Journal of Pharmaceutical & Biomedical Analysis
Vol. 9, No. 8, pp. 605–610, 1991
Printed in Great Britain

0731–7085/91 $3.00 + 0.00

Use of experimental design in the pharmaceutical industry*

NOUNA KETTANEH-WOLD

MDS, Inc., 371 Highland Ave., Winchester, MA 10890, USA

Abstract: Statistical modelling and experimental design (SMED) are essential tools for the development and understanding of complicated products and processes. SMED allows efficient experimentation in which all or a large subset of factors are together varied over a set of experiments, in contrast to the traditional approach of varying only one at a time.

An overview of the SMED methodology and the generalization of statistical design to multivariate design is presented. The following examples illustrating the use of these methods are discussed: (1) use of factorial designs to improve drug solubility; (2) testing the robustness of an analytical method; and (3) use of multivariate design to select the solvent in analytical method development.

Keywords: *Modelling; statistical experimental design; multivariate design; robust method development; optimization; pharmaceutical applications; analytical chemistry applications.*

Introduction

Measurements and experiments are made in pharmaceutical and biomedical analysis to determine concentrations of drug compounds and metabolites in organs and tissues of animals and humans as a function of dose, time, and other factors. Analytical measurements in pharmaceutical quality control involve measuring impurity levels, rate of degradation, stability, etc.

Chemical analysis requires analytical methods that are accurate, selective and precise. In practice, this involves developing methods of work-up and analysis (e.g. HPLC, FT–IR, etc.) that optimize recovery, chromatography peak separation, etc. In addition, the robustness of the method is essential, i.e. the results of the method should not be significantly affected by small variations of the controlled factors nor uncontrolled 'environmental' factors, such as ambient temperature, ambient humidity, brand of analytical reagents, etc.

Review of Statistical Experimental Design

All method development and optimization involves experimentation. An interesting question is how to make this experimentation achieve the stated objective as accurately and efficiently (i.e. with few experiments) as possible. The answer is easiest to understand if we see experimentation as exploring a space defined by the factors that are changed during the investigation (Fig. 1). In this space, the experimental region is defined by the lower and upper levels of each factor. One experiment corresponds to a point in this region, and for each such point, the value(s) of one or several responses, Y (e.g. percentage recovery, peak separation), is measured, as shown in Fig. 2.

To understand how the factors X affect the responses Y, a map of the experimental region is constructed, which describes how the value(s) of the response(s), Y, varies when the values of the factors, X, are changed.

In order to obtain the best map, i.e. a map that best approximates the variation of the responses as a function of the factors, it is necessary to carry out a set of experiments that are representative of the region. In 1925 Fisher solved the problem of efficiently selecting a set of best experiments [1]. From this work grew the concept of statistical experimental design (SED) [1, 2].

Experimental design involves selecting the best set of points in the experimental space

*Presented at the Symposium on "Chemometrics in Pharmaceutical and Biomedical Analysis", November 1990, Stockholm, Sweden.

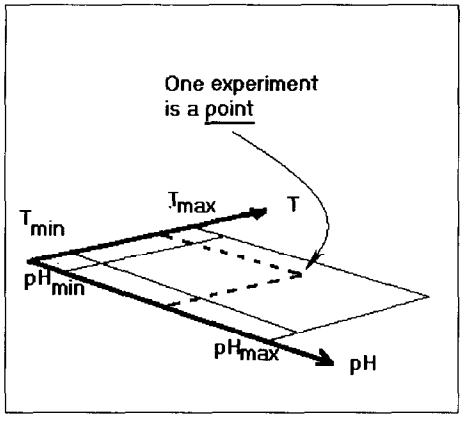

Figure 1
Two-space of two factors (pH and T). Each experiment is a point in this space.

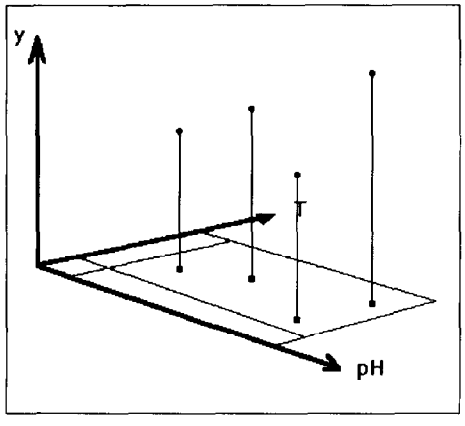

Figure 2
Two factors (x) and one response (y) form a three-dimensional space or a 2-space (X) + a 1-space (Y).

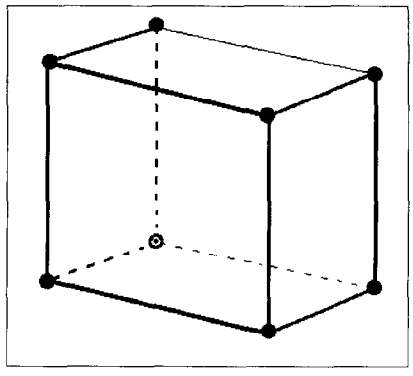

Figure 3
Full factorial design in three factors corresponds to the eight corners of a cube.

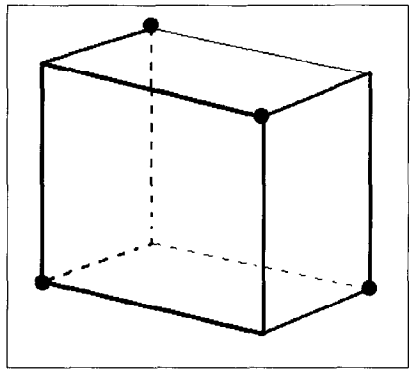

Figure 4
Fractional factorial design is a balanced selection of a fraction of a full factorial design, of the corners of a hypercube.

under the following constraints: (1) the size and shape of the experimental region; (2) the number of desired experimental runs, N; and (3) the type of model used for constructing the 'map', with the objective of obtaining a precise and accurate map of the response.

Orthogonal designs (factorial designs)

With the orthogonal designs of Fisher [1, 2], an efficient strategy for screening many factors to find the dominant factors was developed. These include the full and fractional factorials at two levels (see Figs 3 and 4) and the multi-level fractional factorials (family of Latin squares). With these balanced orthogonal designs, the estimated factor effects are uncorrelated with each other. These designs can also be used to estimate interactions between factors.

As a typical example of fractional factorial design in pharmaceutical analysis, consider a problem encountered in the development of a method for the analysis of a drug. The objective is to improve the solubility of a drug. The goal of the first screening phase, is to identify the factors affecting solubility. Four factors were identified as possibly being important (1) quantity of solvent (butanol); (2) quantity of detergent (sodium dodecyl sulphate; SDS), (3) ionic strength, and (4) quantity of enhancer.

A linear model including interactions (1 + 4 + 6 terms) was proposed, and a fractional factorial design with 16 runs + 3 centre points was used. After the solubility (y) for the 19 different experimental conditions given by the design was measured, multiple regression was used to fit the model to the data. One strong outlier was found in a less interesting part of the region. After removal of this outlier, a linear model was found to approximate well

the response. The variation of the response in the region showed no curvature and no important interactions. Only the quantity of solvent and the ionic strength had large effects.

Orthogonal designs

A drawback of the fractional factorial designs in the screening of many factors is the large increase in the number of runs; from 8 to 16 to 32 to 64. If, for instance, one wishes to screen 16 factors, a fractional factorial in 32 runs is needed, while if the number of screened factors is 15, a 16-run design would suffice (albeit barely).

Plackett–Burman (PB) in 1950 developed orthogonal designs with N being multiples of four, e.g. 12, 20, 24, 28, with which one can screen up to 9, 17, 21, and 25 factors, respectively [2]. Hence, with 16 factors to be screened, a PB design with 20 runs would suffice.

The drawback of PB designs is that they are resolution III designs, and allow only for the estimation of main effects (two-factor interactions are confounded with each other and with main effects). Their advantage is their efficiency for screening many variables. They are therefore good designs for robustness testing.

Robustness testing is an interesting application of screening designs, providing information on those controlled factors that have the greatest influence on the variability (i.e. lack of robustness) of a process. For these applications, linear models without interactions are usually sufficient because of the very small ranges of the factors. PB designs are suitable for the investigation of the robustness (i.e. effects of small changes in the factors) of analytical methods around the 'best conditions'. Factors sensitive to small changes (causing large effects) should be 'better controlled' or conditions selected for these factors that are less sensitive.

For each controlled factor one needs to know its nominal value (i.e. the value at which it is controlled according to the method specifications) and the interval within which it can be controlled.

As an example, the testing of the robustness of an HPLC analytical method for aspirin is discussed. The following are the $K = 5$ factors, and the experimental region, selected as the best conditions (nominal values) ± controlled interval of the factors:

acid concentration: 25 ± 2%;
acid type: 1 ± ?? (depends on objective);
flow rate: 1.5 ± 0.1 ml min^{-1};
temperature: 40 ± 5°C;
wavelength: 295 ± 5 nm.

An eight-run Plackett–Burman design with three centre points was used, with the low and high levels = nominal ± controlled interval, and the centre point = nominal levels. A linear model (six terms = five linear terms + constant term) was used to compute effects and find the factors with the largest effects. Changes in these factors cause unacceptable changes in the results of the method. Hence, these factors should be better controlled. Alternatively, one can seek different conditions where the HPLC method is less sensitive to these changes.

Classical response surface modelling (RSM) designs

The third type of designs that are very useful in analytical work are the classical RSM designs, the so-called composite designs, developed around 1950 [2]. They are used for up to five (or in rare cases six) factors. They allow good estimation of the parameters of the quadratic model and the adequacy of the model to be assessed (Fig. 5).

The application of an RSM design in testing the robustness of a method is illustrated by a test of the dissolution of tablets under the following conditions: temperature, 38°C; volume, 700 ml; stirrer speed, 75 rpm; pH, 4.0. The responses (i.e. characteristics) of interest are measurements of the percentage of the content of active ingredient released at 1, 2, 6 and 10 h.

The questions are: is the testing method robust to: (1) changes in pH that occur in the human body; from 1.2 to 6.8; (2) small changes in temperature, e.g. ±1°C due to equipment; (3) small changes in volume and stirrer rotation speed?

Because of the large variation in pH, a quadratic model is chosen. The experimental region selected is: volume, 500–900 ml; temperature, 37–39°C; rotation speed, 50–100 rpm; and pH, 1.2–6.8.

A central composite face (CCF) design was used ($N = 17$ runs). By using partial least squares (PLS, a generalized regression method) [3] to fit the quadratic model to the data, three significant PLS components were

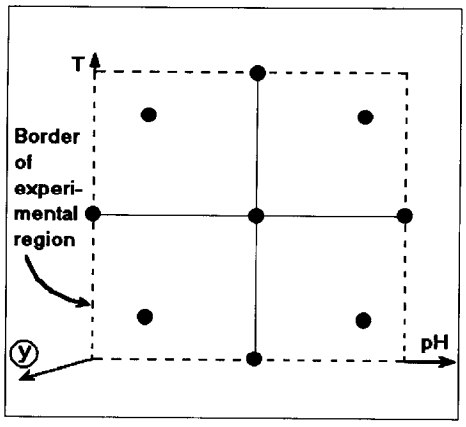

Figure 5
A statistical design (central composite) in two factors. This design supports a quadratic polynomial model:
$y = c_0 - c_1x_1 + c_2x_2 + c_{11}x_1^2 + c_{22}x_2^2 + c_{12}x_1x_2 + \epsilon$.

Table 1

	R^2 (%)	Q^2 (%)
*Y*1 (1 h)	91	72
*Y*2 (2 h)	92	75
*Y*3 (6 h)	89	65
*Y*4 (10 h)	86	50

obtained, explaining 64, 23 and 4% (in all 91%) of the variance of *Y*. The amount of explained (R^2) and predicted (Q^2) variance for the four responses are listed in Table 1 and indicate good fit and predictive power for the quadratic model.

In conclusion, good models were obtained for all four responses. The factors causing largest variation were rotation speed, temperature and pH. The variation of the release of the drug from the tablet after 10 h, as a function of pH, adjusted for the other factors, is curved (i.e. not rectilinear). The release at 10 h is maximum for pH = 4.4. However, the largest variation caused by the variation in the factors was still within acceptable ranges for all four responses. Hence, the analytical method was considered to be robust and required no changes to its conditions.

Mixture designs (1955)

When experiments are made that involve mixtures, special types of designs must be used [8] because of the mixture constraint: Σx_i = constant (usually 1.0). With no additional constraints, i.e. all mixture factors can vary between 0.0 and 1.0, the region is a hyper-simplex (Fig. 6), and the classical mixture

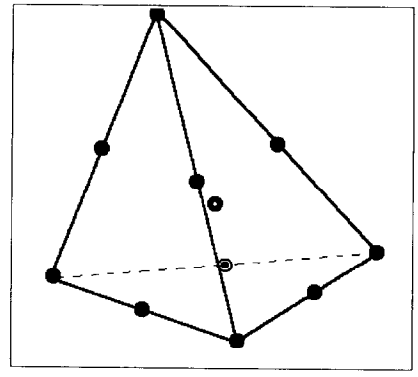

Figure 6
A mixture region is a (hyper)-simplex. A simplex centroid design in four factors is shown here.

designs are: (1) axial designs for screening; and (2) simplex lattice and simplex centroid for RSD (Fig. 6).

Most commonly, however, there are additional constraints, e.g. factors have upper and/or lower levels (i.e. they do not vary between 0 and 100%). For example, the content of glycerine in a cream should be between 0.25 and 0.40% of the total concentration. The experimental region is then an irregular polyhedron volume inside the hyper-simplex (Fig. 7) and D-optimal designs are used (see below) to select a subset of the extreme vertices, edge centres, face centres and the over-all centroid. Richer designs are used for RSMs (in few factors) than for screening of many factors.

D-Optimal designs

This is the last class of designs commonly used in practice. These are computer generated with the objective of maximizing the hyper-volume of the design-points [Det($X'X$)].

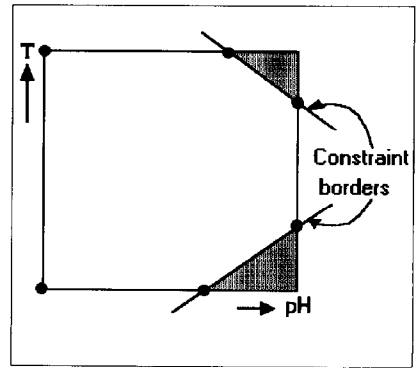

Figure 7
D-Optimal designs put design points on the border of the experimental region at the extreme vertices (intersections of constraint borders and experimental region).

These designs are applicable for constrained regions, for designs with very small numbers of runs, multi-level designs, and whenever there is no classical design.

Taguchi designs

These designs are used when the objective is to optimize response(s) while at the same time minimizing their variability. Taguchi designs have inner and outer arrays, where the latter are used to estimate the variability in each inner array point.

The responses with these designs are called performance statistics. The most commonly used are the mean and standard deviation of the measured y-values of the inner array. The inner arrays are classical designs (e.g. fractional factorials), while the outer arrays are either statistical designs or just repeats under the same conditions.

Taguchi designs are used in manufacturing for developing robust products or processes. Because of the large number of required experimental runs, they are rarely used in the pharmaceutical industry.

Multivariate Design

With most complicated systems, one has many X-variables, many of which cannot be individually manipulated or controlled. A typical case is one in which there is a qualitative change that potentially influences many properties of the system. Examples are selection of solvent, substituent in a chemical compound, batch of raw material, etc. One can then make a multivariate characterization to quantify the qualitative change, and use scales T derived from a multivariate analysis, principal component analysis (PCA) or PLS [3–10], to represent the underlying "real" factors.

Thereafter one tries to span the subspace of the 'latent variables' T by means of a design in these scores, i.e. a multivariate design (MV), followed by a PLS modelling [4, 5] of the resulting data.

The concept of MV design [6] is illustrated with an application from the pharmaceutical industry in which the objective of the study was to improve the solubility of a drug. A first screening experiment carried out with butanol as solvent and SDS as detergent, identified the quantity of solvent and the ionic strength as having effects on the solubility. The settings of the factors yielding optimal solubility were selected.

A second study was undertaken to further improve the solubility of the drug. The question of interest was: can one improve the solubility by selecting a better solvent and detergent? As there is a large number of solvents and detergents available, it is impossible to investigate them all. Consequently three or four solvents should be selected that are representative of the range of solvents. Multivariate characterization of the set of 103 solvents selected, quantifies the variation spanned by the solvents. The method consists of measuring several (10 in this study) relevant properties such as molecular weight, Log P, boiling point, melting point, etc., of every solvent. A PCA analysis is performed on the resulting (103 by 10) matrix. The PCA analysis resulted in two PCs. The two scores $t1$ and $t2$ are the best summary of the 10 measured properties. The score plot of $t1$ vs $t2$ provides a map of the variation of the solvents [7]. The three following solvents were selected from the PC-score map [7] to span the sub-space of solvents with desired properties; butan-1-ol, glycol, and sulphanol. The same method was used to select the following three detergents; SDS, cetyltrimethylammonium bromide (CTAB) and Breen. A statistical experimental design calculation was carried out with the following six factors: ionic strength, quantity of sample, solvent type (3), detergent type (3), quantity of solvent, quantity of detergent. A linear model in all the factors was selected, and a D-Optimal design made with 18 runs + 2 replicates. The measured response was log solubility.

The PLS analysis resulted in 2 PLS components, explaining 90% of Y. Q^2 (PRESS) = 72%. The linear model was found to be a good representation of the variation of the response and showed that butan-1-ol has the largest effect on solubility and was the best of the three solvents. It was recommended that solvents near butan-1-ol on the PC map should also be investigated. Breen was found to be the best detergent. Only mild effects were found for ionic strength and quantity of solvent, these effects being partly masked by the large variation due to the three different solvents.

Experience with modelling and design indicates that in the future there will be increasing demand for more flexible computer generated designs that optimize several criteria: (1)

volume of design |X'X|; (2) 'Bias check points' (interior points); (3) 'enriching points' (richer models supported); and (4) larger candidate sets (and 'continuous').

The increasing complexity of the systems investigated, corresponding to increasing numbers of factors, demand super-saturated designs handling large numbers of factors (considerably greater than 30) and hierarchical strategies to deal with hundreds of factors.

Conclusions

Modelling, design and (multivariate) analysis provide the best available technology for essential elements of research and development. However, the application of these tools is not always easy and automatic, and requires a combination of chemical and statistical knowledge. Modelling, design and multivariate analysis are, for the pharmaceutical industries, an investment. It costs time and money to educate and train the researchers; in return the methods provide an approach that works well in solving research and development problems.

References

[1] R.A. Fisher, *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh (1925).*
[2] G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for experimenters*. Wiley, New York (1978).
[3] H. Wold, in *Systems under indirect observation* (K.-G. Jöreskog and H. Wold, Eds), Vols I and II. North-Holland, Amsterdam (1982).
[4] L. Ståhle and S. Wold, *J. Pharm. Methods* **16**, 91–110 (1986).
[5] A. Höskuldsson, *J. Chemometrics* **2**, 211–228 (1988).
[6] S. Wold, M. Sjöström, R. Carlson, T. Lundstedt, S. Hellberg, B. Skagerberg, C. Wikström and J. Öhman, *Anal. Chim. Acta* **191**, 17–32 (1986).
[7] R. Carlson, T. Lundstedt and C. Albano, *Acta Chem. Scand.* **B39**, 79 (1985).
[8] J.A. Cornell, *Experiments with mixtures*, 2nd edn. Wiley, New York (1990).
[9] S. Wold, U. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).
[10] J.E. Jackson, *A User's guide to Principal Components*. Wiley, New York (1991).

[*Received for review 26 November 1990*]

*The 14th revised and enlarged edition is reprinted in *Statistical Methods, Experimental Design, and Scientific Inference*. Oxford University Publications, Oxford (1990).