SCREENING PROCESS FACTORS IN THE PRESENCE OF INTERACTIONS

Mark J. Anderson
Principal
Stat-Ease, Inc.
Minneapolis, MN 55413
Mark@StatEase.com
www.StatEase.com

Patrick J. Whitcomb
Consultant
Stat-Ease, Inc.
Minneapolis, MN 55413
Pat@StatEase.com
www.StatEase.com

SUMMARY

This talk introduces a new, more efficient type of fractional two-level factorial design of experiments (DOE) tailored for screening of process factors. These designs are referred to as "Min Res IV" because they require a minimal number of factor combinations (runs) to resolve main effects from two-factor interactions (resolution IV). We set the stage for these state-of-the-art minimum-run medium-resolution designs by first reviewing traditional screening approaches – standard two-level fractional factorials (2^{k-p}) and Plackett-Burman designs. To provide an element of realism, we will show how well each screening design identifies main effects and two-factor interactions (2fi's) affecting a hypothetical machine operated by computer numerical control (CNC).

Our ultimate objective is to equip quality professionals with cost-effective statistical tools for making breakthroughs in process efficiency and product quality.

INTRODUCTION

The traditional approach to experimentation requires changing only one factor at a time (OFAT) in a process. However, the OFAT approach cannot detect interactions of factors, a likely occurrence in many industrial processes (Box, 1990). Therefore, quality professionals seeking breakthrough improvements prefer multifactor testing via statistically-based methods known as design of experiments (DOE). The strategy of DOE is flowcharted in Figure 1.

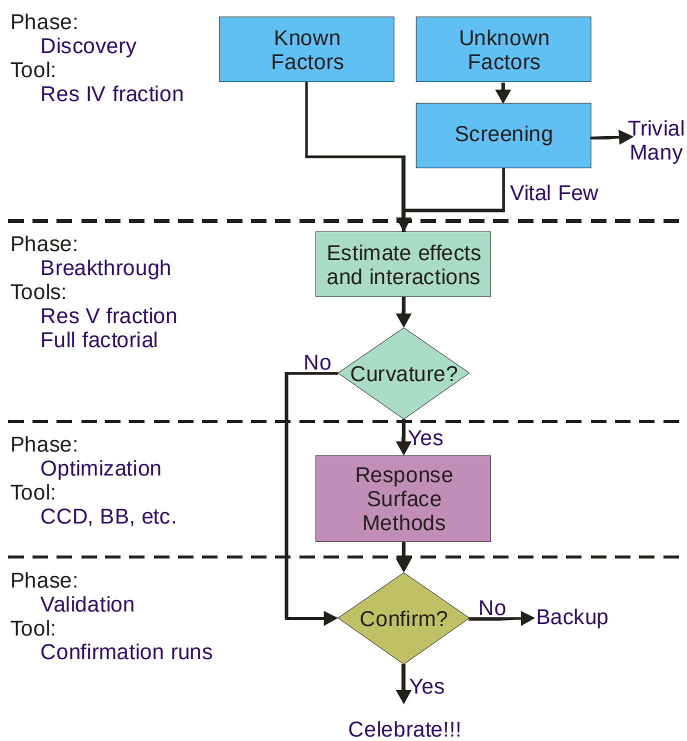

Figure 1: Strategy of Experimentation

It breaks down into four main phases:

1. Discovery: Use screening designs to separate the vital few factors from the trivial many, but temporarily withhold variables known to have an effect.
2. Breakthrough: Follow up by doing an in-depth investigation of the surviving factors, including those previously withheld.
3. Optimization: Generate a "response surface" map and move the process to the optimum location.
4. Validation: Perform confirmatory runs to ensure that the models produce accurate predictions for all responses of interest and that the process remains robust under field conditions.

We will devote this talk to factor screening. The resolution IV two-level, fractional factorial designs employed at this discovery phase are relatively simple, yet incredibly powerful. Mastering this level of DOE puts you far ahead of most technical professionals and paves the way for follow-up breakthroughs and eventually optimization via response surface methods (RSM), such as the central composite design (CCD) and Box-Behnken (BB).

Two-level factorial designs make the best screening tools. If performed properly, they reveal the vital few factors that significantly affect your process. To save on costly runs, experimenters usually perform only a fraction of all the possible combinations. Regardless of how you do it, cutting out runs reduces the ability of the design to resolve all possible effects, specifically the higher-order interactions. Carrying this to an extreme, unwary experimenters fractionate their designs to the point of factor-saturation, such as testing seven factors in only eight runs. Often overlooked is that such designs alias main effects with two-factor interactions (2fi's), achieving only resolution III in statistical terms. Resolution III designs can produce significant improvements, but it's like kicking your PC (or slapping the monitor) to make it work: You won't discover what really caused the failure. Because of their ability to more clearly reveal main effects, resolution IV designs work much better than resolution III for screening purposes. Two-factor interactions remain aliased with each other, but these can be resolved via follow-up experiments (Anderson and Whitcomb, 2001).

The two most popular options for fractional two-level designs are:

- Standard "2^{k-p}"s, where k refers to the number of factors and p is the fraction,
- Plackett-Burman (PB): designs with a multiples of four (4, 8, 12, …).

Layouts for 2^{k-p} fractional designs can be found in DOE textbooks (Anderson and Whitcomb, 2000). These standard two-level designs, which we recommend, provide the choice of 4, 8, 16, 32 or more runs, but only to the power of two. Resolution IV options are readily available from the standard catalog of 2^{k-p} fractional designs. The second option noted above, Plackett-Burman (1946), are two-level designs done in multiples of four runs (as opposed to powers of two). The 12, 20, 24, and 28-run PB designs are of particular interest because they fill gaps in the standard designs. Unfortunately, these particular Plackett-Burman designs have very messy alias structures. For example, the 11 factor in 12 run choice, which is very popular, causes each main effect to be partially aliased with 45 two-factor interactions, thus achieving only resolution III. You can get away with this if absolutely no interactions are present, but this is a very dangerous assumption in our opinion. Of course, you could cut down the number of factors in a given PB design, but the alias structure may not be optimal.

We will see how these two classical screening designs, 2^{k-p} and PB, do on a test case, but first let's introduce a new class of designs, called "Min Res IV," which only require 2k runs (10 runs for k=5 factors, 12 runs for 6 factors, 14 for 7, etc.).

A NEW CLASS OF SCREENING DESIGNS: MINIMUM RUNS WITH RESOLUTION IV

Fractions of standard two-level designs are restricted to negative powers of two (1/2, 1/4, 1/8, etc.). However, it is possible to do "irregular" fractions, that is; not powers of two, and still maintain a relatively high resolution. A prime example of this is the three-quarter (3/4) replication for four factors that produces in essence a resolution V design with only 12 runs. It can be created by identifying the standard quarter-fraction, and then selecting two more quarter-fractions. The resulting matrix estimates all main effects and two-factor interactions aliased only by three-factor or higher interactions, thus making it a viable alternative to the 16-run full factorial.

Aided by a computer-aided algorithm based on a statistical criterion called "D-optimality," the catalog of high-resolution irregular fractions has recently been greatly expanded (Oehlert and Whitcomb, 2002). These designs contain equal numbers of low and high levels for each factor, thus they are described as being "equireplicated." They offer resolution V at or near (+1) minimum runs. In a similar manner, a class of equireplicated irregular fractions can be constructed to be resolution IV, thus making them extremely useful for screening purposes. An example of one such 'Min Res IV' design will be unveiled for the first time in the discussion below. (Templates for other designs in this new class are presented in the Appendix.) As shown in Table 1, the Min Res IV designs compare favorably to the classical alternatives on the basis of required experimental runs.

Factors (k)	# Runs 2^{k-p} Res IV	# Runs Min Res IV	# Runs 2^{k-p} Res III	# Runs PB Res III
5	NA	10	8 (¼)	12
6	16 (¼)	12	8 ($^1/_8$)	12
7	16 ($^1/_8$)	14	8 ($^1/_{16}$)	12
8	16 ($^1/_{16}$)	16	NA	12
9	32 ($^1/_{16}$)	18	16 ($^1/_{32}$)	12
10	32 ($^1/_{32}$)	20	16 ($^1/_{64}$)	12

Table 1: Runs required for screening designs for up to 10 factors

Figure 2 illustrates the layout for the Min Res IV design on five factors. It is a subset of points (shown big and bold below) from the full two-level (2^5) factorial candidate set (numbers inscribed within these points indicate the 2^k standard order).

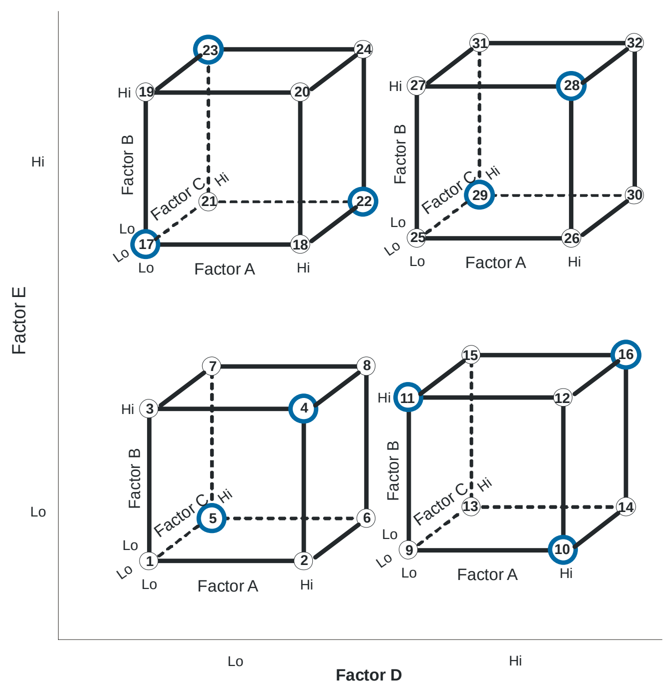

Figure 2: Geometry for Min-Res IV design on five factors

Notice that this design provides an equal number of points at the low versus high levels ('equireplication') along each of the five factor dimensions. For example, there are five points in the two lower cubes versus the ones above – thus factor E is equireplicated. The 10-point design we've illustrated above aliases each of the five main effects only with 3fi's, thus achieving resolution IV. It produces estimates for four 2fi's, but these are aliased with the other six two-factor interactions, thus a follow-up experiment would be required if any 2fi's come out significant. This eventuality is provided for in the strategy of experimentation we outlined previously.

HOW VARIOUS SCREENING DESIGNS PERFORM IN THE PRESENCE OF TWO-FACTOR INTERACTIONS

Let's see how well each screening design identifies main effects and two-factor interactions (2fi's) affecting a hypothetical machine operated by computer numerical control (CNC). (Disclaimer: The response data comes from a simulation designed to generate a thought-provoking analysis, so any resemblance to a real manufacturing process is purely coincidental.) Assume that you encounter a problem with part dimensions on this CNC machine. From prior experience, you know that some subset of these seven factors is likely to be important:

A. x-Axis shift

B. y-Axis shift

C. z-Axis shift

D. Spindle speed

E. Tool vendor

F. Feed rate

G. Fixture height

George Box says in regard to predicting a response based on DOE that "Only God knows the model" and "All models are wrong, but some are useful" (Box, 1976). In this case we are the creators of a predictive model for the part dimension. As shown below, it includes the main effects from four factors (A, C, D & G) and two interactions (AG & CD).

Dimension = 20 – 0.5A + 2.5C + 1D + 1.5G +2.0AG – 1.5CD

We set this up in Design-Expert® software (Helseth, 2000) as a simulation with all other effects ignored (coefficients of zero) and no experimental error (zero standard deviation). Thus any failures to detect effects will be due to the design resolution, not the random error.

Let's first try the seven-factor in eight-run fractional option for a standard two-level design (2^{7-4}). This is a saturated resolution III design, so it offers a maximum number of factors in a minimal number of runs, but at a price: Each main effect is confounded with many two-factor interactions. Figure 3 shows the effects revealed via a half-normal plot of the effects. This is a standard tool for screening the vital few effects that stand out to the right on the x-axis scale of absolute response.

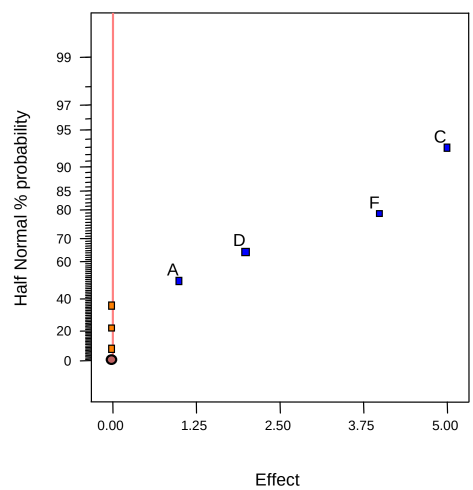

Figure 3: Half-normal plot of effects from Res III 2^{7-4} design

Notice that the insignificant effects (presumably the 'trivial many') line up at zero because there's no error in this simulation. Knowing the true model and the aliases from the design matrix, we can reconcile the observed coefficients, which for a two-level design are simply one-half of the calculated effects. For example, note below that the coefficient for A has an absolute magnitude of 0.5, whereas the effect plotted above is 1.0. These coefficients are arbitrarily labeled – for example, "F" might actually represent the interaction AG, but since main effects tend to occur more often than interactions, we chose the former rather than the latter.

"A" = A = –0.5

"B" = 0

"C" = C = +2.5

"D" = D = +1.0

"E" = 0

"F" = AG = +2.0

"G" = G + CD = +1.5 –1.5 = 0.0

Unfortunately, our assumption that "F" was F was wrong – as we know from the true model (simulated in the software), it's actually AG. Even worse, the true effect of CD cancelled the impact of factor G and thus neither was revealed. That's scary! This case illustrates that Resolution III designs give misleading results in the presence of interactions. Of course you can get around this flaw in experimental design by assuming that your process, product or system exhibits only main effects. However, we think it's unsafe to take such a big leap in faith. For example, the 2^{7-4} resolution III design on the CNC machine missed a significant main effect (G) and incorrectly picked a non-significant effect (F)! This is not a good way to do screening because it's guaranteed to give the wrong answer if interactions exist.

Now let's move on to another design choice (not ours!) for screening: Plackett-Burman (PB). The developers of this fractional two-level factorial design were British statisticians. During World War II they developed experimental plans for testing proximity fuses on bombs. These were simple mechanical devices where no interactions were expected. The PB designs became popular for industrial process development because they identify up to N-1 factors in N experiments, for example; 11 factors in 12 runs. Also, they grow by factors of four rather than powers of two, thus offering more flexibility for dealing with limitations in experimental resources. (Note: the "geometric" PB designs (when N is a power of 2) are equivalent to 2^{k-p} fractional factorials. In such cases, we advise using the latter rather than the former.) Before you get too excited about the possibilities of PB's, look at the aliasing for one main effect, A, in the 7-factor, 12-run design:

"A" = A – 0.333*BC + 0.333*BD + 0.333*BE – 0.333*BF – 0.333*BG – 0.333*CD – 0.333*CE + 0.333*CF – 0.333*CG – 0.333*DE – 0.333*DF – 0.333*DG + 0.333*EF + 0.333*EG – 0.333*FG

As you can see, each main effect in the design is partially aliased with 15 two-factor interactions! For this reason alone, we advise that you avoid use of Plackett-Burman designs. Nevertheless, let's put this 12-run PB to the test on our CNC case study. We've found that in cases like this, where you do not saturate the design matrix with factors, it's best to insert 'dummy' factors to minimize problems caused by partial aliasing. (Our reasons for doing this are technical: PB main effects are orthogonal to each other but not to the two-factor interactions, therefore, if the matrix is not fleshed out with dummy factors, the alias coefficients become greater than one!) Here is our new list of factors:

- A-G: Real factors for CNC machine,
- H, J, K, L: dummy factors

Note that the letter "I" does not appear on the list above because it symbolizes the identity of the design matrix. For all practical purposes you can think of this as a placeholder for the model's intercept – generally calculated from the overall mean response (or grand average).

Figure 4 shows the half-normal plot for the CNC machine simulation we previously studied with the resolution III standard fractional factorial.

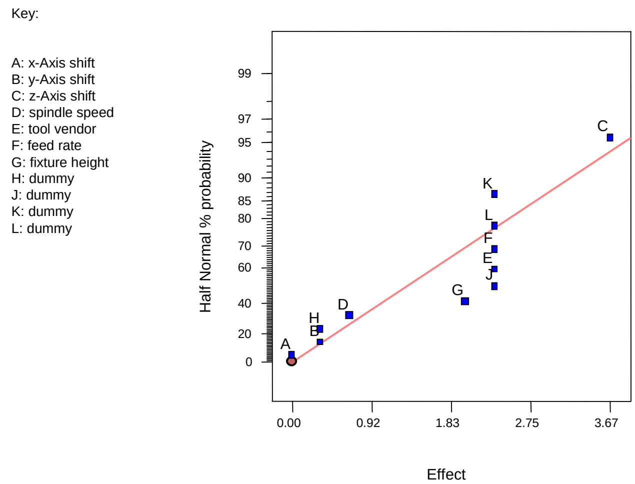

Figure 4: Half-normal plot of effects from Plackett-Burman design

Notice that with the PB design the J, K and L factors evidently create some of the bigger effects. However, these are all dummy factors! Obviously something is not right about this analysis. The problem is the partial aliasing, as shown in the following detail on the effect calculations:

"A" = A – 0.333*CD = –0.5 – 0.333(–1.5) = 0

"B" = –0.333*AG –0.333*CD = –0.333*(2.0) – 0.333*(–1.5) = –0.167

"C" = C – 0.333*AG = +2.5 – 0.333(2.0) = 1.833

"D" = D – 0.333*AG = +1.0 – 0.333*(2.0) = 0.333

"E" = +0.333*AG – 0.333*CD = +0.333*(2.0) – 0.333*(–1.5) = 1.167

"F" = –0.333*AG + 0.333*CD = –0.333*(2.0) + 0.333*(–1.5) = –1.167

"G" = G + 0.333*CD = +1.5 + 0.333*(–1.5) = 1.000

"H" = –0.333*AG – 0.333*CD = –0.333*(2.0) – 0.333*(–1.5) = –0.167

"J" = +0.333*AG – 0.333*CD = +0.333*(2.0) – 0.333*(–1.5) = 1.167

"K" = –0.333*AG + 0.333*CD = –0.333*(2.0) + 0.333*(–1.5) = –1.167

"L" = +0.333*AG – 0.333*CD = +0.333*(2.0) – 0.333*(–1.5) = 1.167

Only effects larger than the largest dummy can be identified as main effects. In this case factor C is all you get from the PB design. What a waste! At this point all you can do is try a fold over, which adds 12 more runs and results in a Res IV design. But by going to a 16-run standard factorial to begin with, you could also get Res IV, but with eight fewer runs than the Plackett-Burman with the follow-up fold-over design.

Here's a summary of how well the PB design did on our hypothetical CNC machine:

- Correctly selected one main effect,
- Missed three significant main effects!

We advise that you not use PBs if you suspect interactions. If you do choose one of these designs, bigger is better because the size of the coefficient in the partial aliasing decreases as the number of runs increases. Also, in cases like the CNC with less than the number of factors needed to saturate the PB design matrix, we suggest you make use of dummy factors.

It will be enlightening to see how we do with a better design – the resolution IV standard 2^{k-p} fractional two-level factorial. This experimental matrix provides estimates of main effects un-aliased with two factor interactions (for example:

“A” = A + BCE + BFG + CDG + DEF), but these 2fi’s remain aliased with each other (such as: “AB” = AB + CE + FG). As can be seen in the half-normal plot below, this resolution IV design correctly identifies all main effects (recall that the model is a function of A, C, D, G main effects, plus the AG and CD two-factor interactions), thus accomplishing the mission of screening.

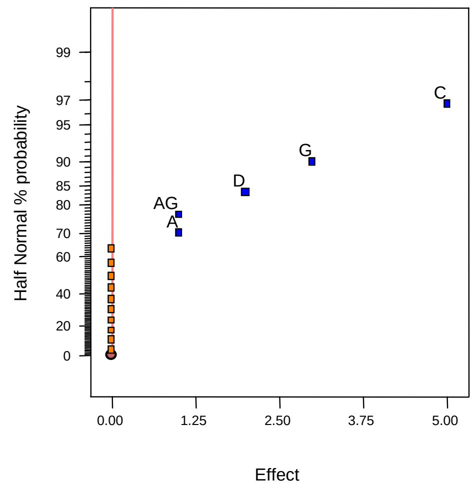

Figure 5: Half-normal plot of effects from Res IV 2^{k-p} design

In the presence of two-factor interactions, only designs of resolution IV (or higher) can ensure accurate screening. Whitcomb and Oehlert (ibid.) recognized that the minimum number of runs for resolution IV design is only two times the number of factors (runs = 2k). They realized that this can offer quite a savings when compared to a regular resolution IV 2^{k-p} fraction. For example, 32 runs are required with 9 through 16 factors to obtain a resolution IV regular fraction. For the same range of factors, the new “Min Res IV” designs from Whitcomb and Oehlert require only 18 to 32 runs, depending on the number of factors. Thus, for nine factors this design provides a savings of 14 runs (32 – 18), but none for 16 factors.

Applying the Min Res IV design for seven factors (layout provided in Appendix) to the CNC machine, we get the following results.

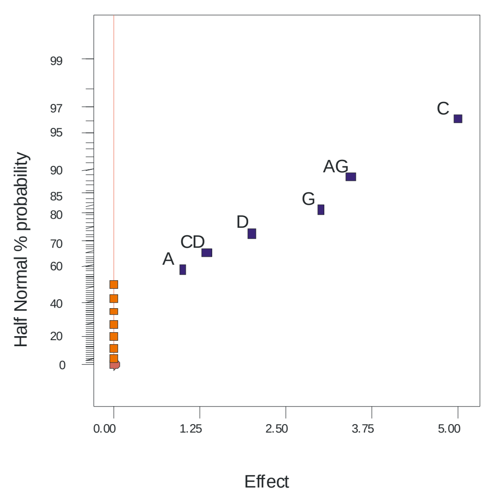

Figure 6: Half-normal plot of effects from Min Res IV design

The min Res IV design correctly selected all main effects! This is to be expected due to the favorable alias structure, for example that shown for the main effect of A (typical):

“A” = A + 0.333*ABC − 0.111*ABD + 0.556*ABE + 0.333*ABF
+ 0.556*ABG − 0.111*ACD + 0.111*ACE + 0.778*ACF
+ 0.333*ACG − 0.111*ADE − 0.111*ADF − 0.111*ADG
+ 0.111*AEF + 0.556*AEG + 0.333*AFG + 0.667*BCD
− 0.222*BCE + 0.222*BCF − 0.444*BCG + 0.444*BDE
+ 0.667*BDF + 0.444*BDG − 0.222*BEF + 0.444*BEG
− 0.444*BFG + 0.889*CDE + 0.222*CDF + 0.667*CDG
− 0.222*CEF − 0.222*CEG + 0.222*CFG + 0.889*DEF
+ 0.444*DEG + 0.667*DFG − 0.222*EFG

Furthermore it turns out that if a two-factor interaction is large enough relative to other 2fi’s (and to the experimental error), it can be correctly identified in some cases. This is an artifact of their partial aliasing, such as that shown below (typical):

“AB” = AB + 0.333*BC − 0.778*BD − 0.111*BE + 0.333*BF
− 0.111*BG+ 0.222*CD − 0.222*CE + 0.444*CF
− 0.667*CG + 0.222*DE + 0.222*DF + 0.222*DG
− 0.222*EF + 0.889*EG − 0.667*FG

However we advise that you use caution accepting any 2fi’s as being an accurate reflection of reality. Always do follow-up studies for this, or any other, type of screening design. Do not use the models from screening for predictive purposes other than to establish a starting point for more in-depth (higher resolution) design of experiments.

CONCLUSION

In this talk we’ve focused on the screening stage of experimentation. In the presence of two-factor interactions, only designs of resolution IV (or higher) can ensure accurate screening. If limited by time or other experimental resources, in most cases (other than eight factors) the new Min Res IV option offers savings in runs over the equivalent 2^{k-p} (standard two-level fractional factorial design). What you do as a result depends on the statistical analysis of the effects:

- Scenario 1 - Nothing significant: Look for other factors that affect your response(s).
- Scenario 2 - Only main effects significant: Change these factors to their best levels.
- Scenario 3 -Two-factor interaction(s) significant: De-alias by performing a semi-fold (detailed in Anderson and Whitcomb, 2001).

By following this strategy you will increase your odds of uncovering breakthrough main effects and interactions at a relatively minimal cost in experimental runs.

REFERENCES

Anderson, M. and P. Whitcomb. 2000. *DOE Simplified, Practical Tools for Experimentation.* Portland, Oregon: Productivity.
Anderson, M. and P. Whitcomb. 2001. How to Save Runs, Yet Reveal Breakthrough Interactions, by Doing Only a Semifoldover on Medium-Resolution Screening Designs. *Proceedings of 55th Annual Quality Congress.* Milwaukee: American Society of Quality.
Box, G.E.P. 1990. George’s Column: Do Interactions Matter? *Quality Engineering.* Vol. 2, No. 3, p365.
Box, G.E.P. 1976. Science and Statistics. *Journal of American Statistical Association*, 71, p791-799.
Helseth, T. J., et al. 2004. *Design-Expert Software, Version 7* (pre-release). Minneapolis: Stat-Ease, Inc.
Li, W and Jeff Wu. 1997. Columnwise-Pairwise Algorithms with Applications to the Construction of Supersaturated Designs. *Technometrics* 39, 1781-179.
Oehlert, G. and P. Whitcomb. 2002. Small, Efficient, Equireplicated Resolution V Fractions of 2k Designs and their Application to Central Composite Designs. *Proceedings of 46th Fall Technical Conference.* American Society of Quality, Chemical and Process Industries Division and Statistic Division; American Statistical Association, Section on Physical and Engineering Sciences.

APPENDIX 1: SELECTED DESIGN TEMPLATES FOR MIN RES IV

Table 1: Min Res IV design for five factors

Std	A	B	C	D	E
1	–	+	+	–	+
2	–	–	–	–	+
3	–	–	+	+	+
4	–	–	+	–	–
5	+	–	+	–	+
6	+	+	+	+	–
7	+	+	–	+	+
8	+	–	–	+	–
9	–	+	–	+	–
10	+	+	–	–	–

Table 2: Min Res IV design for seven factors

Std	A	B	C	D	E	F	G
1	–	–	–	+	+	–	+
2	–	–	+	+	–	+	–
3	–	+	–	+	+	–	–
4	+	–	+	–	+	–	–
5	–	+	+	–	+	+	+
6	–	+	–	+	–	+	+
7	+	–	–	–	+	+	–
8	+	+	+	–	–	+	–
9	+	+	–	–	+	–	+
10	+	–	+	–	–	+	+
11	+	+	+	+	+	+	+
12	+	–	–	+	–	–	–
13	–	+	+	+	–	–	+
14	–	–	–	–	–	–	–

APPENDIX 2: DETAILS ON CONSTRUCTION OF MIN RES IV DESIGNS AND THEIR PROPERTIES

Here's how the Min Res IV designs are constructed:

1. Number of runs in the design matrix equals two times the number of factors: runs = 2k.

2. Use D-optimal search restricted to equireplicated designs, where there are the same number of runs at the high and low level of each factor.

3. Select using the columnwise-pairwise algorithm (Li and Wu, 1997).

Some properties of Min Res IV designs are:

1. Main effects are not aliased with two-factor interactions (2fi's).

2. Partial aliasing occurs among two-factor and higher order interactions.

3. For some designs (not all) the tool of forward-stepwise regression may find the significant two-factor interactions if:

- The non significant 2fi's are small relative to the significant 2fi's.
- The experimental error is small relative to the significant 2fi's.

4. In some cases, the intercept is aliased with 2fi's. For example, in the seven-factor Min Res IV:
"Intercept" = Intercept - 0.333 * BC + 0.333 * BD - 0.333 * BE - 0.333 * CD + 0.333 * CE - 0.333 * DE
This does not interfere with the intent of these designs for use as screening tools that separate the vital few factors from the trivial many. However, aliasing of this sort could pose problems for prediction. As we stated earlier: Always do follow-up studies for this, or any other, type of screening design. Do not use the models from screening for predictive purposes other than to establish a starting point for more in-depth (higher resolution) design of experiments.