Statistical Design of Experiments on Fabrication of Starch Nanoparticles – A Case Study for Application of Response Surface Methods (RSM)

Nadeem Irfan Bukhari,[1]* Simran Kaur,[1] Saringat H. Bai,[2] Yuen Kah Hay,[2] Abu Bakar Abdul Majeed,[3] Yeow Beng Kang,[1] and Mark J. Anderson[4]

[1]*School of Pharmacy International Medical University, Kuala Lumpur, Malaysia*

[2]*School of Pharmaceutical Science, University Sains Malaysia, Penang, Malaysia*

[3]*Faculty of Pharmacy, University Technology Mara, Kuala Lumpur, Malaysia*

[4]*Stat-Ease, Inc., Minneapolis, Minnesota, United States of America*

Abstract

Statistical design of experiments (DoE), a computer-aided optimization technique, provides many benefits:

- identifies critical factors,
- reveals their interactions – synergistic or antagonistic,
- finds ideal process conditions that accomplish the targeted response(s).

The DoE approach facilitates screening of process variables, seeking out robust conditions and testing the recommended settings for ruggedness.

Optimization of a non-linear process response, for example, one that follows a rising ridge in relation to the key factors, requires a more advanced statistical technique called "response surface methods," or RSM.

DoE in all its forms has been widely used in pharmaceutical formulation and research. It requires proper selection of the experiment design and good statistical analysis to produce a useful, validated predictive model.

Screening of available literature revealed a number of potential perils for novices in this methodology:

- inappropriate selection of model,
- inclusion of non-significant terms into the model,
- use of raw response data when a simpler and more precise model would be obtained in a mathematically transformed scale, such as logarithmic.

This leads to over- or under-prediction and thus, reduced reliability of the DoE approach.

Keeping all of these points in view, this report provides education by example – the empirical modeling of a fabrication process for producing nanoparticles from a natural polymer.

Key words: Design of experiments, DoE, response surface methods, RSM, statistical optimization, nanoparticles, process modeling, mixture design, optimal formulation.

Corresponding author: Nadeem Irfan Bukhari, email nadeem_irfan@hotmail.com

Introduction

The traditional approach to drug-design experimentation requires that only one factor at a time (OFAT) be changed while keeping all other variables constant. This OFAT approach has many major flaws, but the two most egregious ones are:

(1) It cannot assess factor interactions, which in pharmaceutical processes must be anticipated,

(2) It covers a small fraction of the total feasible factor space (Anderson, 2005).

Statistical design of experiments (DoE), a matrix-based multifactor method, does measure interaction effects and it encompasses the entire multidimensional experimental region (Anderson and Whitcomb, 2007). Aided by software programmed for this purpose, DoE has become recognized as an important tool for more rapid pharmaceutical process and product development (Bolton and Bon, 2004).

Figure 1 illustrates a generally accepted strategy for DoE aimed at rapid, yet comprehensive, process improvement. It begins with a "Discovery" phase that calls for brainstorming and pilot studies that include range-finding trials. Many experimenters jump too quickly into a test matrix and end up wasting time on the wrong factors with ranges that are either too narrow, thus generating nothing significant, or too wide – going beyond operational limits. Then they must back up and start again by attempting to re-discover a proper starting point for their process and product development.

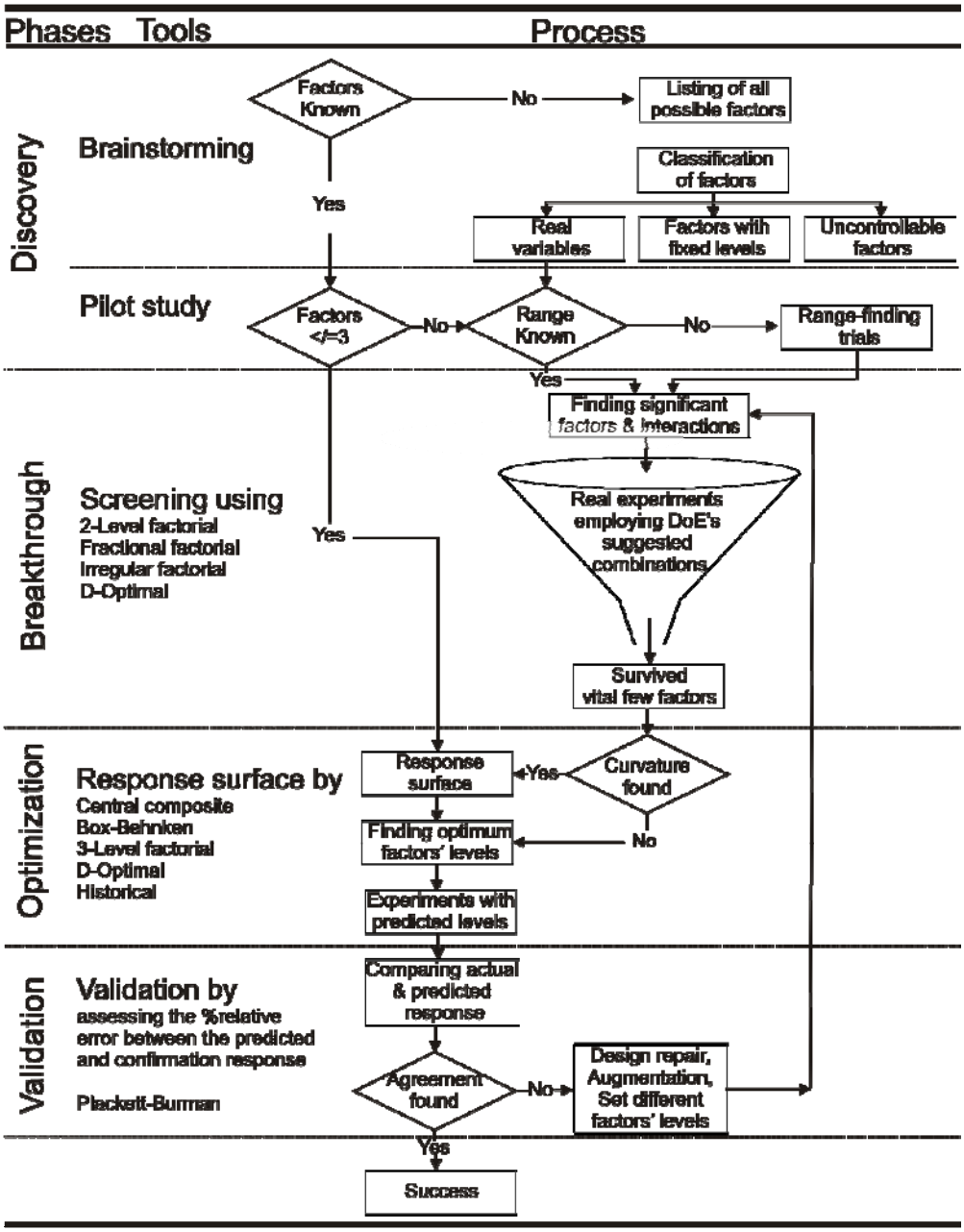

Figure 1: Strategy of experimentation

(Adapted and modifed from Anderson and Whitcomb, 2005)

The "Breakthrough" designs listed in Figure 1 vary all factors simultaneously via cleverly-devised matrices that compute effects with maximal power for predictive modeling. In fact, for a given level of statistical power, modern DoE methods require far fewer experimental runs than the OFAT approach (Anderson, 2005; Lewis, et al, 1999).

Designs listed under the "Optimization" section of Figure 1 are known as response surface methods (RSM), the basis of which is polynomial models fit by least-square regression and confirmed statistically via analysis of variance (ANOVA) (Anderson and Whitcomb, 2005). By deriving empirical, but functionally useful, relationships – typically second-order – between process responses and the critical input variables, RSM facilitates discovery of 'sweet spots' where all requirements can be achieved at lowest cost.

The ultimate goal of DoE is to construct a useful predictive model of all critical response measures of process efficiency and product efficacy. Armed with these polynomial equations, specialized software can apply numerical search algorithms that find the most desirable conditions (Vaughn, et al, 2007). However, this recommendation must be validated via confirmatory tests as detailed in the final stage of the strategy for experimentation outlined in Figure 1. Ideally, the percent relative error between the predicted and validation output falls within ±15% (Wayne, 1991). Also, standard protocols provide templates for establishing the ruggedness of any given system against variations – both internal, for example variations in set points, and external, such as ambient humidity (ASTM, 2007).

Current literature on pharmaceutical dosage-form design and research is replete with the use of DoE. Unfortunately, screening of available literature revealed frequent inappropriate use of the approach, such as:

a) Models not appropriate for their purpose, for example using ones geared for process variables when the application is for formulation,

b) Undermining the model hierarchy by unwise elimination of seemingly insignificant terms,

c) Fitting raw response data that would be modeled more precisely and accurately after transformation by the natural logarithm or other mathematical functions,

d) Predicting responses outside the actual experimental region, in other words, extrapolating the model.

However, the focus of the paper is not to pinpoint these shortcomings of the prior studies, but rather show by good example (the fabrication of nanoparticles) a more statistically-rigorous approach to design and analysis of pharmaceutical experiments.

Experimental

The polymeric nanoparticles were fabricated with phospholipids using probe sonication. After completing initial discovery steps and completing the breakthrough phase in the strategy of experimentation, two variables, the amount of phospholipids (A) and the time of sonication (B) survived as the best candidates for optimization. Particle size was measured as the primary response variable.

To model potentially non-linear effects by response surface methods (RSM), at least three levels of each factor must be tested. The goal of RSM is to fit a second-order equation called a "quadratic model," which for two factors takes the form:

$$\hat{y} = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \beta_{11} A^2 + \beta_{22} B^2$$

where the y under the hat (^) represents the predicted response and the Greek letter betas (β) are the coefficients determined experimentally.

As noted in Figure 1, the top RSM design choice is the central composite design (CCD) because it offers the most flexibility. For example, if a prior two-level factorial experiment encompasses an ideal operating region, it can be built up into a CCD by simply adding more points along the factor axes. Figure 2 illustrates the central composite design structure for this case study. It provides five levels for each factor.

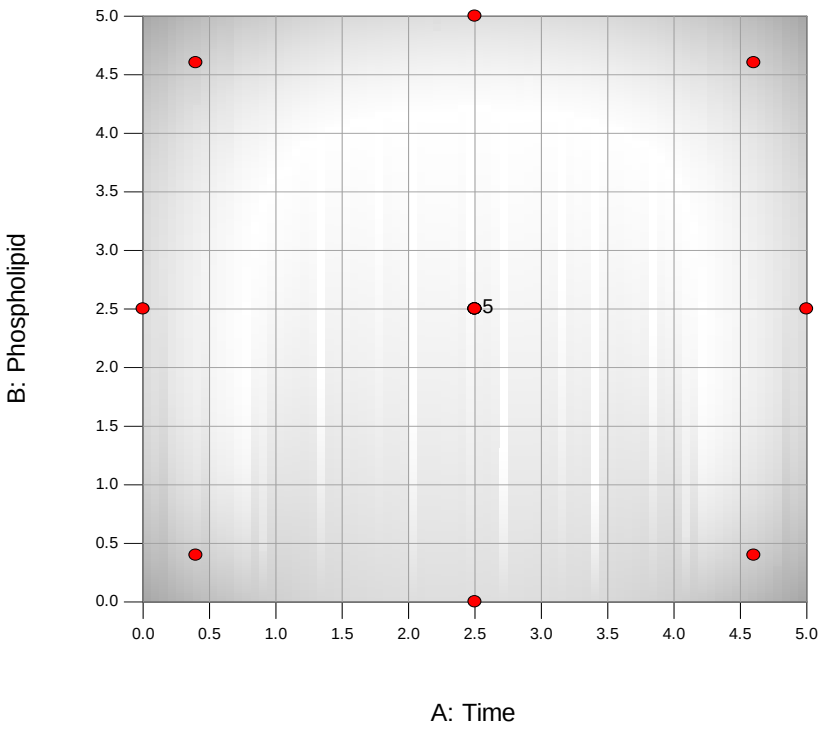

Figure 2: Design layout for nanoparticle RSM experiment

The geometry of the CCD is a square (or cubical in three dimensions) of "factorial" levels, which in this case are:

A. Time (minutes): 0 to 5

B. Phospholipid (percent) 0 to5

These ranges are coded for mathematical purposes as minus one (-1) versus plus one (+1), with four axial points protruding outside of this core region of interest. The distance is usually dictated by design rotatability, that is, equally precise prediction capability at equal distances from the center. By this criterion, a two-factor CCD requires its furthest points be placed at 1.414 coded units, plus or minus. However, more recent research suggests a more practical, less extreme, placement calculated from the fourth root of the number of factors (Anderson and Whitcomb, 2005). Taking the more practical (and modern) route in this case places the points roughly 1.2 coded units out from the center of the experimental region.

The center point in this CCD is replicated five times – a not just re-sampled or re-tested, but completely re-run, thus providing a true measure of error due to natural variations from start to finish. The number of replicates is chosen to provide a broad region where the standard error of prediction remains relatively stable. However, as indicated by the background shading in Figure 2 the outer regions, particularly at the corners, where the response model must be extrapolated, cannot be predicted as precisely. If one of these areas turns out to be desirable, another experiment, perhaps with tighter ranges, would be advisable to provide a proper focus.

The results from the 13-run design shown in Table 1 stem from an actual experiment, which then provided the basis to demonstrate the effectiveness of this modern RSM for producing useful predictive models. The bracketed factor values come from the coded CCD template described above. Actual factors levels were rounded slightly in some cases for convenience sake, but not enough to materially affect the design properties from a statistical view.

Table 1: Design matrix and experimental results

Std Order	Point Type	A: Time (min.)	B: Phospho-lipid (%)	Particle Size (nm)
1	Factorial	0.4 [−1]	0.4 [−1]	526.4
2	Factorial	4.6 [+1]	0.4 [−1]	179.6
3	Factorial	0.4 [−1]	4.6 [+1]	676.6
4	Factorial	4.6 [+1]	4.6 [+1]	186.1
5	Axial	0.0 [−1.2]	2.5 [0]	270.1
6	Axial	5.0 [+1.2]	2.5 [0]	123.6
7	Axial	2.5 [0]	0.0 [−1.2]	198.9
8	Axial	2.5 [0]	5.0 [+1.2}	187.9
9	Center	2.5 [0]	2.5 [0]	163.0
10	Center	2.5 [0]	2.5 [0]	160.0
11	Center	2.5 [0]	2.5 [0]	169.5
12	Center	2.5 [0]	2.5 [0]	146.4
13	Center	2.5 [0]	2.5 [0]	115.0

Analysis

Model selection

Aided by software designed for RSM (Vaughn, et al, 2007), a quadratic model was fitted to the experimental results. However, diagnostics of the residuals, the differences between actual and predicted particle sizes, indicated a significant improvement would be gained by first taking the inverse of the responses (1/y). Transformations of this sort, more often the logarithm, are commonly applied to provide better statistical properties, thus validating a model within the factorial region of the RSM experiment.

The primary statistical tool for identifying the need for transformations, and for pinpointing which one works best, is the Box-Cox plot, which Anderson and Whitcomb (2005) detail in their Chapter 5 appendix. However, in this case the beneficial impact of the inverse is made obvious by the comparison between Figures 3a (left) and 3b (right), which illustrate modeling of particles size with or without response transformation; respectively. Notice how much better the solid curve at the left fits the maximum particle size at the lowest time (with the second factor phospholipid fixed at its center level of 2.5 percent). Furthermore, the dotted confidence band reflects the greater variation at these highest levels of response. An obvious defect of not transforming by the inverse is evident in Figure 3b (plot at the right) by the lower confidence band falling below zero at the highest time. That cannot be!

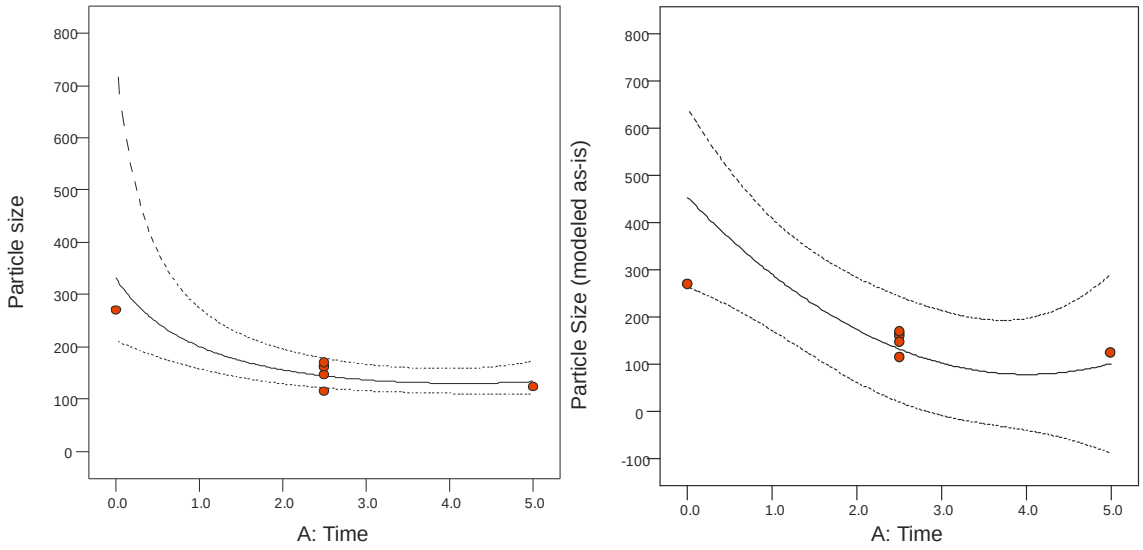

Figure 3a,b: Particle size modeled as inverse (left) and as-is (right) versus time with phospholipid fixed at 2.5%

The full picture of predicted particle size is provided by Figure 4 – the response surface as a function of both factors (time and phospholipid level).

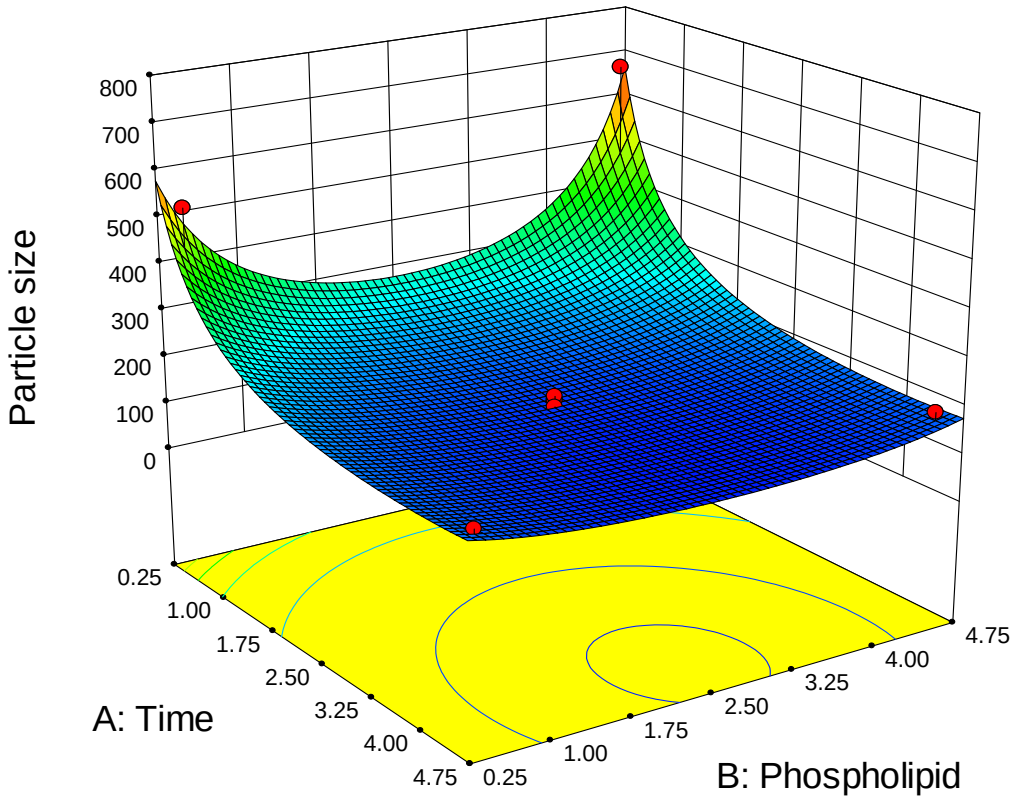

FIgure 4: Response surface for particle size

These predictions are based on the following model (quadratic) in terms of actual factors:

$$1/\hat{y} = 0.000665 + 0.002230\ A + 0.001898\ B - 0.000268\ A^2 - 0.000383\ B^2$$

Statisticians generally model in coded factors to avoid the round-off problems that occur when using actual units of measure, which are more practical for the experimenters (thus, that is what we present here).

The statistical analysis of variance (ANOVA) revealed an overall model p of less than 0.01 – very significant. Lack of fit testing produced a p-value above 0.5, which is very good. The adjusted R^2 came in at 0.74 – highly satisfactory for empirical modeling.

The p value for B exceeds 0.9, which made removal of this term very tempting. However, because B^2 is very significant ($p = 0.01$), removing its parent term B would undermine the hierarchy of the model, thus producing erroneous predictions when converted to actual terms (Peixoto, 1990). Hierarchy is ancestral linkage of effects flowing from main effects (regarded as parents) down through successive generations of higher-order interactions (children). Thus, if two individual interacting factors are included in the model, both of the main effects should also be included. Unwary experimenters, who blindly apply computerized-modeling tools without first educating themselves on RSM, often ignore hierarchy, which leads to unforeseen inaccuracies in their predictive modeling.

A side-benefit of maintaining model hierarchy is that this avoids often misleading impressions that a particular factor is not significant when, indeed, it makes a difference, albeit only at a higher order or in concurrence with other factors (Anderson and Whitcomb, 2007). For example, in this case if term B is removed on the basis of its high p value, one might think that phospholipid level matters little, when, in fact, it greatly affects particle size, as seen in the response surface (Figure 4).

This model can be reduced somewhat by eliminating the insignificant interaction term AB ($p>0.9$). Some statisticians caution against such elimination of

specific terms, advocating instead that modelers maintain the entire quadratic polynomial as standard practice (Myers and Montgomery, 2002). However, in this case by taking out term AB the predicted R^2 increases from 0.47 to 0.62. This calculation-intense statistic is based on the predicted residual sum of squares, called PRESS, which starts by re-fitting the model without the first run. The new model then predicts the first response, from which the residual from the actual result is computed. This process is repeated for each of the remaining runs and the squared residuals are summed into the PRESS. Because the predicted R^2 is based on the PRESS, it provides a more severe test of model adequacy than the adjusted R^2, but if the difference between these two types of R-squareds exceeds 0.20, there may be a problem with either the data or the model (Vaughn, 2007). After removing the AB term, the predictive model for particle size produces reasonably good agreement between the two R^2 statistics – their gap falls within the 0.2 guideline.

Optimization

Ultimately, the goal of the pharmaceutical experimenters in this case was to minimize particle size. Results above 200 nm were unacceptable. Ideally, this size-reduction goal could be exceeded at the lowest possible processing (sonication) time. Given the predicted model detailed above, which diagnostics validated statistically, numerical hill-climbing algorithms can search out most desirable outcomes (Vaughn, 2007). Figure 5 shows the optimal conditions for minimizing particle size with time-reduction added as a desirable outcome.

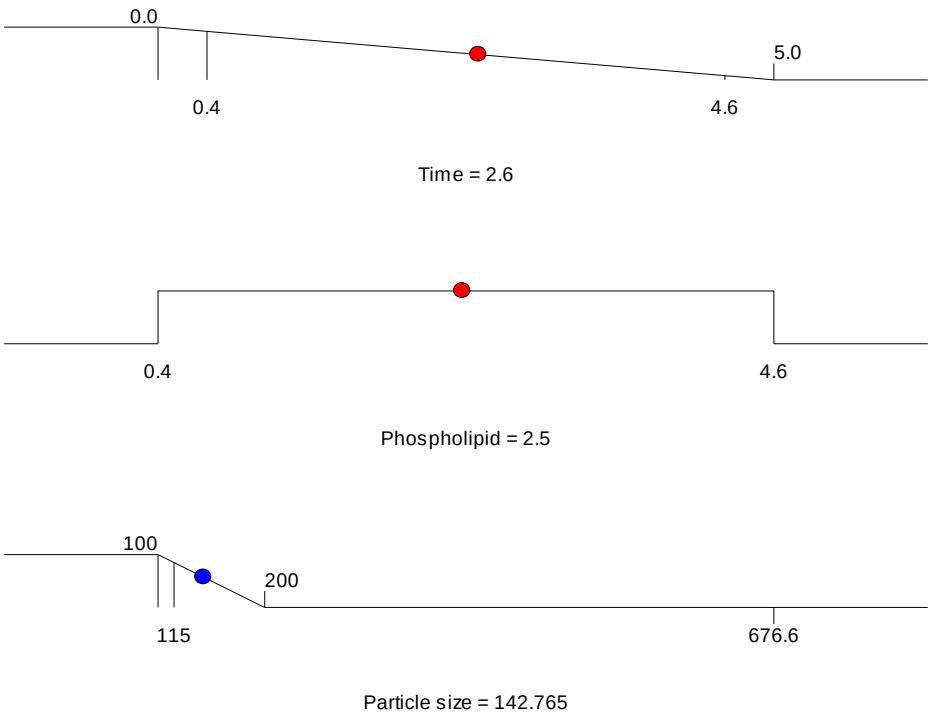

FIgure 5: Most desirable operating conditions for sonication

Figure 6 shows a 3D rendering of desirability.

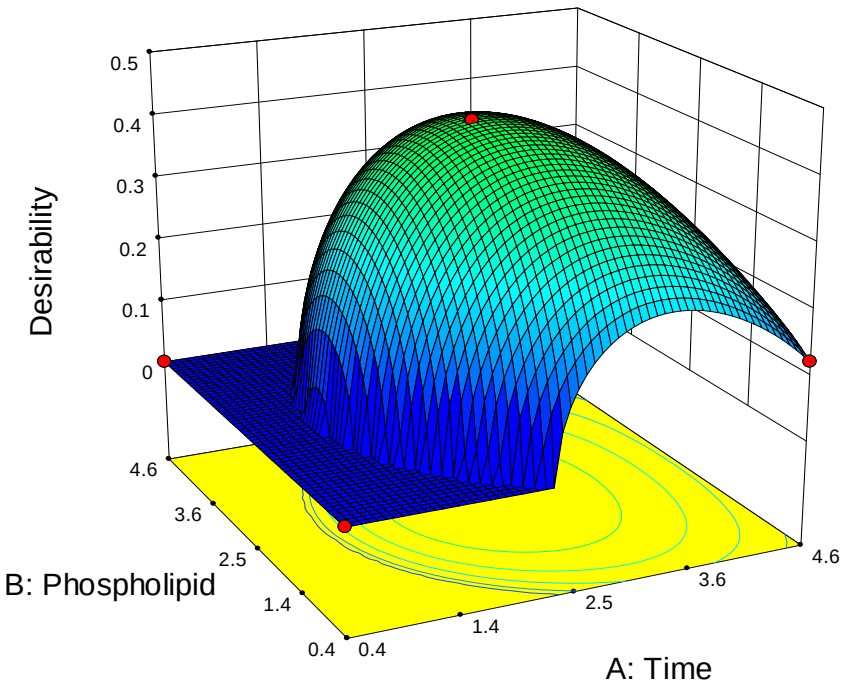

FIgure 6: 3D view of most desirable operating conditions

The suggested setup at 2.60 minutes of sonication with phospholipid at 2.48 percent falls very near the experiment design's center point, which repeatedly produced particles below the desired maximum level of 200 nm. The model predicts a particle size of 142 nm at the optimum with a prediction interval (PI) of 105 to 223.

If longer times can be afforded and thus it no longer becomes desirable to minimize, the optimum setup shifts to 4.2 minutes with 2.48 percent phospholipid. The longer sonication is predicted to reduce particle size down to 130 nm with a PI of 98 to 196.

Conclusion

The case study on pharmaceutical soninication demonstrated the utility of statistical design of experiments (DoE), especially response surface methods, for process optimization. If significant results are achieved, the resulting predictive models enable computer-aided searches for factor combinations that achieve desirable responses at economical operating conditions. The authors hope that this example provides inspiration for other pharmaceutical scientists and engineers to overcome any aversions to statistically-based methods, thus taking advantage of powerful tools for more quickly accomplishing goals for research and development.

Acknowledgements:

The authors thank Dr. Paul Mullenix (SystatS Consulting) for his helpful review of the original draft of this paper.

References

Anderson, M.J. Trimming the FAT out of Experimental Methods. *OE (Optical Engineering) Magazine*, September 2005, p. 29.

Anderson, M.J. and Whitcomb P.J. *DOE Simplified – Practical tools for effective experimentation (2nd Ed)*. Productivity Press, New York, NY, USA, 2007.

Anderson, M.J. and Whitcomb P.J. *RSM Simplified – Optimizing Processes Using Response Surface Methods For Design of Experiments*. Productivity Press, New York, 2005.

ASTM committee. *E 1169 – 07 Standard Practice for Conducting Ruggedness Tests*. 2007. ASTM International, West Conshohocken, PA, USA.

Bolton, S. and Bon C. *Pharmaceutical Statistics – Practical and Clinical Applications (4th Ed)*. Marcel Dekker, Inc. New York. 2004.

Lewis, G.A., Mathieu, R. and Phan-Tan-Luu. *Pharmaceutical Experimental Design*, Marcel Dekker, New York, 1999.

Myers, R. H. and Montgomery, D. C. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 2nd Edition*, John Wiley & Sons, New York, 2002.

Peixoto, J. L. "A Property of Well-Formulated Polynomial Regression Models," *The American Statistician*, V44, No. 1, Feb, 1990.

Vaughn, N.A., et al. *Design-Expert® Software, Version 7.1*. Stat-Ease, Inc., Minneapolis, MN. 2007.

Wayne, W.D. *Biostatistics: Analysis in Health Sciences, 5th Edition*, John Wiley & Sons, New York. 1991