Statistica Sinica **15**(2005), 871-888

# TWO-STAGE GROUP SCREENING IN THE PRESENCE OF NOISE FACTORS AND UNEQUAL PROBABILITIES OF ACTIVE EFFECTS

A. E. Vine[1], S. M. Lewis[1] and A. M. Dean[2]

[1] *University of Southampton and* [2] *Ohio State University*

*Abstract:* Group screening is a technique for examining a large number of factors in order to discover the few factors that have important influences on a measured response. In two-stage group screening, factors are assigned to groups and new "grouped factors" are investigated in a first stage experiment by varying all the factor values within a group simultaneously. The factors within those groups identified as important are then investigated individually in a second stage experiment. This paper describes theory and software that allows investigation of group screening in the presence of unequal-sized groups of factors in the first stage experiment and different probabilities of the various main effects and interactions being important (or active). Examples are given to show how the results can be used in practice to guide the choice of the number and sizes of the groups and to investigate the advantages and disadvantages of different group screening strategies.

*Key words and phrases:* Design of experiments, fractional factorial experiment, interaction, noise factor, screening, simulation.

## 1. Introduction

The process of improvement in the quality of manufactured products requires time-efficient, economical methods of experimentation. A product is of high quality if it achieves a target mean performance and also exhibits little variation in the presence of uncontrolled, or uncontrollable, manufacturing and environmental variability. Factors in an experiment representing uncontrolled sources of variation in the manufacturing process or in product components are known as *noise factors*. The goal of quality improvement experiments is to identify settings of controllable factors (*control factors* or *design factors*) which make the product performance insensitive to the uncontrolled variability of the noise factors. Such experimentation has been used with great success in a wide variety of industries for a moderate number of control and noise factors, for example in environmental engineering (Brickell and Knox (1992)), the automobile industry (Pignatiello and Ramberg (1985)), the electronics industry (Kackar and Shoemaker (1986)),

the food industry (Tuck, Lewis and Cottrell (1993)) and mechanical engineering (Sexton, Dunsmore, Lewis, Please and Pitts (2000)).

In the early stages of experimentation in manufacturing industries, numerous factors are often proposed as possibly influencing the product performance—many more factors than can be accommodated by conventional experimental plans such as classical fractional factorial designs. The technique of *group screening* is capable of exploring the effects of a large number of factors including their interactions. In group screening, the individual factors are placed into groups and a new "grouped factor" is defined to represent each group. Traditional methods of group screening (see, for example, Kleijnen (1987) and Du and Hwang (2000)) ignore the possibility of interactions. For product improvement, this can be a serious disadvantage because the exploration of control×noise interactions is crucial in product improvement (cf., Shoemaker, Tsui and Wu (1991)). In order to include an examination of control×noise interactions at stage 1, the groups are set up so that each consists entirely of control factors or entirely of noise factors. A grouped factor can then be identified as either a *grouped control factor* or a *grouped noise factor*. A first stage experiment is conducted on the grouped factors and the factors in the groups found to be important are investigated individually in a second stage experiment. Two different group screening strategies were examined by Lewis and Dean (2001) and Dean and Lewis (2002). In one strategy, only main effects at the first stage were investigated (*classical group screening*) whilst, in the second, two-factor interactions as well as main effects were examined at stage 1 (*interaction group screening*). The results of these two papers were restricted to the situation in which equal-sized groups are formed and it can be assumed that all control factors have the same probability of being active, as do all the noise factors. The work in the following sections allows considerably wider application of group screening techniques and better use of prior knowledge elicited from subject specialists by permitting investigation of unequal group sizes and/or unequal probabilities of the individual factorial effects being active.

In order to handle this extra complexity, we develop a general framework different from that given by Lewis and Dean (2001). This new framework is somewhat simpler conceptually and allows explicit formulae to be developed for the expected number of effects that need to be examined at stage 2 of the experiment. The results and techniques can easily be applied to the simpler situation of an experiment with no noise factors.

In Section 2, we define the notation and describe the criteria that we adopt for assessing screening strategies. In Section 3, we give general theoretical results

on the expected numbers of effects to be estimated which incorporate unequal probabilities of active effects and differing group sizes. We discuss application of these results and issues related to the selection of numbers of groups, group sizes and screening strategy in Section 4. We give some examples obtained through software implementation of the theory, as well as an application of simulation software which enables an assessment to be made of the number of active effects that may be missed in a given setting.

## 2. Two-stage Group Screening Strategies

We assume that the responses from a first (second) stage experiment may be described as $Y = \tau + \epsilon$ where $\tau$ denotes the vector of effects of each of the combinations of levels of the grouped (individual) factors that are observed in the first (second) stage experiment, and $\epsilon$ denotes a random error variable with mean zero and variance $\sigma^2$.

We define a quantity $\Delta$ $(> 0)$ to be a sufficiently large difference in the responses obtained from two distinct treatment combinations to give an economic advantage when one treatment combination is chosen over the other in the design of the product. In order to compare the estimated main effects and interactions (for individual or grouped factors) with $\Delta$, each factorial contrast is scaled so that its least squares estimator has the same standard deviation as the estimator of the difference between two treatment effects. In the analysis of the data from the first stage experiment, the scaled factorial contrasts for the grouped factors can be compared with $\Delta$ via hypothesis tests, as in equation (2.1) of Lewis and Dean (2001), and a grouped factorial effect is declared *active* if the result of the hypothesis test is significant. Alternative approaches include Bayesian methods similar to those of Box and Meyer (1986), Chipman, Hamada and Wu (1997) and Beattie, Fong and Lin (2002).

In "classical group screening", a grouped control or noise factor is considered to be *active* if the grouped factor is found to have an active main effect in the first stage analysis. An active grouped factor is carried forward to stage 2 of the experiment where the grouping is dismantled and the main effects of all factors within the group are examined individually. All control×control and control×noise interactions among the individual factors involved in the active groups need to be examined also. In "interaction group screening", a grouped control factor is considered to be active if the results of the first stage analysis indicate that the factor has an active main effect or is involved in an active interaction. In contrast, a grouped noise factor is considered active only if it is involved in an active interaction with at least one grouped control factor. At stage

2, the main effects of individual factors in an active group are examined, together with the interactions between control factors in the same group. Control×control and control×noise interactions involving individual factors in *different groups* are examined *only* if their corresponding grouped interactions are deemed active at the first stage.

We consider individual factors $A_{11}, \ldots, A_{1g_1}, A_{21}, \ldots, A_{2g_2}, \ldots, A_{b1}, \ldots, A_{bg_b}$ with two levels each. The high (low) level of a factor is the value at which the largest (smallest) response is thought likely to occur. Where curvature in the response across the range of a factor is anticipated, the value of the factor that corresponds to the anticipated maximum (minimum) response is used as the high (low) level of that factor. An important part of planning an experiment is the elicitation of available knowledge and experience from as many subject specialists as possible about factors that might be investigated, their likely importance and appropriate settings for their levels. A web-based software system (*GISEL*) has been developed to facilitate this elicitation process, see Dupplaw, Brunson, Vine, Please, Lewis, Dean, Keane and Tindall (2004). It also incorporates the two sets of software *gsize* and *gsim* described in this paper. All this software is available at www.maths.soton.ac.uk/staff/Lewis/screen_assemble.

We denote the probability that the main effect of the individual factor $A_{ik}$ is active by $q_{ik}^{(c)}$ if $A_{ik}$ is a control factor and by $q_{ik}^{(n)}$ if $A_{ik}$ is a noise factor. The probability that the two-factor interaction between the individual control factor $A_{ik}$ and individual factor $A_{jl}$ is active is denoted by $q_{ik,jl}^{(cn)}$ when $A_{jl}$ is a noise factor, and by $q_{ik,jl}^{(cc)}$ when $A_{jl}$ is a control factor.

The individual factors are divided into $b$ groups in such a way that the $i$th group, represented by the *grouped factor* $B_i$ $(i = 1, \ldots, b)$, contains individual factors $A_{i1}, A_{i2}, \ldots, A_{ig_i}$. When all factors in the $i$th group are at their high (low) level then grouped factor $B_i$ is at its high (low) level. The relationships between the factorial contrasts for the grouped factors and the individual factors are given by Lewis and Dean (2001), Theorem 1. We denote the $F$ grouped control factors by $B_1^{(c)}, \ldots, B_F^{(c)}$ and the $N$ grouped noise factors by $B_1^{(n)}, \ldots, B_N^{(n)}$. The respective group sizes are denoted by $g_1^{(c)}, \ldots, g_F^{(c)}, g_1^{(n)}, \ldots, g_N^{(n)}$, and the total number of individual control factors and individual noise factors are $n_C = \sum_{i=1}^{F} g_i^{(c)}$ and $n_N = \sum_{i=1}^{N} g_i^{(n)}$, respectively. The probabilities of the grouped main effects being active are denoted by $\rho_1^{(c)}, \ldots, \rho_F^{(c)}, \rho_1^{(n)}, \ldots, \rho_N^{(n)}$.

Suppose that the individual factorial effects are independently active or non-active, that any non-active effect is zero and that the "high" level of each factor

produces the higher response. Then, under these simplifying assumptions,

$$\rho_i^{(c)} = 1 - \prod_{A_{ik} \in B_i^{(c)}} (1 - q_{ik}^{(c)}), \quad k = 1, \ldots, g_i^{(c)}, \ 1 \le i \le F,$$

with an analogous, formulation for $\rho_j^{(n)}$, $1 \le j \le N$. In interaction group screening (IGS), the probability that grouped control×noise interaction $B_i^{(c)} B_j^{(n)}$ is active is

$$\rho_{i,j}^{(cn)} = 1 - \prod_{A_{ik} \in B_i^{(c)}} \prod_{A_{jl} \in B_j^{(n)}} (1 - q_{ik,jl}^{(cn)}),$$

where $k = 1, \ldots, g_i^{(c)}$, $1 \le i \le F$, $l = 1, \ldots, g_j^{(n)}$, $1 \le j \le N$, with corresponding respective probabilities $\rho_{i,k}^{(cc)}$ and $\rho_{j,l}^{(nn)}$ of a grouped control×control interaction $B_i^{(c)} B_k^{(c)}$ and noise×noise interaction $B_j^{(n)} B_l^{(n)}$ being active ($1 \le i < k \le F$, $1 \le j < l \le N$).

Let $p_i^{(c)}$ ($p_j^{(n)}$) be the probability that the analysis of the data from the first stage experiment leads to the main effect of the $i$th control ($j$th noise) factor being declared active, with similar definitions for $p_{i,j}^{(cn)}$, $p_{i,k}^{(cc)}$ ($1 \le i, k \le F; 1 \le j \le N$) for the interactions. If $\Delta$ is large compared with the error standard deviation $\sigma$, if the non-active effects are close to zero and if the effects neither cancel nor accumulate within the groups, then no errors of testing would be made and $p_q^{(x)}$ would be identical to $\rho_q^{(x)}$. Although these ideal conditions cannot be met, $p_q^{(x)}$ should be fairly close to $\rho_q^{(x)}$ under factor sparsity. The theoretical results in Section 3 are given in terms of $p_q^{(x)}$. Simulation studies in terms of $\rho_q^{(x)}$ under factor sparsity show that the approximation is good enough for practical purposes (see also, Lewis and Dean (2001)).

The following practical criteria, which cannot be achieved simultaneously, are important in designing a group screening experiment and will be discussed in Sections 3 and 4.

- *Target-excess criterion*: minimize the probability of exceeding a specified total number of observations.
- *Active effect identification criterion*: maximize the probability of detecting the active individual control main effects, control×noise and control×control interactions.
- *Mean-size criterion*: minimize the total number of observations made on average. This differs from the target-excess criterion because no account is taken of the standard deviation of the size of the second stage experiment.
- *Type I error criterion*: minimize the probability of incorrectly identifying individual factorial effects as active when they are not.

## 3. Experiment Size

The total size of a two-stage group screening experiment depends on the plans used at each stage, the number and sizes of the groups of factors at stage 1, and the number of individual factorial effects that have to be estimated at the second stage. In this section, we formulate the theoretical results needed to examine both classical and interaction group screening under the target-excess and mean-size criteria for given group sizes. Prior knowledge from subject specialists allows formulation of the predictive probability distribution of the number of individual factorial effects that require estimation at the second stage. We assume all interactions involving three or more factors are negligibly small.

### 3.1. Classical group screening

We define $I^{(c)} = (\delta_1^{(c)}, \delta_2^{(c)}, \ldots, \delta_F^{(c)})$ and $I^{(n)} = (\delta_1^{(n)}, \delta_2^{(n)}, \ldots, \delta_N^{(n)})$ to be random indicator vectors with $k$th entry equal to 1 when the corresponding grouped factor $B_k^{(c)}$ or $B_k^{(n)}$ is declared active at stage 1, and 0 otherwise. Then, under classical group screening, the number $S_C^{(c)}$ of individual control main effects and the number $S_C^{(cc)}$ of individual control×control interactions to be investigated at the second stage are random variables as follows:

$$S_C^{(c)} = \sum_{i=1}^{F} g_i^{(c)} \delta_i^{(c)} \quad \text{and} \quad S_C^{(cc)} = \frac{1}{2} S_C^{(c)} (S_C^{(c)} - 1)\,.$$

The experiment only extends to a second stage when at least one grouped control factor is declared active and, therefore, the number of individual noise main effects to be investigated at stage 2 is

$$S_C^{(n)} = \eta_C^{(c)} \sum_{j=1}^{N} g_j^{(n)} \delta_j^{(n)},$$

where $\eta_C^{(c)} = 1$ when $S_C^{(c)} \geq 1$ and 0 otherwise. Since noise×noise interactions are nuisance effects, we adopt the approach of Lewis and Dean (2001) and use the lower bound $S_C^{(n)} - 1$ to represent the number of sets of aliased individual noise×noise interactions at stage 2 when at least one grouped noise main effect is declared active at stage 1. Then, the number $S_C^{(nn)}$ of individual noise×noise interactions to be studied at the second stage is given by

$$S_C^{(nn)} = S_C^{(n)} - \eta_C^{(n)}\,,$$

where $\eta_C^{(n)} = 1$ when $S_C^{(n)} \geq 1$ and 0 otherwise. Finally, the number $S_C^{(cn)}$ of individual control×noise interactions that need to be studied at stage 2 is

$$S_C^{(cn)} = S_C^{(c)} S_C^{(n)}.$$

The total number of effects to be estimated (including the mean) at the second stage under classical group screening can now be expressed as

$$U_{CGS}^{(2)} = S_C^{(c)} + S_C^{(n)} + S_C^{(cc)} + S_C^{(cn)} + S_C^{(nn)} + \eta_C^{(c)}. \tag{3.1}$$

It is clear that, for any given experiment under classical group screening, the total number of effects, $U_{CGS}^{(2)}$, to be estimated at stage 2 is determined by the realisations of the random index vectors $I^{(c)}$ and $I^{(n)}$ alone. The joint probability function of $I^{(c)}$ and $I^{(n)}$ is

$$P(I^{(c)} = I_{t_1}^{(c)}, I^{(n)} = I_{t_2}^{(n)}) = \prod_{i=1}^{F} (p_i^{(c)})^{\delta_{i:t_1}^{(c)}} (1 - p_i^{(c)})^{1-\delta_{i:t_1}^{(c)}} \prod_{j=1}^{N} (p_j^{(n)})^{\delta_{j:t_2}^{(n)}} (1 - p_j^{(n)})^{1-\delta_{j:t_2}^{(n)}},$$

where $1 \leq t_1 \leq 2^F$, $1 \leq t_2 \leq 2^N$, and $\delta_{i:t_1}^{(c)}$ and $\delta_{j:t_2}^{(n)}$ are realisations of the random variables $\delta_i^{(c)}$ and $\delta_j^{(n)}$, respectively. At the first stage of the experiment, only main effects are examined and the number of effects examined (together with the mean) at stage 1 is $U_{CGS}^{(1)} = 1 + F + N$. Hence, for $1 + F + N \leq s \leq 2 + F + N + \binom{n_C+1}{2} + 2n_N + n_C n_N - \eta_C^{(n)}$, the probability under classical group screening that the total number of effects $S_{CGS}$ requiring estimation is $s$ can be expressed as

$$P(S_{CGS} = s) = \sum_{R_{CGS}} P(I_{t_1}^{(c)}, I_{t_2}^{(n)}),$$

where $R_{CGS} = \{(I_{t_1}^{(c)}, I_{t_2}^{(n)}); U_{CGS}^{(2)} = s - (1 + F + N)\}$. We can use (3.1) to calculate the expected total number of effects to be examined $E(S_{CGS}) = U_{CGS}^{(1)} + E(U_{CGS}^{(2)})$ in terms of the probabilities that the grouped main effects will be declared active, as follows.

$$\begin{aligned} &E(S_{CGS}) \\ &= 1 + F + N + \frac{1}{2}E(S_C^{(c)}) + 2E(S_C^{(n)}) + \frac{1}{2}E([S_C^{(c)}]^2) + E(S_C^{(c)} S_C^{(n)}) + E(\eta_C^{(c)}) - E(\eta_C^{(n)}) \end{aligned}$$

$$= 2+F+N+\frac{1}{2}\sum_{i=1}^{F} g_i^{(c)} p_i^{(c)} + 2\Big[\sum_{j=1}^{N} g_j^{(n)} p_j^{(n)}\Big]\Big[1-\prod_{i=1}^{F}(1-p_i^{(c)})\Big]$$
$$+\frac{1}{2}\sum_{i=1}^{F}[g_i^{(c)}]^2 p_i^{(c)}(1-p_i^{(c)}) + \frac{1}{2}\Big[\sum_{i=1}^{F} g_i^{(c)} p_i^{(c)}\Big]^2 + \Big[\sum_{i=1}^{F} g_i^{(c)} p_i^{(c)}\Big]\Big[\sum_{j=1}^{N} g_j^{(n)} p_j^{(n)}\Big]$$
$$-\prod_{i=1}^{F}(1-p_i^{(c)}) - \Big[1-\prod_{i=1}^{F}(1-p_i^{(c)})\Big]\Big[1-\prod_{j=1}^{N}(1-p_j^{(n)})\Big].$$

### 3.2. Interaction group screening

We define two new random vectors $I^{(cc)} = (\delta_{1,2}^{(cc)}, \ldots, \delta_{F-1,F}^{(cc)})$ of length $c_F = F(F-1)/2$ and $I^{(cn)} = (\delta_{1,1}^{(cn)}, \ldots, \delta_{F,N}^{(cn)})$, of length $FN$, where $\delta_{i,k}^{(cc)}$ ($\delta_{i,j}^{(cn)}$) is equal to 1 if the interaction between grouped control factor $B_i^{(c)}$ and grouped control factor $B_k^{(c)}$ (noise factor $B_j^{(n)}$) is declared active at stage 1, and 0 otherwise. The number $S_I^{(cn)}$ of individual control×noise interactions to be examined at the second stage is

$$S_I^{(cn)} = \sum_{i=1}^{F}\sum_{j=1}^{N} g_i^{(c)} g_j^{(n)} \delta_{i,j}^{(cn)}.$$

Since a grouped noise factor is taken forward to the second stage only if it is found to be involved in at least one control×noise interaction, the number of individual noise main effects examined at the second stage is

$$S_I^{(n)} = \sum_{j=1}^{N} g_j^{(n)} \gamma_j^{(n)},$$

where $\gamma_j^{(n)} = 1$ when $\sum_{i=1}^{F} \delta_{i,j}^{(cn)} \geq 1$, and zero otherwise.

In order to count up the number of individual control×control interactions to be examined at the second stage, we begin by counting the number $S_I^{(cc)b}$ of interactions between individual control factors from *different* groups involved in active grouped control×control interactions:

$$S_I^{(cc)b} = \sum_{i=1}^{F-1}\sum_{k=i+1}^{F} g_i^{(c)} g_k^{(c)} \delta_{i,k}^{(cc)}.$$

Since the main effects of, and the interactions between, all the individual control factors within the grouped factors taken forward to the stage 2 experiment are to be examined, we have the following number $S_I^{(cc)w}$ of individual control×control

interactions involving factors *within the same* group and the number $S_I^{(c)}$ of individual control main effects to be examined:

$$S_I^{(cc)w} = \sum_{i=1}^{F} \frac{1}{2}\, g_i^{(c)}(g_i^{(c)} - 1)\gamma_i^{(c)} \quad \text{and} \quad S_I^{(c)} = \sum_{i=1}^{F} g_i^{(c)}\gamma_i^{(c)},$$

where $\gamma_i^{(c)} = 1$ if the $i$th grouped control factor is taken forward to the second stage; that is if

$$\sum_{j=1}^{N} \delta_{i,j}^{(cn)} + \sum_{k=1,\, k\neq i}^{F} \delta_{i,k}^{(cc)} + \delta_i^{(c)} \geq 1\,,$$

and zero otherwise. As in Section 3.1, we use the lower bound $S_I^{(nn)} = S_I^{(n)} - \eta_I^{(n)}$ for the number of individual noise×noise interactions at the second stage, where $\eta_I^{(n)} = 1$ when $S_I^{(n)} \geq 1$ and 0 otherwise. Therefore, the total number of effects to be estimated at the second stage under interaction group screening is

$$U_{IGS}^{(2)} = S_I^{(c)} + 2S_I^{(n)} + S_I^{(cn)} + S_I^{(cc)b} + S_I^{(cc)w} + \eta_I^{(c)} - \eta_I^{(n)}. \tag{3.2}$$

For any given experiment under interaction group screening, the total number of effects, $U_{IGS}^{(2)}$, to be estimated at stage 2 is determined by the realisations of the random index vectors $I^{(c)}$, $I^{(cc)}$ and $I^{(cn)}$. The joint probability function of $I^{(c)}$, $I^{(cc)}$ and $I^{(cn)}$ is

$$\begin{aligned}
&P(I_{t_1}^{(c)}, I_{t_2}^{(cc)}, I_{t_3}^{(cn)}) \\
&= \prod_{i=1}^{F} (p_i^{(c)})^{\delta_{i:t_1}^{(c)}} (1 - p_i^{(c)})^{1-\delta_{i:t_1}^{(c)}} \prod_{i=1}^{F-1} \prod_{k=i+1}^{F} (p_{i,k}^{(cc)})^{\delta_{i,k:t_2}^{(cc)}} (1 - p_{i,k}^{(cc)})^{1-\delta_{i,k:t_2}^{(cc)}} \\
&\quad \times \prod_{i=1}^{F} \prod_{j=1}^{N} (p_{i,j}^{(cn)})^{\delta_{i,j:t_3}^{(cn)}} (1 - p_{i,j}^{(cn)})^{1-\delta_{i,j:t_3}^{(cn)}},
\end{aligned}$$

where $1 \leq t_1 \leq 2^F$, $1 \leq t_2 \leq 2^{c_F}$, $1 \leq t_3 \leq 2^{FN}$ and $\delta_{i:t_1}^{(c)}$, $\delta_{i,k:t_2}^{(cc)}$ and $\delta_{i,j:t_3}^{(cn)}$ are realisations of the random variables $\delta_i^{(c)}$, $\delta_{i,k}^{(cc)}$ and $\delta_{i,j}^{(cn)}$, respectively. At the first stage of the experiment, the number of effects examined (together with the mean) is

$$U_{IGS}^{(1)} = 1 + F + N + c_F + FN + (N - \zeta) = 2N + c_{F+1} + FN + (1 - \zeta),$$

where $\zeta = 0$ if $N = 0$ and 1 otherwise. Hence, for $2N + c_{F+1} + FN + (1 - \zeta) \leq s \leq 2 + 2N + c_{F+1} + FN - \zeta + c_{n_C+1} + 2n_N + n_C n_N - \eta_I^{(n)}$, the probability under

interaction group screening that the total number of effects requiring estimation is equal to $s$ can be expressed as

$$P(S_{IGS} = s) = \sum_{R_{IGS}} P(I_{t_1}^{(c)}, I_{t_2}^{(cc)}, I_{t_3}^{(cn)}),$$

where $R_{IGS} = \{(I_{t_1}^{(c)}, I_{t_2}^{(cc)}, I_{t_3}^{(cn)}); U_{IGS}^{(2)} = s - (2N + c_{F+1} + FN) + (1 - \zeta)\}$. We can use (3.2) to calculate the expected total number of effects to be examined in terms of the probabilities that the various grouped factorial effects will be declared active, as follows.

$$\begin{aligned}
&E(S_{IGS}) = U_{IGS}^{(1)} + E(U_{IGS}^{(2)}) \\
&= 2N + c_{F+1} + FN + E(S_I^{(cn)}) + 2E(S_I^{(n)}) + E(S_I^{(cc)b}) + [E(S_I^{(cc)w}) + E(S_I^{(c)})] \\
&\quad + E(\eta_I^{(c)}) - E(\eta_I^{(n)}) \\
&= 2N + c_{F+1} + FN + \sum_{i=1}^{F}\sum_{j=1}^{N} g_i^{(c)} g_j^{(n)} p_{i,j}^{(cn)} + 2\Big[\sum_{j=1}^{N} g_j^{(n)}\Big]\Big[1 - \prod_{i=1}^{F}(1 - p_{i,j}^{(cn)})\Big] \\
&\quad + \sum_{i=1}^{F-1}\sum_{k=i+1}^{F} g_i^{(c)} g_k^{(c)} p_{i,k}^{(cc)} \\
&\quad + \sum_{i=1}^{F}[g_i^{(c)} \frac{g_i^{(c)} + 1}{2}]\Big[1 - \prod_{j=1}^{N}(1 - p_{i,j}^{(cn)}) \prod_{k=1,\, k\neq i}^{F} (1 - p_{i,k}^{(cc)})(1 - p_i^{(c)})\Big] \\
&\quad - \prod_{i=1}^{F}\prod_{j=1}^{N}\prod_{k=1,\, k\neq i}^{F}\Big[(1 - p_{i,j}^{(cn)})(1 - p_{i,k}^{(cc)})(1 - p_i^{(c)})\Big] + \prod_{i=1}^{F}\prod_{j=1}^{N}(1 - p_{i,j}^{(cn)}).
\end{aligned}$$

## 4. Application of the Theory

The results of Section 3 can be used to explore the effect on $S_{CGS}$ and $S_{IGS}$ of different choices of group sizes, and the two group screening strategies, for different probabilities of the individual main effects and interactions being active. Software *gsize*, within the web based system *GISEL*, has been written to make an exploration feasible by calculating the probability distribution, expected value and standard deviation of each of $S_{CGS}$ and $S_{IGS}$, together with the probability that each of these exceeds a user-specified target. Thus, the software enables a choice to be made for a particular investigation under the mean-size and target-excess criteria described in Section 2.

### 4.1. Comparison of screening strategies

The following example illustrates the use of *GISEL* in comparing the CGS and IGS strategies.

**Example 4.1.** Suppose that there are 17 individual control factors in five groups of sizes 2, 4, 4, 3 and 4, with probabilities of individual main effects being active thought to be (0.6, 0.7), (0.1, 0.1, 0.1, 0.08), (0.05, 0.05, 0.01, 0.02), (0.005, 0.005, 0.005) and (0.005, 0.005, 0.005, 0.005). Suppose there are also six individual noise factors in three groups of sizes 1, 2 and 3 with probabilities of active individual main effects thought to be (0.5), (0.3, 0.4) and (0.01, 0.01, 0.02). Suppose that the value assigned to the probability of an individual control×noise or control×control interaction being active is made dependent on the corresponding main effects probabilities of the factors involved, according to the summary in Table 1.

Table 1. Probabilities of individual interactions being active in terms of the probabilities of the corresponding individual main effects being active.

| Main effects probabilities | $\leq 0.05$ | $0.08, 0.1$ | $\geq 0.3$ |
|---|---|---|---|
| $\leq 0.05$ | 0.0001 | 0.001 | 0.01 |
| 0.08, 0.1 | 0.001 | - | 0.1 |
| $\geq 0.3$ | 0.01 | 0.1 | 0.2 |

Figures 1 and 2 show graphs available from *GISEL* for this example. They indicate that classical group screening (CGS) performs better than interaction group screening (IGS) under both the mean-size and target-excess criteria. However, the CGS strategy often has much worse performance under the active effects identification criterion, as illustrated in Section 4.4. This is typical relative behaviour for the majority of experiments we have considered.

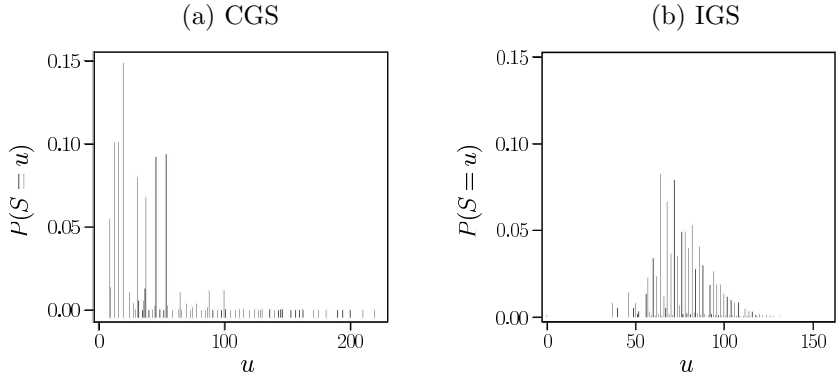


Figure 1. The probability distribution for the total number $S$ of effects requiring estimation under (a) CGS and (b) IGS.

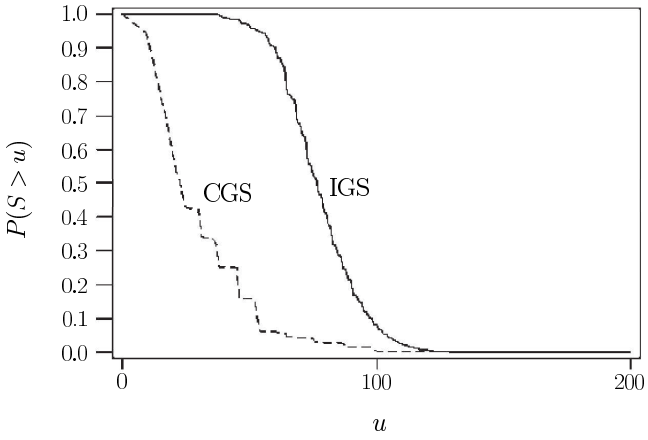


Figure 2. The probability of exceeding a target number of runs $u$ under CGS and IGS.

### 4.2. Assignment of probabilities

Choices for the probabilities of effects being active may be guided by information elicited from subject experts. In practice, these experts are usually more able to assess the probability that a main effect is active than that of an interaction. One way of assigning interaction probabilities is to use an assumption such as *effect heredity* or *relaxed weak heredity*, see Hamada and Wu (1992) and Chipman (1996). Relaxed weak heredity assumes that (i) the individual main effects of the factors in the experiment are active or inactive independently of each other, (ii) that, *conditional on* the status of the main effects of the factors, the interactions are active or inactive independently of each other, and (iii) the probability of an interaction being active depends only on the status of the main effects of the factors involved in the interaction. These assumptions lead to the assignment of the conditional probability that an interaction between two individual control factors, $A_{ik}$ and $A_{jl}$, say, is active:

$$P\left(\zeta_{ik,jl}^{(cc)} = 1 | \, \zeta_{ik}^{(c)} = s, \zeta_{jl}^{(c)} = t\right) = w_{st}^{(cc)}, \qquad s,t = 0,1\,, \tag{4.1}$$

where $\zeta_{ik}^{(c)}$ is an indicator function that takes value 1 when the main effect of $A_{ik}$ is active and 0 otherwise, and $\zeta_{ik,jl}^{(cc)}$ is a similar indicator function for the interaction between $A_{ik}$ and $A_{jl}$. The unconditional probability $q_{ik,jl}^{(cc)}$ is, therefore,

$$\begin{aligned} P\left(\zeta_{ik,jl}^{(cc)}=1\right) &= \sum_{s=0}^{1}\sum_{t=0}^{1} P\left(\zeta_{ik,jl}^{(cc)}=1|\zeta_{ik}^{(c)}=s,\zeta_{jl}^{(c)}=t\right) P\left(\zeta_{ik}^{(c)}=s\right) P\left(\zeta_{jl}^{(c)}=t\right) \\ &= w_{00}^{(cc)}(1-q_{ik}^{(c)})(1-q_{jl}^{(c)}) + w_{01}^{(cc)}(1-q_{ik}^{(c)})q_{jl}^{(c)} + w_{10}^{(cc)}q_{ik}^{(c)}(1-q_{jl}^{(c)}) \\ &\qquad + w_{11}^{(cc)}q_{ik}^{(c)}q_{jl}^{(c)}\,, \end{aligned} \tag{4.2}$$

where $w_{st}^{(cc)}$ is defined in (4.1). We can write a similar formulation for $P(\zeta_{ik,jl}^{(cn)} = 1)$.

The conditional probabilities $w_{ij}^{(..)}$ for control×control and control×noise interactions need to be specified. For any proposed values for $w_{ij}^{(..)}$, it is important to check that the unconditional probabilities are in agreement with factor sparsity. It should be noted that the degree of effect sparsity depends on the choice of $\Delta$. Larger values of $\Delta$ result in fewer effects being labelled as important and this, in turn, translates into smaller values of the conditional probabilities. The software system *GISEL* has a facility that allows the user to assign interaction probabilities via (4.1) and (4.2) and then to explore different choices for the conditional probabilities.

### 4.3. Formation of groups

The degree to which the number of effects to be estimated is affected by (a) the number of groups formed, (b) the selected group sizes and (c) the similarity in assigned probabilities of main effects within the groups, is problem dependent. This issue can be investigated using the theory and software as illustrated in the following example.

**Example 4.2.** Suppose there are six individual control factors with probabilities of active main effects thought to be 0.3, 0.4, 0.5, 0.6, 0.7 and 0.8, and six individual noise factors with main effect probabilities 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0. (A factor whose main effect is assigned probability 0.0 of being active may still be involved in an active interaction). In this example, we selected values of the conditional probabilities in line with factor sparsity in a screening experiment, namely, $\omega_{00}^{(cc)} = 0.005$, $\omega_{01}^{(cc)} = \omega_{10}^{(cc)} = 0.125$, and $\omega_{11}^{(cc)} = 0.25$, with corresponding values for the $\omega_{ij}^{(cn)}$. The unconditional probabilities were calculated through (4.2).

A variety of different groupings into $F = 1$, 2, 3, 4 or 5 groups of control factors and $N = 1$, 2 or 3 groups of noise factors were considered. Without loss of generality, the control (noise) factors were ordered in increasing values of $q_{ik}^{(c)}$ ($q_{ik}^{(n)}$) and labelled 1 to 6 (7 to 12).

The results obtained from *GISEL* are summarized in Table 2. In the table $(a_1, \ldots, a_{n_1}; b_1, \ldots, b_{n_2})$ denotes groups of control factors with sizes $a_1, \ldots, a_{n_1}$ and groups of noise factors with sizes $b_1, \ldots, b_{n_2}$, where the grouping is imposed on the factors in the order 1, 2, 3, 4, 5, 6; 7, 8, 9, 10, 11, 12 for "Similar" (groups containing factors with similar probabilities of an active main effect) and 1, 6, 2, 5, 3, 4; 7, 12, 8, 11, 9, 10 for "Dissimilar" (groups containing factors with dissimilar probabilities of an active main effect).

Table 2. Best and worst values for the number of groups and group sizes for Example 4.2 for similar and dissimilar probabilities of individual main effects being active. (* excluding (6; 1, 5))

| Criterion | Strategy | | | |
|---|---|---|---|---|
| | IGS | | CGS | |
| | Similar | Dissimilar | Similar | Dissimilar |
| max $E(S)$ | 72.98<br>(6; 5,1) | 72.73<br>(6; 6) | 71.65<br>(6; 6) | 71.65<br>(6; 6) |
| min $E(S)$ | 60.02<br>(2,2,2; 2,2,2) | 60.90<br>(2,2,2; 2,2,2) | 39.74<br>(1,1,1,1,2; 2,1,3) | 44.97<br>(1,1,1,2,1; 1,3,2) |
| max $P(S > 65)$ | 0.99<br>(6; 6) | 0.99<br>(6; 6) | 0.99<br>(6; 6) | 0.99<br>(6; 6) |
| min $P(S > 65)$ | 0.30<br>(2,2,2; 2,4) | 0.35<br>(2,2,2; 2,4) | 0.01<br>(1,1,1,1,2; 2,4)* | 0.03<br>(1,1,1,1,2; 1,4,1)* |

For interaction group screening, the best group sizes in terms of minimizing $E(S_{IGS})$, as required by the mean-size criterion, are $(2, 2, 2; 2, 2, 2)$, which correspond to groupings of factors with the following active main effect probabilities:

Similar (0.3, 0.4) (0.5, 0.6) (0.7, 0.8); (0.0, 0.2) (0.4, 0.6) (0.8, 0.1)

Dissimilar (0.3, 1.0) (0.4, 0.8) (0.5, 0.6); (0.0, 1.0) (0.2, 0.8) (0.4, 0.6)

The grouping together of factors with dissimilar main effects probabilities leads to virtually the same value of $E(S_{IGS})$ as that obtained by grouping together similar probabilities (see Table 2). For the target excess criterion, with target 65, these same group sizes (2,2,2; 2,2,2) give probabilities which are very close to those for the best grouping (2,2,2; 2,4). For this example under IGS, this grouping is a good choice, with freedom to allocate factors to groups in the most convenient way.

For classical group screening, there are small differences in the best groupings, for minimising $P(S > 65)$, as can be seen from Table 2. A better grouping (6; 1, 5), for minimising $P(S > 65)$, was excluded because the probability distribution of $S_{CGS}$ for this grouping had 99.5% of the probability concentrated at $s = 65$ and the remainder at $s = 4$. Large groups were uniformly poor choices for both interaction and classical group screening, in line with other examples we have examined.

### 4.4. Cancellation and amalgamation

A study, such as that in Section 4.3, which identifies a small number of promising groupings should be followed by a comparison of the groupings and associated strategies under the active effect identification criterion and the Type I error criterion. This is important because, when directions of effects are unknown or are unable to be lined up in the same way within a group, a small

possibility exists that effects may cancel, resulting in one or more active effects not being identified. On the other hand, when the directions of effects are lined up correctly, there is a small possibility of non-active effects amalgamating to produce a spuriously active grouped effect.

Assessment of different strategies to avoid cancellation or amalgamation can be made by simulation (Dean and Lewis (2002)). The software *gsim* described by these authors has been extended to accommodate unequal-sized groups and the important practical case where the control (noise) factors are divided into two sets such that the main effects of the factors in one set are thought extremely likely to be active but there is less information about the factors in the other set. The extremely likely active factors are assumed by *gsim* within *GISEL* to have main effects whose directions are known. The user can specify that the directions of any number of other main effects (should they be non-zero) are also known. The user can investigate a range of possible proportions of unknown main effects and interactions being active. In approximately the proportions specified, active main effects and two-factor interactions are randomly drawn by *gsim* from a normal distribution with user-specified mean and standard deviation. The non-active main effects and two-factor interactions are randomly drawn from a normal distribution having mean zero and standard deviation $\Delta/3$, where the user specifies $\Delta$ as the change in the response that is regarded as substantial (for more details see Dean and Lewis (2002)). All higher order interactions are set to 0. Random errors are generated from a normal distribution with mean zero and user-specified standard deviation. The user also specifies the required overall significance level and size of the simulation. The software has the facility to read a design for the first stage of experimentation automatically from the table of Russell, Lewis and Dean (2004).

From the simulation studies that we have run, the theoretical results based on the formulae in Section 3 appear to give reasonable approximations to the results that might be expected in practice. Thus, one can select suitable group sizes and grouping strategy using the theoretical results and then examine in depth the apparent best settings via simulation using *gsim*, as illustrated in Example 4.3.

For a particular grouping, CGS typically performs better than IGS under the mean-size and target-excess criteria. This is usually because CGS performs poorly under the active effect identification criterion and hence has a small number of effects requiring estimation at stage 2.

**Example 4.3.** Consider an experiment with 15 individual control factors and 4 individual noise factors. Suppose that the main effects of 7 individual control factors are thought very likely to be active and are assigned probability 1.0. The main effects of the remaining 8 individual control factors are assigned probability 0.2 of being active. There is little information about the 4 individual noise factors

and their main effects are assigned probability 0.3 of being active. The probabilities that individual control×noise, control×control and noise×noise interactions are active are set to 0.07, 0.05 and 0.3 respectively.

A systematic study of all possible groupings with group sizes of two or more, via *gsize*, showed that $\mathrm{E}(S_{IGS})$ ranged from 113 to 175, with $\mathrm{P}(S_{IGS} > 120)$ ranging from 0.27 to 0.98. Equally sized groups tended to give rise to the smaller values of $E(S_{IGS})$ and corresponding standard deviation $sd(S_{IGS})$. We have seen the same pattern in other examples. In this particular example, the smallest group sizes amongst those considered are the best under the mean-size and target-excess criteria, but this is not necessarily true in general. The best of the groupings are listed in Table 3 and any that are viable for practical application can be examined under the active effect identification criterion via *gsim*. Below we give the results of an investigation of the fourth listed grouping in Table 3 with the value of $\Delta$ set to 10.0, the overall Bonferroni significance level to 0.1, and the error distribution variance to 4.0. One possible way of reducing the expected number of effects to be estimated is to hold the levels of the factors in the likely active sets fixed during the experiment since their effects are assumed to be already known. For the current example, $E(S_{IGS})$ would then drop to 49 under the last listed grouping. This reduces the expected resource required but, of course, carries the risk of overlooking interactions between the likely active factors and the other factors screened out at the first stage.

Table 3. Investigation of different control group sizes for interaction group screening. Noise factors are in two groups of size 2.

| 7 v. likely indiv con. | | 8 less likely indiv con. | | $E(S)$ | $sd(S)$ | $P(S > 120)$ | $P(S > 150)$ | $P(S > 180)$ |
|---|---|---|---|---|---|---|---|---|
| gps | sizes | gps | sizes | | | | | |
| 2 | 2,5 | 3 | 2,3,3 | 125.79 | 18.76 | 0.62 | 0.09 | 0.00 |
| 2 | 2,5 | 4 | 2,2,2,2 | 120.85 | 16.42 | 0.52 | 0.04 | 0.00 |
| 2 | 3,4 | 3 | 2,2,4 | 124.45 | 18.56 | 0.60 | 0.08 | 0.00 |
| 2 | 3,4 | 3 | 2,3,3 | 122.18 | 18.23 | 0.54 | 0.06 | 0.00 |
| 2 | 3,4 | 4 | 2,2,2,2 | 117.41 | 15.84 | 0.43 | 0.02 | 0.00 |
| 3 | 2,2,3 | 2 | 4,4 | 124.89 | 18.44 | 0.61 | 0.08 | 0.00 |
| 3 | 2,2,3 | 3 | 2,2,4 | 117.85 | 15.72 | 0.44 | 0.02 | 0.00 |
| 3 | 2,2,3 | 3 | 2,3,3 | 115.69 | 15.34 | 0.38 | 0.01 | 0.00 |
| 3 | 2,2,3 | 4 | 2,2,2,2 | 112.97 | 13.00 | 0.27 | 0.00 | 0.00 |

A first stage design used in the simulation was read by *gsim* from the table of Russell, Lewis and Dean (2004) and had 32 observations. Labelling the five grouped control factors as $A, B$, $C, D$, $E$ and the grouped noise factors as $P$, $Q$, this design aliased the following pairs of two factor interactions: $(AB, CD)$,

$(AC, BD)$ and $(AD, BC)$. The probabilities $q_{i,j}^{(c)}$, $q_{i,j}^{(n)}$, $q_{ik,jl}^{(cc)}$, $q_{ik,jl}^{(cn)}$ and $q_{ik,jl}^{(nn)}$ were set at 0.2, 0.3, 0.05, 0.07 and 0.3, respectively, as above, with the probabilities of active main effects of the seven extremely likely active factors set to 1.0. We note that the simulation software creates actual scenarios, and the closest proportions of active effects that can be achieved are 0.25, 0.25, 0.0476, 0.0666, 0.333, representing two active control main effects out of eight, one active noise main effect, five, four, and two active control×control, control×noise and noise×noise interactions, respectively.

Table 4 shows part of the output from the simulation using the fourth listed grouping in Table 3. The numbers in Table 4 are calculated over 500 data simulations for each of 1,000 effect simulations. Not surprisingly, as the tails of the active effect distribution move away from $\Delta(= 10)$, it can be seen that a lower proportion of active effects is missed. It is also clear that a higher proportion of individual interactions tends to be missed by CGS than by IGS as has happened for most of the other examples that we have investigated.

If the probabilities of factorial effects being active are all set to zero, *gsim* can also be used to assess the probability of selecting effects as active when they are not (the Type I error criterion), see Dean and Lewis (2002) for an example.

Table 4. Simulation results for the proportions of active individual control main effects (cme), noise main effects (nme), control×control interactions (cxc) and control×noise interactions (cxn) that fail to be detected under interaction and classical group screening.

| | Active effect distribution | Proportion missed | | | | Ave. size for stage 2 |
|---|---|---|---|---|---|---|
| | | cme | nme | cxc | cxn | |
| Interaction group screening | N(30, 9) | 0.00 | 0.02 | 0.20 | 0.17 | 91 |
| | N(30,16) | 0.00 | 0.02 | 0.22 | 0.20 | 88 |
| | N(40,16) | 0.00 | 0.00 | 0.16 | 0.13 | 96 |
| | N(50,16) | 0.00 | 0.00 | 0.15 | 0.11 | 97 |
| | N(50,25) | 0.00 | 0.00 | 0.14 | 0.11 | 98 |
| Classical group screening | N(30, 9) | 0.04 | 0.06 | 0.44 | 0.64 | 94 |
| | N(30,16) | 0.04 | 0.09 | 0.44 | 0.67 | 93 |
| | N(40,16) | 0.02 | 0.00 | 0.41 | 0.63 | 99 |
| | N(50,16) | 0.02 | 0.00 | 0.43 | 0.62 | 100 |
| | N(50,25) | 0.03 | 0.00 | 0.42 | 0.62 | 99 |

**Acknowledgement**

This work was supported by EPSRC grants GR/N16754 and GR/R72693 and was undertaken while A. M. Dean was visiting Southampton Statistical Sciences Research Institute. The authors would like to thank William Meeker of Iowa State University for helpful comments.

## References

Beattie, S. D., Fong, D. K. H. and Lin, D. K. J. (2002). A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics* **44**, 55-63.

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

Brickell, J. and Knox, K. (1992). Designed experiments case history—environmental engineering operation of an activated sludge system. In *Understanding Industrial Designed Experiments*, 3rd edition. (Edited by S. R. Schmidt and R. G. Launsby), 8.187-8.194. Air Academy Press, Colorado.

Chipman, H. (1996). Bayesian variable selection with related predictors. *Canad. J. Statist.* **24**, 17-36.

Chipman, H., Hamada, M. and Wu, C. F. J. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics* **39**, 372-381.

Dean, A. M. and Lewis, S. M. (2002). Comparison of group screening strategies for factorial experiments. *Comput. Statist. Data Anal.* **39**, 287-297.

Du, D.-Z. and Hwang, F. (2000).*Combinatorial Group Testing and its Applications.* 2nd edition. World Scientific, Singapore.

Dupplaw, D. P., Brunson, D., Vine, A. E., Please, C. P., Lewis, S. M., Dean, A. M., Keane, A. J. and Tindall, M. J. (2004). A web-based knowledge elicitation system (GISEL) for planning and assessing group screening experiments for product development. *J. Computing and Information Sci. in Engineering* **4**, 218-225.

Hamada, M. and Wu, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *J. Quality Technol.* **24**, 130-137.

Kackar, R. N. and Shoemaker, A. S. (1986). Robust design: a cost-effective method for improving manufacturing processes. *AT&T Technical J.* **65**.

Kleijnen, J. P. C. (1987). Review of random and group-screening designs. *Comm. Statist. Theory Methods* **16**, 2885-2900.

Lewis, S. M. and Dean, A. M. (2001). Detection of interactions in experiments on large numbers of factors (with discussion). *J. Roy. Statist. Soc. Ser. B* **63**, 633-672.

Pignatiello, J. J. and Ramberg, J. S. (1985). Discussion of "Off-line quality control, parameter design, and the Taguchi method", by Kackar, R. N. *J. Quality Technol.* **17**, 198-206.

Russell, K. G., Lewis, S. M. and Dean, A. M. (2004). Fractional factorial designs for the detection of interactions between design and noise factors. *J. Appl. Statist.* **31**, 545-552.

Sexton, C. J., Dunsmore, W., Lewis, S. M., Please, C. P. and Pitts, G. (2000). Semi-controlled experiment plans for improved mechanical engineering designs. *Proc. Instn. Mech, Engnrs. B* **214**, 95-105.

Shoemaker, A. C., Tsui, K. and Wu, C. F. J. (1991). Economical experimentation methods for robust design. *Technometrics* **33**, 415-427.

Tuck, M. G., Lewis, S. M. and Cottrell, J. I. L. (1993). Response surface methodology and Taguchi: a quality improvement study from the milling industry. *J. Appl. Stat.* **42**, 671-681.

School of Mathematics, The University of Southampton, Southampton, SO17 1BJ, U.K.
E-mail: A.E.Vine@maths.soton.ac.uk

School of Mathematics, The University of Southampton, Southampton, SO17 1BJ, U.K.
E-mail: S.M.Lewis@maths.soton.ac.uk

Department of Statistics, The Ohio State University, Columbus, OH 43210-1247, U.S.A.
E-mail: amd@stat.mps.ohio-state.edu

(Received August 2003; accepted October 2004)