Generating Systematic Supersaturated Designs

Dennis K. J. LIN

Department of Statistics
Stokely Management Center
University of Tennessee
Knoxville, TN 37996-0532

Practitioners are routinely faced with distinguishing between factors that have real effects and those whose apparent effects are due to random error. When there are many factors, the usual advice given is to run so-called main-effect designs (Resolution III designs in the orthogonal case), that require at least $k + 1$ runs for investigating k factors. This may be wasteful, however, if the goal is only to *detect* those active factors. This is particularly true when the number of factors is large. In such situations, a supersaturated design can often save considerable cost. A supersaturated design is a (fraction of a factorial) design composed of n observations where $n < k + 1$. When such a design is used, the abandonment of orthogonality is inevitable. This article examines the maximum number of factors that can be accommodated when the degree of the nonorthogonality is specified. Furthermore, interesting properties of systematic supersaturated designs are revealed. For example, such a design may be adequate to allow examination of many prespecified two-factor interactions. Comparisons are made with previous work, and it is shown that the designs given here are superior to other existing supersaturated designs. Data-analysis methods for such designs are discussed, and examples are provided.

KEY WORDS: Equal occurrence; Nonorthogonality; Plackett and Burman design; Random balance design; Stepwise regression.

In a book by Schmidt and Launsby (1991, pp. 8-192–8-199), Curry, Tomick, and Yost described a successful case study for testing and validating an acquired immune deficiency syndrome (AIDS) model. The AIDS model, developed by the U.S. Bureau of the Census, Center for International Research, is a set of approximately 100 deterministic differential equations with over 150 input variables. Each run of the AIDS model requires approximately 15 minutes on a 386 microcomputer. Clearly, an efficient design was needed to investigate which variables in the model are important. From among the 150 input variables, an analysis team selected 97 variables that were of the most interest, using "expert knowledge." A 98-run plan was finally set up from a 100×100 Hadamard matrix (e.g., see Hedayat and Wallis 1978) and the five most "important" variables were identified. As we shall later see, using the results described in this article, 24 runs are sufficient to study *all* 150 variables in this context.

This AIDS-model example corresponds to a typical problem in industry. Routinely, engineers must distinguish between factors that have an actual effect and those factors whose apparent effects are due to random error. Often, the "null" factors are then adjusted to lower cost, and the "nonnull" (active) factors are used to yield better quality results. Many factors can often be listed as possibly important. It is not unusual, however, that among those factors only a small portion are active. In fact, many practitioners believe that the effects are typically *Pareto distributed.* This is sometimes called "effect sparsity." Usually, further investigation on the nonsignificant effects is not of interest. Estimating all effects may be wasteful if the goal is simply to detect those few active factors.

When all factors can be reasonably clustered into several groups, the so-called group screening designs sometimes can be used (e.g., see Watson 1961). Only those factors in groups that are found to have large effects are studied further. Although such methods may be appropriate in certain situations (e.g., blood tests), we are interested in supersaturated designs in which no grouping is made. We study fractions of two-level designs for examining k factors in $n < k + 1$ runs.

Satterthwaite (1959) stands as a pioneer in this area. He suggested constructing such designs at random. Although the idea of *random balance* designs is interesting, the designs themselves are not of maximum efficiency. [See the discussions (Box 1959; Hunter 1959; Kempthorne 1959; Tukey 1959; Youden 1959) that follow Satterthwaite's (1959) article.] Booth and Cox (1962) first examined this problem *systematically* (i.e., with designs that were generated using a specific optimality criterion). They provided seven supersaturated designs obtained via computer search. Recently, Lin (1993) provided a new class of supersaturated designs based on half factions of Hadamard matrices and showed a real-data example using an efficient study of 24 variables in only 14 runs. For other supersaturated design methods and applications, see Rosenberger and Smith (1984), Barnett and Hurwitz (1990), Voelkel (1990), and Lin (1991, 1993).

This article examines the maximum number of factors that can be accommodated when the number of runs is given and when the *degree of nonorthogonality* is specified. The construction method and computer algorithm to search for such systematic supersaturated designs are illustrated in Section 1. Some major results are summarized and comparisons with previous work are made in Section 2. Methods for analyzing data from such designs are discussed in Section 3. Examples are given for illustration in Section 4. Finally, I provide some concluding remarks in Section 5.

1. THE CONSTRUCTION ALGORITHM

When a supersaturated design is needed, the abandonment of orthogonality is inevitable. (Otherwise, these columns would form a set of more than n orthogonal vectors in n-dimensional space.) Because in most familiar contexts lack of orthogonality typically results in lower efficiency, we here seek designs as "near orthogonal" as possible. One simple way to measure the degree of nonorthogonality between two columns, c_i and c_j, is to consider their correlation, $r_{ij} = c_i'c_j/n$. Note that the level values for c_i and c_j are coded as ± 1 here. We denote the largest absolute value of r_{ij} among all pairs of columns for a given design by r, and we desire a minimum value for r ($r = 0$ implies orthogonality). If two designs have the same degree of nonorthogonality, r, we prefer the one in which the number of occurrences of r is a minimum among all $\binom{k}{2}$ pairs. Justification for using this minimax rule for r is given in later sections.

The algorithm, which is illustrated by a flow chart in Figure 1, can be described in the following way. When the run number, n, is specified, the algorithm generates all possible column combinations among the so-called equal-occurrence columns. When n is even, these are $n/2$ +1s and $n/2$ −1s; when n is odd, these are $(n+1)/2$ +1s and $(n-1)/2$ −1s. Permute these signs in every way possible. For example, if $n = 12$, then $12!/(6! \times 6!) = 924$ columns are generated; if $n = 25$, then $25!/(12! \times 13!) = 5{,}200{,}300$ columns are generated. These columns constitute the set of candidates for the design. Next, the algorithm randomizes the order in which these columns are considered as candidates. At each stage, a candidate column enters and its inner products associated with all other columns retained are calculated to check whether the requirement is satisfied (i.e., whether the maximum correlation is less than the prespecified r value). If not, the candidate column is dropped and the search continues. Because of the equal-occurrence property, it can be shown that $c_i'c_j + n \equiv 0 \pmod 4$. Thus the program examines only selected values of $r\,(= c_i'c_j/n)$. Computationally, this algorithm is much simpler than the one given by Booth and Cox (1962). Of course, today's computing facilities are much more powerful than those of the 1960s.

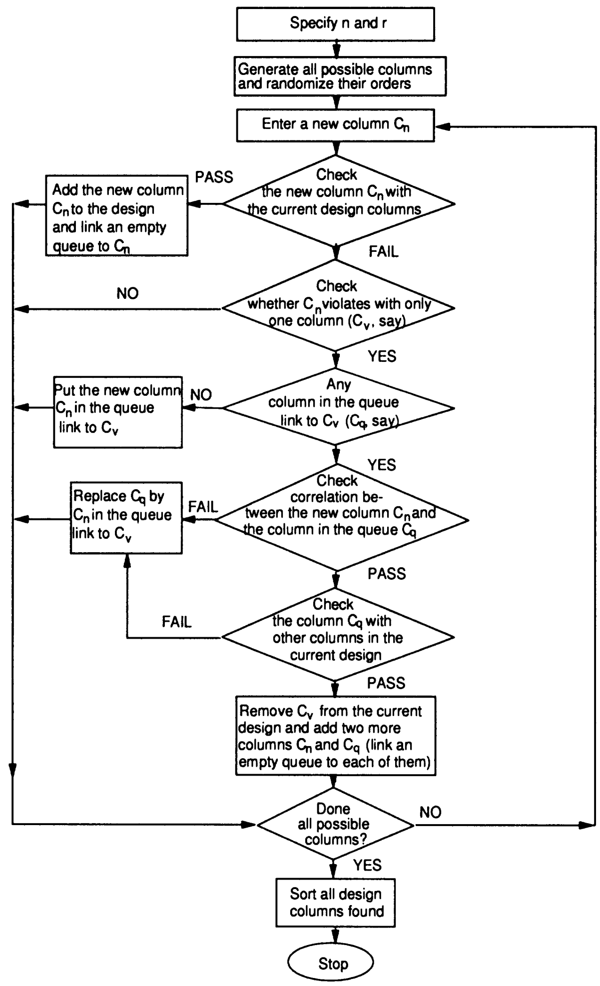

Figure 1. Flow Chart of the Algorithm (one complete loop).

One difficulty for this algorithm arises because the r-value property is not transitive. That is, when columns c_i and c_j produce $c_i'c_j/n \le r$ and columns c_j and c_m produce $c_j'c_m/n \le r$, it does not follow that columns c_i and c_m will necessarily have $c_i'c_m/n \le r$. To handle this, two more features are added to the program. Those candidate columns that meet the requirement for all but one retained column will be saved in a queue. Whenever two columns in the queue link to the same retained columns, the retained columns will be replaced by the two columns in the queue. Note that such an "exchange process" may remove certain columns that have smaller r values so that more columns can be saved to meet the r requirement. Another difficulty with this algorithm is that the extent to which the results depend on the random order of column entry is unknown. To help address this, the program has been rerun with different random orders of entrance for the candidate columns.

Once the design is constructed, the final step of the program is then to sort all columns in *order*. In other words, when fewer columns than the full design are needed, the experimenter can pick the first few columns that are considered to be the best choice in the sense of being the most nearly orthogonal. The criterion used here is the minimization of the average of r_{ij}^2, called mean squared correlation, $\rho^2 = \Sigma r_{ij}^2/\binom{k}{2}$. This criterion is equivalent to the $E(s^2)$ criterion given by Booth and Cox (1962). In fact, $\rho^2 = E(s^2)/n^2$, and this will be discussed in more detail in Section 3.

2. SOME RESULTS AND COMPARISONS

Tables 1 and 2 show the maximum number of factors, $k_{\max}$, that can be accommodated when both n and r are specified for $3 \le n \le 25$ and $0 \le r \le \frac{1}{3}$. (Table 1 is for even n and Table 2 is for odd n.) Because the inherent property of $c_i'c_j + n \equiv 0 \pmod 4$, we tabulate the results using n and $c_i'c_j$ as the column heading and row heading, respectively. We see that for $r \le \frac{1}{3}$ many factors can be accommodated. For fixed n, as the value of r increases, $k_{\max}$ also increases (read Tables 1 and 2 by rows). That is, the larger the nonorthogonality, the more factors can be accommodated. In fact, $k_{\max}$ increases rapidly. Certainly, the more factors accommodated, the more complicated are the biased estimation relationships that occur (as will be shown), leading to more difficulty in data analysis. On the other hand, for fixed r, the value of $k_{\max}$ increases rapidly as n increases. For example, when r is specified as $\frac{1}{3}$, $k_{\max}$ can be as large as 276 for $n = 24$. (As will be shown in Sec. 4, this design can allow one to study 150 variables, as in the AIDS model described in the introduction.) For $r \le \frac{1}{3}$, one can accommodate, at most, 111 factors in 18 runs or 66 factors in 12 runs; for $r \le \frac{1}{4}$, one can accommodate 42 factors in 16 runs; for $r \le \frac{1}{5}$, one can accommodate 34 factors in 20 runs. Provided that these maximal correlations are acceptable, this can lead to an efficient design. Supersaturated designs for $n \le 25$, as well as the computer codes, are all available from the author (e-mail: lin@stat.bus.utk.edu).

Table 1. Maximal Number of Factors Found, k_{max}, as a Function of n and nr, for $3 \le n \le 25$ and $r \le 1/3$: Even n

Number of runs n	Maximum absolute crossproduct, $nr = \|c_i'c_j\|$: 0	2	4	6	8
4	3				
6	—	10			
8	7	—			
10	—	12			
12	11	—	66		
14	—	13	—	113	
16	15	—	42	—	
18	—	17	—	111	
20	19	—	34		
22	—	20	—	92	—
24	23	—	33	—	276

Table 2. Maximal Number of Factors Found, k_{max}, as a Function of n and nr, for $3 \le n \le 25$ and $r \le \frac{1}{3}$: Odd n

Number of runs n	Maximum absolute crossproduct, $nr = \|c_i'c_j\|$: 1	3	5	7
3	3			
5	4			
7	7	15		
9	7	12		
11	11	14		
13	12	14		
15	15	15	37	
17	15	17	50	
19	19	19	33	
21	19	19	34	92
23	23	23	33	94
25	23	23	32	76

The cases $(n, k) = (12, 66)$ and (24, 276) both have $r = \frac{1}{3}$ and deserve special mention. As pointed out by one referee, the $(n, k) = (12, 66)$ design is indeed the 12-run Plackett and Burman (1946) design with all 11 *main-effect columns* plus all other $\binom{11}{2} = 55$ *two-factor interaction columns*. This design was first given by Lin (1991). Following this observation, other Plackett–Burman designs have been investigated. For $n \le 24$, only these two cases ($n = 12, 24$) have such a nice property (i.e., small r). Any two-factor interaction, if prespecified, merely corresponds to one more column in the design array, which can be easily incorporated in the stepwise selection procedure. Indeed, in many industries interactions are too common to be ignored. It is a great benefit of the supersaturated design to generally have plenty of room to examine the many two-factor interactions.

A full display of the new designs found by the algorithm for the case $n = 12$ is given in Table 3. A similar display for the cases $n = 18$ and 24 is given in Tables 4 and 5, for $r = \frac{1}{3}$ and $\frac{1}{4}$, respectively (to save space, only the first 54 columns for the case $n = 18$ were listed). The mean squared correlation of r_{ij}^2, ρ^2, can be easily evaluated. For example, in the design given in Table 3 for $n = 12$ and $k = 66$, among $\binom{66}{2} = 2{,}145$ pairs of all possible correlations, 0 appears 660 times, $-\frac{1}{3}$ appears 495 times, and $+\frac{1}{3}$ appears 990 times. Thus $\rho^2 = [0 \times 660 + (-\frac{1}{3})^2 \times 495 + (+\frac{1}{3})^2 \times 990]/2{,}145 = .077$. The ρ^2 measures the average loss of precision in the estimation of parameters if only two effects exist, and the average is taken over all possible pairs of columns.

Apart from Box's (1959, p. 180) illustrated example for the case $n = 3$, Booth and Cox (1962) first constructed such designs systematically. They provided seven supersaturated designs obtained via computer search ($n = 12$ for $k = 16, 20$, and 24; $n = 18$ for $k = 24, 30$, and 36; $n = 24$ for $k = 30$). It is, thus, a natural class of designs for comparison purposes. The results are given in Tables 6, 7, and 8, pp. 219, 220. Whenever the number, k, of columns is less than the whole design, the first

Table 3. Systematic Supersaturated Designs for n = 12

Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	+	+	+	+	+	−	−	−	−	−	−	+	+	+	+	+	+
3	+	−	+	−	−	+	+	+	−	−	−	+	+	+	−	−	−
4	+	−	−	+	−	−	+	−	+	+	−	+	−	−	+	+	−
5	+	−	−	−	+	+	−	−	+	−	+	−	+	−	+	−	−
6	+	+	−	−	−	−	−	+	−	+	+	−	−	+	−	+	−
7	−	+	+	−	−	+	−	−	+	+	−	+	−	−	−	+	+
8	−	+	−	−	+	−	+	+	+	−	−	−	−	+	+	−	−
9	−	+	−	+	−	+	+	−	−	−	+	+	−	−	+	−	+
10	−	−	−	+	+	+	−	+	−	+	−	−	+	−	−	−	+
11	−	−	+	−	+	−	+	−	−	+	+	−	−	+	−	−	+
12	−	−	+	+	−	−	−	+	+	−	+	−	+	−	−	+	−

Run	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	−	−	−	−	−	+	+	+	+	+	+	+	+	+	−	−	−
3	+	+	+	−	−	−	−	+	−	−	−	−	+	+	+	+	−
4	+	+	−	+	−	+	+	−	−	−	−	−	+	−	+	+	+
5	+	−	−	−	+	+	−	+	+	−	+	−	−	−	+	−	+
6	−	+	−	−	+	−	+	−	+	−	+	−	−	+	−	+	+
7	+	−	+	−	+	−	−	+	−	−	+	+	−	−	−	−	+
8	−	−	+	+	+	−	+	+	−	+	−	+	−	−	+	−	−
9	−	+	+	−	−	−	−	−	+	+	−	−	−	+	−	−	+
10	−	+	−	+	+	−	+	−	+	−	−	+	+	−	+	−	−
11	+	−	−	+	−	+	−	−	−	+	+	−	+	−	−	+	−
12	−	−	+	+	−	+	−	−	−	+	−	+	−	+	−	+	−

Run	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	−	−	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−
3	−	−	+	−	+	−	+	+	−	−	+	+	−	−	+	+	−
4	−	−	−	−	−	+	+	−	−	−	−	−	+	−	+	+	−
5	+	−	−	+	+	−	−	+	+	−	+	−	+	−	+	−	+
6	−	+	+	+	−	−	−	+	−	+	+	−	−	+	−	+	−
7	+	−	+	−	−	+	+	−	−	+	−	+	−	+	−	+	+
8	−	+	−	+	−	+	−	−	+	+	+	+	−	−	−	+	+
9	−	+	−	−	+	−	−	−	+	−	+	−	+	+	+	−	+
10	+	+	+	−	−	−	+	+	+	−	−	+	+	+	−	−	−
11	+	+	−	−	+	+	−	+	−	+	−	+	+	−	−	−	+
12	+	−	−	+	−	−	+	−	+	+	−	−	−	+	+	−	−

Run	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	−	−	−	−	−	+	−	+	+	−	+	+	−	−	+
3	−	−	+	+	−	+	−	−	−	+	+	−	+	−	−
4	+	+	−	−	+	−	+	−	+	−	+	−	−	+	+
5	+	−	−	+	−	−	−	−	+	+	−	+	−	+	−
6	+	−	+	−	+	−	−	+	−	+	−	−	−	+	+
7	−	+	+	−	−	−	+	−	+	+	−	+	−	−	−
8	+	+	−	+	+	+	−	−	−	−	+	−	−	−	−
9	−	+	+	−	+	+	+	−	−	−	−	+	+	−	+
10	−	+	−	−	−	+	−	+	+	−	−	−	+	+	−
11	+	−	+	+	−	−	+	+	−	−	−	−	+	−	+
12	−	−	−	+	+	−	+	+	−	+	+	+	+	+	−

Table 4. Systematic Supersaturated Designs for n = 18 (the first 54 columns)

Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	+	−	+	+	+	+	+	−	−	−	−	−	+	−	−	+	−	+
3	+	+	+	−	+	+	−	+	+	−	−	+	−	+	−	+	+	−
4	+	−	+	+	−	−	−	+	−	+	−	−	−	+	−	−	+	+
5	+	+	−	−	+	−	−	−	−	+	+	+	+	−	+	−	+	+
6	+	−	−	−	−	−	+	−	+	−	−	−	−	+	+	+	−	+
7	+	+	−	−	−	+	+	+	−	−	−	−	+	−	−	+	+	−
8	+	+	+	−	−	+	+	−	+	+	+	−	−	−	−	−	−	+
9	+	−	−	+	−	−	−	+	+	−	+	+	−	−	+	−	−	−
10	−	−	+	−	+	−	+	+	−	+	+	−	+	+	−	−	−	−
11	−	−	−	+	+	+	−	−	+	−	+	−	−	+	+	−	+	−
12	−	+	+	+	−	+	−	−	−	+	−	−	−	+	+	+	−	−
13	−	−	−	+	−	+	+	+	−	+	+	+	−	−	−	+	−	−
14	−	+	+	−	−	−	−	+	+	−	+	−	+	−	+	+	−	−
15	−	+	−	+	+	−	−	−	−	−	+	+	+	+	−	+	+	+
16	−	+	−	+	+	−	+	+	+	+	−	+	+	+	+	−	−	+
17	−	−	−	−	−	+	−	−	+	+	−	+	+	−	−	−	+	+
18	−	−	+	−	+	−	+	−	−	−	−	+	−	−	+	−	+	−

Run	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	+	+	+	+	+	−	−	−	+	+	−	−	+	+	+	+	+	−
3	+	−	−	−	+	+	+	+	−	−	+	−	−	−	+	+	−	+
4	+	+	−	+	−	+	+	−	+	+	+	+	−	−	−	+	−	−
5	+	+	+	−	+	+	+	−	+	−	−	−	−	+	−	−	−	−
6	+	−	−	−	+	−	+	+	−	+	+	+	+	+	−	−	−	−
7	−	−	−	−	−	+	−	+	−	+	−	−	−	−	+	−	+	−
8	−	−	+	+	−	−	−	+	+	−	+	−	−	+	−	−	−	+
9	−	−	+	+	+	+	−	−	−	+	−	+	−	+	+	−	−	+
10	+	−	+	−	+	+	−	−	−	−	+	−	+	−	−	−	+	−
11	+	+	−	+	−	+	−	+	+	−	−	−	−	+	−	+	+	−
12	−	+	+	−	−	+	−	−	−	−	−	+	+	+	+	−	+	+
13	+	−	−	−	−	−	+	−	+	−	+	+	−	+	−	+	+	+
14	−	+	−	+	−	−	+	−	−	+	+	−	+	−	+	+	+	+
15	−	−	+	+	−	−	−	+	−	+	+	−	+	−	−	+	−	−
16	−	−	−	−	−	−	+	+	+	+	−	+	−	−	+	−	+	−
17	−	+	−	+	+	−	−	−	−	−	−	+	+	−	−	−	−	+
18	−	+	+	−	+	−	+	+	+	−	−	+	+	−	+	+	−	+

Run	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	−	−	−	+	−	−	+	+	+	+	+	−	+	+	−	−	+	+
3	−	+	−	−	+	+	+	−	+	−	+	−	−	−	+	+	+	+
4	−	+	+	−	+	−	−	+	−	+	−	+	+	−	−	−	+	−
5	+	−	+	+	−	−	+	−	+	+	+	−	+	−	−	+	−	+
6	−	−	+	+	+	−	+	−	−	+	+	+	−	+	+	+	+	−
7	+	−	−	+	+	+	−	+	−	+	−	+	+	−	+	+	−	−
8	+	+	+	−	+	−	−	+	−	−	+	−	+	+	+	+	−	+
9	−	+	−	+	−	+	+	+	−	−	−	−	−	−	+	−	+	−
10	−	+	+	+	−	+	−	−	+	−	+	+	−	+	+	−	−	−
11	+	−	+	−	−	+	−	−	−	−	−	−	−	+	−	+	+	−
12	+	−	−	−	+	−	+	−	+	−	−	+	−	−	+	−	+	+
13	+	−	−	+	−	−	−	+	+	+	−	−	−	+	−	+	−	−
14	−	−	+	−	−	−	−	−	−	−	+	+	+	+	−	+	+	−
15	+	+	−	+	+	−	+	−	+	−	−	−	+	−	−	−	−	−
16	−	+	−	−	−	+	+	+	−	−	+	+	−	−	−	−	−	+
17	−	−	−	−	+	+	−	+	+	+	−	−	−	+	−	−	−	+
18	+	+	+	−	−	+	−	−	−	+	−	+	+	−	+	−	−	+

Table 5. Symmetric Supersaturated Designs for n = 24, (r = $\frac{1}{6}$)

Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	+	+	+	+	+	+	−	+	+	−	+	+	+	−	+	+	−
3	+	−	−	+	+	−	+	−	+	+	+	+	+	+	−	−	−
4	+	−	+	+	−	+	−	−	+	+	−	+	−	−	−	−	+
5	+	−	−	−	−	+	+	+	−	+	+	+	+	+	+	+	+
6	+	−	+	−	+	+	−	−	+	−	−	−	+	+	−	−	+
7	+	−	−	+	+	+	−	+	−	+	−	−	+	+	−	+	−
8	+	+	−	+	−	+	+	+	+	+	+	−	−	−	+	−	−
9	+	−	+	−	+	−	+	+	+	−	−	+	−	+	+	+	+
10	+	+	−	+	−	−	−	−	−	−	+	−	−	−	+	+	−
11	+	+	+	−	+	−	+	−	−	+	+	−	−	−	−	+	+
12	+	+	+	−	−	−	+	+	−	−	−	+	+	−	+	−	−
13	−	−	+	+	−	+	+	−	+	−	+	−	+	−	−	−	+
14	−	−	−	+	+	−	+	−	−	−	+	+	−	+	+	−	+
15	−	−	+	−	−	+	−	+	−	−	+	−	−	−	−	+	+
16	−	−	+	+	−	−	−	−	−	+	−	+	+	−	+	+	−
17	−	+	−	+	+	−	−	−	+	+	−	−	+	−	+	−	+
18	−	+	−	−	+	+	+	−	−	−	−	−	+	−	−	+	+
19	−	−	−	−	−	−	−	−	+	+	+	−	+	+	+	+	−
20	−	+	−	−	−	−	−	+	+	−	+	+	−	+	−	−	+
21	−	+	+	+	−	−	+	+	−	−	−	−	−	+	−	−	−
22	−	−	−	−	+	−	−	+	+	−	−	+	−	−	−	+	−
23	−	+	−	−	+	+	−	+	−	+	−	−	−	+	+	−	−
24	−	+	+	−	−	+	+	−	−	+	−	+	−	+	−	−	−

Run	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	+	+	−	−	−	−	+	+	+	+	+	+	−	+	+	+
3	−	−	−	+	+	−	+	+	+	−	−	+	+	+	−	+
4	−	+	−	+	+	+	−	−	+	+	+	−	−	−	+	+
5	−	+	+	−	+	−	+	−	−	−	−	−	−	−	+	+
6	+	−	+	−	−	−	+	+	+	+	−	−	−	+	−	−
7	+	−	−	−	+	+	−	−	−	+	+	+	+	−	−	−
8	+	−	+	+	−	−	−	+	−	+	−	+	−	−	−	−
9	+	+	−	+	−	+	+	−	−	−	+	−	−	−	−	−
10	−	+	+	+	+	−	+	−	+	−	+	−	+	+	−	−
11	+	+	+	−	−	+	−	−	−	+	−	−	+	+	+	+
12	+	−	+	+	+	+	−	+	+	−	+	−	−	+	+	+
13	+	+	−	−	+	−	−	+	−	−	−	−	+	−	+	−
14	−	−	+	−	−	+	−	+	+	+	+	−	+	−	−	+
15	−	−	−	+	+	+	+	+	−	+	+	+	−	+	−	+
16	+	−	+	+	−	+	+	+	−	−	−	+	+	−	−	−
17	+	+	−	−	+	−	+	−	−	−	+	+	−	+	−	+
18	−	−	+	+	−	−	−	−	+	−	+	+	−	−	+	−
19	−	−	−	−	−	+	−	−	+	+	−	−	−	+	+	−
20	+	−	+	−	+	+	+	−	+	−	−	+	+	−	+	−
21	−	−	−	+	−	−	+	−	−	+	−	−	−	−	+	+
22	−	+	+	−	+	−	−	+	−	+	−	+	+	+	+	−
23	+	+	−	+	−	+	−	+	+	−	−	−	+	−	−	+
24	−	+	−	−	−	−	−	−	−	−	+	+	+	+	−	−

k columns are selected. Overall, designs found here are better than designs given by Booth and Cox (1962) in the sense of ρ^2 criterion, and therefore $E(s^2)$ as given by Booth and Cox. For cases ($n = 12, k = 24$) and ($n = 24$, $k = 30$), the larger $r = 8/12$ and $r = 8/24$, respectively, should be compared to our designs, for which $r = 4/12$ and $4/24$, respectively. Tables 6, 7, and 8 also give additional details for designs for the cases $n = 12$, 18, and 24.

The designs given by Lin (1993), based on half fractions of Hadamard matrices, have a very nice mathematical structure but can only examine $N - 2$ factors in $N/2$ runs, where N is the order of Hadamard matrix used. It is thus recommended to use such designs when the number of factors is moderate. Moreover, these designs do not control the value of r, and, in fact, large values of r occur in some cases (see Lin 1993, table 2).

Table 6. Comparison With Supersaturated Designs Given by Booth and Cox (1962): n = 12 Runs

n	k	Design	Frequency of r: −2/3	−1/3	0	1/3	2/3	Number of pairs of columns	$\rho^2 \times 100$
12	16	B&C		30	67	23		120	4.90
		Lin		15	73	32			4.35
12	20	B&C		51	75	64		190	6.72
		Lin		27	103	60			5.09
12	24	B&C		68	105	101	2	276	7.13
		Lin		46	135	95			5.67
12	66	Lin		495	660	990		2145	7.69

For random balance designs, ρ^2 is equal to $1/(n-1)$—that is, .091, .059, and .043 for $n = 12$, 18, and 24, respectively. Note that ρ^2 in this case is independent of the number of factors, k. Judged by this criterion, we see that the systematic designs are substantially better than random balance designs. Indeed, these differences get smaller (at a fairly slow pace) when k increases. If we restrict $r \le \frac{1}{3}$ (r can be any number between 0 and 1 in random balance designs) and when the maximum numbers of k found, $k_{\max}$, are used, Tables 9 and 10 show that ρ^2 is .077 ($k = 66$), .052 ($k = 111$), and .04 ($k = 276$), for $n = 12$, 18, and 24, respectively. These values are still superior to those of random balance designs. Moreover, as pointed out by Booth and Cox (1962), because the distribution of r for random balance design is nearly normal, the frequency of occurrence of large values of r would be even more in favor of the systematic designs. Tables 9 and 10 show ρ^2 values for all designs listed in Tables 1 and 2, leading to the same conclusion. In Section 3, we discuss strategies for analyzing data from such systematic supersaturated designs.

3. DATA-ANALYSIS METHODS

Supersaturated designs are particularly useful at the preliminary stage of process understanding at which the number of factors is large and only a few factors are believed to have real effects. Because only a few factors are believed to be active, the basic idea here is to allow a slight bias among all estimated effects (controlled by the value r as will be explained) while the active factors remain identifiable. Thus to be identified any active factor *must* have an effect too large to be masked by the experimental error and the combined effects of unimportant factors. Confirmatory experiments, in general, are recommended to resolve ambiguity. An independent estimate of experimental error, if possible, would be very useful.

Several methods have been proposed to analyze the k effects, given only the $n(<k)$ observations from the random balance design contents, such as plots of responses y versus levels of each factor (e.g., see Anscombe 1959; Budne 1959; Satterthwaite 1959). These methods can also be applied here. Such quick methods provide an appealing straightforward comparison among factors, but it is questionable how much available information can be extracted using them. Thus it is useful to combine several of these methods to have a more satisfying effect. In what follows, three methods will be discussed—normal plotting, stepwise regression, and ridge regression. When the distribution of the size of effects follows a Pareto distribution (only 10–20% of the variables are important), as many practitioners believe, we note that all three methods are able to identify the important factors.

When supersaturated designs are used for screening purposes and the data analyst believes that only a few effects are important, the analysis strategy becomes an outlier-identification problem. The basic problem is to

Table 7. Comparison With Supersaturated Designs Given by Booth and Cox (1962): n = 18 Runs

n	k	Design	Frequency of r: −1/3	−1/9	1/9	1/3	Number of pairs of columns	$\rho^2 \times 100$
18	24	B&C	22	102	96	56	276	4.02
		Lin	17	100	123	36		3.13
18	30	B&C	41	137	144	113	435	4.73
		Lin	33	151	182	69		3.55
18	36	B&C	64	182	203	181	630	5.07
		Lin	55	205	259	111		3.84
18	111	Lin	712	1522	1883	1554	5671	5.32

Table 8. Comparison with Supersaturated Designs Given by Booth and Cox (1962): n = 24 Runs

n	*k*	*Design*	Frequency of r: −1/3	−1/6	0	1/6	1/3	*Number of pairs of columns*	$\rho^2 \times 100$
24	30	B&C	27	53	284	39	32	435	2.09
		Lin		90	111	234			2.06
24	31	Lin		98	116	251		465	2.09
24	276	Lin	1,885	4,838	7,413	6,974	4,541	25,651	4.00

distinguish between apparent effects that are due to noise and effects that are real. Plotting estimated effects on normal probability paper often provides an effective way of identifying real effects. The estimated effect for each factor x_i can be obtained via $\widehat{\beta_i} = \Sigma_j y_j(x_{ji} - \overline{x}_i)/\Sigma_j(x_{ji} - \overline{x}_i)^2$. If an effect is at least three times larger than the overall random-error standard deviation, such an effect can, in practice, virtually always be identified (see Sec. 5). By subtracting these significant effects from the responses, we can identify significant effects sequentially (until the normal plot shows no significant effects). If there are three or more active factors with similar magnitudes, however, the conclusion based on normal plotting can sometimes be misleading. This is mainly due to the fact that they are biased estimates and the correlations among estimated effects play an important role. Note that, for even n, assuming first-order model, we have $E(\widehat{\beta_i}) = \beta_i + \Sigma_{j \neq i} r_{ij} \beta_j$. To reduce the bias for fixed n, one would like to keep the r_{ij} as small as possible. (This is another justification of the design criterion ρ^2.) It becomes obvious that, when there are more than n active factors, supersaturated designs will only allow one to identify *relatively* large effects—a well-known phenomenon in using normal plot technique to identify significant factors.

Suppose that the response y depends on k candidate factors $x_1, x_2, \ldots, x_k$ with the first-order relationship $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$. Employing the usual linear model notation, $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{y}$ is an $n \times 1$ vector of observation, $\mathbf{X}$ is an $n \times (k+1)$ matrix whose jth row is of the form $(1, x_{1j}, x_{2j}, \ldots, x_{kj})$, and $\boldsymbol{\beta}$ is the $(k+1) \times 1$ vector of coefficients to be estimated. Note that for supersaturated designs n is less than k; thus $\mathbf{X}'\mathbf{X}$ is not of full rank. Moreover, the major interest here is to detect those β_i's that are indeed different from 0. One natural way to identify important variables would then be the forward selection procedure. This procedure starts with no variable in the model and first selects the x_i that has the highest correlation with $\mathbf{y}$. Subsequent selections are based on partial correlations, given the variables already selected.

The stepwise selection procedure provides an important modification of forward selection; namely, after each variable is entered through the partial F test, every variable already in the model is examined to check whether it should be removed (e.g., see Draper and Smith 1981, p. 307). The stepwise selection procedure is available in most statistical software packages. This is indeed a powerful and convenient method to identify active factors, provided that the interaction effects are, as assumed, relatively small. Most programs allow the analyst to select the criterion to enter a new variable and also the criterion to remove one. Of course, as with standard stepwise regression, one may reach a "false positive" identification (i.e., misclassify a null factor, and with many candidates to choose from, this risk can be substantial). Conservative significance levels may be called for (see also Westfall, Young, and Lin 1994). Note that screening experiments are usually conducted in

Table 9. The ρ^2 Values for Systematic Supersaturated Designs: $\rho^2 \times 100$ for Designs in Table 1 (even n)

Number of runs n	Maximum crossproduct (nr): 0	2	4	6	8	*Random balance design*
4	0					33.33
6	—	11.11				20.00
8	0					14.29
10	—	4.00				11.11
12	0	—	7.69			9.09
14	—	2.05	—	—		7.69
16	0	—	5.09	—		6.67
18	—	1.23	—	5.33		5.88
20	0	—	4.00	—	—	5.26
22	—	.83	—	3.97	—	4.76
24	0	—	2.11	—	4.00	4.35

Table 10. The ρ^2 Values for Systematic Supersaturated Designs: $\rho^2 \times 100$ for Designs in Table 2 (odd n)

Number of runs n	Maximum crossproduct (nr): 1	3	5	7	*Random balance design*
3	11.11				50.00
5	4.00				25.00
7	2.04				16.67
9	1.23	4.11			12.50
11	.83	4.10	8.45		10.00
13	.59	2.62	7.49		8.33
15	.44	.44	5.16	6.91	7.14
17	.35	1.41	6.34	5.45	6.25
19	.28	.28	3.16	5.32	5.56
21	.27	.27	4.66	4.12	5.00
23	.19	.19	1.96	4.15	4.55
25	.16	.16	2.29	3.14	4.17

the early stages of a study. In this case, "false negatives" (i.e., missing real active factors) are much more serious than false positives. The latter can generally be resolved by subsequent confirmatory runs. The typical significance level $\alpha = 5\%$ has been used in our data analysis.

Ridge regression is another popular method of handling singular $\mathbf{X}'\mathbf{X}$ matrices. Estimated effects are given by $(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1}\mathbf{Z}'\mathbf{y}$, where $\mathbf{I}$ is the $k \times k$ identity matrix and $\mathbf{Z}$ represents the "scaled" $\mathbf{X}$ matrix in correlation form (e.g., see Draper and Smith 1981, pp. 313–323). I have examined many possible situations with simulations, using the λ value suggested by Hoerl, Kennard, and Baldwin (1975), to identify large β_i's. The simulation results show that identification of large-effects-based ridge regression performs well when $\mathbf{X}'\mathbf{X}$ is nearly singular; namely, the rank of $\mathbf{X}'\mathbf{X}$ is close to the size of $\mathbf{X}'\mathbf{X}$. But for cases in which k is much larger than n, ridge regression does not seem to work well.

4. EXAMPLES

As mentioned, a supersaturated design is particularly useful to identify the few (p, say) dominant active factors. For given k and n, the smaller the number of active factors, the easier they can be identified. For example, when only one factor is active, most methods mentioned previously are capable of identifying that effect, as long as the effect is at least three times the overall "pooled" standard deviation (but see Ex. 3 for identifying smaller effects). The extreme case in which there are no active factors (meaning all apparent effects are due to random error) is also studied. It is found, based on the simulation results, that most methods work well in reaching the correct conclusion in such a situation. Alternatively, if the number of active factors is more than the number of runs (i.e., $p > n$), only a few relatively large effects will be identified. Three examples are given to illustrate the usefulness of supersaturated designs.

Example 1. Consider the differential-equation AIDS model discussed in the introduction as a typical screening problem in computer experiments. When the AIDS-prediction computer-model program was used by Curry, Tomick, and Yost, it was in its fledgling state. It contained little more than a basic heterosexual transmission component. It subsequently grew by adding dynamic routines to account for more transmission vectors and modifications of the earlier transmission vectors via new parameters. The mature version of the program had literally hundreds of differential equations and factors. Seitz (1992) incorporated the Curry, Tomick, and Yost experiences and used the updated equations governing those processes known to contribute to the HIV/AIDS pandemics to develop a concrete AIDS computer model called iwgAIDS (the State Department *I*nteragency *W*orking *G*roup *AIDS*). The final (1994) release version of iwgAIDS focuses more than in earlier versions on the known nonlinear age-distribution effects rather than AIDS prevalence and incidence. The program now includes all processes known to account for 5% of observed cases in the world. The inputs to the program presently consist of 138 variables with the output response y = AIDS incidence rate of 100,000 population. (For more details, see Seitz 1992.) Curry et al. generated large numbers of factors (150 as opposed to 138 here) by categorizing ages so that variables might have an instantiation for ages 15–29, 30–45, 45+, and so forth. Each of these counted as a different factor, but the core equations are the same now as they were earlier.

To investigate which factors are most critical, Seitz used the first 138 columns of the design $(n, k) = (24, 276)$ mentioned in Table 1 and obtained the following 24 responses (in order, read as row by row):

22.61	14.26	58.42	24.59	10.28	188.46	22.68	22.90
52.04	381.61	16.22	108.59	98.05	53.13	83.41	13.59
242.96	663.93	57.95	177.49	40.22	52.23	53.50	2,463.24

A stepwise selection procedure was then employed, and the results are summarized in Table 11. We see that 11 factors were identified as "active" factors with $R^2 = 99\%$. The first eight factors (with $R^2 = 91\%$) were selected for further study in Phase II. These eight active factors are as follows:

Factor **#118** = Contact rates for low-risk heterosexual males, age 15–36.67

Factor **#25** = Degree to which HIV-infected urban males introduce infections to their sexual contacts

Factor **#129** = Proportion of high-risk single heterosexual males, age 15–36.67

Factor **#13** = Circumcision HIV infectivity cofactor, males

Factor **#91** = Concurrent-partner rate for high-risk paired heterosexual males, age 15–36.67

Factor **#93** = Concurrent-partner rate for high-risk single females, age 15–36.67

Factor **#86** = Casual-partner-turnover rate for low-risk single heterosexual males, age 15–36.67

Factor **#76** = Casual-partner-turnover rate for high-risk paired females, age 15–36.67

The overall conclusion from this study indicates that the key variables for a generic rural area are (a) contact rate (**#118**), (b) proportion of single males (**#129**), and (c) concurrent-partner rate for paired and single males (**#91** and **#93**). The key variables for a generic urban area are (a) dry sex for males (**#25**), (b) circumcision HIV infectivity for males (**#13**), and (c) casual-partner-turnover rate for single males and for paired females (**#86** and **#76**). Note also that the test environment is strictly heterosexual, so being an infectious male means passing on the infection to a female. The expert judgment of those working with this model is that these sets of variables are quite reasonable.

Recall that the five most important variables identified by Curry et al. in their program are (1) risk of infection, males; (2) risk of infection, females; (3) contact rates,

Table 11. Stepwise Selection for the iwgAIDS Simulation Data

	Entering variables												
Step	118	25	129	13	91	93	86	76	1	101	54	σ	R^2
1	161.78											485.8	.108
	(1.63)												
2	227.53	197.26										455.8	.251
	(2.31)	**(2.00)**											
3	308.24	224.16	215.23									411.7	.418
	(3.23)	**(2.50)**	**(2.40)**										
4	336.19	298.69	224.55	−195.64								369.1	.555
	(3.90)	**(3.47)**	**(2.78)**	**(−2.43)**									
5	422.39	331.01	253.28	−206.42	197.53							315.2	.693
	(5.30)	**(4.44)**	**(3.64)**	**(−2.99)**	**(2.84)**								
6	456.96	411.67	264.80	−233.30	209.05	180.52						256.9	.807
	(6.94)	**(6.25)**	**(4.66)**	**(−4.10)**	**(3.68)**	**(3.18)**							
7	509.99	434.19	282.31	−240.81	176.53	188.03	150.10					206.7	.883
	(9.19)	**(8.13)**	**(6.13)**	**(−5.26)**	**(3.77)**	**(4.11)**	**(3.20)**						
8	509.49	434.19	282.31	−240.81	176.53	188.03	150.10	80.71				187.5	.909
	(10.13)	**(8.96)**	**(6.76)**	**(−5.80)**	**(4.15)**	**(4.53)**	**(3.53)**	**(2.11)**					
9	492.92	427.09	276.79	−238.44	129.19	185.66	160.36	119.11	−115.19			143.4	.951
	(12.71)	**(11.50)**	**(8.65)**	**(−7.50)**	**(3.66)**	**(5.84)**	**(4.91)**	**(3.80)**	**(−3.41)**				
10	473.45	371.53	270.30	−219.86	89.34	167.08	204.01	137.91	−171.60	110.59		100.9	.977
	(17.08)	**(12.48)**	**(11.97)**	**(−9.62)**	**(3.33)**	**(7.31)**	**(7.99)**	**(6.11)**	**(−6.18)**	**(3.91)**			
11	465.12	346.03	267.52	−182.90	129.85	158.64	166.61	135.63	−164.74	132.80	85.58	64.81	.991
	(25.97)	**(17.34)**	**(18.43)**	**(−10.82)**	**(6.64)**	**(10.71)**	**(9.03)**	**(9.34)**	**(−9.20)**	**(7.05)**	**(4.42)**		

NOTE: The numbers given in the table are estimated effects and their *t* ratios.

females; (4) contact rates, males; and (5) increased risk of infection due to cocirculating sexually transmitted diseases, females. The conclusion here has some similarity with that of Curry et al. (to be compared with factors **#118**, **#25**, and **#13**), but we use only 24 runs. Of course, these are only preliminary results from a very complicated computer experiment. The high R^2(= 91%) is somehow dubious. Confirmatory runs are strongly recommended. In Phase II, a 16-run (2_{IV}^{8-4}) confirmatory design is conducted as shown in Table 12 [while all factors except those identified as active in Phase I were set at their low levels ("−")].

Table 12. Design and Results of the Confirmatory Runs

	Design								Response
Run	#118	#25	#129	#13	#91	#93	#86	#76	y
1	−1	−1	−1	1	1	1	−1	−1	171.77
2	1	−1	−1	−1	−1	1	1	−1	118.88
3	−1	1	−1	−1	1	−1	1	−1	110.98
4	1	1	−1	1	−1	−1	−1	−1	187.59
5	−1	−1	1	1	−1	−1	1	−1	321.85
6	1	−1	1	−1	1	−1	−1	−1	254.58
7	−1	1	1	−1	−1	1	−1	−1	257.05
8	1	1	1	1	1	1	1	−1	337.99
9	−1	−1	−1	−1	−1	−1	1	1	267.35
10	1	−1	−1	1	1	−1	−1	1	323.63
11	−1	1	−1	1	−1	1	−1	1	332.84
12	1	1	−1	−1	1	1	1	1	263.61
13	−1	−1	1	−1	1	1	−1	1	120.36
14	1	−1	1	1	−1	1	1	1	185.58
15	−1	1	1	1	1	−1	1	1	190.69
16	1	1	1	−1	−1	−1	−1	1	106.20

Analysis based on these 16 confirmatory runs shows that factors **#129**, **#13**, **#118**, and **#86** are indeed the active factors with $R^2 = 99.8\%$ (see Tables 13 and 14), but factors **#25**, **#93**, and **#76** are moderately active and **#91** is not important at all. It seems that factor **#129** dominates (how much of the population falls toward the "greater sexual activity" end of the continuum), and the same holds for factor **#118** (the number of contacts per person at the "general sexual activity" end of continuum). The casual-partner-turnover rate, factor **#86**, is important because of variation in viral load. That is, it appears that subjects are more infectious right after contracting the virus and more infectious as their immune systems break down and frank AIDS ensues. Thus the casual-partner-turnover rate is a measure of how many different people a carrier might be in contact with during infectious phases (whose time duration is short, vis-à-vis the HIV period itself). If most of the sexually active people were not circumcised and if the lack of circumcision greatly

Table 13. Analysis of Variance for the Confirmatory Runs

Source	df	Sum square	Mean square	F ratio	p value
Regression	4	104,622	26,156	1391.43	.000
Error	11	207	19		
Total	15	104,829			

Table 14. The Fitted Model in Table 13

Predictor	*Coefficient*	*Standard deviation*	*t ratio*	*p value*
Const.	221.934	1.084	204.75	.000
x_{129}	72.928	1.084	67.28	.000
x_{13}	34.558	1.084	31.88	.000
x_{118}	4.351	1.084	4.01	.002
x_{86}	2.682	1.084	2.47	.031

NOTE: $s = 4.336$ and $R^2 = 99.8\%$.

magnified HIV transmission by making infectious males more infectious, this would explain why factor **#13** is dominant. The biology here is that a low-grade infection in uncircumcised males concentrates the immune system response, which might also increase the viral load of HIV in sexual body fluids. The data also suggest some potential interaction effects (e.g., between factors **#129** and **#76** and between factors **#118** and **#76**) and possible transformation. This needs further investigation. The subject-matter conclusions given previously, mainly due to Mr. Seitz, are believed to be useful and previously unknown.

The next two simulation examples show the typical performances of stepwise selection in the analysis of data from supersaturated designs. Because the "true" effects are known, one can be reassured that the conclusion from the procedure is correct. Research on analysis of supersaturated designs is encouraged. Example 2 shows that large effects (three times or more of the overall standard deviation) can be easily identified. Supersaturated designs are indeed reliable and economical. On the other hand, Example 3 shows some results for identifying active factors with only small effects (in the range of 2–3σ), which are known to be difficult in general.

Example 2. Suppose that we want to examine 60 factors using only 12 runs. The first-order model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{60} x_{60i} + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, \ldots, 12$, is assumed, where the 60 columns $x_1, x_2, \ldots, x_{60}$ are chosen to be the first 60 columns from Table 3. Without loss of generality, we assume that $\beta_0 = 30$ and $\sigma^2 = 1$. In all cases, 12 ε_i's were generated from the standard normal distribution and added to $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{60} x_{60i}$ for the specific values of β_i's to create the responses y_i. Based on these y_i's, we would like to identify the relatively large β_i's (i.e., "large" effects). The examples given here have $(n, k) = (12, 60)$. Certainly, the closer k is to n, the easier will be the identification of the active factors.

In this example, we simulate the case of five active factors, three with large magnitudes and two with moderate magnitudes. Apart from the five (randomly) selected β_i's, all other β_i's are set to 0. Figure 2(a) shows the five active factors to be identified. Factors **3**, **7**, and **42** are the three dominating factors, with effects 17σ, 24σ, and 15σ, respectively. Factors **10** and **23** are designated to be moderate and are less likely to be identified (both have effects of 3σ). All other β_i's are assumed to be 0 (i.e., all other factors have no effects). When the 12 ε_i's are added to $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{60} x_{60i}$, the resulting y_i's are

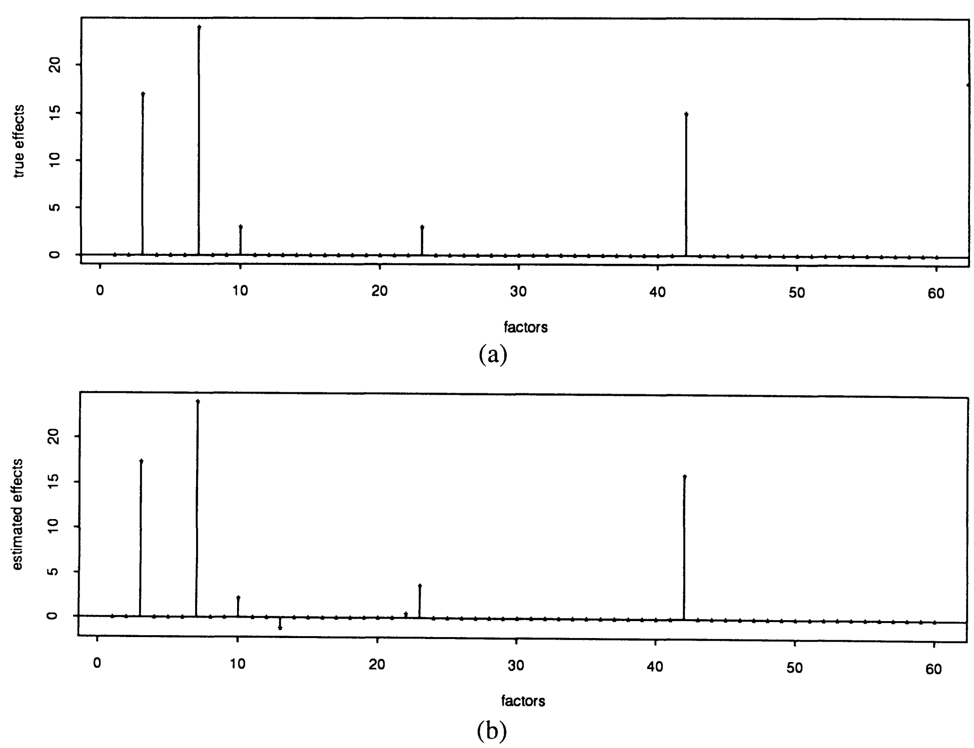

Figure 2. Simulated Example With (n, k) = (12, 60): (a) True Effects to be Identified; (b) Estimated Effects via Stepwise Regression.

Table 15. Successful Identification Rates From Simulation ($n = 12$ case)

Number of active* factors (p)	Number of factors under study (k)	Number of correctly identified factors 0	1	2	3	4
1	15	0	1.000			
	20	0	1.000			
	25	0	1.000			
	30	0	1.000			
2	15	0	0	1.000		
	20	0	0	1.000		
	25	.034	.003	.963		
	30	.068	.005	.927		
3	15	.027	.020	.004	.949	
	20	.071	.029	.001	.899	
	25	.084	.140	.015	.761	
	30	.138	.197	.014	.651	
4	15	.016	.132	.076	.008	.768
	20	.062	.256	.134	.006	.542
	25	.115	.293	.218	.024	.350
	30	.122	.319	.213	.027	.319

*All active factors are with the size in the range of 2–3σ.

92.88, 9.17, 79.24, 29.19, 4.19, 4.07, 7.31, 15.01, 14.48, 4.20, 93.66, and 7.18.

By applying the stepwise procedure, these five active effects are correctly identified with $R^2 = 99.9\%$. Factors **20** and **24** are also listed but with only a small effect to make the R^2 near 100%. These effects are plotted in Figure 2(b) to be compared with Figure 2(a), a very satisfactory result. Really large effects should not be difficult to identify, in general. From this example, we see that supersaturated design can also identify factors with moderate effects. Similar examples with nonnormally distributed errors [specifically, exponential, uniform, and Cauchy distributions; see Lin (1991)] were also simulated, and supersaturated design and stepwise regression methods work successfully.

Example 3. To identify the few "active" factors with absolute effects in the range of 2–3σ from many factors is, in general, difficult. In this example, the typical performance of the stepwise selection procedure is shown. Take the 12-run case (Table 3) as a typical example. Consider once again the first-order model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma = 1)$, $i = 1, 2, \ldots, 12$. Given k and p, sizes of effect in the range 2–3σ were uniformly generated and assigned to the p randomly selected β_i's, while all other β_i's were set to 0. The ε_i's were generated to create y_i's as in Example 2. The stepwise selection procedure was then performed. For each specific value of k and p, 2,000 cases were simulated. The successful identification rates (= percentages of cases that correctly identify those active factors) were tabulated in Table 15 for $p = 1, 2, 3, 4$ and $k = 15, 20, 25, 30$.

For the cases $p = 1, 2$, and 3, the stepwise selection procedure works reasonably well, even in such difficult cases as these. For $p = 4$, the successful identification rates are only moderate and get worse as k increases. This seems to suggest that besides those large effects that can be easily identified, as in Example 2, a stepwise selection procedure is only capable of detecting a few effects in the range 2–3σ. Thus, if a process has many factors with effects of size 2–3σ, the effect sparsity assumption that there are a *few dominant active* factors is no longer valid. A supersaturated design may not be an adequate experimental design in such a situation. Most saturated main-effect designs will also fail to work and higher experimental costs seem inevitable.

5. CONCLUDING REMARKS

Consider a two-level k-factor design in n observations with maximum correlation r. Given any two of the quantities (n, k, or r), this article discusses what value can be achieved for the third quantity. Apart from the property that $nr_{ij} + n \equiv 0 \pmod 4$, no other theoretical implication is currently available. The related theoretical implications need more investigation, in particular for $n < k$, whose designs as given here are apparently new. Furthermore, the largest k found here are not guaranteed to be maximum, but they, nevertheless, are sufficiently large for practical use.

Booth and Cox (1962) showed that if k effects are included for analysis with design vectors selected from the design matrix, then the average variance of an estimated effect (β) is approximately $\sigma^2[1 + (k-1)\rho^2]/n$, where σ^2 is the error variance. They also pointed out that this quantity is likely to be seriously underestimated unless the nonzero true effects are all very large. The power curves for various k ($k = 11, 15, 20, 30, 40, 50, 60$) and $n = 12$ when the true model is $E(y) = \beta_0 + \beta_1 x_1$ are given in Figure 3. Note that the curve for $k = 11$ is a saturated orthogonal case. When the ratio β/σ is larger than 2, the power is close to 1 for all k. In other words, when the largest effect is significantly different from 0 and the amount of this effect is three times (or more) larger than the overall standard deviation obtained by treating all other factors as null, such an effect can always be identified.

We see that average variance increases as k increases; namely, certain true effects that can be identified for a moderate k may no longer be recognized as the number of factors, k, increases. The obvious conclusion from this is that (for given n) the larger the number of factors under investigation, the larger an effect must be to be identified. On the other hand, when n and k are specified, this average variance only depends on ρ^2. An optimal design in this sense will seek to minimize ρ^2, as I have done by sorting my design columns.

Designs given here are constructed for the degree of nonorthogonality that the experimenter is willing to accept (the criterion of r). For two designs with the same r, I prefer the one in which the number of r is a minimum (the criterion of ρ^2). Such an idea was originated by Booth and Cox (1962). Tukey (1959), however, suggested the concept of *elongation* as an indicator of quality of

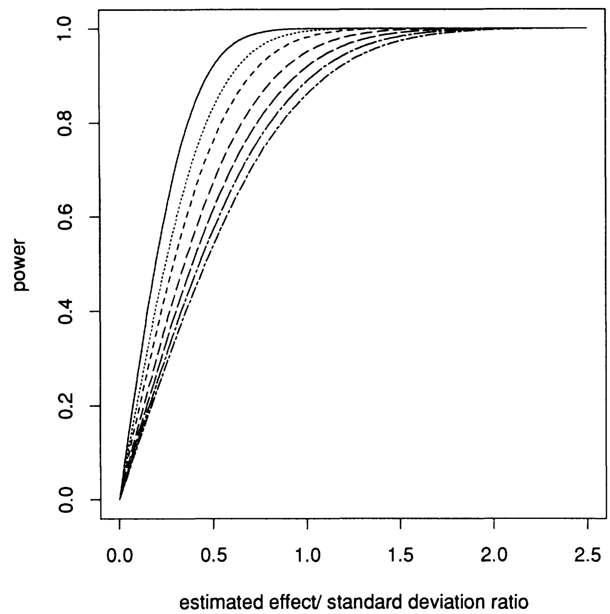

Figure 3. Power Curves for Various k, N = 12 Case With One Significant Factor: ———, k = 11; · · · ·, k = 15; - - - -, k = 20; - - - ·, k = 30; - - -, k = 40; — - —, k = 50; - - - - -, k = 60.

confounding. Specifically, for any two columns of signs, the elongation depends on the sum of fourth powers of cosines of angles, and this is another possible criterion to use to construct a supersaturated design. For equal-occurrence classes (the occurrences of high and low levels are differed by at most 1), the difference between ρ^2 and elongation is not substantial.

If the assumption of Pareto distribution for the size of effects is true, the supersaturated design can easily identify those few largest effects. In general, large effects should not be difficult to identify using any data-analysis methods. In this case, a supersaturated design can save considerable costs. To detect effects with magnitudes in the range of 2–3σ in the presence of many factors, however, is a very difficult task. Moreover, to study many factors in a relatively few runs, the data-analysis methods are very sensitive to outlying observations—a difficulty that the practitioner should bear in mind. Common advice to practitioners for such an outlier problem is to replicate the experiment with the factors combination that yields the maximal (or minimal) response. Such an "outlier confirmatory" run, if possible, is recommended.

ACKNOWLEDGMENTS

I am grateful to the editor, an associate editor, and two referees for their extensive and constructive comments and to Lynn Landry (University of Tennessee) and Stan Young (Glaxo), who gave many comments on early versions of this article that have led to major improvements in the presentation. Thanks also go to Steven Seitz, University of Illinois, who kindly provided the simulation AIDS model (iwgAIDS), evaluated all necessary computations, and thoroughly commented on the data-analysis results. Without his help, Example 1 in Section 4 would not be possible. This research was partially supported by a Professional Development Award, the University of Tennessee; a Visiting Scientist Fellowship, Mathematical Sciences Department, IBM Watson Research Center; and the National Science Foundation via Grant DMS-9204007.

[Received March 1992. Revised October 1994.]

REFERENCES

Anscombe, F. J. (1959), "Quick Analysis Methods for Random Screening Experiments," *Technometrics*, 1, 195–209.

Barnett, E. H., and Hurwitz, A. M. (1990), "Near Orthogonality: Systematic Supersaturated Designs—Extension and Application," in *Transactions, The 46th Annual Rochester ASQC Conference, University of Rochester*, Milwaukee: American Society for Quality Control, pp. 27–42.

Booth, K. H. V., and Cox, D. R. (1962), "Some Systematic Supersaturated Experimental Designs," *Technometrics*, 4, 489–495.

Box, G. E. P. (1959), Discussion of "Random Balance Experimentation," by F. E. Satterthwaite and "The Application of Random Balance Designs," by T. A. Budne, *Technometrics*, 1, 174–180.

Budne, T. A. (1959), "The Application of Random Balance Designs," *Technometrics*, 1, 139–155; response to discussion, 192–193.

Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis*, New York: John Wiley.

Hedayat, A., and Wallis, W. D. (1978), "Hadamard Matrices and Their Applications," *The Annals of Statistics*, 6, 1184–1238.

Hoerl, A. E., Kennard, R. W., and Baldwin, K. F. (1975), "Ridge Regression: Some Simulations," *Communications in Statistics*, 4, 105–123.

Hunter, J. S. (1959), Discussion of "Random Balance Experimentation," by F. E. Satterthwaite and "The Application of Random Balance Designs," by T. A. Budne, *Technometrics*, 1, 180–184.

Kempthorne, O. (1959), Discussion of "Random Balance Experimentation," by F. E. Satterthwaite and "The Application of Random Balance Designs," by T. A. Budne, *Technometrics*, 1, 159–166.

Lin, D. K. J. (1991), "Systematic Supersaturated Designs," Working Paper 264, University of Tennessee, College of Business Administration.

——— (1993), "A New Class of Supersaturated Designs," *Technometrics*, 35, 28–31.

Plackett, R. L., and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 303–325.

Rosenberger, J. L., and Smith, D. E. (1984), "Interruptible Supersaturated Two-Level Designs," *Communications in Statistics*, 13, 599–609.

Satterthwaite, F. E. (1959), "Random Balance Experimentation," *Technometrics*, 1, 111–137; response to discussion, 184–192.

Schmidt, S. R., and Launsby, R. G. (1991), *Understanding Industrial Designed Experiments*, Colorado Springs: Air Academy Press.

Seitz, S. T. (1992), *iwgAIDS User's Manual*, Urbana; Merriam Laboratory, University of Illinois.

Tukey, J. W. (1959), Discussion of "Random Balance Experimentation," by F. E. Satterthwaite and "The Application of Random Balance Designs," by T. A. Budne, *Technometrics*, 1, 166–174.

Voelkel, J. G. (1990), "Supersaturated Designs in Quality Improvement," unpublished paper presented at the 34th Annual Fall Technical Conference, Richmond, VA., October 18.

Watson, G. S. (1961), "A Study of the Group Screening Methods," *Technometrics*, 3, 371–388.

Westfall, P. H., Young, S. S., and Lin, D. K. J. (1994), "Inference Issues in the Analysis of Supersaturated Designs," unpublished manuscript.

Youden, W. J. (1959), Discussion of "Random Balance Experimentation," by F. E. Satterthwaite and "The Application of Random Balancex Designs," by T. A. Budne, *Technometrics*, 1, 157–159.