WORLD JOURNAL OF PHARMACEUTICAL RESEARCH
SJIF Impact Factor 8.074
Volume 8, Issue 5, 865-877. Review Article ISSN 2277– 7105

DESIGN OF EXPERIMENT: A NOVEL AND SYSTEMATIC APPROACH OF DEVELOPING FORMULATIONS IN PHARMACEUTICALS

Mahima Mathur and V. Kusumdevi*

Department of Pharmaceutics, Al-Ameen College of Pharmacy, Near Lalbagh Main Gate, Hosur Road, Bangalore-560027. Karnataka, India.

Article Received on
15 Feb. 2019,
Revised on 06 Mar. 2019,
Accepted on 27 Mar. 2019
DOI: 10.20959/wjpr20195-14711

***Corresponding Author**
V. Kusumdevi
Department of Pharmaceutics, Al-Ameen College of Pharmacy, Near Lalbagh Main Gate, Hosur Road, Bangalore-560027. Karnataka, India.

ABSTRACT

Quality by design (QbD) is a new modern perspective towards the qualitative pharmaceutical development. Many pharmaceutical companies have used several Quality Management System (QMS) for instance ISO 9001. This is a systematic approach to design and develop the pharmaceutical formulations and manufacturing processes that ensures the predefined product quality. In order to achieve this, DoE should be adopted, which confirms a structured and organized method for determining the relationships among factors affecting a process and its output. A DOE evaluates effects of the design factors on manufacturability and Critical Quality Attributes (CQAs) of the final product, thereby establishing a design space to validate desired CQAs, which includes critical material and process parameters. It has been suggested that DoE can offer returns that are four to eight times greater than the cost of running the experiments in a fraction of time when compared to one-factor-at-a-time (OFAT) experiments. In the present review, the authors have reviewed various design of experiments required for the screening and optimization of the various factors. They have mentioned that it also helps to indicate the various resources and assists in managing expectations from a study's outcome. DoE studies in support of QbD are often a delicate balance between delivering defined, high-quality products and meeting predetermined time, labor, and financial constraints.

KEYWORDS: Quality by design (QbD), Design of experiment (DoE), Process analytical technologies (PAT), One factor at a time (OFAT), Quality risk management (QRM), Critical

process parameters (CPP), Analysis of variance (ANOVA).

INTRODUCTION

Statistical design of experiments (DoE) is a powerful approach to optimize chemical processes.[1] Over the past 11 years the application of DoE to pharmaceutical process development has greatly increased[2], as measured by the volume of DoE-related publications. The changing R&D environment, recent technological advances, fickle drug substance regulatory environment and increased implementation of green chemistry principles are some of the reasons for this upward trend in DoE implementation. Overcoming the perceived barriers in implementing DoE in the process chemistry setting can lead to wider adoption of this powerful tool. This concept of applying statistical analysis during the planning stages of research rather than at the end of experimentation was introduced by Sir Ronald Fisher in the beginning of the twentieth century. When statistical thinking is applied from the design phase, it enables to build quality into the product, by adopting Deming's profound knowledge approach, comprising system thinking, variation understanding, theory of knowledge, and psychology. The traditional optimization approach, varying one factor/variable at a time (OFAT, also called OVAT)[3] suffers from many drawbacks. First, it rarely uncovers the optimal conditions in large part due to the fact that the outcomes are highly dependent on the starting point. Secondly, the OFAT approach is unable to separate the “noise” (the inherent run-to-run variation of a system) of a reaction from actual improvement unless a significant number of reactions are repeated using the same conditions.[2,3]

In the early stages, the pharmaceutical industry heavily focused on blockbuster drugs, and formulation development was mainly performed by One Factor at a Time (OFAT) studies, rather than implementing Quality by Design (QbD) and modern engineering-based manufacturing methodologies. Among various mathematical modeling approaches, Design of Experiments (DoE) is extensively used for the implementation of QbD in both research and industrial settings. The systematic approach inherent in DoE eliminates researcher bias and often leads to reaction conditions that one had not considered previously. But the most significant advantage of using DoE is the ability to quickly detect how interactions between factors can affect product yield and quality. Also, by simultaneously varying parameters, DoE approach can be more efficient than that achieved by the traditional approach of varying one factor/variable at a time.[4] For instance, if OFAT approach was used to investigate the influence of three factors on a reaction (temperature, concentration, and reagent

stoichiometry) eight experiments would be required; but more information could be generated through four experiments in a half-factorial DoE. Running a DoE may seem daunting initially since the number of experiments to be run is defined at the beginning, unlike the traditional approach. But even if the effort to perform the multiple reactions and assays engendered by a DoE is time-consuming, the *quality* and thoroughness of the information obtained outweighs the effort. In QbD, product and process understanding is the key enabler of assuring quality in the final product. Knowledge is achieved by establishing models correlating the inputs with the outputs of the process.[1] Unlocking the full potential of a DoE is predicated on systematically proceeding through a prescribed work flow. This includes defining the objectives of the experiment, followed by defining the factors and variables along with their ranges using existing process knowledge. This leads to the assigning high and low settings for the factors selected for inclusion in the study. These settings need to be relevant, achievable and practical. The end product of this exercise is creation of the *design space*. The third step is to define the responses, which are measurable outcomes of the process. Accuracy and reproducibility of experimental technique and assays are of paramount importance in order to achieve meaningful results. The next step involves selection of the experimental design which depends on the objective, number of factors and resources available. The most common are screening designs in which *qualitative* information about the relevant factors is obtained. These designs also enable the factors to be ranked in order of extent of impact on the response. Screening designs are frequently employed to weed out the irrelevant factors to enable focusing on the most relevant factors in a second DoE design, also known as optimization designs.[4] The subsequent step involves the generation of reaction worksheet and data collection from the software used. Ideally, the reactions are all performed at the same time but often this is impractical, thus it is best have the software randomize the run order as a means of mitigating any systemic bias or error. It is critical to perform the reactions under identical experimental and analytical conditions. Further, the results are analyzed by using a mathematical model. It gives the data such as adjusted and predictive R-squared terms and a theoretical set of reaction conditions within the design space.[4, 5]

Advantages of implementing DOE includes better innovation due to the ability to improve processes, more efficient technology transfer to manufacturing, reduced batch failures, greater regulator confidence of robust products, risk- based approach and identification, innovative process validation approaches and greater product consistency.[5]

There are different types of variables which are usually involved in DoE. Design variables are factors with controlled variations which are also called as design variables or factors. Design variable is completely defined by its name and type.[6] Response variables are non- designed variable and are usually measured as the outcome. Non-controllable variable are second type of non designed variables which may have an influence on the response variables and cannot be controlled or reliably fixed to a value e.g. Air humidity or Temperature. Continuous variables have a numerical value and can be measured quantitatively. E.g. Temperature, Concentration of ingredients. In this, variations are usually set within predefined range which goes from the lower to the higher. More levels between the extremes may be specified if the values are to be studied more specifically, If only 2 levels are specified-other necessary levels will be computed automatically. Category Variables are non continuous variables are category variables. Their level can be named, but not measured quantitatively. Binary variables are special types of category variables that have only 2 levels-dichotomous. When concentrations of various components are required to be mixed in different ratios and cannot be varied independently from each other, a special type of design is used which is known as the mixed design and the variables used for the same re known as mixed variables. In a mixed design, one may also want to investigate the effects of variations in some other design variables which are not the components of the mixture. Such variables are called process variables. E.g. Temperature, Stirring rate, type of solvent, amount of catalyst.[6]

Types of design of experiments

Full-Factorial Design

Factorial experiments with two-level factors are used widely because they are easy to design, efficient to run, straightforward to analyze, and full of information. A full factorial design contains all possible combinations of a set of factors. This is the most fool proof design approach, but it is also the most costly in experimental resources. The full factorial designer supports both continuous factors and categorical factors with up to nine levels. Factorial designs with only two-level factors have a sample size that is a power of two (specifically 2 where f is the number of factors). When there are three factors have a sample size that is a power of three.[2]

$N = {}_{L}k$

Where k = number of variables, L = number of variable levels, N = number of experimental trials.

Fractional Factorial Design

Specific cases- with 2 level variables (continuous with upper and lower levels, and /or binary variables)-one can define fraction of full factorial design. It enables the investigation of as many design variables are chosen full factorial design with fewer experiments. These designs might be set up by selecting half the experimental runs of the original design. An example of fractional factorial design is given here, where four design variables-A, B, C, D are used and the lower and upper levels are coded '-'and '+'respectively.[7] First the full factorial design is built with only 3 variables A, B & C (2^3) as shown in table 1.

Table 1: Full factorial design is built with only 3 variables.

Experiment	A	B	C
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+

Full Factorial Design 2^3

In the table 2, additional columns are generated, which are computed from the products of the original 3 columns A, B, C given in table 1. These additional columns represent the interactions between the design variables, which are the crux of full factorial design 2^3.[7]

Table 2: Description of full factorial design 2^3

Experiment	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

Another concept involved in full factorial design is confounding, which refers to some effects cannot be studied independently of each other. Hence resolution of full factorial design is based on this factor also. The 3 most common cases are as follows.

- Resolution III Design : Main effects are confounded with two-factor interaction

- Resolution IV Design: Main effects are free of confounding with two-factor interaction, but two factor interactions are confounded with each other.
- Resolution V Design: Main effects and two-factor interactions are free of confounding with each other; however some two factor interactions are confounded with three factor interactions.

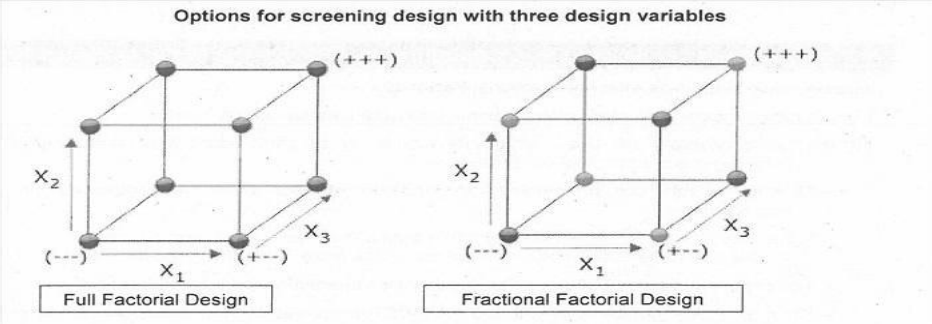


Figure 1: Description of full factorial design and fractional factorial design.

Plackett-Burman (PB) designs

Plackett-Burman (PB) designs are used for screening experiments because, in a PB design, main effects are, in general, heavily confounded with two-factor interactions. The PB design in 12 runs, for example, may be used for an experiment containing up to 11 factors. In this experimental objective should be the study of Main effects only. It is very economical as they require only 1 to 4 more experiments than the number of design variables.[8] The interaction between the factors is considered negligible. A simple example of placket burman design in given in table 3.

Table 3: Plackett-Burman design for 12 runs and up to 11 two-level factors.

Run	A	B	C	D	E	F	G	H	I	J	K
1	+	-	+	-	-	-	+	+	+	-	+
2	+	+	-	+	-	-	-	+	+	+	-
3	-	+	+	-	+	-	-	-	+	+	+
4	+	-	+	+	-	+	-	-	-	+	+
5	+	+	-	+	+	-	+	-	-	-	+
6	+	+	+	-	+	+	-	+	-	-	-
7	-	+	+	+	-	+	+	-	+	-	-
8	-	-	+	+	+	-	+	+	-	+	-
9	-	-	-	+	+	+	-	+	+	-	+
10	+	-	-	-	+	+	+	-	+	+	-
11	-	+	-	-	-	+	+	+	-	+	+
12	-	-	-	-	-	-	-	-	-	-	-

Taguchi Orthogonal Array Model

Taguchi has envisaged a new method of conducting the design of experiments which are based on well defined guidelines. This method uses a special set of arrays called orthogonal arrays. These standard arrays stipulate the way of conducting the minimal number of experiments which could give the full information of all the factors that affect the performance parameter. The crux of the orthogonal arrays method lies in choosing the level combinations of the input design variables for each experiment. While there are many standard orthogonal arrays available, each of the arrays is meant for a specific number of independent design variables and levels. For example, if one wants to conduct an experiment to understand the influence of 4 different independent variables with each variable having 3 set values (level values), then an L9 orthogonal array might be the right choice. The L9 orthogonal array is meant for understanding the effect of 4 independent factors each having 3 factor level values. This array assumes that there is no interaction between any two factors. While in many cases, no interaction model assumption is valid, there are some cases where there is a clear evidence of interaction. A typical case of interaction would be the interaction between the material properties and temperature. The orthogonal arrays have the following special properties that reduce the number of experiments to be conducted.[9]

Response Surface Design

Response surface design is a set of advanced design of experiments (DoE) techniques that help in better understanding and optimize response. Response surface design methodology is often used to refine models after important factors using screening designs or factorial designs are determined. The difference between a response surface equation and the equation for a factorial design is the addition of the squared (or quadratic) terms that allows the model curvature in the response, making them useful for understanding or mapping a region of a response surface. Response surface equations model how changes in variables affect a response of interest. It also helps in finding the levels of variables that optimize a response and selecting the operating conditions to meet specifications. There are two main types of response surface designs namely central composite designs and box behnken designs.[10]

Central Composite designs (CCD)

Central Composite designs can fit a full quadratic model. They are often used when the design plan calls for sequential experimentation because these designs can include information from a correctly planned factorial experiment. A central composite design is the

most commonly used response surface designed experiment. Central composite designs are a factorial or fractional factorial design with center points, augmented with a group of axial points (also called star points) which helps in estimating the curvature. A central composite design efficiently estimate first- and second-order terms and can model a response variable with curvature by adding center and axial points to a previously-done factorial design. Central composite designs are especially useful in sequential experiments.[11]

As the position of the response surface optimum is not known; it is ensured that the prediction error is same for any point at the same distance from the centre of the design, which is known as rotability.[7] Depending on the constraints of the experiments and the accuracy to achieve, select the appropriate CC design using the following table 5.

Table 4: Different types of CCD.

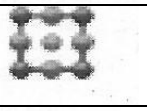
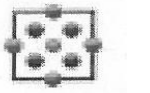
Circumscribed Central Composite Design (CCD)	
Faced Central Composite Design (CCF)	
Inscribed Central Composite Design (CCI)	

Table 5: Different designs with different number of levels with accuracy of estimates.

Design	Number of levels	Uses point outside high and low variables	Accuracy of estimates
Circumscribed	5	Yes	Good over entire design space
Inscribed	5	No	Good over central subset of the design space
Face centered	3	No	Fair over entire design space and poor for the quadratic coefficients
Box Behnken	3	No	Good for the entire design space, more certainly on the edge of the design area

Box-Behnken designs

A Box-Behnken design is a type of response surface design that does not contain an embedded factorial or fractional factorial design. Box-Behnken designs usually have fewer design points than central composite designs, thus, it is less expensive to run with the same number of factors. It can efficiently estimate the first- and second-order coefficients; however, it can't include runs from a factorial experiment. Box-Behnken designs always have

3 levels per factor, unlike central composite designs which can have up to 5. Also unlike central composite designs, Box-Behnken designs never include runs where all factors are at their extreme setting, such as all of the low settings. Box-Behnken designs have treatment combinations that are at the midpoints of the edges of the experimental space and require at least three continuous factors.[12] The following figure 2 shows a three-factor Box-Behnken design. Points on the diagram represent the experimental runs that are done.

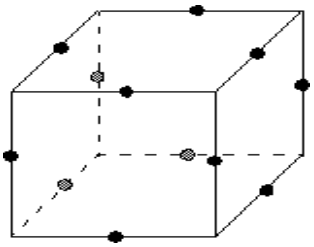


Figure 2: Three-factor Box-Behnken design.

These designs allow efficient estimation of the first- and second-order coefficients. Because Box-Behnken designs often have fewer design points, this can be less expensive to perform than central composite designs with the same number of factors. However, because it is not have an embedded factorial design, it is not suited for sequential experiments. Box-Behnken designs can also prove useful if there are safe operating zone for the process. Central composite designs usually have axial points outside the "cube." These points may not be in the region of interest, or may be impossible to conduct because it is beyond safe operating limits. Box- Behnken designs do not have axial points, thus, it can be ensured that all design points fall within the safe operating zone. Box-Behnken designs also ensure that all factors are not set at their high levels at the same time.[13]

Choice of experimental design

The most important part of a DoE process, choosing an appropriate experimental design, is critical for the success of the study. The choice of experimental design depends on a number of aspects, including the nature of the problem and/or study (e.g., a screening, optimization, or robustness study), the factors and interactions to be studied (e.g., four, six, or nine factors, and main effects or two-way interactions), and available resources (e.g., time, labour, cost, and materials). Using previous knowledge of a product or previous experiments to identify possible interactions among the input process parameters before performing an experiment also plays a key part in selecting an appropriate experimental design.[14]

Statistical Analysis and softwares

Once data have been collected according to the chosen design, the results should be analyzed using statistical methods so that objective conclusions can be drawn. Many software packages are available to assist, including those that help users choose a design to those that perform statistical analysis, report results, and generate a mathematical model. One such model is the ANOVA, which a statistical method is based on the F- test to assess the significance of model terms. Once the appropriateness of those terms and the overall model satisfies an ANOVA check, the next step is to determine what cannot be modelled. This is done by residual analysis.[14]

Good DoE software helps users follow the regressive modeling approach. It should guide them in carefully choosing model terms on the basis of graphical tools and statistics, and it should verify a model and its significance based on statistics in addition to verifying unaccounted residuals. Graphical tools play a key part in understanding and presenting statistical analysis results, hence, smart way to diagnose, analyze, predict, and present the results in two and three dimensions should be known. A systematic application of DoE facilitates the identification of CPPs and their relationship to CQAs, leading to the development of a design space. In combination with quality risk management (QRM) and process analytical technologies (PAT), these help companies maintain good manufacturing control and consistency, ultimately guaranteeing the quality of their drug products [15]. The summary of all the design of experiments discussed in this paper is given in table 6.

Table 6: Summary of the designs available.

Types of design	Screening	Factor influence	Optimization	Field of use	No. of designs available
Full factorial design	X	X		Study the effect of a lower number of design variables independently from each other, including interaction terms. The only design that allows for categorical variables with 3 or more levels	2-9
Fractional factorial design	X	X		Depending on the number of variables choose the study lower order effects independently from each other, or create a screening design aimed at finding the most important main effects among	3-13

				many	
Plackett-Burman Design	X			Economical, alternative to factorial design, studies main effects only	8-35
Taguchi orthogonal array	X			Study main effects only. Can be used only for screening	2-7
Central composite design			X	Can find optimal levels of design variables by adding more experiments to a full fractional design. All design variables must be continuous	2-6
Box- Behnken design			X	An alternative to central composite designs, when the optimum response is not located at the extremes of the experimental region and when previous results from a factorial design are not available. All design variables must be continuous	3.-6
D-optimal Design	X	X	X	Some design variables have multilinear constraints, and design is not orthogonal. Must be analysed with Partial least Squares Regression	2-9
Axial (Mixture) design	X			Contains mixture variables only, design region is simplex. Only linear (first order) effects can be bound	3-30
Simplex-Lattice (Mixture) Design	X	X	X	Contains mixture variables only, design region is simplex. Tuneable lattice degree (order)	3-6 (9 if linear only)
Simplex Centroid (Mixture) design			X	Contains mixture variables only, design region is simplex	3-6

CONCLUSION

It is an essential requirement by the regulatory bodies all over the world to provide the data of experimentation in a systematic manner. The most appropriate way of demonstrating the same in given by implementing the design of experiments. This has not only defined the statistical optimization for pharmaceutical scientist in formulating the product with optimum characteristics but also provide solutions to larger-scale manufacturing problems, which occasionally arise. The authors have tried to collectively understand and interpret the aspects of various DoEs in order to come out with favourable outcomes, economically and effectively.

REFERENCES

1. Politis S, Colombo P, Colombo G, Rekkas D. Design of experiments (DoE) in pharmaceutical development. Drug development and Industrial Pharmacy, 2017; 43(6): 889-901.
2. Chowdary K.P.R., Ravi Shankar K., Optimization of pharmaceutical product formula-tion by factorial designs: case studies, Journal of pharmaceutical research, 2016; 15(4): 105-109.
3. Weissman S. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications.*Org. Process Res. Dev*, 2015; *19* (11): 1605–1633.
4. Ranga S et al. A Review on Design OF Experiments (DOE). International Journal of Pharmaceutical and Chemical Sciences, 2014; 3(1): 216-224.
5. Satish LK, Smita R and Rohit V. Quality by design: Facilitate A Robust Pharmaceutical Process. Journal of Pharmacy Research, 2011; 4: 2714- 2743.
6. Shivare M, Mc Creath G, Practical consideration for DOE Implementation in Quality by design, Bioprocess, Technical, 2010.
7. Gujral G, Kapoor D, Jaimini M. An updated review on design of experiment (DOE) in Pharmaceuticals. Journal of Drug Delivery and Therapeutics, 2018; 8(3): 147-152.
8. Noushine B and Naghmeh H. Formulation and optimization of captopril sublingual tablet using D Optimal design. Iranian journal of pharmaceutical research, 2008; 7: 259-267.
9. Kaacker RN, Lagergren ES, Filliben JJ. Taguchi's Orthogonal Arrays are classical designs of experiments. Journal of Research and National Institute of standard Technology, 1991; 96(5): 577–591.
10. Wang Y, Deng L, Fan Y. Preparation of Soy-Based Adhesive Enhanced by Waterborne Polyurethane: Optimization by Response Surface Methodology. Advances in Materials Science and Engineering Volume 2018, Article ID 9253670, 8 pages https://doi.org/10.1155/2018/9253670
11. Asadzadeh F, Maleki-Kaklar M, Soiltanalineiad N, Shabani F. Central Composite Design Optimization of Zinc Removal from Contaminated Soil, Using Citric Acid as Biodegradable Chelant. Scientific Reports, 2018; 8; https://doi.org/10.1038/s41598-018-20942-9.
12. Sharma V, Kumar V. Application of Box-Behnken design and response surface methodology for multi-optimization of laser cutting of AA5052/ZrO_2metal−matrix composites, 2016; https://doi.org/10.1177/1464420716642619

13. Ye G, Ma L, Li L, Liu J, Yuan S, Huang G. Application of Box–Behnken design and response surface methodology for modeling and optimization of batch flotation of coal. International Journal of Coal Preparation and Utilization, 2017; https://doi.org/10.1080/19392699.2017.1350657
14. Role of Statistics in Pharmaceutical Development Using Quality-by-Design Approach – an FDA Perspective by Chi-wan Chen, PhD and Christine Moore, Ph.D. office of New Drug Quality Assessment CDER/FDA.
15. Sumit K and Shikha T. A Quantitative Approach for Pharmaceutical Quality by design patterns. Invention Rapid: Pharm Analysis & Quality Assurance, 2012; 4: 1-8.