DESIGN OF EXPERIMENTS

1. Introduction

Obtaining valid results from a test program calls for commitment to sound statistical design. In fact, proper experimental design is often more important than sophisticated statistical analysis. Results of a well-planned experiment are often evident from simple graphical analyses. However, the world's best statistical analysis cannot rescue a poorly planned experimental program. The main reason for designing an experiment statistically is to obtain unambiguous answers to questions of major interest at a minimum cost. The need to quantify relationships, learn about interactions among variables, and to measure experimental error are a few of the added reasons for designing an experiment statistically.

Many chemists and engineers think of experimental design mainly in terms of standard plans for assigning treatments to experimental units, such as the Latin square, factorial, fractional factorial, and central composite (or Box) designs. These designs are described in books, such as those summarized in the general references of this article, and catalogued in various reports and articles. Additionally, numerous commercial software packages are available for generating such experimental designs, as well as to aid the experimenter in analyzing the resulting data. Important as such formal plans are, the final selection of test points represents only the proverbial tip of the iceberg, the culmination of a careful planning process. For this reason, we will place particular emphasis on the *process* of designing an experiment.

Statistically planned experiments are characterized by (*1*) the proper consideration of extraneous variables; (*2*) the fact that primary variables are changed together, rather than one at a time, in order to obtain information about the magnitude and nature of interactions and to gain improved precision in estimating the effects of these variables; and (*3*) built-in procedures for measuring the various sources of random variation and for obtaining a valid measure of experimental error against which one can assess the impact of the primary variables and their interactions.

A well-planned experiment is often tailor-made to meet specific objectives and to satisfy practical constraints. The final plan may or may not involve a standard textbook design. If possible, a statistician knowledgable in the design of experiments should be called in early, made a full-fledged team member, and be fully apprised of the objectives of the test program and of the practical considerations and constraints. He or she may contribute significantly merely by asking probing questions. After the problem and the constraints have been clearly defined, the statistician can evolve an experimental layout to minimize the required testing effort to obtain the desired information—or, as a minimum, review a plan that may have been developed by a practitioner. However, designing an experiment is often an iterative process requiring rework as new information and preliminary data become available. With a full understanding of the problem, the statistician is in an improved position to respond rapidly if last-minute changes are required, to help experimenters gain understanding in a sequential manner, and to provide meaningful analyses of the experimental results, including statements about the statistical precision of any estimates.

Some of this article previously appeared in Refs. (1) and (2). Reference (3) proposes a systematic approach to planning a designed experiment. References (4–7) illustrate many of the considerations discussed here and introduce a few others. References (8) and (9) provide two other surveys dealing with the design of experiments, including more detailed discussions than given here of specific designs and their analysis. Reference (10) provides a relatively technical overview of experimental design concepts and plans. In addition, the books listed and described in the general references give information concerning the technical aspects of experimental design.

2. Purpose and Scope of the Experiment

Designing an experiment is like designing a product. Every product serves a purpose; so should every experiment. This purpose must be clearly defined at the outset. It may, for example, be to optimize a process, to estimate the probability that a component operates properly under a given stress for a specified number of years, to maximize robustness of a new chemical formulation to variability in raw materials or end use, or to determine whether a new drug or medical treatment is superior to an existing one. An understanding of this purpose is important in developing an appropriate experimental plan.

In addition to defining the purpose of a program, one must decide on its scope. An experiment is generally a vehicle for drawing inferences about the real world, as expressed by some, usually quantitative, response (or performance) variable. Since it is highly risky to draw inferences about situations beyond the scope of the experiment, care must be exercised to make this scope sufficiently broad. For example, when developing a new product, experimental studies are often conducted on a smaller (eg, lab) scale, with the objective of scaling up successful results to production. This is true whether the "product" is a thermoplastic resin, food product, or a new drug. Thus, if the results are material dependent, the material for fabricating experimental units must be as representative as possible of what one might expect to encounter in production. If the test program were limited to a single batch of raw material, the conclusions might be applicable only to that batch, irrespective of the sample size. Similarly, in deciding whether or not a processing factor, such as temperature, should be included as an experimental variable to compare different formulations, it must be decided whether the possible result that one formulation outperforms another at a particular temperature would also apply for other temperatures of practical interest. If not, one should consider including temperature as an experimental variable. In any case, one need keep in mind that the statistical inferences one can draw from the experiment apply only to the range of conditions under which the experiment was conducted.

3. Experimental Variables

An important part of planning an experimental program is the identification of the controllable or "independent" variables (also known as factors) that affect the

response and deciding what to do about them. The decision as to how to deal with each of the candidate variables can be made jointly by the experimenter and the statistician. However, identifying the variables is the experimenter's responsibility. (Tools such as the fishbone diagram and brainstorming can be used to facilitate the process.) Controllable or independent variables in a statistical experiment can be dealt with in four different ways, as described next. The assignment of a particular variable to a category often involves a trade-off among information, cost, and time.

3.1. Primary Variables. The most obvious variables are those whose effects on the expected or mean performance of the response variable(s) are to be evaluated directly; these are the variables that, most likely, created the need for the investigation in the first place. Such variables may be quantitative, such as catalyst concentration, temperature, or pressure, or they may be qualitative, such as method of preparation, catalyst type, or batch of material.

Quantitative controllable variables are frequently related to the response variable by some assumed statistical relationship or model. The minimum number of conditions or levels per variable is determined by the form of the assumed model. For example, if a straight-line relationship can be assumed, two levels (or conditions) may be sufficient; for a quadratic relationship a minimum of three levels is required. However, it is recommended that additional points, above the minimum, be included so as to allow assessment of the adequacy of the assumed model (see RESPONSE SURFACE DESIGNS).

Qualitative variables can be broken down into two categories. The first consists of those variables whose specific effects on the mean response are to be compared directly; eg, comparison of the impact on average performance of two proposed preparation methods or of three catalyst types. The required number of conditions for such variables is generally evident from the context of the experiment. Such variables are sometimes referred to as fixed effects or Type I variables.

The second type of qualitative variables are those whose individual contributions to variability or noise in the responses are to be evaluated. The specific conditions of such variables are generally randomly determined. Material batch is a typical example. Usually, one is not interested in the behavior of the specific batches per se that happened to have been selected for the experiment. Instead, one wishes to quantify the variation in performance, as measured by the variance (or, its square root, the standard deviation) caused by differences among batches. The batches used in the experiment are selected randomly (or as close to randomly as is practically feasible) from a large population of batches. It is desirable to have a reasonably large sample (eg, five or more batches) in order to obtain an adequate degree of precision in estimating the variability in response attributable to such variables. These variables are generally referred to as random effects or Type II variables. Differentiation between fixed and random effect variables is an important consideration both in the design of the experiment, and in the analysis of the resulting data.

When there are two or more variables, they might interact with one another, ie, the effect of one variable upon the response depends on the value of the other variable. Figure 1 shows a situation where two noninteracting variables, preparation type and temperature, independently affect time to rupture,
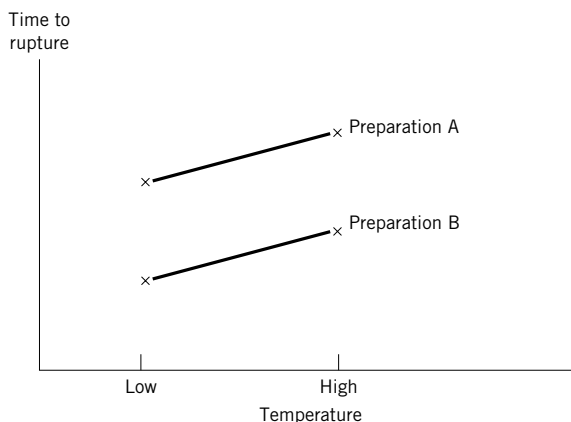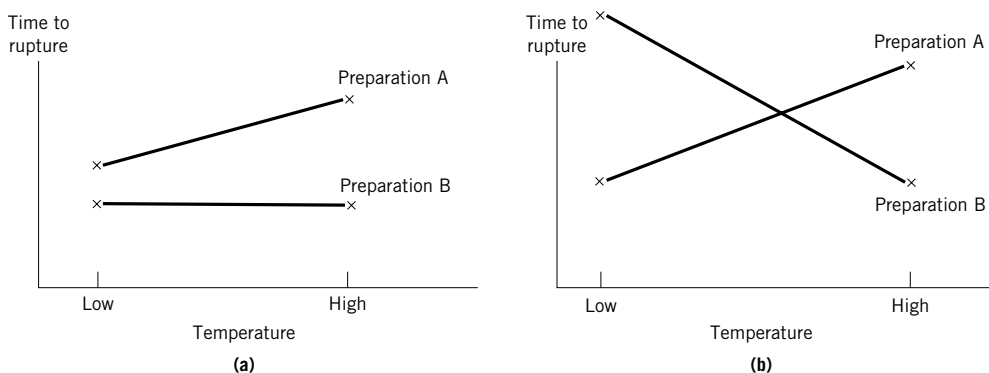
Fig. 1. Situation with no interaction between variables.

ie, the effect of temperature on time to rupture is the same for both preparation types. In contrast, Figure 2 shows two examples of interactions between preparation and temperature.

An important purpose of a designed experiment is to obtain information about interactions among the primary variables which is accomplished by varying factors simultaneously rather than one at a time. Thus in Figure 2, each of the two preparations would be run at both low and high temperatures using, for example, a full factorial experiment (see FORMAL EXPERIMENTAL PLANS).

3.2. Background Variables and Blocking.   In addition to the primary controllable variables there are those variables, though not of primary interest, that cannot, and perhaps should not, be held constant in the experiment. Such variables are present in real world situations, where they introduce added variability in the response, and where, unlike the experiment, they generally cannot be controlled. They are often referred to as "noise variables". Typical examples are day-to-day, operator-to-operator, machine-to-machine, and batch-to-batch

Fig. 2. Situation with interactions between variables, where in (**a**) an increase in temperature is beneficial for preparation A but does not make any difference for preparation B, and (**b**) an increase in temperature raises time to rupture for preparation A but decreases it for preparation B.

variability. (Some of these variables, in fact, might be the same as the ones that we had identified in the earlier discusssion as Type II primary variables. The difference is that, previously, we were interested in assessing their impact on response variability, per se. Now our major concern is, not so much in their evaluation, but just to ensure that they do not "mess up" our evaluation of the primary variables).

In particular, it is crucial that such background, or noise, variables are separable from (in technical terms, not "confounded" with) the primary variables in the experiment. For example, if preparation A were run only on day 1 and preparation B only on day 2, it would not be possible to determine how much of any observed difference in response between the two preparations is due to normal day-to-day process variation.

Background variables may be introduced into the experiment in the form of experimental blocks. An experimental block represents a relatively homogeneous set of conditions within which different conditions of the primary variables are compared. For example, if one expects day-to-day variation in the response, a day might be taken as an experimental block over which different conditions of the primary variables are compared.

A specific example of blocking arises in the comparison of wear for different types of automobile tires. Tire wear may vary from one automobile to the next, irrespective of the tire type, because of differences among automobiles, variability among drivers, and so on. Assume, for example, that for the comparison of four tire types (A, B, C, and D), four automobiles (1, 2, 3, and 4) are available. A poor procedure would be to use the same type of tire on each of the four wheels of an automobile varying the tire type among automobiles, as in the following tabulation:

Automobile			
1	2	3	4
A	B	C	D
A	B	C	D
A	B	C	D
A	B	C	D

Such an assignment is undesirable because the differences between tires cannot be separated from the differences between automobiles in the subsequent data analysis. Separation of these effects can be obtained by treating automobiles as experimental blocks and randomly assigning tires of each of the four types to each automobile as follows:

Automobile			
1	2	3	4
A	A	A	A
B	B	B	B
C	C	C	C
D	D	D	D

The above arrangement is known as a randomized block design.

The symmetry of the preceding example is not always found in practice. For example, there may be 6 tire types under comparison and 15 available automobiles. Tires are then assigned to automobiles to obtain the most precise comparison among tire types, using a so-called incomplete block design. Similar concepts apply if there are two or more primary variables, rather than one, as was the case in the tire example.

A main reason for running an experiment in blocks is to ensure that the effect of a background variable does not contaminate evaluation of the effects of the primary variables. Blocking, moreover, removes the effect of the blocked variables from the experimental error also, thus allowing more precise estimation of the experimental error and, as a result, more precise estimates of the effects of the primary variables. Finally, in many situations, the effect of the blocking variables on the response can also be readily evaluated, an important added bonus for blocking.

In some situations, there may be more than one background variable whose possible contaminating effect is removed by blocking. Thus in the automobile tire comparison, the differences between wheel positions may be of concern in addition to differences between automobiles. In this case, wheel position might be introduced into the experiment as a second blocking variable. If there are four tire types to be compared, this might be done by randomly assigning the tires of each of the four types according to the following plan, known as a Latin square design:

Wheel position	Automobile			
	1	2	3	4
1	A	D	C	B
2	B	A	D	C
3	C	B	A	D
4	D	C	B	A

In this plan, the effects of both automobile and wheel position are controlled by blocking. However, it should be kept in mind that for the Latin square design, as for various other blocking plans, it is generally assumed that the blocking variables do not interact with the primary variable to be evaluated.

3.3. Uncontrolled Variables and Randomization. A number of further variables, such as ambient conditions (temperature, pressure, etc), can be identified but not controlled, or are only hazily identified or not identified at all but affect the results of the experiment. To ensure that such uncontrolled variables do not bias the results, randomization is introduced in various ways into the experiment to the extent that this is practical.

Randomization means that the sequence of preparing experimental units, assigning treatments, running tests, taking measurements, etc., is randomly determined, based, eg, on numbers selected from a random number table. The total effect of the uncontrolled variables is thus lumped together as unaccounted variability and part of the experimental error. The more influential the effect of such uncontrolled variables, the larger the resulting experimental error, and the more imprecise the evaluations of the effects of the primary variables.

Sometimes, when the uncontrolled variables can be measured, their effect can be removed statistically from experimental error. In addition, note that measurement error on the response variable (and, sometimes, on the controllable variables), is a further, often small, component of experimental error.

Background variables could also be introduced into the experiment by randomization, rather than by blocking techniques. Thus in the previous example, the four tires of each type could have been assigned to automobiles and wheel positions completely randomly, instead of treating automobiles and wheel positions as experimental blocks. This could have resulted in an assignment such as the following:

Wheel position	Automobile			
	1	2	3	4
1	B	D	B	D
2	A	D	D	A
3	C	C	A	D
4	B	B	C	A

Both blocking and randomization generally ensure that the background variables do not contaminate the evaluation of the primary variables. Randomization sometimes offers the advantage of greater simplicity compared to blocking. However, under blocking the effect of a background variable is removed from the experimental error (as well as being measurable), whereas under randomization it usually is not. Thus the aim might be to remove the effects of the one or two most important background variables by blocking, while counteracting the possible contaminating effect of others by randomization.

3.4. Variables Held Constant.   Finally, some variables should be held constant in the experiment. Holding a variable constant limits the size and complexity of the experiment but, as previously noted, can also limit the scope of the resulting inferences. The variables to be held constant in the experiment must be identified and the mechanisms for keeping them constant defined. The experimental technique should be clearly specified at the outset of the experiment and closely followed.

4. Experimental Environment and Constraints

The operational conditions under which the experiment is to be conducted and the manner in which each of the factors is to be varied must be clearly spelled out.

All variables are not created equal; some can be varied more easily than others. For example, suppose that oven rack position and baking temperature are two experimental variables in a study to determine recommended baking instructions for a new cake mix. A change in oven rack position can be made in a matter of seconds. Changing the oven temperature, however, adds costly time to each experimental run, since additional time is required to reach the

target temperature. Also, in many experiments it is easier to change pressure than it is to change temperature—due to the difference in time that it takes to stabilize at the new condition. In such situations, completely randomizing the sequence of testing is impractical. However, basing the experimental plan on convenience alone may lead to ambiguous and unanalyzable results. For example, if the first one-half of the experiment is run at one temperature and the second half at another temperature, it may not be possible to tell whether the observed difference in response results from the difference in temperature or from some other factors that varied during the course of the experiment such as raw material (cake batter), ambient conditions (humidity), or operator technique. Thus the final experimental plan must be a compromise between cost and information. The experiment must be practical to run, yet still yield statistically valid results (11,12).

Practical considerations enter into the experimental plan in various other ways. In many programs, variables are introduced at different operational levels. For example, in evaluating the effect of alloy composition, oven temperature, and varnish coat on tensile strength, it may be convenient to make a number of master alloys with each composition, split the alloys into separate parts to be subjected to different heat treatments, and then cut the treated samples into subsamples to which different coatings are applied. Tensile strength measurements are then obtained on all coated subsamples.

Situations such as the preceding arise frequently in practice and are referred to as split-plot experiments. The terminology results from the agricultural origins of experimental design, eg, a farmer needed to compare different fertilizer types on a plot of land with varying normal fertility. A characteristic of split-plot plans is that more precise information is obtained on the low level variables (varnish coats in the preceding example) than on the high level variables (alloy composition). The split-plot nature of the experimental environment, if present, is important information, both in the planning and in the analysis of the experiment.

4.1. Prior Knowledge.   The experimenter should draw on subject matter expertise when selecting the response variables, design variables, and variable ranges for experimentation. When selecting the response variable, for example, there may be prior knowledge regarding the need for data transformation. For selection of variable ranges, prior knowledge is often available concerning the expected response outcome at certain experimental conditions. For example, some combinations of conditions might be known to yield poor results or might not be attainable with the equipment available, or worse yet, could result in endangering the plant. Furthermore, all proposed conditions in the experiment need to make sense. For example, a misguided proposed experiment had a condition that required a resistance of 50 Ω for a circuit without resistors. Clearly, unreasonable or hazardous conditions must be omitted from the experimental design, irrespective of whether they happen to coincide with a standard statistical pattern. Thus the experiment must be adjusted to accommodate reality and not the reverse. This can be achieved by adjusting variable ranges to obtain a feasible design space, or alternatively, choosing an experimental design plan that accomodates asymmetric constraints in the design space (see OPTIMAL DESIGNS).

4.2. Response Variables. A clear statement is required of the performance characteristics or dependent variables to be evaluated as the experimental response. Even a well-designed experiment fails if the response cannot be measured properly. Frequently, there are a number of response variables; for example, tensile strength, yield strength, percent elongation, etc. It is important that standard procedures for measuring each variable be established and documented. Sometimes the response is on a semiquantitative scale; eg, material appearance may be in one of five categories, such as outstanding, superior, good, fair, and poor. In this case, it is particularly important that standards that define (and perhaps illustrate) each of these categories are developed initially, especially if judgments are to be made at different times and perhaps by different observers. Consideration need also be given to ensuring that the response variables are measured consistently, and in a well-defined manner, throughout the course of the experiment.

4.3. Types of Repeat Information. The various ways of obtaining repeat results in the experiment need to be specified. Different information about repeatability is obtained by (*1*) taking replicate measurements on the same experimental unit, (*2*) cutting a sample in half at the end of the experiment and obtaining a reading on each half, and (*3*) taking readings on two samples prepared independently of one another, eg, on different runs, at the same target conditions. Often, there is greater homogeneity among replicate measurements on the same sample than among measurements on different samples. The latter reflect the random unexplained variation of repeat runs conducted under identical conditions. A skillfully planned experiment imparts information about each component of variability, if such information is not initially available, and uses a mixture of replication and repeat runs to yield the most precise information for the available testing budget. The way in which such information is obtained must also be known to perform a valid analysis of the results.

4.4. Preliminary Estimates of Repeatability. Initial estimates of repeatability should be obtained before embarking on any significant test program. Such information may be available from previous testing; but if not, a variance component study should be conducted. Its purpose is to validate that all critical experimental variables have been identified. Additionally, this study will quantify the experimental noise that can be expected due to (*1*) an inability to reproduce target conditions of critical variables (known as reproducibility) and (*2*) nonhomogeneity of the samples or measurement error (known as repeatability). The first step in conducting the variance component study is to collect samples from preliminary runs at different times, under supposedly identical conditions. Each of these samples is then tested multiple times (or in the case of a destructive test method, homogenous subsamples are tested from each of the original samples). A variance component analysis of the resulting data decomposes and quantifies the amount of variability in each response variable due to differences between experimental runs versus test variation. If the results show that the "run-to-run" variability is high, then one might conclude that the important variables that affect the results have not been identified, and further research may be needed before the proposed experiment can commence. Alternatively, if the measurement variability is unsatisfactory, then it may be appropriate to improve precision by taking multiple measurements of the response

variable and use the average of these measurements as the prime response in the subsequent analyses of the experimental results. The process of establishing the precision of a measurment process in advance of conducting the experiment has become known as gage R&R (reproducibility and repeatability) studies.

4.5. Consistent Data-Recording Procedures. Clear procedures for recording all pertinent data from the experiment must be developed and documented. These should include provisions not only for recording the values of the measured responses and the desired experimental conditions, but also the experimental conditions that actually occurred, if these differ from those planned. It is generally preferable to use the values of the actual, rather than the aimed-at, conditions in the statistical analysis of the experimental results. For example, if a test was supposed to have been conducted at 150°C but was run at 148.3°C, the latter temperature would be used in the analysis. (In addition, it is also often instructive to conduct an analysis based upon the planned variable settings so as to compare the estimated experimental error standard deviations. The difference in the resulting two estimates reflects the variability in the response that is introduced due to not meeting the desired settings.)

In experimentation with industrial processes, equilibrium should generally be reached before the responses are measured. This is particularly important when complex chemical reactions are involved. The values of other variables that might affect the responses should also be recorded, if possible. For example, although it may not be possible to control the ambient humidity, its value should be measured if it might affect the results. In addition, variations in the factors to be held constant, special happenings (eg, voltage surges), and other unplanned events should be recorded. The values of such covariates can be factored into the statistical analysis (possibly by performing a so-called *covariance analysis*), thereby reducing the unexplained variability or experimental error. If the covariates do indeed have an effect, this leads to more precise evaluations of the primary variables. Alternatively, such covariates may be related to the unexplained or residual variation that remains after the statistical analysis of the experimental results, using graphical or other techniques.

5. Stagewise Experimentation

Contrary to popular belief, a statistically planned experiment does not require all testing to be conducted at one time. As the eminent statistician, George Box and his associates have pointed out repeatedly [eg, (13)], the design of experiments is a catalyst for the general scientific learning process. Thus, much experimentation should be sequential, involving, for example, stages of 8−20 runs. This permits changes to be made in later tests based on early results and allows preliminary findings to be reported. For example, an experiment to improve the properties of a plastic material involved such variables as mold temperature, cylinder temperature, pressure, ram speed, and material aging. The experiment was conducted in three stages. After the first stage, the overall or main effects of each of the variables on the experimental responses were evaluated; after the second stage, interactions between pairs of variables were analyzed, and after the third stage nonlinear effects were assessed. Each stage involved about a

month of elapsed time, and management was periodically apprised of progress. If unexpected results had been obtained at an early stage, eg, poor results at one of the selected ram speeds, the general plan for the later stages of the experiment might be changed.

Whether or not to conduct an experiment in stages depends on the program objectives and the specific situation; a stagewise approach is recommended when units are made in groups or one at a time and a rapid feedback of results is possible. Running the experiment in stages is also attractive in searching for an optimum response, because it might permit moves closer to the optimum from stage to stage. On the other hand, a single-stage experiment may be desirable if there are large start-up costs at each stage or if there is a long waiting time between the start of the experiment and the measurement of the results. This is often the case in many agricultural experiments and also when the measured variable is product life.

If the experiment is conducted in stages, precautions must be taken to ensure that possible differences between the stages do not invalidate the results. Appropriate procedures to compare the stages must be included, both in the test plan and in the statistical analysis. For example, some baseline standard test conditions, known as controls, may be included in each stage of the experiment.

For further discussion of stagewise experimenation, and its role as a catalyst to learning by the scientific method, see (Ref. 13). Also Refs. 14–16 and the book by Myers and Montgomery (see *General References*) discuss the concept of evolutionary operations—a stagewise approach for seeking an optimum directly on an operating manufacturing process or pilot line.

6. Other Considerations

Many other questions must be considered in planning the experiment:

1. What is the most meaningful way to express the controllable or independent variables? For example, should current density and time be taken as the experimental variables, or are time and the product of current density and time the real variables impacting the mean response? Judicious selection of the independent variables often reduces or eliminates interactions between variables, thereby leading to a simpler experiment and analysis. Also interrelationships among variables need to be recognized. For example, in an atomic absorption analysis, there are four possible variables: air-flow rate, fuel-flow rate, gas-flow rate, and air/fuel ratio, but there are really only two independent variables.

2. What is a proper experimental range for the selected quantitative controllable variables? Assuming a linear relationship between these variables and performance, the wider the range of conditions or settings, the better, usually, are the chances of detecting the effects of the variable. However, the wider the range, the less reasonable is the assumption of a linear or other simple relationship between the experimental variables and the response variables. Also, one generally would not want to conduct experiments appreciably beyond the range of physically or practically useful

conditions. The selection of the range of the variables depends in part on the ultimate purpose of the experiment: is it to learn about performance over a broad region (eg, to establish a "response library", that provides information of what to expect under a variety of possible experimental conditions), or is it to search for an optimum condition? A wider range of experimentation would be more appropriate in the first case than in the second.

3. What is a reasonable statistical model, or equation form, to approximate the relationship between the independent variables and each response variable? The more complex the assumed model, the more runs are usually required in the experiment in order to fit the model. We shall consider this topic further in the discussion of response surface designs.

4. What is the desired degree of precision of the statistical estimates and conclusions based upon the analyses of the experimental results? The greater the desired precision, the larger is the required number of experimental runs. Statistical precision is, most frequently, quantified by a statistical confidence interval. Such an interval expresses, eg, the uncertainty in the estimated mean value of the response variable for a specified set of conditions, in the estimated coefficients of a fitted statistical model, or in the estimated experimental error standard deviation.

5. Are there any benchmarks of performance? If so, it might be judicious to include these conditions in the experiment in order to compare the results with those from past experience (and seek an explanation so as to remove the causes of differences larger than what one might expect from random variation).

6. What statistical techniques are required for the analysis of the resulting data, and can these tools be rapidly brought to bear after the experiment has been conducted (see Statistical Tools)?

7. Formal Experimental Plans

The test plan is developed to best meet the goals of the program. This might involve one of the standard plans developed by statisticians and practitioners. Such plans are described in varying detail in numerous texts on the design of experiments. In Table 1, we provide a comprehensive guide to such texts, indicating where to find more information about specific designs or design features, and, therefore, present these plans only briefly here. As already suggested, in practice, one frequently would use combinations of such plans in a stagewise approach—eg, a factorial experiment conducted in blocks or a central composite design using a fractional factorial base.

7.1. Blocking Designs.   As previously described, blocking is used to remove the effect of extraneous variables from experimental error. Well-known blocking designs include randomized block designs and balanced incomplete block designs to remove the effects of a single extraneous variable, Latin and Youden square designs to remove the effects of two extraneous variables, and Greco-Latin and hyper-Latin square plans to remove the effects of three or more extraneous variables.

Table 1. **Summary of Experimental Design Texts**[a]

| General reference (see text) | Subject coverage | | | | | | | Number of pages | Emphasis on calculations and analysis | Approximate number of references | Examples | Technical level[b] | Applications emphasis[c] |
	Blocking designs	Factorial designs	Fractional factorial designs	Response surface designs	Split-plot situations	Nested situations	Taguchi approach						
Anderson and McLean	M–H	M–H	M–H	M–H	H	M–H	N	418	L–M	150	M	I	G/I
Berger and Maurer	L	M	L–M	M	L	M	M	480	M	55	L–M	E–I	G/I
Box, Hunter and Hunter	M	M	L–M	M–H	N	L	N	653	L	150	M–H	E	G/I
Cox	L–M	L–M	L–M	L	M	N	N	308	N	100	M	E	G
Daniel	N–L	H	H	L	L–M	N	N	294	L–M	100	M	I	I
Davies and co-workers	M–H	M–H	M–H	M–H	N	L–M	N	656	L–M	50	M	I	I
Del Vecchio	L	M	L–M	L–M	N	N	M	175	L	11	M	E	I
Diamond	L	L–M	M–H	L–M	N	L–M	N	423	M	20	M	E–I	I
Hicks and Turner	M–H	M–H	M	L–M	M–H	M–H	M–H	565	H	100	M	I	G/I
Kuehl	M–H	M–H	M	M–H	M–H	M–H	L–M	666	M	240	M	I	G
Lipson and Sheth	N	L	L	N	N	N	N	518	M	100	M	E	I
Lochner and Matar	L–M	M	M	L	N	N	H	241	L	50	M–H	E	I
Mason, Gunst and Hess	M	M–H	M	M	M	M–H	N–L	692	M–H	120	M	E–I	I
Mendenhall	L–M	L–M	L–M	L–M	L	L–M	N	464	L–M	100	M	I	G/I
Moen, Nolan and Provost	M	M–H	M	L	L	M–H	L–M	414	L	30	M–H	E	I
Montgomery	M–H	M–H	M–H	M–H	M–H	M–H	M	684	M–H	160	M	I	I
Myers and Montgomery	L	L–M	L–M	H	M	N	M–H	798	M–H	255	M	I	I
Robinson	L	L–M	L–M	L–M	L	L–M	L–M	265	N–L	100	H	I	G/I
Schmidt and Launsby	L	M	M	M	N	L	M	767	L–M	100	H	E–I	I
Wu and Hahamada	M–H	M–H	H	M–H	L	L–M	M–H	630	M–H	150	M	I	G/I

[a] N is none; L, little; M, moderate; and H, heavy; see General References for descriptions.

[b] E is elementary, no background in statistics needed; and I is intermediate, assumes one or two elementary statistics courses.

[c] G is general; I, industrial/engineering/scientific.

7.2. Factorial and Fractional Factorial Designs. These very popular designs apply for two or more primary independent variables. Factors are varied simultaneously rather than one at a time, so as to obtain information about interactions among variables and to obtain a maximum degree of precision in the resulting estimates. In complete factorial plans, all combinations of conditions of the independent variables are run. For example, a $3 \times 3 \times 2 \times 2$ complete or full factorial design requires running all 36 combinations of four variables at three, three, two, and two conditions, respectively.

A fractional factorial design is often used when there is a large number of combinations of possible test points, arising from many variables or many conditions per variable or both, and it is not possible or practical to run all combinations. Instead, a specially selected fraction is run. For example, a $(1/2)2^6$ fractional factorial plan is one where there are six variables each at two conditions, resulting in a total of 64 possible combinations, but only a specially selected one-half, or 32, of these combinations are actually run. The selection of test points is based upon the specific information that is desired from the experiment. For example, for a $(1/2)2^6$ fractional factorial plan one can independently estimate the impact of all two-factor interactions, but it is assumed that higher order interactions (ie, interactions among three or more factors) are negligible. The larger the number of primary variables, the greater is the degree of fractionation that is possible.

Fractional factorial plans are especially useful for screening purposes when it is desired to find those variables that have the greatest impact (over the specified experimental region). An additional advantage of full factorial and fractional factorial designs is that by providing a comprehensive scanning of the experimental region they can often identify, without formal analyses, a small number of test conditions that appear to provide especially favorable results. The region around these conditions would then be explored further in subsequent experimentation.

The most frequently used fractional factorials are for situations in which all variable are at two conditions (or levels). Test plans (providing specific test points) for this situation, together with their properties, are provided in most texts on the design of experiments (see Table 1) and by computer programs for generating experimental designs (see COMMERCIAL SOFTWARE). A popular family of two-level fractional factorial designs, known as Plackett-Burman plans, permit a large number of variables to be evaluated in a small number of test runs and is discussed in many standard texts. Next in popularity are fractional factorial plans with all variables at three levels. In addition, some situations, especially ones involving qualitative experimental variables, call for variables at a varying number of levels. These are discussed in more advanced texts.

Detailed discussions of fractional factorial designs are provided in books by Daniel, and, at a more advanced level, by Wu and Hamada, (see *General References*). Reference (17) provides a comprehensive catalogue of fractional factorial designs, including three- and mixed-level plans and their properties.

7.3. Response Surface Designs. These designs apply when one is dealing principally with quantitative independent variables (the x's), such as temperature and pressure, that one wishes to relate to the response variable(s) (y or y's), assuming some statistical model. When the form of this model is known

from physical considerations, then one would want to use that model form (if it is not too complicated). As already suggested, when the form of the relationship is not known, as is frequently the case, one instead approximates it by a polynomial (Taylor series approximation) model. A first-order linear model for k independent variables is expressed by the simple relationship

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \varepsilon \tag{1}$$

where ε represents the model error, which results from any unexplained behavior in y, and is usually assumed to be normally distributed with constant variance. This model error can result from experimental noise, exclusion of important x's in the model, or the need for higher order terms in the model. The results of the experiment are used to estimate the unknown β's using least-squares regression analysis (see STATISTICAL TOOLS).

However, the model that is most frequently used in practice is the second-order model

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum_{i} \sum_{<j=2}^{k} \beta_{ij} x_i x_j + \varepsilon \tag{2}$$

which includes linear (x_j), curvature (x_j^2), and interaction ($x_i x_j$) terms.

An intermediate model might include the interaction (or cross-product) terms, but not the quadratic ones.

The choice of an appropriate model merits considerable reflection, based on both physical and empirical considerations. As already suggested, a linear model may provide a reasonable approximation over a narrow range, but a second-order, or more complex, relationship might be called for over a broader range. Also, often based upon physical considerations, an improved, or simpler, representation might be gained by transforming the original x's to some other scale (such as taking logarithms or reciprocals). In addition, transformations of the y's is also frequently beneficial.

Two-level full and fractional factorial designs can be used to explore first-order response surface models. Three level full factorial and some fractional factorial designs apply for second-order models. However, these may require more experimental points than are feasible from a practical viewpoint. For this and other reasons, specialized designs have been developed for response surface analysis, especially for second-order models. Most popular among these are the so-called central composite (or Box) designs and the Box-Behnken designs.

Response surface designs are discussed in many books on experimental design (see Table 2), and in detail in the book by Myers and Montgomery, (see *General References*). In addition, various recent developments are reviewed in (18).

7.4. Optimal Designs.   Another type of experimental design, used principally for response surface exploration, and favored by some statisticians, are the so-called "optimal designs". Such designs are generally computer-generated and have properties such as minimizing the variance of the predictions over a

Table 2. **Commercial Software Comparison**[a]

	Design-ease 6.0.6	Design-expert 6.0.5	DOE Pro XL 3.0	ECHIP	JMP 4.0.4	Minitab 13	STATGRAPHICS plus 5 (quality and design)
Design feature							
blocking	H	H	N/A	H	H	H	H
randomization	H	H	L	H	H	H	H
Design Type							
full and fractional factorials	H	H	H	H	H	H	H
central composite	N/A	H	H	M	H	H	H
Box-Behnken	N/A	H	M	H	H	H	H
mixture	N/A	H	N/A	H	L	M	M
mixture-process	N/A	H	N/A	H	N/A	M	N/A
Taguchi orthogonal arrays	H	H	H	H	H	H	H
optimal designs	L	L	N/A	M	N/A	N/A	N/A
Analysis methods							
graphical tools	H	H	H	H	H	H	H
ANOVA	M	M	H	H	H	H	H
regression analysis	M	H	H	H	H	H	H
Software features							
ease of using design features	H	H	M	H	L	M	H
general purpose software?	No	No	No	No	Yes	Yes	Yes
price[b] for single user license (in U.S. dollars)	$495	$995	$199	$1495	$895[c]	$1195	$999
vendor (see text for more information)	Stat-Ease, Inc.	Stat-Ease, Inc.	Air Academy Associates	ECHIP, Inc.	SAS Institute Inc.	Minitab Inc.	Manugistics, Inc.

[a] Table key: "N/A" = feature not available; "L" = limited capability; "M" = moderate capability; "H" = high capability.

[b] The quoted price is for commercial (nonacademic) users. An academic discount is available from most vendors. Quoted prices are as of March, 2002.

[c] ECHIP license must be renewed annually, at a cost of $150, which includes software upgrades. Other products are purchased, with some additional cost to upgrade to newer version.

specified "design region" (the so-called G-optimal design). The best-known design of this type is the D-optimal design, which "minimizes the volume of the joint confidence region on the vector of regression coefficients" (see book by Montgomery in *General References*).

Optimal designs are of particular value in situations when a full factorial, fractional factorial, Box, Box-Behnken (or other "symmetrical-type") design is not practical or desirable. Thus, in various applications, some points in the design space defined by, say, a full factorial experiment, might be uninteresting, or inappropriate. This might be because, based upon prior knowledge, the results at that condition are fully predictable, or, perhaps, known to be inferior. Another possibility might be that a particular experimental test condition might be potentially hazardous. Sometimes, such undesired points can be avoided by an appropriate redefinition of the experimental variables. If this is not possible, the resulting design region might be asymmetrical. For this case, optimal designs might be especially useful.

In addition, optimal designs are employed for situations in which, based upon process knowledge, one wishes to fit a nonstandard model (ie, one that assumes something other than a first- or second-order polynomial model with normally distributed errors), or in which one is limited to running a specified number of test points. The book by Montgomery (see *General References*) provides further information and a specific example. Additional detail is provided in the book by Myers and Montgomery (see *General References*). Optimal designs can be generated by the software package ECHIP, and, in a more limited manner, by Design-Ease and Design-Expert (see Commercial Software).

7.5. Mixture Designs. These designs arise when the variables under consideration are the ingredients of a product that must add to 100%, for example in an experiment to learn how to optimize the baking of a cake, the relative impact of each of the ingredients (baking powder, shortening, flour, sugar, milk, water, etc) might need to be assessed. Often, an experimenter will want to study *both* mixture and nonmixture variables. In the previous example, one may wish to assess the effect of varying bulk density of the flour or the fat content of the milk, in addition to the relative percentages of the cake ingredients. These designs are referred to as *mixture-process designs*. Both mixture and mixture-process designs are especially relevant in the evaluation of chemical processes. Most standard texts do not discuss these plans. However, detailed expositions are provided in (19) and (20).

7.6. The Taguchi Approach. An eminent Japanese engineering professor and advocate for quality improvement, Genichi Taguchi, has been an important proponent of planned experimentation in product design, development, and evaluation. In fact, the influence of Taguchi and his followers has been an important factor in popularizing the use of experimental design in industry. Taguchi has especially promoted "orthogonal arrays", a family of experimental designs closely related to fractional factorial plans.

Taguchi has also stressed the role of experimental design in identifying the process conditions that show the greatest consistency, or are most robust, in the face of variability in manufacturing conditions and customer use. The goal is to achieve similar and desirable product performance despite uncontrollable variation due to such noise variables as incoming raw materials, and the ambient

conditions under which a product is to be used. For example, an automobile battery needs to work well, irrespective of whether a car is driven in Florida or Minnesota.

In particular, Taguchi has introduced experimental plans that help identify the conditions (ie, settings of the primary variables) that minimize the variability in the noise variables, thus elevating the role of such variables from nuisance factors whose impact needs merely to be neutralized by appropriate blocking, etc, in the experimental plan.

Many classical blocking designs, such as incomplete block and Latin square designs assume that there is no interaction between the primary variables and the noise variables and that the variability generated by the noise variables is the same for all conditions of the primary variables. These plans can, therefore, not be used for identifying the conditions of the primary variables that minimize variability. Instead, Taguchi has proposed "robust parameter designs" for this purpose, including the concept of "inner and outer array" plans. In such designs, the noise variables (outer array) are evaluated for a series of conditions of the primary variables (inner array). Typically, the inner array is in the form of a fractional factorial plan, while the outer array is a full factorial plan. These designs have been criticized due to the fact that, although they provide the desired information about the noise variables at various conditions of the primary variables, they tend to provide limited information about interactions among primary variables (due to the fractionization of the inner array). Also, though appropriate for some "split-plot" situations, these plans frequently lead to a large number of test conditions. Instead, some "combined array" experimental plans have been proposed to meet Taguchi's basic goal, but in a more efficient manner; see the book by Montgomery (*General References*) for a more detailed discussion.

Taguchi has also proposed various methods of data analysis and other concepts, some of which are controversial. These include the concept of combining the mean and variability in the response variable in the form of the "signal/noise ratio," which is to be maximized. This concept has been criticized on the grounds that it might be more meaningful to evaluate the mean response and the variability separately, rather than combining them into a single metric.

A series of articles that describe, illustrate, and critique Taguchi methods is provided in (21). Most texts on design of experiments published or republished since around 1990 contain discussions of the Taguchi approach. In addition, various books devoted exclusively (or principally) to Taguchi methods, including Taguchi's approach to experimental design have recently appeared [(22−26) and Lochner and Matar, see *General References*].

8. Combinatorial Chemistry

Technology now allows us to search vast experimental regions to identify materials with desired properties. This is the realm of "combinatorial chemistry", also known as high throughput screening (or "finding a needle in a haystack" experimentation). In fact, combinatorial chemistry "is now practical because of the convergence of low-cost computer systems, reliable robotic systems, sophisticated

molecular modeling, statistical experimental strategies, and database software tools" (27).

Early applications of combinatorial chemistry were targeted at identifying potentially promising drugs by pharmaceutical companies. However, in recent years, similar approaches are being used by chemical and other companies for new product discovery and development. Frequently, the goal is to establish a library of performance responses for different material combinations. This library can then be searched to rapidly develop a material that satisfies customer specifications, possibly in some optimum manner—after, perhaps, a relatively small amount of validation testing. Alternately, this information may be used proactively to develop a new material with favorable properties that out-perform the competition, or to find one "miracle combination".

Combinatorial chemistry for materials improvement has been made possible by the ability to make up and assess, possibly, hundreds of vials a day. However, the "experimental space" that one wishes to explore may consist of millions of possible combinations of conditions. For example, in one typical application (27), the experimental space consisted of the 2,916,000 possible combinations of

- Three qualitative formulation variables with 20, 20, and 10 conditions, respectively.
- Three quantitative formulation variables, each with three conditions.
- Three quantitative process variables, each with three conditions.

Clearly, even with 100 vials per day, one can explore only a small fraction of the possible conditions in the time that is generally available for such explorations. Thus, one turns to the design of experiments and related methods to help in making the testing maximally efficient. This has, in itself, generated a variety of strategies for defining test points for combinatorial chemistry applications, see (27–30). Among these, only (27) discusses statistical plans, such as fractional factorial designs, in some detail.

General design of experiment concepts have much applicability for combinatorial chemistry. For example, stage-wise experimentation, and the resulting sequential learning process, discussed previously, is especially relevant in such assessments. In an initial stage, one might consider the entire response surface to identify potential regions that warrant further exploration. In subsequent testing, one would home in on such promising regions with various stages of further testing.

9. Statistical Tools for the Analysis of Designed Experiments

We have so far in this article made little mention of statistical tools for the analysis of experiments for two reasons. First, as stated at the outset, we feel that the proper design of the experiment is more important than its analysis. Second, various books on experimental design, including those in the General References, provide ample descriptions of these methods. Thus, we provide only a brief summary. In particular, the analysis of most designed experiments involves a combination of three major types of tools: the analysis of variance, regression analysis,

and graphical analysis. We will comment briefly on each. In so doing, we reiterate that in stagewise experimentation such analyses are interspersed with the experimental runs and, in fact, provide important inputs on how to proceed in the next stage of testing.

The analysis of variance is, probably, the most frequently used tool for analyzing the results of a designed experiment. It is a formal statistical method that breaks down the total observed variability in a response variable into separate parts attributable to each of the sources of variation, based on an assumed statistical model. Thus, the analysis of variance isolates the variation due to the main effects of each of the experimental variables and their interactions and assess the "statistical significance" of each of these sources of variation. This provides useful information about the relative importance in the experiment of different effects and their interactions.

The results of an analysis of variance should, however, not be overinterpreted because "statistical significance" does not necessarily imply practical significance. In particular, a statistically significant result is one that is unlikely to have occurred due to chance alone. It does not, per se, measure the magnitude of the impact of the associated effect. For Type II variables, the analysis of variance also provides insightful estimates, in the form of so-called "variance components", of the variability attributable to individual sources of variation.

Regression analysis allows one to fit a relationship between a series of experimental variables (x's) and a response variable (y), which is especially relevant for data from response surface designs. For such designs, regression analysis allows one, for example, to estimate the unknown β's in an assumed model, such as the previously discussed second-order model, together with their associated confidence bounds. This will permit prediction of an expected response and/or variability for that response, also with statistical confidence bounds, for a specified combination of the experimental variables, based upon the assumed model. In fact, regression analysis has been referred to as "the workhorse of statistical data analysis".

Sometimes, when one is searching over a vast experimental space, as is frequently the case, for example, in combinatorial chemistry, the standard first-or second-order polynomial models (even after transformations) are frequently inadequate to relate the formulation and process variables to the material response variables. Thus, one might need to supplement standard statistical methods (such as regression analysis and analysis of variance) with more advanced tools [such as MARS (multiple adaptive regression splines) (31)], to obtain, at least approximate, smoothed representations of the response surface.

After a model is fitted, it is frequently desired to exercise it to find an optimum condition, or set of experimental conditions. This would be the set of conditions that minimizes or maximizes the response variable when one is considering the average response, or the condition that minimizes the variability (or some combination of the two); the book by Myers and Montgomery (see *General References*) provides further discussion.

We strongly encourage graphical analyses to supplement the more formal statistical analyses, and, on occasion, to take their place. To quote our former colleague Wayne Nelson, one good graphical display is often worth a thousand statistics. For example, incisive multiple dimensional plots of the experimental

results, or of the fitted regression analysis, can frequently be highly informative. Also, half-normal probability plots (see the book by Daniel in *General References*) provide information similar to a formal analysis of variance concerning the relative importance of the individual sources of variation, but in graphical form.

Finally, we note that improved methods for the analysis of designed experiments, frequently capitalizing on increased computations capabilities, continue to be developed. A typical example is the use of Bayesian methods in finding the active factors in fractionated screening experiments (32). Such new methods are published in *Technometrics, The Journal of Quality Technology*, and other technical publications.

10. Multiple Response Variables

The preceding discussion has assumed, by and large, that we are dealing with a single response variable, although, in practice, this frequently is not the case. As already mentioned, one may be interested in both the average and the variability of the response variable (leading Taguchi to propose the controversial signal/noise ratio, to combine these two variables into one for the purpose of analysis). Moreover, often one is concerned with two or more performance variables, such as, for example, the viscosity, tensile strength, ductility, etc, of a plastic material. Multiple response variables complicate the analyses as well as the stagewise development of the experimental plan, especially if we are seeking an optimum experimental region (versus just developing a response library). The problem of multiple response variables is discussed in the books by Wu and Hahmada and by Myers and Montgomery (see *General References*). The tools of chemometrics also have applicability in such situations.

11. Commercial Software for the Design and Analysis of Experiments

Numerous computer software packages have been developed to help generate experimental designs, [see (33) for an evaluation of some early packages]. These vary greatly in price, computer platform, and level of technical sophistication assumed of the user. They are especially useful in performing the mechanical tasks associated with the design and analysis of experiments, such as the generation of a matrix of test points and the statistical analysis of the resulting data. Computer software is needed for determining the optimum plans, discussed above, that are often relevant for special situations (see also COMPUTER-AIDED ENGINEERING). However, while almost all of these packages have features for generating full and fractional factorial, and response surface designs and the orthogonal array plans advocated by Taguchi, only a few can generate optimal designs.

Seven software packages were evaluated for their capabilities for the generation of experimental designs and for the analysis of the resulting data, as well as their ease of use. This information, together with current cost, is summarized in Table 2. This tabulation by no means provides an exhaustive list of currently available commercial software packages. However, it does include the best

known ones, and describes offerings with varying depths of capabilities and costs, ranging from current single-user commercial license fees of $199 to $1495.

Note that four of the packages are specialized in the sense that their major purpose is the design and analysis of experiments. In contrast, the other three offerings (JMP, Minitab, and Statgraphics) are well-known general statistical packages that are already in place in many facilities and that, among their many features, have capabilities for the design and analysis of experiments (the license fee shown in Table 2 is for the entire package).

12. Some Historical Background

Sir Ronald Fisher is generally regarded as the "founding father" of the design of experiments, and wrote an early book on the subject (34). As previously indicated, many of the initial applications were in agriculture. Application to industry, in general, and the chemical industry, in particular, was recognized after World War II, and was spurred on by important work by George Box, Cuthbert Daniel, and Stuart Hunter, among others. Starting in the late 1970s, the previously mentioned work by Genichi Taguchi and his associates generated extensive interest in experimental design in industry (as well as some controversy). Most recently, the Six Sigma initiative (35) has resulted in focusing a great deal of attention by some large corporations on the design of experiments. In fact, the statistical design of experiments is one of the key items in the Six Sigma "toolbox".

13. Acknowledgments

BIBLIOGRAPHY

"Design of Experiments" in *ECT* 3rd ed., Vol. 7, pp. 526–538, by G. J. Hahn, General Electric Co.; in *ECT* 4th ed., Vol. 7, pp. 1056–1071, by Gerald J. Hahn, General Electric Company; "Design of Experiments" in *ECT* (online), posting date: December 4, 2000, by Gerald J. Hahn, General Electric Company.

CITED PUBLICATIONS

1. G. J. Hahn, *Chemtech* **5**, 496 (Aug. 1975); **5**, 561 (Sept. 1975).
2. G. J. Hahn, *J. Quality Technol.* **9**, 13 (Jan. 1977).
3. D. E. Coleman and D. C. Montgomery, *Technometrics* **35**, 1 (Feb. 1993), with discussion.
4. F. X. Mueller and D. M. Olsson, *J. Paint Technol.* **43**, 54 (May 1971).
5. G. J. Hahn, *Technometrics* **26**, 19 (Feb. 1984).
6. C. D. Hendrix, *Chem. Technol.* **16**, 228 (1986).

7. J. S. Hunter, *Chem. Technol.* **17**, 187 (1987).

8. J. S. Hunter, in J. M. Juran, ed., *Quality Control Handbook*, 5th ed., McGraw-Hill Book Co., New York, 1998, Chapt. 47.

9. F. G. Tingey, in I. M. Kolthoff and P. J. Elving, eds., *Treatise on Analytical Chemistry*, Pt. 1, Wiley-Interscience, New York, 1972, Chapt. 106.

10. D. M. Steinberg and W. G. Hunter, *Technometrics* **26**, 71 (May 1984), with discussion.

11. G. J. Hahn, *Chem. Technol.* **7**, 630 (Oct. 1977).

12. G. J. Hahn, *Chem. Technol.* **8**, 164 (Mar. 1978).

13. G. E. P. Box and P. Y. T. Liu, *J. Quality Technol.* **31**, 1 (1999).

14. G. E. P. Box and N. R. Draper, *Evolutionary Operation: A Statistical Method for Process Improvement*, John Wiley & Sons, New York, 1969.

15. G. J. Hahn, *Chem. Technol.* **6**, 204 (March 1976).

16. G. J. Hahn, *Chem. Technol.* **6**, 343 (May 1976).

17. G. J. Hahn and S. S. Shapiro, *A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments*, General Electric Co. TIS Report 66C165, General Electric Co., Schenectady, N.Y., May 1966.

18. R. H. Myers, *J. Quality Technol* **31** 1 (1999).

19. R. D. Snee, *J. Quality Technol.* **3**, 159 (Oct. 1971).

20. J. A. Cornell, *Experiments with Mixtures*, 3rd ed., John Wiley & Sons, Inc., New York, 2002.

21. *Qual. Relia. Eng. Int.* **4**, 193 (Aug. 1988).

22. T. B. Barker, *Engineering Quality by Design: Interpreting the Taguchi Approach*, Statistcs Textbooks and Monographs, No. #113, Marcel Dekker, New York 1990.

23. W. Y. Fowlkes and C. M. Creveling, *Engineering Methods for Robust Product Design: Using Taguchi Methods in Technology and Product Development*, Prentice-Hall PTR, Upper Saddle River, N.J., 1995.

24. M. S. Phadke, *Quality Engineering Using Robust Design*, Prentice-Hall, Englewood Cliffs, N.J., 1989.

25. R. K. Roy, *Design of Experiments using Taguchi Approach: 16 Steps to Product and Process Improvement*, John Wiley & Sons, Inc., New York, 2001.

26. H. P. Wynn and N. Logothetis, *Quality through Design: Experimental Design, Off-Line Quality Control and Taguchi's Contributions*, Oxford Series on Advanced Manufacturing, No. 7, Oxford University Press, New York, 1990.

27. J. N. Cawse, *Acc. Chem.Res.* **34**, 213 (2001).

28. M. W. Deem, *A Statistical Mechanical Approach to Combinatorial Chemistry*, in *Advances in Chemical Engieering*, Academic Press, San Diego, 2002.

29. T. X. Sun, *Biotechnol. Bioeng. Combinatorial Chemis.* **61**, 193 (1999).

30. D. Wolf, O. V. Buyevskay, and M. Baerns, *Appl. Catalysis A General* **200**, 63 (2000).

31. J. H. Friedman, *Ann. Sta.* **19**, 1 (1991).

32. G. E. P. Box and R. D. Meyer, *J. Quality Technol.* **25**, 94 (1993).

33. G. J. Hahn and C. B. Morgan, *Chem. Technol.* **18**, 664 (Nov. 1988).

34. R. A. Fisher, *The Design of Experiments*, Oliver and Boyd, London, 1935.

35. M. Harry and R. Schroeder, *Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's Top Corporations*, Currency/Doubleday, New York, 2000.

GENERAL REFERENCES AND SOME RECOMMENDATIONS

This section provides a subjective summary and evaluation of books that mainly deal with the application of experimental design to scientific, industrial, and general situations. Books directed principally at educational, psychological, or related applications, and

those dealing mostly with the theory of experimental design, or principally with the analysis of experimental data, are omitted. Most of the books presume that the reader has had at least one introductory course in statistics. Items in quotation marks are taken directly from the book, generally from its preface. Table 1 provides a comprehensive summary of the subject coverage, technical level, applications emphasis, etc of these books. In addition, such journals as CHEMTECH, The Journal of Quality Technology, and Technometrics (in increasing order of complexity) periodically carry articles about experimental design.

The book by Box, Hunter and Hunter (1978) is, perhaps, the one that has become best known among practitioners. It provides a good starting point for those with limited knowledge of statistics, although its 653 pages may be a little foreboding. The current and only edition, issued in 1978, does not cover various topics that have received significant attention in recent years, such as the Taguchi approach and mixture designs. However, a new edition is scheduled for 2003. Basic-to-intermediate offerings are also provided in the recent books by Berger and Maurer (2002), and by Schmidt and Launsby (2000).

The most elementary, and, as a consequence, least in-depth, overview discussion is provided in the 175 page introductory book by Del Vecchio (1997). At the other end of the spectrum, those who wish an in-depth and topical, but still applications-oriented, discussion might want to turn to the 684 page volume by Montgomery (2001), or the 630 page book by Wu and Hamada (2000).

Those interested in a particular type of experimental design might want to turn to a specialized text, as suggested in the discussion in this article and in Table 1. Thus, the classic text by Daniel (1976) is especially strong on fractional factorial designs, and Myers and Montgomery (2002) specializes in response surface experiments.

V. L. Anderson and R. A. McLean, *Design of Experiments—A Realistic Approach*, Marcel Dekker, New York, 1974. This book provides an extensive exposition of experimental design at a relatively elementary level. It includes most of the standard material, as well as detailed discussions of such subjects as nested and split-plot experiments. Restrictions on randomization receive special emphasis. Published in 1974, this book does not include subsequent developments.

P. D. Berger and R. E. Maurer, *Experimental Design with Applications in Management, Engineering and the Sciences*, Duxbury, Belmant, CA2002. The goal of this book is to cover "the most important and commonly used methods in the field of experimental design." It claims "to minimize the amount of mathematical detail, while still doing full justice to mathematical rigor..." and "to provide an intuitive understanding of the principles." Thus, it is intermediate between books that provide a simple introductory exposition and those that try to tell it all. An unusual feature is the concluding chapter that provides a personalized assessment of the literature (mainly other books) on experimental design. The other 14 chapters include ones on nested or hierarchal designs and on Taguchi methods, as well as more standard discussions on factorial, fractional factorial and response surface designs.

G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons, Inc., New York, 1978 2nd ed. scheduled for 2003). This book, by three eminent practitioners, has become one of the most popular works on the subject. It "is an introduction to the philosophy of experimentation and the part that statistics plays in experimentation". It provides a practically motivated introduction to basic concepts and methods of experimental design. It "is written for those who collect data and try to make sense of it", and gives an "introduction to those ideas and techniques" that the authors have found especially useful. Statistical theory is introduced as it becomes necessary. Readers are assumed to have no previous knowledge of the subject; "the mathematics needed is

elementary". The book includes numerous examples and case studies and provides appreciable detail about many elementary and a few more advanced designs. It is, however, an introductory treatment; therefore, more advanced situations are not discussed. Heavy emphasis is placed on the use of graphical methods for analyzing experimental results. In contrast, the more computationally involved analysis-of-variance tools receive relatively little attention. The four parts of the book deal with (*1*) comparing two treatments, (*2*) comparing more than two treatments, (*3*) measuring the effects of variables, and (*4*) building models and using them. Published in 1978, the first edition of this book does not include subsequent developments.

D. R. Cox, *Planning of Experiments*, John Wiley & Sons, Inc., New York, 1958. (Reprinted in Wiley Classic Series, 1994.) This book provides a simple survey of the principles of experimental design and of some of the most useful experimental schemes. It tries "as far as possible, to avoid statistical and mathematical technicalities and to concentrate on a treatment that will be intuitively acceptable to the experimental worker, for whom the book is primarily intended". As a result, the book emphasizes basic concepts rather than calculations or technical details. Chapters are devoted to such topics as 'Some key assumptions", "Randomization", and "Choice of units, treatments, and observations". This book has become a classic (as reflected by its republication). Published in 1958, it does not include subsequent developments.

C. Daniel, *Applications of Statistics to Industrial Experimentation*, John Wiley & Sons, Inc., New York, 1976. This book is based on the personal experiences and insights of the author, an eminent practitioner of industrial applications of experimental design. It provides extensive discussions and concepts, especially in the areas of factorial and fractional factorial designs. "The book should be of use to experimenters who have some knowledge of elementary statistics and to statisticians who want simple explanations, detailed examples, and a documentation of the variety of outcomes that may be encountered." Some of the unusual features are chapters on "Sequences of fractional replicates" and "Trend-robust plans", and sections entitled, "What is the answer? (what is the question?)", and "Conclusions and apologies". Published in 1976, this book does not include subsequent developments.

O. L. Davies and co-workers, *The Design and Analysis of Industrial Experiments*, 2nd ed., Hafner, New York, 1967; reprinted by Longman, New York, 1978 (out of print). This book, which is a sequel to the authors' basic text *Statistical Methods in Research and Production*, is directed at industrial situations and chemical applications. Three chapters are devoted to factorial experiments and one chapter to fractional factorial plans. A lengthy chapter (84 pp.) discusses the determination of optimum conditions and response surface designs, which are associated with the name of George Box, one of the seven co-authors. Theoretical material is presented in chapter appendices. This book was one of the earliest to reflect chemical and industrial applications. Published in 1978, it does not include subsequent developments.

R. J. Del Vecchio, *Understanding Design of Experiments*, Hanser/Gardner Publications, Inc., Cincinnati, Ohio, 1997. This short, introductory paper-back volume, sponsored by the Society of Plastics Engineers, claims to be "presented in ... a simplified and non-conforming format and at a level where it can ... provide a jumping-off place for those who may wish to move on to the many weightier publications available .... The goals are to explain the basics underlying designed experiments, supply instructions on how to use several families of convenient designs ... and provide some overviews on assorted subtopics ...". The 23 short chapters move from such introductory topics as "What are Designed Experiments?" and "How Hard are Designed Experiments to Use?" to more advanced topics such as "Basics of Mixture Designs", "Taguchi's Contributions", and "Computer Programs for Design of Experiments". Thus, it surveys a broad area, but covers each topic only at an introductory level. For example, the

discussion of fractional factorial plans is limited to the most popular two- and three-level designs. Only limited references are provided.

W. J. Diamond, *Practical Experimental Designs for Engineers and Scientists*, 3rd ed., Van Nostrand Reinhold, New York, 2001 (1st ed., 1981). "This book is for engineers and scientists with little or no statistical background who want to learn to design efficient experiments and to analyze data correctly . . . . The emphasis is on practical methods, rather than on statistical theory". The discussion is quite detailed in some areas, eg, fractional factorial designs based on Hadamard matrices, and scanty in others. Some unusual features are chapters on "Random Strategy Experimentation" and "Project Engineer's Game". The third edition does not seem to be a significant update from the second edition, published in 1989. Thus, there is little discussion of modern computer software and there appear to be no references later than 1988.

C. R. Hicks and K. V. Turner, Jr., *Fundamental Concepts in the Design of Experiments*, 5th ed., Oxford, New York, 1999. (1st ed., 1963) "The primary purpose of this book to present the fundamental concepts in the design of experiments using simple numerical problems, many from actual research work . . . . The book is written for anyone engaged in experimental work who has a good background in statistical inference. It will be most profitable reading to those with a background in statistical methods including analysis of variance". This work provides an intermediate coverage of most basic experimental designs. Recent editions "provide more emphasis on the use of computer outputs as part of the analysis". However, the authors continue to emphasize formulas for hand calculation.

R. O. Kuehl, *Design of Experiments: Statistical Principles of Research Design and Analysis*, 2nd ed., Pacific Grove, Calif, 2000 (1st ed. 1994). The objective of "this text is to present the principles of statistical design and analysis for comparative scientific studies to graduate students in the experimental sciences and applied statistics". As such, it is a reasonably complete, intermediate level, text directed mainly at graduate students in statistics. The 17 chapters include four on block designs, and ones on experiments to study variances, split-plot designs, repeated measures designs, and cross-over designs, as well as the more standard chapters on factorial, fractional factorial, and response surface designs.

C. Lipson and N. J. Sheth, *Statistical Design and Analysis of Engineering Experiments*, McGraw-Hill, New York, 1972 (out of print). "This book is written in a relatively simple style so that a reader with a moderate knowledge of mathematics may follow the subject matter. No prior knowledge of statistics is necessary". Appreciably more discussion is devoted to statistical analysis than to the planning of experiments. Some relatively nonstandard subjects for an introductory text, such as accelerated experiments, fatigue experiments, and renewal analysis are also included. Published in 1972, this book does not include subsequent developments.

R. H. Lochner and J. E. Matar, *Designing for Quality: An Introduction to the Best of Taguchi and Western Methods of Statistical Experimental Design*, Quality Resources, White Plains, N.Y., and ASQ Quality Press, Milwaukee, Wis., 1990. "The purpose of this book is to show engineers with little or no previous exposure to experimental design how to use statistically designed experiments to improve products and processes. A prerequisite . . . is an appreciation for the concept of random variation, as is gained through a typical statistics course or by using Statistical Process Control (SPC) in a manufacturing environment". As its subtitle suggests, this book merges a discussion of concepts introduced or popularized by G. Taguchi, with those of classical experimental design.

R. L. Mason, R. F. Gunst, and J. L. Hess, *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*, John Wiley & Sons, Inc., New York, 1989. "The intended audience of this book consists of two groups. The first group covers

a broad spectrum of practicing engineers and scientists, including those in supervisory positions .... This audience includes those who have little formal training in statistics .... The second group ... is students in introductory statistics courses ... in which statistical experimental design and the analysis of data are the main topics .... College algebra is the only prerequisite." This book, in addition to discussing the design of experiments, provides a comprehensive introduction to modern statistical methods. Following two introductory chapters, the four parts of this book deal with describing variability, experimental design, analysis of designed experiments, and fitting data.

W. Mendenhall, *Introduction to Linear Models and the Design and Analysis of Experiments*, Duxbury Press, Belmont, Calif., 1968 (out of print). This book provides an introduction to basic concepts and the most popular experimental designs without going into extensive detail. In contrast to most other books, the emphasis in the development of many of the underlying models and analysis methods is on a regression, rather than an analysis-of-variance, viewpoint. Published in 1968, this book does not include subsequent developments.

R. D. Moen, T. W. Nolan, and L. P. Provost, *Improving Quality through Planned Experimentation*, McGraw-Hill, New York, 1991. "This book is about planned experimentation to improve quality." It aims to promote "a substantial increase in the number of people who will use these methods" by requiring "a lower level of mathematical and statistical sophistication than was previously necessary .... Taguchi's contributions to the application of experimentation to the design of products and processes have been integrated ... (and) the book is written to be compatible with W. E. Deming's viewpoint of analytic studies". Graphical, rather than formal statistical, methods for data analysis are presented.

D. C. Montgomery, *Design and Analysis of Experiments*, 5th ed., John Wiley & Sons, Inc., New York, 2001. (1st ed., 1976) This "introductory textbook dealing with the design and analysis of experiments ... is intended for readers who have completed a first course in statistical methods". It provides a detailed treatment of standard experimental plans and techniques for analyzing the resulting data, and has become one of the most popular books on the subject. Five chapters are devoted to factorial and fractional factorial designs, and added chapters deal with experiments with random factors and nested and split-plot designs. The revised editions include an introduction to mixture experiments and the Taguchi approach, and give greater emphasis to computer analyses.

R. H. Myers and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd ed., John Wiley & Sons, 2002 (1st ed., 1995). This book deals principally "with the exploration and optimization of response surfaces"—an important applications area for design of experiments. However, it also provides some of the fundamentals of experimental design. The book has been used "in a graduate-level course on Response Surface Methodology for statisticians, engineers, and chemical/physical scientists", as well as in "industrial short courses and seminars for individuals with a wide variety of technical backgrounds". Six of fourteen chapters are devoted to response surface designs and methodology, but there are also chapters on two level factorial and on two-level fractional factorial designs. The book also contains two chapters on mixture experiments and introductory chapters on robust parameter design and on continuous online process improvement with evolutionary operation (EVOP).

G. K. Robinson, *Practical Strategies for Experimentation*, John Wiley & Sons, 2000. This book claims to have "a wider scope but less mathematical depth. It gives anecdotes and checklists of suggestions about experimenting .... It presents an approach to experimentation on a strategic level. It describes broadly how to proceed and what questions to think about .... " For example, the nine chapters include ones on "Clarify the Objective", "Decide on a Strategy", and "Revisit the Objective". Although the book is written

on a relatively elementary level, it is assumed that readers are familiar with the standard experimental designs discussed in most other texts on the subject. Instead, the book leverages the author's experiences and insights dealing with the practical and strategic aspects of experimentation. Thus, one example begins with "This is an example of how an experimental programme can fail badly due to inadequate consultation".

S. R. Schmidt and R. G. Launsby, *Understanding Industrial Designed Experiments*, 4th. ed., Air Academy Press, Colorado Springs, Co. 2000 (1st ed., 1988). "This text is written for managers, engineers, scientists and technicians. It is intended to bridge the gap between texts that are too mathematical and those that attain simplicity by omitting important concepts." It is made up of seven relatively lengthy chapters on foundations (stressing basic concepts and steps); conducting simple experimental designs and analysis; design types; statistical techniques; analysis of experimental data; Taguchi philosophy, design and analysis; and optimization and advanced response surface methods. These are followed by a 280 page chapter that presents 31 unaltered case studies and example topics, submitted by different individuals. The prime emphasis of this book is on industrial and quality improvement applications. The text includes a disk for the student version of the DOE KISS (Keep It Simple Statistically) software.

C. F. J. Wu and M. Hamada, *Experiments: Planning, Analysis, Parameter Design Optimization*, John Wiley & Sons, New York, 2000. This book, written by two authorities in the field, is directed at the serious student who wishes complete and up-to-date knowledge. Thus, it provides a detailed exposition of a large variety of experimental designs, as well as some of the "newest discoveries". Even though the "intended readership ... includes general practitioners as well as specialists," this book will likely be used most extensively in graduate courses in experimental design for statisticians. It will also serve as a useful reference. Fractional factorial designs, beyond two-level designs are discussed in detail. Fore example, there is a chapter on complex aliasing. The 13 chapters also include two on (Taguchi) robust parameter designs and ones on experiments for improving reliability and experiments with nonnormal data. Each chapter concludes with a useful "Practical Summary". As noted in a recent review in *Technometrics*, two downsides of the book are its lack of coverage of mixture designs (in light of its extensive coverage of other current important topics) and its limited discussion of statistical software packages for the design and analysis of experiments.

SOFTWARE VENDOR INFORMATION

Air Academy Associates 1650 Telstar Drive, Suite 110, Colorado Springs, CO 80920 Phone: (719) 531−0777 E-mail: suellen@airacad.com, Web: http://www.airacad.com

ECHIP, Inc.724 Yorklyn Road, Hockessin, DE 19707-8733 Phone: (302) 239-5429 E-mail: support@echip.com, Web: http://www.echip.com

Manugistics, Inc. 2115 East Jefferson St., Rockville, MD 20852 Phone (800) 592-0050 E-mail: sgsales@manu.com, Web: http://www.statgraphics.com

Minitab Inc. 3081 Enterprise Drive, State College, PA 16801-3008 Phone: (814) 238-3280 E-mail: sales@minitab.com, Web: http://www.minitab.com

SAS Institute Inc. JMP Software, SAS Campus Drive, Cary, NC 27513 Phone: (919) 677-8000 E-mail: jmpsales@sas.com, Web: http://www.sas.com

Stat-Ease, Inc. 021 East Hennepin Avenue, Suite 191, Minneapolis, MN 55413-2723 Phone: (612) 378-9449 E-mail: Stat-Ease, Inc., Web: http://www.statease.com

GERALD J. HAHN
ANGELA N. PATTERSON
General Electric