

# What to Look for in Statistical Software for the Pharmaceutical Industry

Statistical methods are becoming increasingly important for the pharmaceutical industry. The FDA and other regulatory and standard-setting organizations are moving swiftly to establish Quality by Design (QbD) guidance relevant to the needs of pharmaceutical manufacturing. The FDA suggests the use of design of experiments (DOE) because “it provides a structured, organized method for determining the relationship between factors affecting a process and the response of that process.”

While it is possible to perform DOE with general statistical software, most users in the pharmaceutical industry should look for software designed specially for DOE because it is generally much easier for non-statisticians to use. This article identifies a number of key capabilities that pharmaceutical companies should look for in selecting DOE software that will best meet their needs.

## QbD and DOE background

QbD revolves around the concept of the design space, which is defined by the FDA as the “multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have demonstrated to provide assurance of quality. Working within the design space is not considered as a change. Movement out of the design space is considered to be a change and would normally initiate a regulatory post approval change process. Design space is proposed by the applicant and is subject to regulatory assessment and approval.”

The FDA recommends that the design space be determined through a combination of three methods. The first principles approach, combining experimental data and mechanistic knowledge, is used to model and predict process performance. DOE is used to determine the impact of multiple factors and their interaction. Finally scale-up correlation is used to translate operating conditions between different scales or types of equipment.

Compared to the traditional one-factor-at-a-time (OFAT) method, DOE drastically reduces the number of runs required to determine the optimal value of each factor. This is done by varying the values of all factors in parallel. This approach determines not just the main effects of each factor but also the interactions between the factors, which is often the key to big breakthroughs. To best meet the needs of pharmaceutical industry engineers and scientists, DOE software should provide the following attributes and features:

## Designed for use by technologists as opposed to statisticians

DOE software should eliminate the need for statistical expertise on the part of the users by walking the user through the complete process. For example, the software might

prompt the user to enter the factors and responses and select the type of experiment while providing information that will help the user pick the best type. The software will then generate a randomized list of experimental runs. As each run is completed in the order given, the results are entered into the software. The software then generates tabular and graphical data that helps define the region where quality product is produced.

### Wide range of experimental design choices

There are many different types of experimental designs, each of which offers advantages for certain applications. DOE software should enable the user to easily choose from a range of experimental designs. For example, mixture experiments are useful in many pharmaceutical applications. A typical mixture experiment might be used to investigate the effect of changing the proportions of polymer, drug and three excipients on four responses in a sustained release tablet based on a hydrophilic polymer.

### Put bounds on the parameters to reduce the number of experimental runs

It should be possible to enter bounds on parameters in order to avoid wasting time performing experiments that technologists know are not practical for various reasons. DOE software should make it easy to define the design space by entering low and high values for components. In the case of the mixture experiment mentioned above we will define the limits of the components for our hydrophilic tablet mixture design as shown in the table below.

Name	Low	High
Lactose	5	42
Phosphate	5	47
Cellulose	5	52
Polymer	17	25
Drug	1	2
Total = 100 wt %		

### The software should automatically set up the experimental design

In this case, we will assume that a quadratic polynomial, which includes second order terms for curvature, will adequately model the responses. We will pick:

- 15 model points using the optimality criteria
- 5 lack of fit points using distance as the criteria
- 5 replicates using the optimality criteria

There are four responses

Name	Units
T(50%)	h
Shape	factor
Hardness	kP
Friability	%

The software automatically designs an experiment with 25 runs as shown in the table below.

Select	Std	Run	Component 1 A:Lactose wt %	Component 2 B:Phosphate wt %	Component 3 C:Cellulose wt %	Component 4 D:Polymer wt %	Component 5 E:Drug wt %	Response 1 t(50%) h	Response 2 shape factor	Response 3 hardness kP	Response 4 friability %	
	18	1	30.000	5.000	46.000	17.000	2.000	4.531	1.576	8.042	0	
		13	2	5.000	24.333	52.000	17.000	1.667	7.563	1.517	6.613	0.7824
		21	3	22.667	40.000	10.333	25.000	2.000	7.978	1.542	8.856	1.896
		16	4	17.000	47.000	17.000	17.000	2.000	6.018	1.679	7.224	1.39
		14	5	14.167	36.417	24.667	23.000	1.750	9.163	1.34	8.39	0
		15	6	17.000	47.000	17.000	17.000	2.000	4.661	1.685	6.098	0.5227
		23	7	33.333	12.000	27.667	25.000	2.000	9.207	1.399	9.472	0
		5	8	42.000	29.667	5.000	22.333	1.000	7.887	1.414	8.303	4.24
		22	9	5.000	47.000	21.000	25.000	2.000	10.45	1.532	10.54	0
		1	10	42.000	20.000	20.000	17.000	1.000	6.269	1.316	8.24	0
		2	11	22.333	5.000	52.000	19.667	1.000	6.14	1.167	6.48	1.471
		17	12	30.000	5.000	46.000	17.000	2.000	5.612	1.413	9.174	0
		9	13	29.667	47.000	5.000	17.000	1.333	6.327	1.659	6.09	4.337
		12	14	5.000	24.333	52.000	17.000	1.667	8.265	1.549	6.483	0
		6	15	5.000	32.000	37.000	25.000	1.000	14.12	1.046	3.971	0
		19	16	42.000	30.000	5.000	21.000	2.000	3.938	1.172	6.2	4.279
		8	17	20.167	15.417	40.167	23.000	1.250	8.989	1.235	4.132	0.8854
		4	18	42.000	29.667	5.000	22.333	1.000	8.416	1.475	6.43	3.776
		24	19	29.000	5.000	39.000	25.000	2.000	7.637	1.378	9.829	0
		3	20	5.000	47.000	26.000	21.000	1.000	11.61	1.26	9.436	1.085
		7	21	29.667	33.417	16.667	19.000	1.250	8.181	1.489	7.352	1.765
		25	22	8.667	12.333	52.000	25.000	2.000	9.836	1.301	5.085	0.3608
		11	23	23.333	25.833	28.333	21.000	1.500	8.233	1.392	7.559	0.422
		20	24	42.000	30.000	5.000	21.000	2.000	4.559	1.286	7.785	4.551
		10	25	42.000	5.000	29.333	22.333	1.333	5.693	1.266	8.686	1.624

Design of experiment with 25 runs

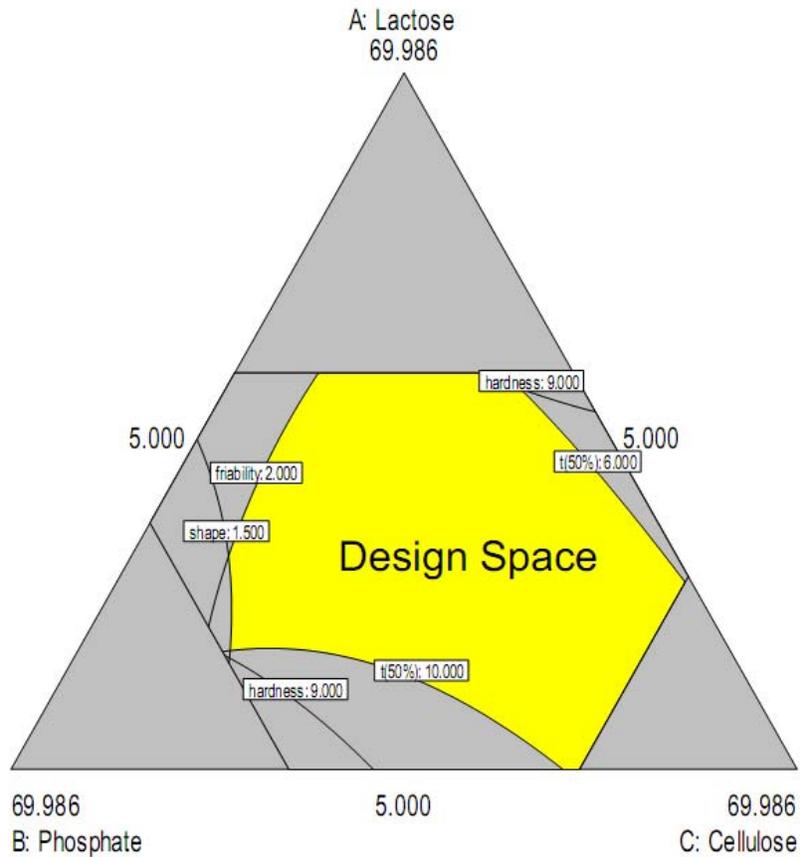
### Ability to optimize the factors based on a combination of responses

State-of-the-art DOE software allows the user to optimize the values of the factors based on the goals for each of the responses. DOE software typically allows for five possible goals in constructing an index of desirability: maximum, minimize, target, in range and equal to. Desirabilities range from zero to one for any given response. The program will combine individual desirabilities into a single number and then search for the greatest overall desirability. A value of 1 provides the ideal case. For example, in the chart below, the desirability of T(50%) is 1 at a value of 8 and the desirability ramps down to 0 as it moves from the target value to the low and high values of 6 and 10 respectively.

Response	Goal	Low	High
T(50%)	Target = 8	6	10
Shape	Minimize	6	1.5
Hardness	Target=6	4	9

Friability	minimize	1	2
------------	----------	---	---

Now we will optimize the model using the criteria shown in the table above for two drug dosages: regular strength = 1 wt% and extra strength at 2 wt%. The chart below shows the optimized values for each factor superimposed on an overlay plot of the factors analyzed in the study. The overlay window shows the design space which indicates the various combinations of the factors that will provide results within the acceptable range.



DOE software maps out design space

**Framing of design space to account for confidence prediction and tolerance intervals**

