Overview

Fractional factorial design

Richard F. Gunst[1]* and Robert L. Mason[2]*

Fractional factorial designs are among the most important statistical contributions to the efficient exploration of the effects of several controllable factors on a response of interest. Fractional factorials are widely used in experiments in fields as diverse as agriculture, industry, and medical research. A key feature of fractional factorials that is not shared by more ad hoc methods for reducing the size of experiments is that the statistical properties are known in advance of experimentation. Consequently, an experimenter can investigate alternatives that enable the goals of the experiment to be met with the least cost, shortest time, or most effective use of resources. On occasion, an experimenter might decide not to conduct an experiment as originally planned once the statistical properties of the design are known. This article highlights the fundamental concepts, design strategies, and statistical properties of fractional factorial designs.. *WIREs Comp Stat* 2009 1 234–244

INTRODUCTION

Fisher[1] (p. 94) introduced complete factorial designs (all combinations of the factor levels are included) in the first edition (1935) of his classic work on experimental design with the following statement: '... in a wide class of cases an experimental investigation, at the same time as it is made more comprehensive, may also be made more efficient if by more efficient we mean that more knowledge and a higher degree of precision are obtainable by the same number of observations.' Fisher elaborated in elegant detail the basis for recommending complete factorial designs by citing three advantages for doing so: greater efficiency, greater comprehensiveness, and a wider inductive basis (p. 101ff). Box and Hunter[2] (p. 313) appear to apologize for the richness of results with complete factorial experiments in their classic article on the construction of fractional factorial designs: '... 2^k estimates can be obtained from 2^k runs and when k is large the wealth of such estimates becomes almost an embarrassment.' This statement actually reflects a dilemma many investigators face and it contributes to the disparagement of complete factorial experiments: the need to examine all combination of a number of design factors when there are not enough resources to do so.

The concern just raised applies to any experiment wherein its execution requires nontrivial resources consisting of personnel, equipment, time, capital, logistics, or the exclusive dedication of natural resources or intellectual capacity. The reduction of the size of a complete factorial experiment and the subsequent development of fractional factorial experiments have historically been driven by experiments in agriculture or manufacturing, but increasingly the same pressures are being felt in modern research areas as diverse as computer-aided design, software development, medical imaging and genetic coding experiments, environmental impact studies involving new facilities or technologies, and climate change research on the effects of atmospheric chemical composition. In addition, global competitive pressures that demand rapid answers or solutions drive the need for careful consideration of reducing the magnitude of proposed experiments while simultaneously seeking to ensure the greatest generality of conclusions drawn from the results of the experiments.

VALUE OF FRACTIONAL FACTORIALS

Fractional factorial experiments use known properties of the design to selectively reduce the size of an experiment while at the same time limiting the tradeoff of critical information that might be lost by not conducting a comprehensive investigation of all possible combinations of the levels of the factors of interest. Rather than hope that ad hoc methods of reducing the size of an experiment might lead

*Correspondence to: rgunst@smu.edu; robert.mason@swri.org

[1]Department of Statistical Science, Southern Methodist University, Dallas, TX 75275–0332, USA

[2]Southwest Research Institute, San Antonio, TX 78228–0510, USA

DOI: 10.1002/wics.027

to optimal results, fractional factorial experiments allow informed decisions to be made about the consequences of reducing the size of an experiment. It is this informed, knowledge-based decision on the final construction of the experiment that forms the scientific basis for recommending fractional factorial designs over alternative methods for reducing the size of an experiment.

Economic Use of Resources

Key pressures that drive the need for efficient fractional factorial experiments are time and limited resources, which preclude the use of complete factorials. Much of the seminal work on fractional factorials was a result of the length of time needed for agricultural experiments, which is an entire growing season. Coupled with differences in soil fertility and composition, homogeneous plots of ground on which to experiment were limited.

In recent decades, similar pressures were experienced by manufacturing industries. For example, the automobile and petroleum industries face a dilemma when attempting to improve engine design or fuel blends in order to increase fuel economy or reduce regulated emissions. Requisite testing using complete factorials to ensure that engine design or fuel blending changes are appropriate for all vehicle manufacturers or petroleum refineries can require excessive commitments of equipment, personnel, test facilities, and time.

Modern scientific research using emerging technologies can also result in prohibitively expensive resource commitments for complete factorial experiments. The innovative use of magnetic imaging for the detection of differences in brain functioning is increasingly the focal point of medical research because such experiments are noninvasive; e.g., no radioactive chemicals need to be injected into patients. Magnetic imaging equipment and time devoted to individual patients in magnets are both limited. The equipment and highly trained scientists needed to conduct experiments are sometimes scarce and always expensive. Moreover, patient time in the magnets must be limited because of the need to remain as motionless as possible for the duration of the brain scans and the difficulty some patients have to the confinement.

Apart from the necessary need to conserve resources in many experiments, even when ample resources are available, complete factorial experiments might waste those resources. In general, most of the resources needed to conduct large complete factorial experiments are used in the estimation of high-order *interaction effects*. Interactions are joint effects of the factors that cannot be explained by the individual effects of each of the factors alone, the *main effects* of the factors. For example, if one experiments with 8 factors, each having 2 values or *levels*, a complete factorial design would consist of $2^8 = 256$ combinations of the factor levels. Consequently, if there are no repeat test runs, the experiment size would be $n = 256$. The 255 degrees of freedom (d.f.) for estimating the factor effects would be apportioned into 8 d.f. for factor main effects, 28 d.f. for two-factor interaction effects, 56 d.f. for three-factor interaction effects, and the remaining 163 d.f. for the higher order interaction effects—effects that are likely to be nonexistent or negligible. Approximately 64% (163/255) of the experimental resources would be used to estimate these high-order interaction effects.

A more subtle implication of experiments like the previous one that have so many test runs is that the average for each level of each factor is calculated over 128 of the 256 test runs. In many industrial experiments where the precision is high (variability is low), averages of a large number of test runs can result in differences between factor-level averages being declared statistically significant when the differences are so small that they are regarded as inconsequential to the response being studied; i.e., the differences are statistically significant but of no meaningful importance to the study.

Identifying Critical Factors

In many experimental settings, it is not desirable or feasible to assess all factors and their joint effects; rather, it is only the dominant factors that need to be or can be identified. For example, a chemical production process might have to temporarily cease operation in order to experiment with several factors that are intended to increase the yield of the process. Because of the major loss of income and the cost of experimentation, interest might center on only the factors that can individually demonstrate substantial increases in yield. A complete factorial experiment is not needed to achieve this goal.

Similarly, it is often wise not to plan a comprehensive experiment that involves a large number of factors of interest. Such an experiment presupposes that most or all of the factors are important contributors to changes in the response variable and that they contribute jointly; i.e., that individual factor effects and their interactions are statistically significant and meaningful. If this type of knowledge is not relatively certain, it is often preferable to conduct a much smaller experiment in which the goal is simply to identify the dominant individual factors. Once the dominant factors are identified, informed decisions can be made about whether to proceed with further experiments involving only the dominant factors or to terminate experimentation because sufficient knowledge had been gained about them. Each of these settings benefits from carefully designed fractional factorial experiments.

Screening designs (e.g., Refs 3,4) which are small fractions of complete factorials are ideal for accomplishing the goal of identifying critical factors. There is an important trade-off with screening designs: the individual factor effects are assumed to be dominant and not due to biases attributable to unevaluated joint factor effects. In spite of this very serious concern, screening designs are in widespread use and in some situations, like the chemical processing plant example, there might not be another feasible option.

The ideal setting in which critical factors can be identified in a fractional factorial without concern over possible biases from unevaluated joint factor effects is when it is known that some or all of the interaction factor effects do not exist or are negligible. In a fuel blending study, it might be known from previous experiments that critical engine factor effects do not change when fuel properties change. If so, interaction effects between the engine factors and the fuel factors do not exist. Experiment sizes can be reduced without concern over biases if the proper fractional factorial designs are used. In many settings, selective fractional factorial designs are used in which biases of individual factor effects or joint effects involving two or three factors only involve unevaluated higher order interaction effects involving four or more factors. It is common to assume that these high-order effects do not exist or are negligible relative to the individual factor effects and the low-order interaction effects.

One must be cautious about arbitrarily assuming the nonexistence of joint factor effects. If knowledge about the existence of individual or joint factor effects is unavailable, one cannot simply design a fractional factorial experiment as if the effects were known to be nonexistent. As Box and Hunter[2] (p. 318) caution, one should only '*tentatively entertain* the possibility of negligible interactions and try to check assumptions as the evidence unfolds.'

Managing Knowledge Gains

Many experiments are not isolated but are part of a process of discovery wherein a series of experiments is conducted. Unlike screening experiments, one might wish to ultimately be able to evaluate all the individual factor effects and most or all of the joint factor effects; however, by evaluating partial results as experimentation proceeds, one might be able to introduce new factors or discard ones that show no evidence of an effect. In addition, as individual factor effects are evaluated at a preliminary stage, decisions about the desirability to evaluate joint factor effects can be made. From a purely pragmatic perspective, a lengthy experiment should allow periodic reports of progress to be available for supervisory oversight or approval.

All of these requirements mandate that there be some ability to perform a large sequence of experimental runs in much smaller groups or 'blocks' of runs. The advantage of doing so using selected fractions of the entire sequence of runs is that meaningful evaluations of factor effects can be conducted and reported as each block is completed. Consequently, informed decisions about continuing, terminating, or changing the experiment can be made as each block of test runs is analyzed.

BASIC PRINCIPLES IN DESIGNING FRACTIONAL FACTORIALS

While this article cannot comprehensively cover all the varieties of and methods for constructing statistically designed fractional factorial experiments, the key principles can be discussed and illustrated. That is the purpose of this section.

Aliases

The first essential principle that needs to be understood is that any fraction of a complete factorial experiment introduces aliasing among the factor effects. Two or more factor effect are said to be *aliases* of one another if an estimate of one effect simultaneously includes the effects of one or more of the other factors. This is similar to stating that the effects are correlated, but in this setting the correlation is due to the design and not due to inherent correlations among the values of the response variable.

Table 1 shows the coded levels of three factors, each of which is to be experimented with at two levels: −1 represents one level and +1 represents the other level. In the last column of the table random variables denoting the response values are listed. The effect of any factor main effect (A, B, C) or interaction (AB, AC, BC, ABC) is the difference of two averages, the average of the responses corresponding to +1 levels and the average of the responses corresponding to −1 levels. For example, the main effect for Factor C is the

TABLE 1 | Complete Factorial Design in Three Factors, Two Levels Each

Run	A	B	C	AB	AC	BC	ABC	Response
1	−1	−1	−1	+1	+1	+1	−1	y_1
2	−1	−1	+1	+1	−1	−1	+1	y_2
3	−1	+1	−1	−1	+1	−1	+1	y_3
4	−1	+1	+1	−1	−1	+1	−1	y_4
5	+1	−1	−1	−1	−1	+1	+1	y_5
6	+1	−1	+1	−1	+1	−1	−1	y_6
7	+1	+1	−1	+1	−1	−1	−1	y_7
8	+1	+1	+1	+1	+1	+1	+1	y_8

difference in the averages for the even-numbered and the odd-numbered test runs. Under the assumptions that the main effects and interaction effects are each constant (each changes only the mean response) and the response errors are independent with constant variance, one can show that all the estimated effects from the design in Table 1 are pairwise statistically independent of one another. This is because the vectors of −1 and +1 in the columns of Table 1 are *contrasts* (the −1 and +1 elements in each column sum to 0) that are *mutually orthogonal* (their vector inner products are 0).

Compare the implications of Table 1 with those from the half-fraction shown in Table 2. This table eliminates every even-numbered test run from Table 1. Once again, factor effects are differences of average responses. Now, however, the main effect for Factor A is exactly equal to and opposite in sign from the interaction AC; i.e., A is aliased with −AC. Every effect in Table 2 is aliased with another effect, including the main effect for C which is aliased with the estimate of the overall mean response. Consequently, every difference in averages actually measures the difference of two effects; e.g., $A - AC$. Had the half-fraction with the odd-numbered test runs been deleted, every difference of averages for an effect would be estimating the sum of two effects; e.g., $A + AC$.

Every fractional factorial design requires the aliasing of some or all of the factor effects. It is the unscientific selection of fractional factorial designs

TABLE 2 | Half-fraction of a Complete Factorial Design with Three Factors

Run	A	B	C	AB	AC	BC	ABC	Response
1	−1	−1	−1	+1	+1	+1	−1	y_1
3	−1	+1	−1	−1	+1	−1	+1	y_3
5	+1	−1	−1	−1	−1	+1	+1	y_5
7	+1	+1	−1	+1	−1	−1	−1	y_7

TABLE 3 | Half-fraction of a Complete Factorial Design, Main Effects Can be Estimated

Run	A	B	C	AB	AC	BC	ABC	Response
2	−1	−1	+1	+1	−1	−1	+1	y_2
3	−1	+1	−1	−1	+1	−1	+1	y_3
5	+1	−1	−1	−1	−1	+1	+1	y_5
8	+1	+1	+1	+1	+1	+1	+1	y_8

that can lead to misleading, even incorrect, conclusions about factor effects. In contrast, it is precisely the careful selection of which fraction is used that can lead to efficient experimentation without the aliasing of important effects. Table 3 contains another half-fraction of the complete factorial. Notice that the vectors in the columns for the main effects are orthogonal to one another. For this design, each main effect can be estimated independently of one another. Each main effect is aliased with exactly one two-factor interaction, and the three-factor interaction is aliased with the estimate of the mean response. If it is known that the interaction effects are small or nonexistent, this experiment can permit the estimation of the three main effects using only half the test runs of the complete factorial. This is an example of the basic principle behind the statistical design of fractional factorial experiments: choose fractions of a complete factorial that will permit estimation of main effects and important interaction effects if higher order interaction effects are known or suspected to be negligible.

The key to selecting appropriate fractional factorial designs is to identify *design generators*, which are ordinarily high-order interactions, that divide the test runs for a complete factorial into fractions that have desirable properties. A design generator for factors that have two levels is of the form I = ABC. . . where ABC. . . is an interaction. If a half fraction of a complete factorial is desired, the best design generator is the highest-order interaction; e.g., for five factors I = ±ABCDE, where + or − is chosen randomly. Once a design generator is selected, all aliases among the factor effects can be determined by multiplication on the left and right of the design generator equation by the factor effects. When doing so, use the convention that for any factor A, AI = A and A^2 = I. Using this convention, the complete aliasing pattern for the half fraction with I = ABCDE is as follows:

A = BCDE	AB = CDE	BD = ACE
B = ACDE	AC = BDE	BE = ACD
C = ABDE	AD = BCE	CD = ABE
D = ABCE	AE = BCD	CE = ABD
E = ABCD	BC = ADE	DE = ABC

This aliasing pattern indicates that main effects are aliased with four-factor interactions, and two-factor interactions are aliased with three-factor interactions. Since the five-factor interaction is part of the design generator, it is aliased with the estimate of the response mean just like ABC was aliased with the estimate of the response mean in Table 3.

The importance of knowing the aliasing pattern is well illustrated by this example. If three-factor or higher order interaction effects are nonexistent or negligible relative to main effects and two-factor interaction effects, this half-fraction design is capable of estimating all the important effects using only half the number of test runs of a complete factorial.

Often the aliases are reported as A + BCDE, B + ACDE, etc. to emphasize the combined effects of the aliased components. The sums would be replaced by differences if the design generator was I = −ABCDE. The design generators for the half fractions in Tables 2 and 3 are, respectively, I = −C and I = ABC.

Higher order fractions for factorials in which all the factors have two levels are most efficiently constructed using $(\frac{1}{2})^p$ fractions, where p is an integer. For these fractions, p design generators must be specified and simultaneously satisfied. The choice of design generators must be carefully considered because with higher order fractions each effect is aliased with $2^p - p - 1$ other effects. For example, a quarter-fraction of a five-factor experiment in which all the factors have two levels could use the two design generators I = ±ABD and I = ±ACE with the sign in each randomly selected. The complete aliasing pattern is derived from these two generators and their symbolic product, the *implicit design generator*, (±ABD)(±ACE) = ±BCDE. The complete set of aliases can be found from the symbolic product $(I \pm ABD)(I \pm ACE) = I \pm ABD \pm ACE \pm BCDE)$. Thus, aliased effects are A ± BD ± CE ± ABCDE, etc.

Comprehensive discussions of aliasing, selection of design generators, and the related concept of the *resolution* of a fractional factorial design can be found in Mason et al.,[3] Montgomery[4] Box et al.,[5] and Cochran and Cox.[6]

Listing the Test Runs

After design generators are selected for a fractional factorial experiment, listing the factor-level combinations is readily accomplished. Using the quarter-fraction discussed in the last section, suppose the design generators are I = ABD and I = −ACE. From these two design generators, the levels of the last two factors can be obtained from the first three: D = AB, E = −AC. Also observe that a complete factorial in the five factors requires $2^5 = 32$ combinations of the factor levels, so that a quarter-fraction only requires eight test runs. These fractions are often succinctly written as 2^{k-p} fractions, where k is the number of two-level factors. Hence $2^{5-2} = 8$ test runs, exactly as many as are included in a complete factorial for the first three factors. This is the key to the listing of the test runs: write a complete factorial in the first $k - p$ factors and calculate the levels of the remaining factors using the design generators.

TABLE 4 | Test Runs for a Quarter-fraction of a Complete Factorial Design with Five Factors

Run	A	B	C	D = AB	E = −AC
1	−1	−1	−1	+1	−1
2	−1	−1	+1	+1	+1
3	−1	+1	−1	−1	−1
4	−1	+1	+1	−1	+1
5	+1	−1	−1	−1	+1
6	+1	−1	+1	−1	−1
7	+1	+1	−1	+1	+1
8	+1	+1	+1	+1	−1

The first three columns in Table 4 contain coded factor levels for a complete factorial in the factors A, B, C. The last two columns contain the calculated factor levels for the last two factors, where each level is the product of the factor levels for the interaction with which it is aliased. If an estimate of the experimental error variance is needed, some of the rows would be randomly selected and repeated. After this total number of test runs is listed, the coded factor levels would be replaced by the actual factor levels and the run order would be randomized. Note that in some experiments there is no run order; each 'Run' is actually an assignment of the factor-level combination to an *experimental unit* such as a plot of ground in an agricultural experiment.

Further details on the individual test runs for fractional factorial designs and comprehensive examples can be found in the books referenced in the bibliography. A number of statistical software programs are available for the construction of fractional factorial designs for two-level factors, including SAS/QC®[7] software and Minitab®[8] software.

Factors with Three or More Levels

Statistical designs involving only two-level factors were discussed in detail in previous sections because they are so important when designs contain more than a small number of factors. There are many reasons for this. First, the minimum size of a complete factorial design is the product of the numbers of levels of the individual factors. Even with only two levels for each of k factors, the minimum size of a complete factorial is 2^k. Second, many experiments are used for screening purposes in order to identify a small number of factors for more comprehensive study. In these circumstances, two levels for each factor are initially tested; for quantitative factors, usually these levels are near the extremes of the feasible values.

From efficiency and estimation perspectives, a very critical third reason that factors at two levels are so important is that designs in two-level factors can be fractionated with fewer test runs than factors that have more levels. This is because factors that have two levels have a single contrast in the responses that is used to calculate an effect. Factors that have three or more levels need more than one contrast to estimate each factor effect. For example, the observed effects (i.e., changes) on the response for a factor with three levels are all functions of $\bar{y}_1 - \bar{y}_2$ and $\bar{y}_1 - \bar{y}_3$; e.g., $\bar{y}_2 - \bar{y}_3 = (\bar{y}_1 - \bar{y}_3) - (\bar{y}_1 - \bar{y}_2)$. Thus, when constructing fractional factorials for designs involving three-level factors, one or more of the contrasts for each factor effect may be aliased with one or more of the contrasts for several other factor effects; i.e., each effect might be fully or partially aliased with other effects. This can make the assessment of the effects of each factor quite complicated.

One of the best $\frac{1}{3}$ fractions of a complete factorial design with three three-level factors has the following aliasing pattern[4] (p. 362):

$$I + ABC_1, A + BC_1 + ABC_2, B + AC_1 + ABC_3,$$
$$C + AB_1 + ABC_4, AB_2 + AC_2 + BC_2$$

where each main effect (A, B, etc.) has two d.f. and each subscripted interaction (AB_1, ABC_1, etc.) represents two of the d.f. for that interaction. Unless all the interaction effects are nonexistent or negligible, this design would alias or partially alias all the effects of interest. Small fractions involving large numbers of three-level factors can enable main effects and low-order interactions to be aliased with only high-order interactions but the numbers of test runs for such designs would remain high. It is often preferable to conduct screening experiments with fractions of two-level factorial designs and then more comprehensive experiments involving only a few dominant factors that can have more than two levels.

On occasion, fractional factorials that involve two-level factors and a few factors whose numbers of levels are a power of 2 can be effective when a small number of factors must be tested at more than two levels. Consult the references for details on such designs.

SCREENING DESIGNS

Very popular designs for screening a large number of factors are the *saturated main effects designs*. These highly fractionated designs consist of only $n = k + 1$ test runs and are used to investigate only the main effects of k factors.[3,4,9] A consequence is that if repeat tests are not included in these designs, the experimental variance cannot be estimated. Examples of saturated main effects designs include a 2^{3-1} experiment for investigating the main effects of three factors using only four test runs, a 2^{7-4} experiment for seven factors using only eight test runs, and a 2^{15-11} experiment for 15 factors using only 16 test runs. Table 3 is an example of a saturated main effects design. When the number of test runs is less than the main effects d.f., the design is labeled as *supersaturated*.[9] Supersaturated designs can be heavily biased and are not recommended unless an experiment consists of a large number of factors, many of which are suspected to have negligible effects on the response.

An example of a saturated main effects design is shown in Table 5. The generators for this 2^{7-4} design are I = ABD, I = ACE, I = BCF, and I = ABCG (Ref 3, Table 7A.1, p.263). Listing a complete factorial in the first three factors, the test runs in Table 5 are generated from them using the interactions indicated in the column headings.

The aliasing pattern can be derived from the design generators and their symbolic products by multiplying (I + ABD)(I + ACE)(I + BCF)(I + ABCG) = I + ABD + ACE + BCF + ABCG + BCDE + ACDF + CDG + ABEF + BEG + AFG + DEF + ADEG + BDFG + ABCDEFG. If only the aliasing pattern among the main effects and two-factor interactions are of interest, the aliases are as follows:

TABLE 5 | Screening Experiment for Seven Factors, Main Effects Can be Estimated

Run	A	B	C	D = AB	E = AC	F = BC	G = ABC
1	−1	−1	−1	+1	+1	+1	−1
2	−1	−1	+1	+1	−1	−1	+1
3	−1	+1	−1	−1	+1	−1	+1
4	−1	+1	+1	−1	−1	+1	−1
5	+1	−1	−1	−1	−1	+1	+1
6	+1	−1	+1	−1	+1	−1	−1
7	+1	+1	−1	+1	−1	−1	−1
8	+1	+1	+1	+1	+1	+1	+1

$$\begin{gathered} A + BD + CE + FG \\ B + AD + CF + EG \\ C + AE + BF + DG \\ D + AB + CG + EF \\ E + AC + BG + DF \\ F + BC + AG + DE \\ G + CD + BE + AF \end{gathered}$$

This aliasing pattern clearly indicates that main effects can only be estimated if all two-factor interactions are nonexistent or negligible; hence, this is a saturated (seven factors, eight test runs) main effects design.

A *Plackett–Burman design*[10] is a special type of a screening design. Highly popular and widely used for decades, these designs have numbers of test runs that are a multiple of 4. When the number of test runs is a power of 2, the Plackett–Burman designs are equivalent to 2^{k-p} saturated main effects designs. For example, with an experiment size of $n = 4$, the Plackett–Burman design would be equivalent to the 2^{3-1} fractional factorial experiment for three factors with four test runs, which is shown in Table 3. When the sample size is between powers of 2, such as $n = 12$, 20, or 24, the Plackett–Burman design provides a more efficient design than the usual saturated main effects design since it requires a smaller experiment size. An excellent description of Plackett–Burman designs is contained in Ref 9, including references to an extensive list of different types of such designs.

Saturated designs, like many types of fractional factorial designs, are *orthogonal arrays*. This means that in the list of test runs the levels of each factor occur an equal number of times with each of the levels of the other factors. Orthogonal arrays are desirable because they ensure that the estimated factor effects are statistically independent. Some collections of orthogonal arrays that are used as saturated main effects designs are called *L arrays*, so named because many of them are based on Latin square or Greco–Latin square designs.[3] As with other saturated designs, they have very few test runs and thus they are highly efficient. Nevertheless, they have been criticized because the designs are not always ones that maximize the resolution of the designs (e.g., Ref 9).

FOLDOVER DESIGNS

Foldover designs are important supplements to highly fractionated designs. They are very effective in resolving aliasing issues when an analysis of a fractional factorial experiment indicates that some aliased effects are statistically significant. For example, suppose the main effects for factors B, E, and G are judged to be statistically significant in an experiment in which the design shown in Table 5 is used. Because the aliasing pattern is known—one of the benefits of using the principles and procedures discussed in this article—one might become concerned over the results. Observe that even if factors B, E, and G are the only ones that affect the response, they might not do so only as main effects. The aliasing pattern indicates that each of these three main effects are aliased with two-factor interactions involving the other two factors: B + EG, E + BG, and G + BE. This is precisely the setting in which foldover designs can be effective.

Rather than conducting a new experiment using a different design in which main effects are not aliased with two factor interactions, conduct a second saturated main effects experiment in which the signs on all the factors are reversed; i.e., for each test run switch the levels on each of the factors. Table 6 shows this foldover design for the screening experiment in Table 5.

Since the signs on the factors are reversed in Table 6 compared to those in Table 5, the aliasing pattern has the signs on all effects main effects and interactions that have an odd number of letters reversed also. For the aliasing pattern between the main effects and two-factor interactions, the aliasing pattern for the foldover design becomes as follows:

$$\begin{gathered} -A + BD + CE + FG \\ -B + AD + CF + EG \\ -C + AE + BF + DG \\ -D + AB + CG + EF \\ -E + AC + BG + DF \\ -F + BC + AG + DE \\ -G + CD + BE + AF \end{gathered}$$

Equivalently, all the signs on the above aliases could be reversed; e.g., A−BD−CE−FG, etc.

The importance of the sign reversals is that for the combined 16-run design each of the columns for main effects is now orthogonal to each of the two-factor interaction columns with which it was aliased;

TABLE 6 | Foldover Experiment for the Seven Factors in Table 5

Run	A	B	C	D = AB	E = AC	F = BC	G = ABC
1	+1	+1	+1	−1	−1	−1	+1
2	+1	+1	−1	−1	+1	+1	−1
3	+1	−1	+1	+1	−1	+1	−1
4	+1	−1	−1	+1	+1	−1	+1
5	−1	+1	+1	+1	+1	−1	−1
6	−1	+1	−1	+1	−1	+1	+1
7	−1	−1	+1	−1	+1	+1	+1
8	−1	−1	−1	−1	−1	−1	−1

i.e., main effects are no longer aliased with any of the two-factor interactions. The two-factor interactions remain aliased with one another as indicated in the above aliasing patterns. It is now possible to analyze each of the main effects and the aliased two-factor interactions to determine whether it is indeed the three main effects that individually affect the response or perhaps it is two of the three factors and their interaction.

Designs of this type can be formed from any fraction of a factorial experiment where there is a need to separate the effects of the individual factors and the interactions among the factors. In each situation, certain signs of factors or factor combinations are reversed depending on the objective of the experiment. For example, if the levels for only one of the factors are reversed in the foldover design, the combined experiment is sometimes referred to as a *single-factor foldover design*. In this design, the main effect of one factor and all the two-factor interactions involving that factor are un-aliased with the other main effects and two-factor interactions. The other main effects and two-factor interactions remain aliased with one another.

Other alternative foldover designs exist that have equally valuable properties for breaking aliases (e.g., Refs 11–14). Some Plackett–Burman screening designs include effects that have a more complicated aliasing structure than a fractional factorial design constructed as in the Basic Principles section. Information about *Plackett–Burman foldover designs* can be found in Miller and Sitter.[15]

Finally, Montgomery[4] (p. 314) discusses an interesting example of an experiment in which a foldover design was used to determine whether main effects or aliased interactions were causing significant results like those posed in the 2^{7-4} fractional factorial design discussed above. He also points out that if the significant effects in the illustration above are A (not G), B, and E, a foldover design augmenting the original design might not be needed since each of the main effects A, B, and E is not aliased with the two-factor interactions involving the other two of these three factors.

BLOCKING THE TEST RUNS

Blocks of test runs are very effective in experiments in which test runs in a block are more uniform or homogeneous than test runs that occur in different blocks. Blocks have long been prominent in agricultural experiments in which plots of ground in a field (block) were more homogeneous with respect to soil nutrients, moisture content, etc. than plots of ground in different fields. In this setting, blocks as physical units (groups of plots) and not the levels of factors that are of primary interest in the experiment. Over the last half century or so, designs with blocks were adapted to industrial experiments in which blocks contained other types of physical units that increase variability because they are not all uniform. Block designs are now also used in situations where there are no physical units. Often, uncontrollable test environments can contribute to the variation of the responses: e.g., when test runs must be made over several days and equipment set-up or uncontrollable ambient conditions like humidity are likely to cause differences in responses. Additional examples of blocking include grouping the test runs into different time periods of testing, different laboratories, different groups of subjects, or different batches of raw materials.

When the block size is not large enough to accommodate all factor combinations in every block, an *incomplete block design* in which subsets of the factor-level combinations are included in different blocks is often a solution. In this type of design, planned aliasing of selected factor effects with the block effects can be used to reduce the required size of the block while still allowing for the factor effects of interest to be evaluated. The fraction of the experiment to include in each block will depend on the size of the block. For example, to design a complete factorial for a 2^k experiment with incomplete blocks, generally the most efficient block sizes are a power of two. In a design having four factors, each consisting of two levels, if the block size is $2^{4-1} = 8$, a half-fraction consisting of eight test runs could be included in one block and the test runs for the other half-fraction of eight runs could be included in the other block. To determine which test runs to include in each block, the block effects must be aliased with a factor effect, preferably the highest-order interaction, ABCD. As with half fractions, the design generator is $I = \pm ABCD$. Since this is a complete factorial experiment, however, one block would contain the test runs corresponding to $+ABCD$, and the other block would contain the test runs corresponding to $-ABCD$. Unlike a half fraction, all factor effects can be estimated in this design except for ABCD and the block effect, which are aliased with one another.

Fractional factorial experiments can also be conducted in incomplete blocks. This often happens when the number of factors is large, the number of required test runs remains large even after selecting a suitable fractional factorial design, and the number of test runs that can be conducted in homogeneous blocks is not sufficient to include all the test runs in

TABLE 7 | Complete Factorial Experiment in an Incomplete Block Design

Block 1	Block 2	Block 3	Block 4
$I = +ABD$	$I = +ABD$	$I = -ABD$	$I = -ABD$
$I = +ACE$	$I = -ACE$	$I = +ACE$	$I = -ACE$
$(I = +BDCE)$	$(I = -BDCE)$	$(I = -BDCE)$	$(I = +BDCE)$

one block. When a fractional factorial is conducted in incomplete blocks, there are two types of aliases: the aliases of factor effects with one another from the fractional factorial and additional aliases of the block effects with the factor effects chosen to assign test runs to blocks. Blocking generators for 8-run and 16-run fractional factorial designs are given in Bisgaard[16] and selected designs for the case where the range of the number of factors is 5–11 are given in Mason et al.,[3] and Box et al.[5] Designs for the special case where the blocks are of size two, such as in paired comparisons or in foldover designs of pairs of runs, are discussed in Draper and Guttman.[17]

A very important application of blocking fractional factorials occurs when experiments must be conducted in several phases. This is most frequently the case when experiments are expensive and a high priority is placed on periodic evaluations of the test results in order to determine whether to continue or terminate the experiment. Complete or fractional factorials can then be partitioned into blocks of runs so that increasingly informative analyses of the factor effects can be made after each block of runs is completed.

To illustrate how fractional factorials can be conducted in incomplete block designs, suppose it is desired to conduct a factorial experiment in five factors, each of which has two levels. The minimum number of test runs for a complete factorial experiment is $2^5 = 32$. Suppose further it is required that periodic evaluations of the test results be available due to the length of time required for each test run. This could occur if the effects of five fuel properties must be gauged by driving each of several vehicles several thousand miles and a single test run results in emissions, toxics, and fuel economy measurements for one vehicle. One incomplete block design would consist of blocks having eight test runs using the design generators I = ±ABD = ±ACE (= ±BDCE, the implicit design generator). This generator was used to design the quarter-fraction that was discussed in Test Runs section. The incomplete blocks are shown in Table 7.

The experiment would then consist of the following phases.

- Phase 1: Randomly select a block of test runs; e.g., Block 3
  - Main effects unaliased with one another
  - Main effects aliased with interactions
- Phase 2: Select another block with the same sign as BCDE, Block 2
  - Main effects unaliased with one another and two-factor interactions
  - Factor effects aliased: I = $-BCDE$
  - Block effect aliased with ABD and ACE
- Phase 3: Randomly select another block of test runs; e.g., Block 4
  - Main effects unaliased with one another and two-factor interactions
  - Some interactions partially aliased
  - Block effects aliased with ABD, ACE, and BCDE
- Phase 4: Run the last block of test runs, Block 1
  - Only block effects aliased: Blocks with ABD, ACE, and BCDE

Observe that after the first phase of the experiment, information on the main effects is available, although the main effects are aliased with two-factor interactions. After phase 2, main effects are unaliased with one another and with two-factor interactions. Some two-factor interactions are aliased with one another. This would be an important juncture in the experiment when decisions could be made about continuing, terminating, or modifying the experiment because one could at least evaluate main effects without concern that two-factor interaction aliases would bias the results. Finally, note that when the experiment is completed, only three interactions are aliased and these are aliased with the block effects.

A drawback to this incomplete block design is that if there are no block effects, it is not necessarily the best design. For example, after two phases of the experiment are run, a better design would be one in which ABCDE was one of the design generators since all main effects and two-factor interactions are unaliased with one another if there are no block effects. In addition, if one is specifically interested in fractions other than $(\frac{1}{2})^p$, more efficient designs may exist than blocks constructed from $(\frac{1}{2})^p$ fractions. For example, there are classes of special designs for $\frac{3}{4}$ fractions of complete factorials. These special designs allow more effects to be unaliased than the usual construction of fractional factorials (see Refs 18,19).

Throughout this discussion of block designs, it has been assumed that there are no interactions between blocks and factor effects. This is because blocks are ordinarily assumed to be random contributors to the variability of the response. If block effects are believed to interact with factor effects, they should be treated as another factor for the purpose of designing the experiment. Also, when constructing block designs, separate randomizations of the test runs to experimental units or to test run sequences, whichever is appropriate, within each block should be part of the design.

ADDITIONAL DESIGN CONSIDERATIONS

Most of the designs discussed in this article can be and should be, if possible, augmented with repeat test runs; i.e., factor-level combinations that are run more than once. Not all the combinations need to be repeated but, within the constraints of budget and time, some should. Repeat tests provide the best estimates of experimental error variation and are unaffected by any aliasing. The partial aliasing of factor effects for designs in which each factor-level combination is not repeated an equal number of times is ordinarily so small that it can be ignored.

Randomization is important in all statistical designs so that some protection can be obtained against unknown sources of bias. When the test runs in an experiment cannot be run in a completely random order, such as when the levels of factors are hard to change, the restrictions imposed can result in a design actually being a complete or fractional factorial in a special incomplete block design that is similar to a *split-plot design*; (see, for example, Ref 3).

REFERENCES

1. Fisher RA. *The Design of Experiments*. 8th ed. New York: Hafner; 1966.
2. Box GEP, Hunter JS. The 2^{k-p} fractional factorial designs, part I. *Technometrics* 1961, 3(3):311–351. doi:10.2307/1266725.
3. Mason RL, Gunst RF, Hess JL. *Statistical Design and Analysis of Experiments*. 2nd ed. New York: Wiley; 2003. doi:10.1002/0471458503.
4. Montgomery DC. *Design and Analysis of Experiments*. 6th ed. New York: Wiley; 2005.
5. Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters*. 2nd ed. New York: Wiley; 2005.
6. Cochran WG, Cox GM. *Experimental Design*. New York: Wiley; 1957.
7. *SAS/QC® Software*. Cary, NC: SAS Institute, Inc.; 2008.
8. *Minitab® Statistical Software*. State College, PA: Minitab, Inc.; 2007.
9. Ryan TR. *Modern Experimental Design*. New York: Wiley; 2007, doi:10.1002/0470074353.
10. Plackett RL, Burman JP. The design of optimum multifactorial experiments. *Biometrika* 1946, 33(4):305–325. doi: 10.1093/biomet/33.4.305.
11. John PWM. Time trends and factorial experiments. *Technometrics* 1990, 32(3):275–282. doi:10.2307/1269104.
12. Montgomery DC, Runger GC. Foldovers of 2^{k-p} resolution IV experimental designs. *J Qual Technol* 1996, 28(4):446–450.
13. Li H, Mee RW. Better foldover fractions for two-level fractional factorial designs. *Technometrics* 2002, 44(3):278–283. doi:10.1198/004017002188618464.
14. Li W, Lin DKJ. Optimal foldover plans for two-level nonregular orthogonal designs. *Technometrics* 2003, 45(4):142–149. doi.org/10.1198/004017003188618779.
15. Miller A, Sitter RR. Using the folded-over 12-run Plackett-Burman design to consider interactions. *Technometrics* 2001, 43(1):44–55. doi:10.1198/00401700152404318.
16. Bisgaard S. Blocking generators for small 2^{k-p} designs. *J Qual Technol* 1994, 26(4):288–296.
17. Draper NR, Guttman I. Two-level factorial and fractional factorial designs in blocks of size two. *J Qual Technol* 1997, 29(1):71–75.
18. Diamond WJ. *Practical Experimental Designs for Engineers and Scientists*. Belmont, CA: Wadsworth; 1981.
19. John PWM. *Statistical Design and Analysis of Experiments*. Philadelphia, PA: SIAM; 1998.

FURTHER READING

Block Designs

Bingham DR, Sitter RR. Design issues in fractional factorial split-plot experiments. *J Qual Technol* 2001, 33(1): 2–15.
Kulahci M, Bisgaard S. The use of Plackett-Burman designs to construct split-plot designs. *Technometrics* 2005, 47(4): 495–501. doi:10.1198/004017005000000427.
Loeppky JL, Sitter RR. Analyzing unreplicated blocked or split-plot fractional factorial designs. *J Qual Technol* 2002, 34(3): 229–243.
McLeod RG, Brewster JF. The design of blocked fractional factorial split-plot experiments. *Technometrics* 2004, 46(2): 135–146. doi:10.1198/004017004000000176.
Sitter RR, Chen J, Feder M. Fractional resolution and minimum aberration in blocked 2^{n-k} designs. *Technometrics* 1997, 39(4): 382–390. doi:10.2307/1271502.

Foldover Designs

Li W, Lin DKJ, Ye KQ. Optimal foldover plans for two-level nonregular orthogonal designs. *Technometrics* 2003, 45(4): 347–351. doi:10.1198/004017003000000177.
Mee RW, Peralta M. Semifolding 2^{k-p} designs. *Technometrics* 2000, 42(2): 122–134. doi:10.2307/1271444.
Webb S. Non-orthogonal designs of even resolution. *Technometrics* 1968, 10(2): 291–299. doi:10.2307/1267044.

Three-quarter fractions

Addelman S. Irregular fractions of the 2^n factorial experiments. *Technometrics* 1961, 3(4): 479–496. doi:10.2307/1266555.
John PWM. Three-quarter replicates of 2^4 and 2^5 designs. *Biometrics* 1961, 17(2): 319–321. doi:10.2307/2527996.
John PWM. Three quarter replicates of 2^n designs. *Biometrics* 1962, 18(2): 172–184. doi:10.2307/2527455.
John PWM. Augmenting 2^{n-1} designs. *Technometrics* 1966, 8(3): 469–480. doi:10.2307/1266694.
John PWM. Some non-orthogonal fractions of 2^n designs. *J R Stat Soc [Ser B]* 1969, 31(2): 270–275.
Peter PWM. Blocking of $3(2^{n-k})$ designs. *Technometrics* 1964, 6(4): 371–376. doi:10.2307/1266092.

Alternative Designs

Bingham D, Sitter RR. Fractional factorial split-plot designs for robust parameter experiments. *Technometrics* 2003, 45(1): 80–89. doi:10.1198/004017002188618725.
Miller A, Sitter RR. Using folded-over nonorthogonal designs. *Technometrics* 2005, 47(4): 502–513. doi:10.1198/004017005000000210.
Xu H, Deng L-Y. Moment aberration projection for nonregular fractional factorial designs. *Technometrics* 2005, 47(2): 121–131. doi:10.1198/004017004000000617.