9
An Overview of Group Factor Screening

MAX D. MORRIS

The idea of using a group screening procedure to identify the important or active factors using a small designed experiment was described by Watson (1961) and is now applied in a variety of areas of science and engineering. Watson's work built on the earlier ideas of Dorfman (1943) for screening pooled samples of blood in order to identify diseased individuals using minimal resources. Generalizations and extensions of Watson's technique have been developed by a number of authors who have relaxed some of the stringent assumptions of the original work to make the methods more widely applicable to real problems. An overview of some of the proposed screening strategies is presented, including the use of several stages of experimentation, the reuse of runs from earlier stages, and screening techniques for detecting important main effects and interactions.

1 Introduction

In experimental programs involving large numbers of controllable factors, one of the first goals is the identification of the subset of factors that have substantial influence on the responses of interest. There are at least two practical reasons for this. One is the empirical observation that, in many important physical systems, much or most of the variability in response variables can eventually be traced to a relatively small number of factors—the concept of *effect sparsity* (see, for example, Box and Meyer, 1986). When this is true, it is certainly sensible to "trim down" the problem to the factors that are "effective" or "active" before detailed experimentation begins. Even when effect sparsity does not hold, however, it is obvious that careful experimentation involving many factors simply costs more than careful experimentation involving a few. Economic and operational reality may necessitate experimentation in phases, beginning with attempts to characterize the influence of the apparently most important factors while conditioning on reasonable fixed values for other possibly interesting factors. Hence, small *factor screening experiments* are performed not to provide definitive estimates of parameters but to identify the parameters that should be estimated first.

Following Dorfman's (1943) description of an analysis for screening physically pooled samples of blood, the idea of using *group screening* strategies to design factorial experiments was carefully described by Watson (1961) whose analysis is the starting point for much of the subsequent research in this area. As a matter of historical record, Watson credited W. S. Conner as having suggested the idea of group factor screening to him.

This overview begins with a description of Watson's treatment (Section 2) and then briefly discusses a number of modifications and generalizations that have been presented by others (see Kleijnen, 1987, and Du and Hwang, 2000, for additional reviews). Section 3 discusses strategies involving more than two stages and a variety of other issues, including the reuse of runs. Multiple grouping strategies and screening for interactions are discussed in Sections 4 and 5, respectively. The intent of this chapter is not to offer a complete review of all that has been done in the area, but to give the reader a sense of some things that can be done to make group factor screening more applicable in specific situations.

2 Basic Group Factor Screening

Watson began his description of the technical problem in the following way.

> Suppose that f factors are to be tested for their effect on the response. Initially we will assume that
>
> (i) all factors have, independently, the same prior probability of being effective, $p\,(q = 1 - p)$,
> (ii) effective factors have the same effect, $\Delta > 0$,
> (iii) there are no interactions present,
> (iv) the required designs exist,
> (v) the directions of possible effects are known,
> (vi) the errors of all observations are independently normal with a constant known variance, σ^2,
> (vii) $f = gk$ where $g =$ number of groups and $k =$ number of factors per group.
>
> These stringent assumptions are made only to provide a simple initial framework. . . . Actually they are not as limiting as they appear.

The two-level experimental designs that Watson goes on to describe partition the individual factors into the groups referenced in point (vii). In each experimental run, all factors in the same group are either simultaneously at their high values or simultaneously at their low values. In other words, the level of the group factor dictates the level of all individual factors within the group.

Watson's seven points might be restated in the common modern language of linear models (although admittedly with a loss of some elegance and intuitive simplicity) by saying that the collection of n observed responses may be written in matrix notation as

$$\boldsymbol{Y} = \mathbf{1}\beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ are vectors containing the n responses and the n error variables, respectively, $\mathbf{1}$ is a vector of 1s, $\boldsymbol{\beta}$ is a vector of the f main effect parameters, and $\boldsymbol{X}$ is an $n \times f$ matrix, and where

1. For each element of $\boldsymbol{\beta}$, independently,

$$\begin{aligned} \beta_i &= \Delta/2 \quad \text{with probability } p, \\ &= 0 \qquad \text{with probability } q = 1 - p, \end{aligned}$$

   for some $\Delta > 0$ (Watson's (i) and (ii)),
2. $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$ with known σ^2 (Watson's (vi)),
3. $\boldsymbol{X}$ contains f columns, and has the (i, j)th element equal to

   $+1$ if the jth individual factor is at its "high" level in the ith run
   -1 if the jth individual factor is at its "low" level in the ith run

   (Watson's (iii) and (v)), and
4. $\boldsymbol{X}$ may be written in partitioned form as

$$\boldsymbol{X} = (z_1 \cdot \mathbf{1}' | z_2 \cdot \mathbf{1}' | \ldots | z_g \cdot \mathbf{1}'), \tag{2}$$

   where each z is an $n \times 1$ vector and each $\mathbf{1}'$ is a $1 \times k$ vector, and

$$(\mathbf{1} | \boldsymbol{Z}) = (\mathbf{1} | z_1 | z_2 | \ldots | z_g)$$

   is of full column rank (Watson's (iv) and (vii)).

Section 2.1 contains an example with explicit quantities for several of the expressions listed in these items.

The distinguishing characteristics of basic group factor screening, from the perspective of linear models, are the probabilistic assumption about the value of the parameter vector $\boldsymbol{\beta}$ and the grouped-column restriction on the form of the model matrix $\boldsymbol{X}$. The intent of group screening is to learn important characteristics of the system using fewer (often, far fewer) runs than would be required in a "conventional" experiment. When data are lacking, inference requires that assumptions must be made, and Watson's points (i) and (ii) provide a practical and often reasonable basis for approximate interpretation of the system. In fact, these assumptions are "not as limiting as they appear" (see Watsons's text above); the group screening technique often works quite well in cases where at least some of them are violated.

The generation of the matrix $\boldsymbol{X}$ from $\boldsymbol{Z}$ is a statement of the group design strategy. Physical factors are initially confounded in groups of size $k > 1$ so as to construct artificially a reduced statistical model (which can be estimated based on the desired smaller sample size). By this intentional confounding, the investigator abandons the goal of estimating individual elements of $\boldsymbol{\beta}$, focusing instead on grouped parameters representing estimable functions in the original problem. Practically, the problem can be restated as being

$$\boldsymbol{Y} = \mathbf{1}\beta_0 + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_g]'$ is a $g \times 1$ vector and $\boldsymbol{Z}$ is an $n \times g$ matrix, and where γ_i is the sum of the subset of elements of $\boldsymbol{\beta}$ associated with individual factors in the ith group. These grouped parameters can, therefore, be written as

$$\gamma_i = A_i \Delta / 2,$$

where A_i is a binomial random variable with parameters k and p, and A_i are independent for $i = 1, 2, 3, \ldots, g$. In the first stage, the focus is on determining which of these grouped parameters is nonzero.

The second stage or "follow-up" experiment is designed to examine only those individual factors included in groups that appear to be effective under the assumptions of the model, those for which $\gamma_i > 0$ based on Watson's assumption (ii). The decisions concerning whether groups are effective could, for example, be based on individual z- (σ known) or t- (σ estimated) tests for each grouped parameter. If Watson's assumption (ii) is taken seriously, these would logically be one-sided tests ($H_0 \colon \gamma_i = 0$, $H_A \colon \gamma_i > 0$), but a two-sided test ($H_0 \colon \gamma_i = 0$, $H_A \colon \gamma_i \neq 0$) is more robust against the failure of this assumption and so is often preferred in practice. Factors included in groups that do not appear to be effective are fixed at constant values in the second stage experiment. Watson applied the word "effective" only to individual factors. Here I substitute the more popular term "active", and extend this use to say that a group factor is active if it includes at least one active factor.

Hence, a *two-stage* screening experiment as described by Watson requires a predetermined number n of runs in the first stage and a random number M of runs in the second. The distribution of M and the effectiveness of the screening program (that is, success in labeling individual factors as "active" or "not active") depend on characteristics that can be controlled, at least to some degree:

- value of k,
- value of n,
- the specific form of the stage 1 decision rule for each group; for example, selection of a significance level α for a z- or t-test,

and characteristics that cannot:

- value of p,
- value of Δ,
- value of σ^2.

Given values of p and Δ/σ, values of k, n, and α can be selected to produce desirable results, such as small expected sample size or small probability of misclassifying individual factors. As with other statistical design problems, the obvious goals are generally in conflict; smaller $n + E(M)$ generally corresponds to a larger expected number of misclassifications of at least one kind, and so some degree of compromise between expense and performance is required. Watson derived expected values of the number of runs required and the number of factors misclassified, as functions of the parameters. These expressions may be used to evaluate the performance of alternative sampling plans.

2.1 Example

To demonstrate the strategy, suppose $f = 50$ experimental factors are to be screened and that a decision is made to do this by forming $g = 7$ groups composed of $k = 7$ factors (as factors 1–7, 8–14, and so on), with the 50th factor added as an "extra" to the 7th group. If σ has a known value of 4, say, replicate runs will not be needed and an orthogonal 8-run, 2-level design can be used in stage 1. For example, the 8-run Plackett and Burman (1946) design has design matrix

$$\boldsymbol{Z} = \begin{pmatrix} + & + & + & - & + & - & - \\ - & + & + & + & - & + & - \\ - & - & + & + & + & - & + \\ + & - & - & + & + & + & - \\ - & + & - & - & + & + & + \\ + & - & + & - & - & + & + \\ + & + & - & + & - & - & + \\ - & - & - & - & - & - & - \end{pmatrix}.$$

The expanded $\boldsymbol{X}$ matrix for individual factors would be comprised of the columns of $\boldsymbol{Z}$, each repeated 7 times (or 8 for $\boldsymbol{z}_7$).

Suppose now that the true parameter values are as displayed in Table 1. These values do not exactly correspond to Watson's assumptions; in particular, (ii) is violated. Still, if model (3) is fitted to the data via least squares, estimates of the elements of $\boldsymbol{\gamma}$ will be independent and normally distributed, each with a standard error of $\sqrt{2}$, and with means $E(\hat{\gamma}_1) = 6$, $E(\hat{\gamma}_2) = 0$, $E(\hat{\gamma}_3) = -4$, $E(\hat{\gamma}_4) = 0$, $E(\hat{\gamma}_5) = 0$, $E(\hat{\gamma}_6) = 3$, and $E(\hat{\gamma}_7) = 1$. It is likely that groups 1 and 6 will be declared active, because z statistics based on $\hat{\gamma}_1$ and $\hat{\gamma}_6$ will probably be unusually large. Group 3 would also be detected with high probability, but only if a two-sided test is used; if Watson's working assumption (v) is taken seriously, this would not be the indicated procedure, but the uncertainty associated with many real applications would make two-sided testing an attractive modification. Group 7 might be detected as active, but the probability of this is reduced, for either a one- or two-sided test, by the partial "cancellation" of individual effects of opposite sign

TABLE 1. Individual parameters, grouped as in a first-stage screening experiment, for the example in Section 2.1

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
$\beta_1 = 0$	$\beta_8 = 0$	$\beta_{15} = 0$	$\beta_{22} = 0$	$\beta_{29} = 0$	$\beta_{36} = 3$	$\beta_{43} = 0$
$\beta_2 = 0$	$\beta_9 = 0$	$\beta_{16} = -4$	$\beta_{23} = 0$	$\beta_{30} = 0$	$\beta_{37} = 0$	$\beta_{44} = 0$
$\beta_3 = 2$	$\beta_{10} = 0$	$\beta_{17} = 0$	$\beta_{24} = 0$	$\beta_{31} = 0$	$\beta_{38} = 0$	$\beta_{45} = -2$
$\beta_4 = 0$	$\beta_{11} = 0$	$\beta_{18} = 0$	$\beta_{25} = 0$	$\beta_{32} = 0$	$\beta_{39} = 0$	$\beta_{46} = 0$
$\beta_5 = 0$	$\beta_{12} = 0$	$\beta_{19} = 0$	$\beta_{26} = 0$	$\beta_{33} = 0$	$\beta_{40} = 0$	$\beta_{47} = 0$
$\beta_6 = 0$	$\beta_{13} = 0$	$\beta_{20} = 0$	$\beta_{27} = 0$	$\beta_{34} = 0$	$\beta_{41} = 0$	$\beta_{48} = 3$
$\beta_7 = 4$	$\beta_{14} = 0$	$\beta_{21} = 0$	$\beta_{28} = 0$	$\beta_{35} = 0$	$\beta_{42} = 0$	$\beta_{49} = 0$
						$\beta_{50} = 0$
$\gamma_1 = 6$	$\gamma_2 = 0$	$\gamma_3 = -4$	$\gamma_4 = 0$	$\gamma_5 = 0$	$\gamma_6 = 3$	$\gamma_7 = 1$

in this group—a violation of Watson's assumption (ii). Mauro and Smith (1982) examined the effect of factor cancellation on screening in the more extreme case where parameters associated with active factors all have the same absolute value, but not the same sign. They presented tables and graphs summarizing a numerical study of performance. Expected degradation in performance is modest when both p and the factor group sizes are small, because the probability is minimal that active factors having main effects of opposite signs mask each other. However, performance degradation becomes a more important issue as either p or the factor group sizes increase. Both the expected number of runs in the second stage of the experiment and the expected number of correctly identified active factors are minimized when, other things being equal, the proportion of active factors with positive main effects equals the proportion with negative main effects. One interesting conclusion from the study is that the factor group size leading to the minimum number of runs is the same whether or not the main effects of the active factors have the same signs.

In the present example, if each of groups 1, 3, 6, and 7 were to be declared active, a follow-up experiment in the associated 29 individual factors might be undertaken to complete the screening process. If an orthogonal design is used, this will require $M = 32$ new runs. This ignores the possibility that some of the original 8 runs might be "reused" which is discussed later. Regardless of the specific orthogonal two-level plan used, the standard error of each parameter estimate will be $\sigma/\sqrt{\text{sample size}} = 4/\sqrt{32}$, so all active factors are likely to be discovered if two-sided testing is used. The expected number of "false positives" will depend on the significance level selected. The total number of runs used here is $n + M = 40$, compared to the 52 that would be required by a minimal orthogonal design for all 50 factors.

3 Strategies Involving More Than Two Stages

Most group screening experiments are sequential because the specific form of the second stage design depends upon the analysis of data collected at the first stage. Many other sequential plans are possible. I briefly describe a few strategies that can be viewed as generalizations of (and, in some cases, improvements over) basic group screening.

3.1 Multiple Stage Screening and Sequential Bifurcation

Perhaps the most obvious extension of Watson's basic screening strategy is to grouped factor plans involving more experimental stages. Patel (1962) described multiple stage screening as follows.

- in stage 1, group the f factors into g_1 groups of size $k_i = f/g_1$;
- for each group found apparently active at stage 1, the k_1 factors are grouped into g_2 groups of size $k_2 = f/(g_1 g_2)$;
- ...

- for each group found apparently active at stage $s-1$, the k_{s-1} factors are grouped into g_s groups of size $k_s = f/(g_1 g_2 \cdots g_{s-1} g_s)$;
- for each group found apparently active at stage s, all factors are individually examined.

Here, s refers to the number of screening stages not counting the final follow-up in which each of the remaining factors is examined individually. Hence, Watson's description is of a 2-stage procedure characterized by $s = 1$.

Patel offered an analysis based on assumptions that the experimental design used at each stage is of minimal size (and so contains one more run than the number of group factors being investigated) and that $\sigma^2 = 0$. Under these conditions, he showed that values of g_i which minimize the expected number of total runs required are approximately

$$g_1 \approx f p^{s/(s+1)}, \qquad g_i \approx p^{-1/(s+1)}, \quad i = 2, \ldots, s,$$

and that, when these numbers of equal-sized groups are used, the expected number of runs required by this multiple stage procedure is approximately

$$(s+1) f p^{s/(s+1)} + 1.$$

These expressions can, in turn, be used to determine an optimal number of stages for a given value of p, at least under the idealized assumptions of the analysis. So, for example, if p were taken to be $6/50$ in the example of Section 2.1, these expressions would yield rounded values of 17 groups and 36 runs, respectively, for the two-stage plans of Watson ($s = 1$). If one additional stage is added ($s = 2$) then the values would be 12 groups for the first stage and 2 groups for the second stage with a total of 37 runs, indicating that the addition of a stage does not improve the expected number of required runs in this case.

Multiple stage screening could also be defined without the requirement of successive splitting of each apparently active group. Instead, the factors included in active groups at one stage could be randomly assigned to smaller groups at the next stage without imposing this constraint.

Bettonvil (1995) discussed a particular form of the multiple group screening idea called *sequential bifurcation*, in which

1. all factors are included in a single group in the first stage, $k_1 = f$, and
2. the factors from an apparently active group at any stage are divided into two subgroups of equal size in the next, $k_{i+1} = k_i/2$.

Chapter 13 reviews sequential bifurcation in more detail.

3.2 Orthogonality and Reuse of Runs

The expected number of runs required by a sequential screening plan depends, sometimes heavily, on (1) whether a response value may be used only once in the analysis immediately following the experimental stage in which it is acquired or may be reused in subsequent analyses, and (2) whether the experimental designs

TABLE 2. Example of experimental runs added at each stage and used in each analysis: sequential bifurcation

		Factor								Data used in analysis following stage			
Run	Added at stage	1	2	3	4	5	6	7	8	1	2	3	4
1	1	−	−	−	−	−	−	−	−	•	•	•	•
2	1	+	+	+	+	+	+	+	+	•	•		
3	2	+	+	+	+	−	−	−	−		•	•	
4	3	+	+	−	−	−	−	−	−			•	•
5	4	+	−	−	−	−	−	−	−				•

used at each stage are required to be orthogonal. Neither Patel (1962) nor Bettonvil (1995) required orthogonality in the designs that they described. This is because each author initially motivated his design for situations in which $\sigma^2 = 0$, and so the usual statistical arguments for precision associated with orthogonality are not relevant. However, Bettonvil allowed the reuse of as many runs as possible from stage to stage, resulting in further reduction in the required number of runs. For example, Table 2 presents a sequence of experimental runs that would be made using sequential bifurcation as described by Bettonvil in an experiment in which there are 8 factors, only the first is active, and no mistakes are made in testing. After stage 1, runs 1 and 2 are used to test all 8 factors as a single group. After stage 2, runs 1, 2, and 3 are used to test group factors (1, 2, 3, 4) and (5, 6, 7, 8); hence runs 1 and 2 are reused. Similarly, following stage 3, runs 1, 3, and 4 are used to test group factors (1, 2) and (3, 4) and, following stage 4, runs 1, 4, and 5 are used to test individual factors 1 and 2. In the analysis following each of stages 2, 3, and 4, the response values of two runs from previous stages are incorporated in the analysis.

In comparison, Table 3 displays a similar description of how Patel's multiple screening would evolve in the same situation. An initial group of all the factors ($g_1 = 1, k_1 = f$) is used, followed in subsequent stages by groups that are half

TABLE 3. Example of experimental runs added at each stage and used in each analysis: multiple stage (Patel •, modified ○)

		Factor								Data used in analysis following stage			
Run	Added at stage	1	2	3	4	5	6	7	8	1	2	3	4
1	1	−	−	−	−	−	−	−	−	•	•	•	•
2	1	+	+	+	+	+	+	+	+	•	○		
3	2	+	+	+	+	−	−	−	−		•	○	
4	2	−	−	−	−	+	+	+	+		•		
5	3	+	+	−	−	−	−	−	−			•	○
6	3	−	−	+	+	−	−	−	−			•	
7	4	+	−	−	−	−	−	−	−				•
8	4	−	+	−	−	−	−	−	−				•

TABLE 4. Example of experimental runs added at each stage and used in each analysis: multiple stage with orthogonal plans at each stage and no reuse of runs

		Factor								Data used in analysis following stage			
Run	Added at stage	1	2	3	4	5	6	7	8	1	2	3	4
1	1	−	−	−	−	−	−	−	−	•			
2	1	+	+	+	+	+	+	+	+	•			
3	2	−	−	−	−	−	−	−	−		•		
4	2	+	+	+	+	+	+	+	+		•		
5	2	+	+	+	+	−	−	−	−		•		
6	2	−	−	−	−	+	+	+	+		•		
7	3	−	−	−	−	−	−	−	−			•	
8	3	+	+	+	+	−	−	−	−			•	
9	3	+	+	−	−	−	−	−	−			•	
10	3	−	−	+	+	−	−	−	−			•	
11	4	−	−	−	−	−	−	−	−				•
12	4	+	+	−	−	−	−	−	−				•
13	4	+	−	−	−	−	−	−	−				•
14	4	−	+	−	−	−	−	−	−				•

the size of their predecessors. Patel assumed that only the first run (all factors at the low level) is reused at each stage. Hence runs 1, 3, and 4 are used for testing group factors (1, 2, 3, 4) and (5, 6, 7, 8) following stage 2; runs 1, 5, and 6 are used for testing group factors (1, 2) and (3, 4) following stage 3; and runs 1, 7, and 8 are used for testing factors 1 and 2 following stage 4. All eight runs are unique, but all are not strictly necessary for the estimation of the group effects required in the screening strategy. In this example, the analysis of Patel's design can be modified somewhat to allow the reuse of two runs at each stage, as indicated by the open circles in Table 3, but this modification does not change the result of the tests if the assumptions actually hold and there are no random errors in observations.

When σ^2 is not negligible, the benefits of run reuse and nonorthogonal saturated designs are not so clear-cut. Then, the reuse of runs makes the analysis of performance more complicated because it introduces dependencies between test statistics at each stage. Some duplication of runs (rather than reuse) would allow estimation, or at least a check on the assumed value, of σ^2. Furthermore, as noted by Watson, when observations include error, many investigators would be more comfortable with the more efficient orthogonal designs. Table 4 shows a sequence of runs for a modified version of multiple stage screening using minimal orthogonal designs at each stage and allowing no reuse of runs from previous stages. This design requires more runs than either Patel's multiple group procedure or Bettonvil's sequential bifurcation, but it provides more precise estimates of group factor parameters in stages 2–4 when there is random error. Furthermore, it allows for the option of blocking to correct for stage effects, or 6 degrees of freedom (after the last stage) for estimating σ^2 if blocking is not needed. In this example,

the strategy based on orthogonal designs with maximum reuse of runs from stage to stage would be equivalent to the modified multiple stage plan in Table 3.

The decision to use orthogonal plans or to allow nonorthogonal plans, and the decision to allow or disallow the reuse of runs from stage to stage, are related operational issues. They individually and jointly affect the performance of a screening plan and the complexity of calculations required to assess analytically that performance. Depending on characteristics of the application, such as the degree of measurement error and the need to account for block effects in sequential experimentation, either or both may be important considerations.

3.3 Stepwise Screening

Odhiambo and Manene (1987) introduced a stepwise screening plan featuring sequential testing of individual factors after the first (grouped) experiment. After an initial stage as described by Watson, a new experiment is undertaken for each apparently active group in which individual factors are tested one by one until one of them is found to be active. At that point, any remaining factors (not yet individually tested) are tested together as a group, and depending on the result of that test, all are labeled as not active or subjected to further individual examination as following the initial stage. The sequential process of individual tests and group tests, following the discovery of individual active factors, continues until all factors are classified. A schematic of how this might develop in a hypothetical example is given in Figure 1.

This strategy can offer some additional efficiency if some initial groups contain only one active factor, because this factor may be discovered early in follow-up testing and the remaining factors eliminated in one further group test. This occurs in the second and third initial groups in Figure 1. Such efficiency may not necessarily occur, however. For example, identification of the active factors in group 6 (individual factors 26–30) in Figure 1 requires 6 follow-up runs.

Odhiambo and Manene presented a performance analysis of stepwise screening that assumes $\sigma^2 > 0$, where statistical tests are fallible even if all assumptions are correct. They derived expected values of the number of runs required, the number of factors mistakenly classified as active, and the number of factors mistakenly classified as not active, in terms of p, f, k, and the significance level and power of the tests used. These expressions are fairly complicated and are not repeated here, but Odhiambo and Manene also provide simpler approximations that are appropriate for small values of p.

4 Multiple Grouping Strategies

Sequential group screening methods can lead to substantial test savings; loosely speaking, the more sequential a procedure, in terms of the number of decision points, the greater is the potential for reduction in the expected number of runs required. However, there are settings in which such approaches are operationally impractical, for example, where execution of each run takes substantial time but

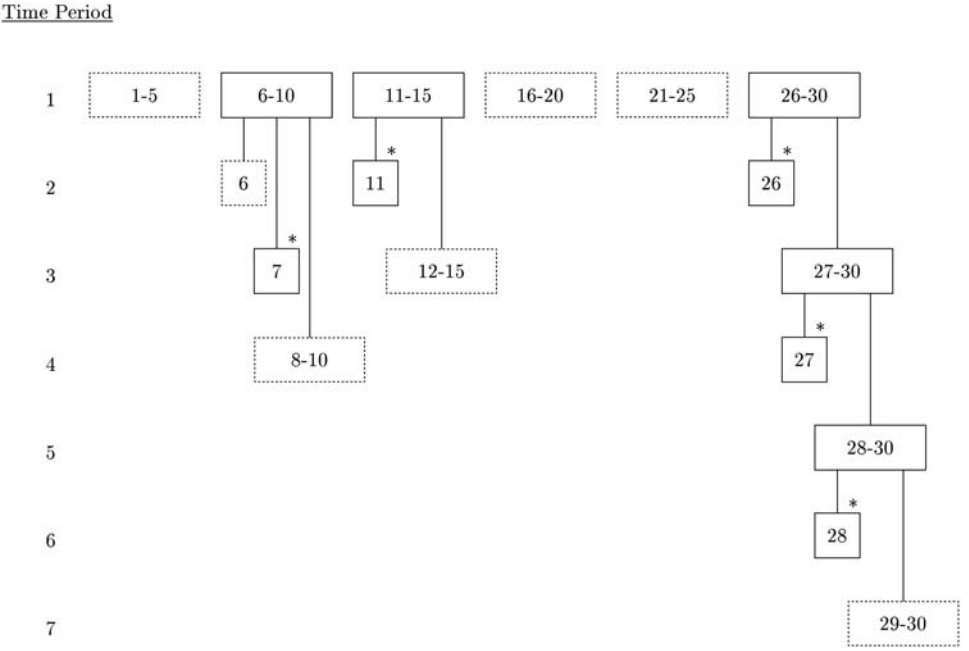

FIGURE 1. Example of analyses and decisions made in a stepwise experiment. The numbers in the boxes refer to factors and each box represents a test of the indicated group or individual factor. Dashed and solid boxes indicate tests in which factors are determined to be not active and active, respectively; asterisks indicate points at which individual active factors are discovered.

many runs can be executed simultaneously as a "batch". Nonsequential procedures based on assigning each object/factor to more than one group were discussed by Federer (1987) for the blood screening problem addressed by Dorfman, and by Morris (1987) for factor screening. These "multiple grouping" methods can in some cases, attain some of the savings of sequential approaches although requiring only one or two temporal sets of tests.

The first, and sometimes only, stage of a multiple grouping screening experiment can be thought of as r simultaneous applications of Watson's original concept, in which the factor groups are defined "orthogonally" in the different applications. Hence $f = 48$ factors might be organized in $g_{(1)} = 3$ type-1 groups of size $k_{(1)} = 16$ factors, and $g_{(2)} = 4$ type-2 groups of size $k_{(2)} = 12$ factors, such that the intersection of any group of type 1 with any group of type 2 contains 4 factors. This arrangement is depicted graphically in Figure 2. The individual factors followed up in the second stage are those for which all types of groups are apparently active. So, for example, if only the first group of type 1 (containing factors 1–16) and the first group of type 2 (containing factors 1–4, 17–20, and 33–36) are declared active, only factors 1, 2, 3, and 4 would be examined in the follow-up experiment. If intersections contain only one factor each, the second stage may be eliminated or used for verification purposes.

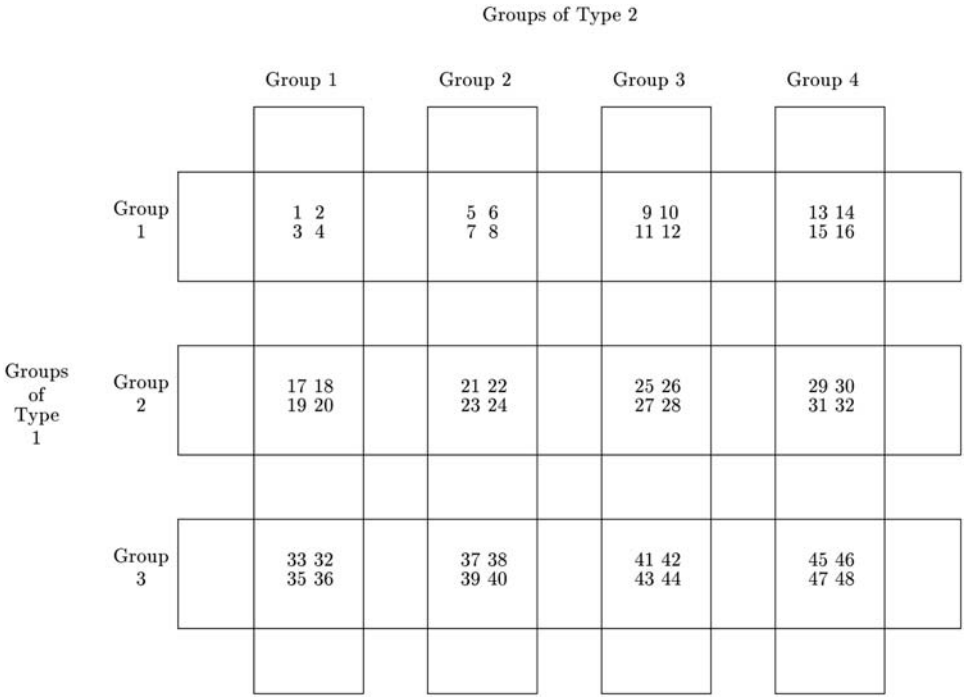

FIGURE 2. Example of individual factor assignments in a multiple grouping screening experiment involving factors labeled $1, 2, \ldots, 48$.

Morris discussed the construction of minimal experimental designs for such procedures, assuming $\sigma^2 = 0$. In most practical applications, performance tradeoffs involving run reuse and orthogonality, as discussed above, would need to be addressed.

5 Interactions

All discussion up to this point is predicated fairly seriously on Watson's assumption (iii), that is, the assumption that factors do not interact. However, suppose now that some two-factor interactions do exist so that, as distinct from equation (3), the model is

$$\boldsymbol{Y} = \mathbf{1}\beta_0 + \boldsymbol{Z}_1\boldsymbol{\gamma}_1 + \boldsymbol{Z}_2\boldsymbol{\gamma}_2 + \boldsymbol{\epsilon}, \tag{4}$$

where

- $\boldsymbol{Z}_1$ and $\boldsymbol{\gamma}_1$ are as described as $\boldsymbol{Z}$ and $\boldsymbol{\gamma}$ before (representing main effects for factor groups), and
- $\boldsymbol{Z}_2$ is the appropriate model matrix for a set of two-factor interactions, elements of the vector $\boldsymbol{\gamma}_2$.

For experiments in which $\mathbf{Z}_1$ is of full column rank, it is well known that if model (3) is used as the basis for analysis, least squares estimation is biased by the nonzero elements of $\boldsymbol{\gamma}_2$:

$$E(\hat{\boldsymbol{\gamma}}_1) = \boldsymbol{\gamma}_1 + (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Z}_2\boldsymbol{\gamma}_2. \tag{5}$$

This issue is addressed, for instance, in the discussion of model coefficient aliasing in books on response surface analysis such as Myers and Montgomery (2002). It is clear that this aliasing can introduce serious problems into decision processes based upon the realized estimates of model coefficients.

5.1 Avoiding Bias Due to Interactions

In a classic reference, Box and Hunter (1961) noted that "foldover" designs comprised of pairs of runs that are "mirror images" of each other, for example:

$$(+\,+\,-\,-\,-\,+\,-)$$
$$(-\,-\,+\,+\,+\,-\,+),$$

eliminate the aliasing between odd- and even-order effects, and so allow unbiased estimation of main effects even when two-factor interactions exist. Resolution IV main effects plans comprised of foldover run pairs require at least twice as many runs as factors—the operational cost of this benefit. Bettonvil (1993) noted that the sequential bifurcation strategy can be modified to avoid aliasing of main effects with two-factor interactions by adding foldover pairs of runs, rather than individual runs, at each step; similar modifications could certainly be made to the other strategies mentioned here.

5.2 Modeling Interactions

Often interest lies not in simply eliminating the bias from main effect estimates, but also in identifying the interactions that are nonzero. The goal here is screening the effects (main effects and two-factor interactions together) rather than the factors (assuming only main effects are present). Dean and Lewis (2002) and Lewis and Dean (2001) discussed the use of Resolution V designs (which allow estimation of main effects and two-factor interactions) in the group factors in the first stage of a two-stage screening study. These designs use more runs than the Resolution III plans (main effects only) typically used in screening experiments. However, they allow estimation of

- group main effects (the sum of all individual main effects for the group), and
- two-group interactions (the sum of all individual two-factor interactions with one factor in each of two different groups).

Individual two-factor interactions for pairs of factors within the same group are aliased with the intercept and so are not part of any informative estimable combination. In the second stage, the model of interest contains:

- all individual main effects for factors included in an apparently active group main effect,
- all individual two-factor interactions for pairs of factors included in a single group with an apparently active group main effect,
- all individual two-factor interactions for pairs of factors, one each from two groups with an apparently active group interaction, and
- any additional individual main effects required to make the model fully hierarchical.

For example, to screen the main effects and two-factor interactions associated with 20 individual factors, 5 groups of 4 factors each might be formed (say, with factors 1–4 in group 1, and so on). Each 2^{5-1} half-replicate associated with the defining relation $I = \pm ABCDE$ is a resolution V design that supports estimation of grouped main effects and two-factor interactions. Suppose that only the main effect associated with group 1 and the two-factor interaction associated with groups 1 and 2 appear to be active in the first stage. Then the individual-factors model used in the second stage would contain:

- an intercept,
- main effects for factors 1–4, because the group 1 main effect is active,
- two-factor interactions for all pairs of factors 1–4, because the group 1 main effect is active,
- two-factor interactions involving one factor from group 1 (1, 2, 3, and 4) and one factor from group 2 (5, 6, 7, and 8), because the interaction for groups 1 and 2 is active, and
- main effects for factors 5–8, so that the model is hierarchical.

The motivating context for this work is robust product design, where each factor is labeled as either a *control* factor or *noise* factor. The distinction between these factors leads to somewhat different effect classification rules and allows the use of group designs of total resolution less than V when some interactions are not of interest. See Lewis and Dean (2001) and Vine et al. (2004) for details, as well as a description of software to evaluate interaction screening designs; the software is available at www.maths.soton.ac.uk/staff/Lewis/screen_assemble/group_screening.html.

6 Discussion

The essential characteristic of group screening for factors is the intentional confounding of main effects at various experimental stages, with the aim of reducing the number of runs required to identify those factors that are most important. The number of possible variations on the original theme described in Watson's (1961) paper is nearly limitless. The degree to which runs may be reused, the decision as to whether orthogonal designs should be required at each stage, and modifications to allow consideration of models that include interactions have been briefly considered here.

The formulae for the expected numbers of runs and misclassified factors derived in some of the referenced papers are somewhat complicated, but they are useful in understanding how alternative screening designs and procedures can be expected to perform under simple assumptions. When less stringent assumptions can be made, more elaborate decision rules can be considered. In other circumstances for which classical analysis is difficult, expected performance of competing plans may more easily be evaluated by numerical simulation studies that mimic the screening process. Randomly generated "realities" (such as the number and magnitude of active effects) can be generated, results of each screening strategy/plan applied to the simulated experiment, and those strategies with the best statistical properties (such as smallest expected number of runs or misclassified factors) can be identified. An investigator facing a specific factor screening problem, with specific requirements for replication, blocking, and the possibility that some combination of Watson's working assumptions may be inappropriate, can experiment numerically with the ideas discussed in the literature in order to understand the most relevant performance characteristics of alternative strategies.

References

Bettonvil, B. (1995). Factor screening by sequential bifurcation. *Communications in Statistics—Simulation and Computation*, **24**, 165–185.

Box, G.E.P. and Hunter, J.S. (1961). The 2^{k-p} fractional factorial designs, part I. *Technometrics*, **3**, 311–351.

Box, G.E.P. and Meyer, R.D. (1986). An analysis of unreplicated fractional factorials. *Technometrics*, **28**, 11–18.

Dean, A.M. and Lewis, S.M. (2002). Comparison of group screening strategies for factorial experiments. *Computational Statistics and Data Analysis*, **39**, 287–297.

Dorfman, R. (1943). The detection of defective members of large populations. *Annals of Mathematical Statistics*, **14**, 436–440.

Du, D-Z. and Hwang, F. (2000). *Combinatorial Group Testing and its Applications*, second edition. World Scientific, Singapore.

Federer, W.T. (1987). On screening samples in the laboratory and factors in factorial investigations. *Communications in Statistics—Theory and Methods*, **16**, 3033–3049.

Kleijnen, J.P.C (1987). Review of random and group-screening designs. *Communication in Statistics—Theory and Methods*, **16**, 2885–2900.

Lewis, S.M. and Dean, A.M. (2001). Detection of interactions in experiments on large numbers of factors. *Journal of the Royal Statistical Society B*, **63**, 633–672.

Mauro, C.A. and Smith, D.E. (1982). The performance of two-stage group screening in factor screening experiments. *Technometrics*, **24**, 325–330.

Morris, M.D. (1987). Two-stage factor screening procedures using multiple grouping assignments. *Communications in Statistics—Theory and Methods*, **16**, 3051–3067.

Myers, R. and Montgomery, D. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, second edition. Wiley, New York.

Odhiambo, J.W. and Manene, M.M. (1987). Step-wise group screening designs with errors in observations. *Communications in Statistics—Theory and Methods*, **16**, 3095–3115.

Patel, M.S. (1962). Group-screening with more than two stages. *Technometrics*, **4**, 209–217.

Plackett, R.L., and Burman, J.P. (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–325.

Vine, A.E., Lewis, S.M., and Dean, A.M. (2004). Two-stage group screening in the presence of noise factors and unequal probabilities of active effects. *Statistica Sinica*, **15**, 871–888.

Watson, G.S. (1961). A study of the group screening method. *Technometrics*, **3**, 371–388.