearrangement. Use the second set of randomly selected random numbers with their ranks to select the column numbers of each altitude-level from the first modified design layout and rearrange them in the second modified design layout respectively.

Altitude numbers should be randomized in a similar way using the same procedure for further rearrangement. Use the third and final set of randomly selected random numbers with their ranks to select the altitude numbers from the second modified design layout to further rearrange them in this ranking order which is the final design layout plan of the experiment.

Once the experiment is designed properly, the depending study variable is simulated under the following linear model. Each term in the model, plus the model as a whole, is tested for its ability to account for variation in the dependent variable.

$$Y_{ijks} = (i+j+k) * t + s, \quad i, j, k, t = 1, 2, \dots, n$$

(Serial numbers of rows, columns, altitudes and treatments in the initial plan) &

$$s = 1, 2, \dots, m$$

(Serial numbers of the sampling units in the blocks)

is the simulated model to demonstrate the efficiency of the UKCD design using the illustration described below:

## 2.2 Illustration

Clinical study simulation is simulation used for clinical study management training and analysis. It is often used for supposing analysis and supporting decision-making in real clinical studies. In this illustration, simulation is developed using above defined model for the operation of Phase III multi-center clinical trial for 320 cardiac patients conducted in four districts like, District-1 (A1), District-2 (A2), District-3 (A3), and District-4 (A4) in four age groups like, 1-15 (R1), 16-30 (R2), 31-45 (R3), 46-60 (R4) at four centers like, Center-1 (C1), Center-2 (C2), Center-3 (C3) and Center-4 (C4) to compare and find the difference among the four treatments like treatment A, treatment B, treatment C and Placebo D. The depending study variable is cardiac enzyme level of cardiac patients.

Simulation is the imitation of the operation of a real-world process. This simulated model is the imitation of the operation of phase III multi-center clinical trial study using four areas, age groups and centers as confounding factors for comparing four treatments in Cardiac patients. Inputs for treatment A, treatment B, treatment C and placebo D have been given scores from 1 to 4 respectively. This simulation is the imitation to show treatment A is better than treatment B, treatment B is better than treatment C, treatment C is better than placebo D. In this simulated model, confounding factors are given scores from 1 to 4 according to the natural order of their levels, which can offer the alternative explanations for the experimental results. Randomization of UKCD design minimizes or eliminates these three confounding factors. It allows

making inference about the relationship between depending variable “cardiac enzyme level of cardiac patients” and independent variable “treatments” as per simulation perfectly. Using ANOVA, it is going to be revealed whether these treatments are differing significantly as per simulation or not for simulated data in UKCD.

The 320 cardiac patients are blocked randomly into three factors such as, 4 age-groups, 4 centers and 4 districts in such a way that four treatments randomly allocated to sixty four blocks which have five cardiac patients in each block. All treatments are equally and randomly allocated in sixty four blocks. So each treatment can be replicated in sixteen blocks (in fact eighty sampling units) as per UKCD. UKCD is a three factor experiment. It is assumed that there is no interaction between treatments, age-groups, centers and districts. A requirement of UKCD is that the number of cardiac patients (size 5) in each block is equal but no need to equal to the number of treatments (size 4). Each treatment contains one level of each factor.

This design of experiment is discussed empirically with the depending study variable exhibiting linear trend through the simulated model. The data for depending study variable that appear in the final layout (Table 1 to Table 4) are the samples of cardiac patients output that is, cardiac enzyme level of cardiac patients.

## 3. Results

### 3.1 Analysis

The designing of experiment and the analysis of obtained data are inseparable. If the experiment is designed properly keeping in mind the question, then the data generated is valid and proper analysis of data provides the valid statistical inferences. If the experiment is not well designed, the validity of the statistical inferences is questionable and may be invalid [4].

In this illustration, ANOVA is demonstrated with the independent factors such as treatments, age-groups, centers and districts (each of which has four levels) for a depending study variable like, cardiac enzyme level of cardiac patients. It examines effects of the factors, simultaneously tests for the differences among four levels of the factors. F- test tells us there is significant difference among means (Table 5).

The relative size of the two variations like (Treatments, Error), (Age-groups, Error), (Centers, Error) and (Districts, Error) is used to indicate whether the respective observed difference among the treatments, among the age-groups, among the centers and among the districts is real or due to chance. The tests of between subjects effects help to determine the significance of factors. However, they do not indicate how the levels of factors differ. The post hoc tests show the differences in model-predicted means for each pair of factor levels. To find which treatments are different, multiple comparisons test should be used to detect where the differences lie (Table 6).

In this case, there appear to be significant differences in all the treatments, that is, Treatment-A, Treatment-B and Treatment-C are significantly different from Placebo-D. The post hoc tests imply that outcome of treatments at curing cardiac patients to get well soon is correctly found because lesser cardiac enzyme level of cardiac patients significantly better than placebo for all the treatments as per simulation. That is, treatment A is better than treatment B, treatment B is better than treatment C, treatment C is better than placebo D. Treatment-A is significantly different from other two treatments and reflecting lesser cardiac enzyme level of cardiac patients.

#### 4. Discussions

ClinicalTrials.gov currently lists 236,008 studies with locations in all 50 States and in 195 countries. Clinical studies can take place in many locations, including medical centers, clinics etc. One of the common reasons for conducting clinical studies is evaluating one or more interventions for treating a disease, disorder etc.

A multi-center clinical trial is conducted at more than one medical center or clinic. Most of the large multi-center clinical trials are Phase III trials.

Generally experimental designs are used in the clinical research. In this paper, experimental design UKCD is demonstrated with the hypothetical data for the multi-center clinical trial approach. In UKCD, the experimental units are grouped according to the factors. Between treatments, between rows, between columns and between altitudes variations are eliminated from the within variation. Hence three more effects (like as three block effects) are removed from the experimental error compared to normal design. So the error variance can be considerably reduced in UKCD compared to CRD, RCBD and LSD.

A common method is CRD, where participants are assigned to groups at random. A second method is RCBD, where participants are divided into homogeneous blocks (for example, with age groups) before being randomly assigned to groups. A third method is LSD, where participants are divided into three way randomly allocated homogeneous blocks (for example, with age groups and districts).

Distribution of data in final design layout plan.

**Table 1: District-1**

Age-groups	Centers			
	C1	C2	C3	C4
1-15	D	C	B	A
	13	13	11	7
	14	14	12	8
	15	15	13	9
	16	16	14	10
	17	17	15	11
16-30	C	B	A	D
	13	11	7	29
	14	12	8	30
	15	13	9	31
	16	14	10	32
31-45	A	D	C	B
	6	25	22	17
	7	26	23	18
	8	27	24	19
	9	28	25	20
	10	29	26	21
46-60	B	A	D	C
	13	8	33	28
	14	9	34	29
	15	10	35	30
	16	11	36	31
	17	12	37	32

**Table 2: District-2**

Age-groups	Centers			
	C1	C2	C3	C4
1 -15	B	A	D	C
	9	6	25	22
	10	7	26	23
	11	8	27	24
	12	9	28	25
16-30	A	D	C	B
	6	25	22	17
	7	26	23	18
	8	27	24	19
	9	28	25	20
31 -45	C	B	A	D
	19	15	9	37
	20	16	10	38
	21	17	11	39
	22	18	12	40
46-60	D	C	B	A
	29	25	19	11
	30	26	20	12
	31	27	21	13
	32	28	22	14
	33	29	23	15

**Table 3: District-3**

Age-groups	Centers			
	C1	C2	C3	C4
1 -15	A	D	C	B
	6	25	22	17
	7	26	23	18
	8	27	24	19
	9	28	25	20
16-30	D	C	B	A
	25	22	17	10
	26	23	18	11
	27	24	19	12
	28	25	20	13
31 -45	B	A	D	C
	15	9	37	31
	16	10	38	32
	17	11	39	33
	18	12	40	34
46-60	C	B	A	D
	25	19	11	45
	26	20	12	46
	27	21	13	47
	28	22	14	48
	29	23	15	49

**Table 4: District-4**

Age-groups	Centers			
	C1	C2	C3	C4
1 -15	C	B	A	D
	19	15	9	37
	20	16	10	38
	21	17	11	39
	22	18	12	40
16-30	B	A	D	C
	15	9	37	31
	16	10	38	32
	17	11	39	33
	18	12	40	34
31 -45	D	C	B	A
	33	28	21	12
	34	29	22	13
	35	30	23	14
	36	31	24	15
46-60	A	D	C	B
	10	41	34	25
	11	42	35	26
	12	43	36	27
	13	44	37	28
	14	45	38	29

UKCD- ANOVA

**Table 5: Depending study variable: cardiac enzyme level of cardiac patients**

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Model	181380 <sup>a</sup>	13	13952.308	2001.569	.000
Treatments	22500	3	7500.000	1075.935	.000
Age-groups	2500	3	833.333	119.548	.000
Centers	2500	3	833.333	119.548	.000
Districts	2500	3	833.333	119.548	.000
Error	2140	307	6.971		
Total	183520	320			

a. R Squared = .988 (Adjusted R Squared = .988)

**Table 6: Depending study variable: cardiac enzyme level of cardiac patients:Post hoc test- Tukey HSD**

(I) Treatments	(J) Treatments	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
A	B	-7.50*	.417	.000	-8.58	-6.42
	C	-15.00*	.417	.000	-16.08	-13.92
	D	-22.50*	.417	.000	-23.58	-21.42
B	A	7.50*	.417	.000	6.42	8.58
	C	-7.50*	.417	.000	-8.58	-6.42
	D	-15.00*	.417	.000	-16.08	-13.92
C	A	15.00*	.417	.000	13.92	16.08
	B	7.50*	.417	.000	6.42	8.58
	D	-7.50*	.417	.000	-8.58	-6.42
D	A	22.50*	.417	.000	21.42	23.58
	B	15.00*	.417	.000	13.92	16.08
	C	7.50*	.417	.000	6.42	8.58

\*. The mean difference is significant at the 0.05 level.

The major advantages of LSD are to minimize or eliminate confounding variables, which can offer alternative explanations for the experimental results. Allow us to make inferences about the relationship between independent and dependent variables. Reduce variability, to make it easier for us to find differences in treatment outcomes [4].

The major disadvantages of LSD are

1) Incomplete three-way layout.

It is assumed as a three way but the treatments are grouped into replication in two-ways once in rows and in columns only. Analysis is done with the missing blocks.

2) The analysis of data in LSD is conditional.

In the sense, it depends on which Latin square is used for allocating the treatments.

3) In LSD, the experimental material is divided into rows and columns, each having the same number of experimental units which is equal to the number of treatments [4].

That is, the number of experimental units in each row and column is limited to the number of treatments which seldom exceeds five. So, it is difficult to handle large sample size.

All these disadvantages are overcome by the new experimental design UKCD.

## 5. Conclusions

To reveal the efficiency of this design and its good randomization, it reflects the results perfectly as per the data simulated through linear model.

It includes larger number of participants, different geographic locations, wider range of population groups and the ability to compare results among centers, all of which increase the generalizability of the multi-center clinical trial study. This design will be useful for large multi-center clinical trials.

In many cases, efficacy will vary significantly between population groups with different backgrounds, like, genetic, environmental, ethnic, etc. This design can properly evaluate efficacy by adopting these groups in dispersed trial approach.

Research is an area of methodological innovation. This design describes how to perform in a broader and innovative way towards generalization of the study. This meaningful design conveys with a deeper consideration of its real-life environment, whether particular drug which has been invented is beneficial or not in the treatment of particular disease.

This design may be useful in thousands of clinical trials like, phase III, multicenter, randomized, double-blind,

placebo-controlled study to evaluate the efficacy, safety, and tolerability of drug in the acute treatment of several diseases like, migraine, abdominal cancer, acute respiratory distress syndrome, adrenal cancer and TB etc.

Analysis of the randomized data of UKCD is explained through ANOVA. This approach is comparatively easy and provides a good randomization, for example, in the above illustration; it has been demonstrated with four treatments randomly allocated to sixty four blocks which have five sampling units in each block. Also provides good replication, since all treatments are equally and randomly allocated in sixty four blocks. So each treatment can be replicated in sixteen blocks (in fact eighty sampling units).

It also provides local control by adopting the following way. Firstly, it forms the homogeneous blocks naturally and cyclically then randomly allocating treatments using the process of randomization in rows, columns and altitudes. The error variance will be much smaller because of homogeneous blocks; some amount of variances of between rows, between columns and between altitudes will be parted away from the actual error variance due to the difference among them. It also provides completeness, since analysis is made without missing blocks and number of sampling units in the blocks is equal but no need to equal to the number of treatments in turn it is possible to handle large sample size.

Distribution of the randomized data in UKCD is simulated through the linear model. The hypothetical data generated through this model towards explaining the UKCD design demonstrates its ease and practical utility. UKCD is a good experimental design refers to a plan for assigning experimental units to treatment conditions. This design serves all the three purposes of a good experimental design like, causation, control and variability [4].

1. It allows the experimenter to make causal inferences about the relationship between independent variables and a dependent variable.
2. It allows the experimenter to rule out alternative explanations due to the confounding effects of extraneous variables (i.e., variables other than the independent variables).
3. It reduces variability within treatment conditions, which makes it easier to detect differences in treatment outcomes.

More research with real life data for validation of the above experimental design with relevant order will strengthen the scope of this design.

## Acknowledgements

Author is very much grateful to Dr. Manoj V. Murhekar, Director-in-charge, NIE, ICMR for his continuous encouragement in the development of this paper.

## References

- [1] Das, M.N. and Giri, N.C. (2006). Design and Analysis of Experiments. Delhi New Age International (P) Ltd. New: 2<sup>nd</sup> edition.
- [2] Leslie Kish. (1987). Statistical Design For Research. John Wiley & Sons: P33-
- [3] Montgomery, D.C. (2006). Design and Analysis of Experiments. Wiley India: 5<sup>th</sup> Edition.
- [4] Shalabh. (2013). MTH513 Analysis of variance. Teaching courses, lecture notes-4, Experimental designs and their analysis [online] available by Chapter4. anova-experimental-design-analysis.pdf-: Adobe Acrobat Reader Dc.
- [5] Uthayakumaran, N. (2015). Multi-label systematic sampling design: An unbiased estimator in the presence of linear trend. International journal of science and research: Vol. 4, Issue 8, 1343-1345.
- [6] Uthayakumaran, N. (2015). Multi-circular systematic sampling design. International journal of science and research: Vol. 4, Issue 9, 29-32.
- [7] Uthayakumaran, N. (2015). Weights proportional to size (WPS) Multi-systematic sampling design. International journal of science and research: Vol. 4, Issue 9, 834-837
- [8] Uthayakumaran, N. (2015). Cuboidal systematic sampling method: An unbiased estimator in the presence of linear trend. International journal of statistics: Vol. 39, Issue 2, 1139-1145.
- [9] Uthayakumaran, N. and Venkatasubramanian, S. (2015). An alternate approach to multistage sampling: UV Cubical circular systematic sampling method. International journal of statistics and applications: 5(5), 169-180.
- [10] Vishwanth Reddy, S. (2013). ANOVA & experimental designs slide shares. Statistical analysis [online], available at <http://www.slideshare.net/vishwanth555/anova-experimental-designs>.

## Author Profile



**Dr. N. Uthayakumaran** received his Ph.d. (Statistics) from Madras University, Chennai, Tamilnadu, India. Presently he is a Principal Technical Officer at NIE, ICMR, Chennai and has published many research articles in Internationally Reputed Journals.