No-confounding Designs of 20 and 24 Runs for Screening Experiments

And a Design Selection Methodology

by

Brian Stone

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2013 by the
Graduate Supervisory Committee:

Douglas Montgomery, Chair
Rachel Silvestrini
John Fowler
Connie Borror

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

Nonregular screening designs can be an economical alternative to traditional resolution IV 2^{k-p} fractional factorials. Recently 16-run nonregular designs, referred to as no-confounding designs, were introduced in the literature. These designs have the property that no pair of main effect (ME) and two-factor interaction (2FI) estimates are completely confounded. In this dissertation, orthogonal arrays were evaluated with many popular design-ranking criteria in order to identify optimal 20-run and 24-run no-confounding designs. Monte Carlo simulation was used to empirically assess the model-fitting effectiveness of the recommended no-confounding designs. The results of the simulation demonstrated that these new designs, particularly the 24-run designs, are successful at detecting active effects over 95% of the time given sufficient model effect sparsity. The final chapter presents a screening design selection methodology, based on decision trees, to aid in the selection of a screening design from a list of published options. The methodology determines which of a candidate set of screening designs has the lowest expected experimental cost.

DEDICATION

I dedicate this dissertation to my beautiful and talented wife, Robin, and my incredible children, Jackson and Austin. Your love and support throughout this experience has meant everything to me. Robin, nothing I could write would properly summarize how much you have done to make this accomplishment possible.

ACKNOWLEDGMENTS

I would like to give a heartfelt thanks to my PhD committee for their expertise, guidance and friendship during the course of my studies. A special thank you goes to my advisor, Dr. Douglas Montgomery, first for your guidance and friendship and second for all of the fascinating DOE "war stories." I learned as much from your tales of experimental triumph as I did from your fascinating lectures. Thank you, Dr. Rachel Silvestrini, for carefully reviewing my work and helping to keep my writing clear and focused. Thank you, Dr. Fowler, for being an example of leadership outside of the classroom. I hope one day I can say that I gave as much to the academic community as you have. Thank you, Dr. Connie Borror, for your support, encouragement, and insightful questions that helped me see this research in a bigger context.

I want to thank all of the brilliant students who joined me in this challenging endeavor. To spend so much time with sharp and talented minds has been a real pleasure. A special thanks to Edgar Hassler for sharing your vast theoretical knowledge and your clever sense of humor. Jerry Huff, thanks for always being there at study groups and for sharing your wisdom about the applied side of industrial engineering. Gina Dumkrieger, thank you for your friendship and outstanding work as the President of our INFORMS student chapter – it was great to be a part of your leadership team.

Finally I would like to thank the Industrial Engineering department for allowing me to conduct my research at ASU. This is an incredible and inspiring institution and my experiences here will never be forgotten.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

Table Page

Table Page

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

When an experimenter is unable to utilize a resolution V screening design due to budget or resource constraints, a resolution IV 2^{k-p} fractional factorial design is often used as an alternative. A disadvantage of these designs is that two-factor interactions (2FI) are completely confounded with other candidate 2FI. If fully confounded 2FI are found to be active, the experimenter is forced to assume the additional cost of running a foldover experiment to disambiguate the experimental results. As a solution to this problem, Jones and Montgomery (2010) identified a class of 16-run nonregular designs, referred to as no-confounding designs, which have the property that no pair of main effect (ME) and 2FI estimates are completely confounded. An important advantage of these screening designs is that when analyzed with techniques such as stepwise regression, the no-confounding design can estimate all of the model terms without the need for additional experimentation - an advantage which can save considerable experimental resources.

This dissertation introduces 20-run and 24-run no-confounding designs as additional alternatives to resolution IV 2^{k-p} fractional factorials. As with the 16-run no-confounding designs, no pair of ME or 2FI are completely confounded, allowing an experimenter to avoid foldover experiments. Compared to the 16-run designs, designs of 20 runs and 24 runs have more capability to accurately screen effects for models of one to eight terms. These designs also provide more accurate coefficient estimates for the cost of only four and eight additional runs respectively. In addition to providing experimenters economical screening design options, this dissertation introduces a

screening designs selection methodology used to identify which of a candidate set of screening designs has the lowest expected experimental cost given uncertainty about the model that must be fit. This methodology provides an objective approach to the usually subjective process of selecting a screening design.

The no-confounding designs are identified through an evaluation, using Monte Carlo simulation, of orthogonal arrays (OA). Formally, an OA is a $N \times m$ matrix whose ith column contains s_i levels. The OA has strength t if, within any t columns, every t-tuple of levels appears equally often. The general notation is $OA(N; s_1, \ldots, s_m; t)$. In the case of pure level arrays where each column has the same number of levels then the notation is shortened to $OA(N; s^m; t)$. Most OAs are categorized in the experimental design literature as nonregular designs, defined as factorial designs which produce partially confounded effect estimates.

A literature survey of nonregular designs is provided in Chapter 2. The chapter discusses five aspects of nonregular designs: design evaluation criteria; design construction methods; design projection properties; nonregular design analysis methods; and recent articles in the literature concerning nonregular designs for screening experiments. The background material in Chapter 2 gives important details, particularly about design-ranking criteria, that will help the reader understand the theoretical foundation for assessing and selecting screening designs from the catalog of orthogonal arrays in Chapters 3 and 4.

Chapter 3 presents the research conducted to identify the 20-run no-confounding designs. The search for these designs involved the completely enumerated catalog of non-isomorphic orthogonal arrays of 6-12 columns provided by Sun, Li et al. (2002).

The OAs were evaluated and ranked using a set of 24 design criteria. Standard criteria such as G-aberration, projection estimation capacity (PEC), and $E(s^2)$ are included in the set. Other criteria included unique points in k-factor projections and average variance inflation factors (VIF).

The Pareto efficient set of designs, as defined by the 24 design criteria, was of reasonable cardinality such that it was feasible to conduct an empirical model-fitting analysis of each design. The empirical analysis was conducted using Monte Carlo simulation which randomly generated experimental response data and automated the model fitting process. To assess the performance of a design, the simulation was repeated 5000 times, each time with unique response data.

The designs were evaluated on two metrics: the percentage of simulations which resulted in an estimated model with at least the active effects (no type II errors), and the percentage of simulations which resulted in an estimated model with all of the active effects and no inactive effects (no type I or type II errors). These metrics were called the *% at least correct* and the *% correct*. The Monte Carlo simulation results were used to give an empirical characterization, in terms of the *% at least correct* and the *% correct,* of the model fitting capability of the optimal 20-run no-confounding designs for models composed of every possible combination of 2-5 ME and 0-2 2FI. The 20-run designs are shown to be extremely effective at identifying at least the active effects – the most important quality of a screening design - when there is at most four ME and one 2FI.

In addition to empirically assessing orthogonal arrays, the Monte Carlo simulation was also used to assess model fitting methods. Many variants of stepwise regression were evaluated to determine which was the most effective. The results in this dissertation

corroborate the recommendation in Shinde (2012) to use a two-stage forward stepwise regression when analyzing no-confounding designs.

Chapter 4 discusses the process used to identify the 24-run no-confounding designs. The approach used to identify those designs involved an empirical assessment of minimum G-aberration designs, maximum PEC designs, and designs created by two variants of a column-exchange search algorithm. The lack of a completely enumerated catalog of $\mathrm{OA}(24; 2^{\mathrm{m}}; 2)$ for $m > 7$ necessitated a search algorithm. The column-exchange algorithms used in this research differed from previous search algorithms in that they were not restricted to the columns of Hadamard matrices.

Several methods were used to generate columns and efficiently search the space of balanced, orthogonal 2-level designs. The algorithm was also modified to search for non-orthogonal designs to explore whether sacrificing some column orthogonality might lead to a better over-all design. It was discovered that the minimum G-aberration 24-run designs performed better in terms of the *% at least correct* and the *% correct* metrics than any other published or algorithmically generated design, despite the fact that no 6-factor projection can estimate every term in a 6-factor full model. The algorithms did find designs whose values of *% at least correct* and *% correct* were within 4 percentage points of the minimum G-aberration designs, but could estimate the full 6-factor model in every projection. The 24-run no-confounding designs detected every active effect over 95% of the time for response models composed of up to five ME and two 2FI. These results were consistent when investigating any number of 6-12 factors.

The 20-run and 24-run screening designs introduced in this dissertation give an experimenter an expanded list of screening designs options. These screening designs are

alternatives to the more traditional options: the 2^{k-p} fractional factorials and the Plackett-Burman designs. Definitive Screening Designs have also been introduced recently by Jones and Nachtsheim (2011). These designs incorporate a third level to allow for the unambiguous estimation of quadratic effects. With so many screening design options, it can be difficult for an experimenter to make an informed choice.

In Chapter 5 a screening design selection methodology is presented that helps an experimenter select a design based on which one has the lowest expected cost. Often an experimental budget must be established prior to the experiment. However, when experimenting in stages, it becomes more difficult to provide an initial estimate of the total cost of the experiment. This is because the total number of runs in each stage of experimentation - usually the key driver of experimental cost - depends on the number of significant main effects (ME), two-factor interactions (2FI) and quadratic effects (QE) required to adequately model the response variable.

The total number of experimental runs is heavily influenced by the choice of screening design. The screening design determines aspects of the data such as whether active effects are completely confounded with other effects and whether quadratic effects can be estimated in the screening phase. These aspects affect how many experimental runs are required in the later stages of experimentation.

The expected cost screening design selection methodology extends the previous research on cost-constrained experimental design in three important ways. First, while the previous methods optimized the construction of a one-stage experimental design, this method minimizes the expected start-to-finish experimental cost of a multi-stage experiment involving a screening design, fold over runs, and axial runs. Second, instead

of using an algorithm to select cost-optimal runs for a fractional factorial design, our methodology is a process for design selection from a candidate list of cataloged screening designs. The methodology can be used with any traditional or modern experimental design, such as a definitive screening or no-confounding design. Finally, an expanded scope of three cost sources is considered: the cost of the experimental runs; the cost of building a model that omits significant model terms or includes superfluous model terms; and the cost of building a model with inaccurate estimates of the regression coefficients.

The screening design with the lowest expected cost is reported for all feasible combinations of model probability and cost penalty values. This facilitates a sensitivity analysis that informs the experimenter of how robust the optimal screening design is to changes in the probability and cost penalty assumptions. Instead of making predictions or assumptions about specific probabilities or inaccurate model cost penalties, the experimenter has the easier task of predicting a region of probabilities and penalties.

Chapter 6 summarizes the results of this research into 20 and 24-run screening designs and the designs selection methodology. It ties together commonalities between the no-confounding designs of different dimensions and highlights the important differences. The chapter concludes with avenues for future research related to this topic.

As a note to the reader, the chapters in this dissertation are written so that they can be read separately from the rest of the document. Some material from the introduction as well as background information from Chapter 2 are included in Chapters 3-5. The goal is to avoid situations where the reader must refer to other chapters in order to understand the discussion in the chapters pertaining to the research.

Chapter 2

LITERATURE SURVEY

Experimental designs which produce effect estimates that are partially confounded with other effects are known as nonregular designs. Sun and Wu (1993) introduced the term "nonregular design" when studying Hadamard matrices of order 16. There are many circumstances in which nonregular designs have been shown to be preferable alternatives to 2^{k-p} fractional factorials in the screening design process. Much of the work in this dissertation involves the evaluation of nonregular designs in order to find more alternative screening designs with specific properties.

This chapter highlights important research concerning nonregular designs that relates to the process of identifying the 20-run no-confounding designs in Chapter 3 and the 24-run no-confounding designs in Chapter 4. Section 2.1 discusses seven of the most popular ways to assess and rank nonregular designs. Section 2.2 presents nonregular design construction methods. Section 2.3 describes the projection properties of nonregular designs and how these properties make them effective screening designs. Section 2.4 discusses three analysis methods used to analyze experiments conducted with nonregular designs. The chapter concludes with Section 2.5 which highlights two recently-introduced screening designs that are rapidly gaining popularity: the no-confounding designs and the definitive screening designs.

A more formal definition of nonregular designs are those designs whose alias matrix contains elements other than ± 1 or 0. The alias matrix is computed as follows: Suppose the proposed linear model is

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$$

Where $\boldsymbol{y}$ is the $n \times 1$ vector of responses, $\boldsymbol{X_1}$ is the $n \times p_1$ model matrix, $\boldsymbol{\beta}_1$ is the $p_1 \times 1$ vector of model coefficients, and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector of independent and normally distributed random errors with mean 0 and variance σ^2. Then the least squares estimate of $\boldsymbol{\beta}_1$ is

$$\widehat{\boldsymbol{\beta}}_1 = (\boldsymbol{X}_1{}'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}$$

Let $\boldsymbol{X_2}$ be a $n \times p_2$ model matrix containing variables that were not included in the original model matrix and let $\boldsymbol{\beta}_2$ be the $p_2 \times 1$ vector of model coefficients associated with those variables. Suppose the true model is

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

Then it can be shown that

$$E(\widehat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + (\boldsymbol{X}_1{}'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2$$

$$= \boldsymbol{\beta}_1 + \boldsymbol{A}\boldsymbol{\beta}_2$$

The matrix $\boldsymbol{A} = (\boldsymbol{X}_1{}'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$ is known as the alias matrix (Montgomery (2012).

The first screening designs were nonregular designs introduced by Plackett and Burman (1946), who provided a collection of 2-level and 3-level designs to investigate $\mathrm{N} - 1$ factors with N experimental runs, where N is a multiple of four. All columns in the design matrix are pair-wise orthogonal, meaning main effects are completely de-aliased, but every main effect is partially aliased with every two-factor interaction not involving itself. Hamada and Wu (1992) showed that the use of Plackett-Burman designs need not be restricted to screening for main effects. With proper analysis, two-factor interactions could be detected as well. This revelation prompted further research into nonregular designs.

Cheng and Wu (2001) proposed a modification of the traditional response surface methodology in which screening and response surface exploration are conducted with a single design. An important condition for this approach is that the single design has good low-dimensional projection properties since it is not known a priori which factors are important. The following section reviews research attempting to identify optimal or superior nonregular designs in terms of projection properties and orthogonality of effects.

2.1 Evaluating Nonregular Screening Designs

Regular designs, such as the 2^{k-p} fractional factorials, have been evaluated by using the minimum aberration criteria proposed by Fries and Hunter (1980). Designs with minimum aberration have the most favorable alias relationships among all designs of equivalent dimension. A minimum aberration design has the minimum number of words in the defining relation that are of minimum length. For a given design of resolution R, these designs minimize the number of main effects (ME) aliased with $(R-1)$-factor interactions and the number of two-factor interactions (2FI) aliased with $(R-2)$-factor interactions Montgomery (2012). Section 2.1.1 discusses research concerning how to evaluate nonregular designs with a criterion that is analogous to minimum aberration for regular fractional factorials.

2.1.1 G-aberration

The first to generalize the concept of minimum aberration for nonregular fractional factorials was Deng and Tang (1999). Their paper introduced the concept of generalized resolution and generalized minimum aberration (minimum G-aberration). For regular designs, minimum G-aberration reduces to the traditional minimum aberration criterion.

In order to define minimum G-aberration, Deng and Tang first developed the idea of J-characteristics and the confounding frequency vector (CFV). Suppose there is an $n \times m$ design matrix $D = (d_{ij})_{n\times m}$ with $d_{ij} = \pm 1$. Let $Z_m = \{1, \dots, m\}$ be an index set of the columns of D. For every subset s of Z_m, define $J_s(D) = \sum_{i=1}^{n} \prod_{j\in s} d_{ij}$. The collection of J_s values is known as the J-characteristics of design D. Let $n = 4t$ and f_{kj} be the frequency of k column combinations that give $|J_s| = 4(t + 1 - j)$ for $j = 1, \dots, t$, where $|J_s|$ is the absolute value of a J-characteristic. The confounding frequency vector of design D is defined to be

$$F(D) = [(f_{21}, \dots, f_{2t}), (f_{31}, \dots, f_{3t}), (f_{41}, \dots, f_{4t}), \dots, (f_{m1}, \dots, f_{mt})]$$

Note that $(f_{11}, \dots, f_{1t})$ is simply the frequency of non-zero column sums of D in reverse order, which for a balanced two-level design are all zero since all column sums are zero.

Let $f_l(D_1)$ and $f_l(D_2)$ be the lth entries of $F(D_1)$ and $F(D_2)$, respectively. Let r be the smallest integer such that $f_r(D_1) \neq f_r(D_2)$. If $f_r(D_1) < f_r(D_2)$ then D_1 has less G-aberration than D_2. If no design has less G-aberration than D_1, then D_1 has minimum G-aberration. Recently, Lin, Sitter et al. (2012) used G-aberration to catalog and rank two-level nonregular fractional factorial designs of 32 and 40 runs.

2.1.2 G_2-Aberration

Tang and Deng (1999) introduced a second design evaluation concept known as G_2-aberration. Unlike G-aberration, this metric penalized average non-orthogonality of effects more than the maximum non-orthogonality. As was the case for the G-aberration, minimum G_2-aberration reduces to the traditional minimum aberration criterion for regular designs. Minimum G_2-aberration designs are also derived from the J-characteristics. Let $A_k(\mathrm{D}) = N^{-2} \sum_{|s|=k} |J_k|^2(s; D)$. The generalized wordlength

pattern is defined to be $A_1(\mathrm{D}), \ldots, A_m(\mathrm{D})$). For two designs D_1 and D_2, let r be the smallest integer such that $A_r(D_1) \neq A_r(D_2)$. If $A_r(D_1) < A_r(D_2)$ then D_1 has less G_2-aberration than D_2. If no design has less G_2-aberration than D_1, then D_1 has minimum G_2-aberration.

2.1.3 Generalized Minimum Aberration

Tang and Deng (1999) had extended the idea of minimum aberration to two-level nonregular designs, but they did not provide an obvious extension to designs of three or more levels. This extension was accomplished independently by two groups of researchers. Xu and Wu (2001) proposed generalized minimum aberration (GMA) as the generalization of G_2-aberration to asymmetrical (mixed level) designs. Ma and Fang (2001) independently developed *minimum generalized aberration* which is a special case of GMA from Xu and Wu (2001). Fang, Zhang et al. (2007) provided optimal conditions for GMA designs, and used the conditions to create an effective sub-design selection algorithm to find GMA designs.

2.1.4 The $E(s^2)$

Even before minimum aberration had been developed to compare regular fractional factorials, researchers searched for a way to compare supersaturated designs. A supersaturated design is a factorial design with n rows and k effect columns with $k > n$. Booth and Cox (1962) introduced $E(s^2)$ as a solution to the problem of evaluating supersaturated designs. Let $\mathbf{x}_\mathrm{i}$ be the ith column of the $n \times k$ model matrix $\boldsymbol{X}$. Then

$$\mathrm{E}(\mathrm{s}^2) = \sum\nolimits_{\mathrm{i}<\mathrm{j}} \left(\mathbf{x}_\mathrm{i}`\mathbf{x}_\mathrm{j}\right)^2 \Big/ \binom{\mathrm{k}}{2}$$

Minimizing $E(s^2)$ is equivalent to minimizing the sum of squared off-diagonal elements of the correlation matrix of the model matrix. Lin (1993) was the first to rejuvenate interest in the metric. He constructed supersaturated designs of size $(n, m) = (2t, 4t - 2)$ using half fractions of Hadamard matrices.

The $E(s^2)$ criterion was used to construct and evaluate supersaturated designs in Wu (1993), Lin (1995), Nguyen (1996), Tang and Wu (1997), Cheng (1997), and Liu and Zhang (2000). Wu (1993) used Hadamard matrices to develop supersaturated designs of 12 and 20 runs to investigate up to 66 and 124 factors respectively. Lin (1995) examined the maximum number of factors that can be investigated with a supersaturated designs of size N given a pre-specified degree of non-orthogonality between effects. Nguyen (1996) generalized the work in Lin (1993) by discovering a method of constructing supersaturated designs with cyclic BIBDs. Tang and Wu (1997) also generalized the work in Lin (1993) by using Hadamard matrices to construct supersaturated designs. Cheng (1997) presented a unified synthesis of the optimality results from Lin (1993) and Tang and Wu (1997). Liu and Zhang (2000) present a general algorithm of constructing $E(s^2)$-optimal supersaturated designs from cyclic BIBDs. They compare their results to work in Nguyen (1996) and Tang and Wu (1997). Jones and Montgomery (2010) used $E(s^2)$ to identify which projections of the Hadamard matrices of order 16 were optimal no-confounding designs.

2.1.5 Minimum Moment Aberration

A new approach for comparing nonregular and supersaturated designs based on the dissimilarity of the rows in a design matrix was developed by Xu (2003). He defined a new criterion called minimum moment aberration which minimized the power moments

of the number of coincidences between runs. For an (N, s^n) design $D = [r_{ij}]_{N\times n}$ and a positive integer t, the tth power moment is defined to be

$$K_t(D) = [N(N-1)/2]^{-1} \sum_{1\le i<j\le N} [\delta_{ij}(D)]^t$$

where the number of coincidences between the ith and jth rows of a design D is

$$\delta_{ij}(D) = \sum_{k=1}^{n} \delta(r_{ik}, r_{jk})$$

Inside the summation, $\delta(x, y)$ is the Kronecker delta function function, which equals 1 if $x = y$ and 0 otherwise. Minimum moment aberration created a link between design theory and algebraic coding theory since $n - \delta_{ij}(D)$ is the Hamming distance between the ith and jth rows in a design matrix.

Xu showed minimum moment aberration was a good surrogate for minimum G_2-aberration, generalized minimum aberration and $E(s^2)$. In addition this metric was less expensive in terms of computation time and proved to be useful in several design construction algorithms. Xu and Deng (2005) developed moment aberration projection to rank and classify general nonregular designs. They discovered the surprising result that the minimum G-aberration criterion and the moment aberration projection criterion are not equivalent for two-level designs.

2.1.6 Projection Estimation Capacity

Another general way to evaluate designs is by model robustness, often measured by some variant of estimation capacity. Estimation capacity was first introduced by Sun (1993) and measures the number of estimable models containing all of the ME evaluated in a design, and a given number of q 2FI's. Estimation capacity was used by Li and

Nachtsheim (2000) and Li (2006) to create model robust designs for screening experiments.

Loeppky, Sitter et al. (2007) introduced a variant of estimation capacity known as projection estimation capacity (PEC). Instead of the potential model space involving all ME in the screening design, the model space for PEC involves all q ME in q-factor projections of the original design plus all of the associated 2FI. Using the completed catalog of non-isomorphic $\mathrm{OA}(16; 2^{\mathrm{m}}; 2)$ and $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$, Loeppky et al. were able to identify optimal 16-run and 20-run designs in terms of PEC. They used the complete catalog of non-isomorphic Hadamard matrices of orders 24 and 28 to identify 24 and 28-run PEC designs.

The PEC of experimental designs is important when screening all of the effects in a full ME and 2FI model. Schoen (2010) looked at OAs of strength 3 with full estimation capacity (the ability to estimate all terms in a full model) and discussed when these OAs perform better than an optimal design.

2.1.7 Cost-constrained Screening Designs

Another important method of evaluating designs, which has been applied to both regular and nonregular designs, is based on design cost. Early work on design evaluation based on budgetary constraints includes Neuhardt, Bradley et al. (1973) and Neuhardt, Bradley et al. (1978) who examined the problem of selecting the lowest cost 2^{k-p} fractional factorial experiment based on the assignment of a cost c_i to each "cell" (factor level combination) in the full factorial experiment. Neuhardt, Bradley et al. (1978) discuss a dynamic programming approach to obtaining an optimal cost fractional factorial based on cell costs. See Mount-Campbell and Neuhardt (1980) and Mount-Campbell and

Neuhardt (1982) for discussions concerning cost-optimal 3^{n-r} fractional factorial designs and $3^{m-r}2^{n-s}$ fractional factorials respectively.

Pignatiello (1985) introduced a procedure to determine cost-optimal p^{n-r} fractional factorials (for prime p) that simultaneously considered cell costs and the ability to estimate a rank ordered list of specified main effects and two-factor interactions. Rafajlowicz (1989) developed an algorithm to find cost-efficient fractional factorials based on a specified information matrix.

Tack and Vandebroek (2001) were the first to simultaneously consider resource costs and run-transition costs as criteria for constructing cost-optimal designs. In Tack and Vandebroek (2002) they extend their cost-efficient and time trend-resistant optimal design approach to block designs. In Tack and Vandebroek (2004) the work is extended to situations where there is a budgetary constraint. Additional optimal design research was conducted by Park, Montgomery et al. (2006), who developed a genetic algorithm to create cost-constrained designs with good G-efficiency.

2.2 Nonregular Design Construction Methods

Advancing computing capability in the past 15 years has led to significant research in the construction of nonregular designs. Many construction methods utilize Hadamard matrices which are special cases of orthogonal arrays. Earlier research involved searching through projections of Hadamard matrices to find designs that minimized various design criteria. Research has shifted to the more general class of orthogonal arrays as a result of the complete enumeration of the non-isomorphic OAs of certain dimensions. The next subsections detail the significant work in nonregular design construction.

2.2.1 Hadamard Matrices

A Hadamard matrix is a square matrix whose entries are either ± 1. A Hadamard matrix with n rows (columns) is said to be of order n. The columns of a Hadamard matrix H are mutually orthogonal such that $H^T H = nI_n$. A large online repository of Hadamard matrices is maintained by Neil Sloane of AT&T Shannon Labs and is located at http://www2.research.att.com/~njas/hadamard/. This website includes all Hadamard matrices of orders N up through $N = 28$, and at least one of every order N up through $N = 256$.

The orthogonality of the columns and the fact that their entries are ± 1 made Hadamard matrices the basis of much nonregular design research. Tang and Wu (1997) generalized the work in Lin (1993) by using Hadamard matrices to construct supersaturated designs. Li and Wu (1997) presented a class of supersaturated design construction algorithms called column-wise pairwise exchange algorithms. A key contribution was applying the row-exchange concept of the k-exchange algorithms proposed by Johnson and Nachtsheim (1983) to a column-exchange concept with designs based on Hadamard matrices. Ingram (2000) developed a column exchange algorithm based on the selection of columns from Hadamard matrices, called the excursion-at-target algorithm, to construct minimum G-aberration designs of 24 runs for 9-23 factors. Belcher-Novosad and Ingram (2003) improved their excursion-at-target algorithm to identify minimum G-aberration designs from Hadamard Matrices of order 28.

2.2.2 Orthogonal Arrays

Orthogonal arrays are an important tool to facilitate factorial experimentation, not only in industrial contexts but in a growing number of other disciplines as well. SAS

maintains a library of 117,000 orthogonal arrays (of strength two) up through 144 runs at http://support.sas.com/techsup/technote/ts723.html. SAS also provides a set of free macros for making orthogonal arrays and D-efficient non-orthogonal designs.

Orthogonal arrays were introduced by Rao (1947) and included Plackett-Burman designs as a special case. Formally, an OA is a $\mathrm{N} \times \mathrm{m}$ matrix whose ith column contains $\mathrm{s_i}$ levels. The OA has strength t if, within any t columns, every t-tuple of levels appears equally often. The general notation is $\mathrm{OA}(\mathrm{N}; \mathrm{s_1}, \ldots, \mathrm{s_m}; \mathrm{t})$. In the case of pure level arrays where each column has the same number of levels then the notation is shortened to $\mathrm{OA}(\mathrm{N}; \mathrm{s^m}; \mathrm{t})$. Orthogonal arrays include both regular and nonregular designs. Regular designs of resolution R are orthogonal arrays of strength $t = R - 1$. Cheng (1980) proved that for the main effects model, orthogonal arrays are universally optimal. Box and Tyssedal (1996) defined a design to be of projectivity p if the projection onto every subset of p factors contains at least a 2^p full factorial design. An orthogonal array of strength t is therefore of projectivity t. Hedayat, Sloane et al. (1999) gives a detailed presentation of the theory and application of orthogonal arrays.

Sun, Li et al. (2002) developed an algorithm for sequentially constructing non-isomorphic regular and nonregular orthogonal arrays. Two OAs are said to be isomorphic if one design can be obtained from the other by permuting rows, columns or relabeling factor levels. The algorithm was used to obtain a complete catalogue of $n \times k$ two-level orthogonal designs for $n \in \{12, 16, 20\}$ and $k \in \{2,3, \ldots, (n-1)\}$. The catalog was used to identify generalized minimum aberration designs. Zhang, Li et al. (2011) present and prove an equivalence relationship between orthogonal arrays and generalized

difference matrices. Schoen and Mee (2012) classified all two-level orthogonal arrays of strength 3 up to 48 runs.

2.2.3 Constructing Designs Using Design Evaluation Criteria

Tang and Deng (2003) constructed minimum G-aberration designs, including designs that were not projections of Hadamard matrices, for 3-5 factors and any run size n that is a multiple of four. Butler (2003) provided construction results that identified minimum G_2-aberration $n \times m$ designs where $n \in \{16, 24,32,48,64,96\}$ and $m \geq (n/2) - 2$. Butler (2004) provided construction results that identified minimum G_2-aberration designs for many of the cases with N =16, 24, 32, 48, 64 and 96 runs. Li, Deng et al. (2004) developed an algorithm to search for minimum G-aberration designs of 4-6 factors and $n \in \{16, 20,24,32,36\}$ and found designs that were not projections of Hadamard matrices. Butler (2005) used a design construction approach based on subsets of saturated orthogonal arrays to produce GMA designs.

Li (2006) presents screening designs for 12, 16 and 20 runs that are optimal with respect to full estimation capacity. Stufken and Tang (2007) classified all two-level orthogonal arrays with $\mathrm{t} + 2$ factors and strength t for any run size. Loeppky, Sitter et al. (2007) ranked two-level orthogonal design based on the number of estimable models containing a subset of k main effects and all $\binom{\mathrm{k}}{2}$ of their associated two-factor interactions. They presented catalogs of useful designs with 16, 20, 24, and 28 runs. Bulutoglu and Margot (2008) classified some orthogonal arrays of strength 3 up to 56 runs and of strength 4 up to 144 runs for up to eleven factors.

2.3 Projection Properties

When using screening designs, experimenters rely on the principle of effect sparsity – a concept that implies relatively few of the factors in the initial experiment will be active. Thus screening designs are often analyzed in terms of their projections onto small subsets of factors. Box and Hunter (1961) were the first to present a rationale for examining the projection properties of screening designs. This paper also presented the idea of design resolution, which is related to the projections of a design. A design has resolution R if no p-factor effect is aliased with another effect containing less than $R - p$ factors. For designs of resolution R, the projection onto any R factors results in either a 2^R full factorial or a replicated 2^{R-1}. The projection properties of nonregular designs are complex due to the partial aliasing of factors. Recall from Section 2.2.2 that Box and Tyssedal (1996) defined a design to be of projectivity p if the projection onto every subset of p factors contains at least a 2^p full factorial design.

Lin and Draper (1992) looked at all projections of 12, 16, 20, 24, 28, 32, and 36-run Plackett-Burman designs and found that all projections either contain a 2^3 factorial or replicated 2^{3-1} fractional factorial or both. Cheng (1995) showed that as long as the run size N is not a multiple of 2^{t+1}, an $OA(N; 2^m; t)$ with $m \geq t + 2$ has projectivity $t + 1$. The paper further showed that as long as the run size N of an $OA(N; 2^m; 2)$ is not a multiple of 8, its projection onto any four factors allows the estimation of all the main effects and two-factor Interactions when the higher-order interactions are negligible. Cheng (1998) proved that as long as the run size N of an $OA(N; 2^m; 3)$ is not a multiple of 16, its projection onto any five factors allows the estimation of all the main effects and two-factor interactions. Wang and Wu (1995)

introduced the concept of the hidden projection property, which is a design's ability to estimate interactions without relying on geometric projection. This property is the result of complex aliasing. Bulutoglu and Cheng (2003) showed that the hidden projection property also holds for Paley designs of sizes greater than eight, including those with run sizes that are multiples of eight. Tang (2001) evaluated the minimum G_2-aberration designs and demonstrated that they have good low-dimensional projection properties. Ai and Zhang (2004) showed that generalized minimum aberration designs have good low-dimensional projection properties. Cheng (2006) reviews the projection properties of factorial designs and how projection relates to factor screening.

2.4 Nonregular Screening Design Analysis

Hamada and Wu (1992) showed that the use of Plackett-Burman designs need not be restricted to screening for main effects. With proper analysis, two-factor interactions could be detected as well. This important discovery initiated a large volume of research in the area of nonregular screening design analysis. The next subsections highlight three popular analysis techniques: stepwise regression, the Dantzig selector, and LASSO.

2.4.1 Stepwise Regression

The Hamada and Wu (1992) design analysis strategy uses a forward stepwise regression procedure that is based on effect sparsity and effect heredity. Forward stepwise regression is a model building technique that starts from a void model and model terms are added if the p-value for the t-statistic of that effect is below a threshold. At each step a variable is added from a candidate list that improves the model the most according to some criteria. When effect heredity is enforced, if a two-factor interaction

or quadratic term is added to the model then all corresponding main effects are automatically added to the model.

The p-value approach mentioned above is one method to evaluate and compare potential models. Another class of model evaluation methods seeks to identify models that make the best trade-off between bias and variance. These criteria all have the general form

$$\left\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|_{l_2}^2+\Lambda\sigma^2\left\|\widehat{\boldsymbol{\beta}}\right\|_{l_0},\qquad \left\|\widehat{\boldsymbol{\beta}}\right\|_{l_0}\coloneqq\left|\{i:\hat{\beta}_i\neq 0\}\right|$$

According to Candes and Tao (2007), model selection criteria including the Akaike Information Criterion (AIC), the Bayesian Information criterion (BIC) and Mallow's C_p can be written in this form by choosing appropriate values of Λ. The AIC, based on the concept of information entropy, was introduced in Akaike (1974) as a measure that balances the reward for goodness of fit with a penalty for overfitting. The Akaike's information criterion corrected (AICc) discussed in Hurvich and Tsai (1989) utilizes a bias correction for small sample sizes making it a valuable statistical tool when building models from screening designs.

2.4.2 LASSO

The least absolute shrinkage and selection operator (LASSO) was introduced by Tibshirani (1996) as a method to estimate $\boldsymbol{\beta}$ by minimizing an objective function that included both the sum of squares of the residuals and the l_1 norm of $\widehat{\boldsymbol{\beta}}$ as a penalty term. According to Tibshirani (2011), the LASSO method increased in popularity due to the development of efficient computation algorithms such as the least angle regression (LARS) algorithm in Efron, Hastie et al. (2004). Efron et al. produced another version of

LARS that improved on forward stage-wise regression. Zhao and Yu (2006) showed that Lasso is a computationally feasible model selection technique and provided an almost necessary and sufficient condition for it to achieve consistent selection.

2.4.3 Dantzig Selector

A variant of LASSO called the Dantzig selector, introduced in Candes and Tao (2007), is a method to estimate $\widehat{\boldsymbol{\beta}}$ when the number of model terms p is greater than the number of observations n. The method uses a convex program with an objective function that minimizes the complexity of $\widehat{\boldsymbol{\beta}}$ via a l_1-norm. Since the convex program can be recast as a linear program, this model building technique is efficient even for large values of n and p.

2.5 Recent Developments in Nonregular Screening Designs

Two important developments in nonregular screening designs in the recent literature are the no-confounding designs and the definitive screening designs. In some situations, these designs offer important advantages over the more traditional 2^{k-p} fractional factorials and Plackett-Burman designs.

2.5.1 No-confounding Designs

No confounding designs are a class of 16-run nonregular fractional factorial designs introduced by Jones and Montgomery (2010) which have the defining feature that no main effect is completely confounded with any two-factor interaction. For many of these designs it is also true that no pair of two-factor interactions are completely confounded.

These 16-run designs for investigating six to ten factors are projections of the five orthogonal designs for screening up to 15 factors in 16 runs introduced in Hall (1961).

The no-confounding design for k factors is the optimal combination of k Hall matrix columns that minimize $E(s^2)$.

The lack of complete confounding of model effects presents a significant cost savings opportunity to an experimenter. If effect sparsity holds and there are no quadratic terms required in the model, analysis with a model building technique such as step-wise regression can produce an accurate model without additional experimentation beyond the screening design. This can result in a significant savings in experimental costs.

2.5.2 Definitive Screening Designs

Definitive screening designs are a class of three level nonregular screening designs introduced by Jones and Nachtsheim (2011) that allow the estimation of all possible full quadratic models involving any three factors, as long as at least six factors are screened. To screen k factors, the design uses $2k + 1$ runs in a structure that includes one center run and k pairs of runs called fold over pairs. The first run in each fold over pair has one factor at the zero level and the other $k - 1$ factors have levels of ± 1. The second run has the factor levels of the first run after multiplication by -1. For every factor under investigation, one fold over pair is added to the design. Table 1 from Jones and Nachtsheim (2011) illustrates this structure.

Table 1. General Design Structure for m Factors

Foldover Pair	Run	Factor Levels				
	(i)	$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$\cdots$	$x_{i,m}$
1	1	0	± 1	± 1	$\cdots$	± 1
	2	0	∓ 1	∓ 1	$\cdots$	∓ 1
2	3	± 1	0	± 1	$\cdots$	± 1
	4	∓ 1	0	∓ 1	$\cdots$	∓ 1
3	5	± 1	± 1	0	$\cdots$	± 1
	6	∓ 1	∓ 1	0	$\cdots$	∓ 1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
m	$2m-1$	± 1	± 1	± 1	$\cdots$	0
	$2m$	∓ 1	∓ 1	∓ 1	$\cdots$	0
Center point	$2m+1$	0	0	0	$\cdots$	0

For these designs, main effects are orthogonal to other main effects as well as quadratic effects. In addition, there is no complete confounding of any pair of second-order effects. The alias structure between two-factor interactions and other two-factor interactions, as well as between quadratic effects and two-factor interactions is complex.

Chapter 3

NO-CONFOUNDING DESIGNS OF 20 RUNS

3.1 Introduction

When an experimenter is unable to utilize a resolution V screening design due to budget or resource constraints, a resolution IV 2^{k-p} fractional factorial design is often used as an alternative. With these designs, the experimenter must accept the risk of finding active two-factor interactions (2FI) which are completely confounded with other candidate 2FI. In this situation there is additional cost in running a foldover experiment to disambiguate the experimental results. As a solution to this problem, Jones and Montgomery (2010) identified 16-run nonregular designs, referred to as no-confounding designs, which have the property that no pair of main effect (ME) and 2FI estimates are completely confounded. An important advantage of these screening designs is that, when the response is determined by a small number of ME and 2FI effects, the no-confounding design can estimate all of the model terms without the need for additional experimentation beyond the screening design - an advantage which can save considerable experimental resources.

The 16-run no-confounding designs were discovered via an exhaustive search of orthogonal arrays (OA) of $N = 16$ rows and m columns for $m \in \{1, \dots, 15\}$. Likewise, the search for 20-run no-confounding designs also used a completely enumerated catalog of non-isomorphic $OA(20; 2^m; 2)$ for $m \in \{6, \dots, 12\}$ factors provided by Sun, Li et al. (2002). OAs are an important tool to facilitate factorial experimentation, not only in industrial contexts but in a growing number of other disciplines as well. Formally, an OA is a $N \times m$ matrix whose ith column contains s_i levels. The OA has strength t if,

within any t columns, every t-tuple of levels appears equally often. The general notation is $OA(N; s_1, \ldots, s_m; t)$. In the case of pure level OAs where each of the m columns has the same number of levels, the notation is shortened to $OA(N; s^m; t)$. OA **A** is non-isomorphic to OA **B** if no permutation of rows, columns, or factor levels in **A** results in a copy of **B**. Most OAs are categorized in the experimental design literature as nonregular designs, defined as factorial designs which produce partially confounded effect estimates.

The economic advantages of the 16-run no-confounding designs were established in the original journal article, but there are limitations to the designs. Unpublished work in Shinde (2012) reported the degradation of quality metrics related to model fitting, such as the frequency of type I and type II model building errors, when the number of model terms was greater than four. One solution to mitigate this problem is to use a design with more runs.

The research into 20-run no-confounding designs was motivated by the desire to provide researchers a no-confounding design option with better model-estimation performance for regression models of five to seven terms. This chapter identifies optimal 20-run no-confounding designs and discusses the process of identifying these OAs based on the empirical results of a Monte Carlo simulation. The simulation results are used to characterize the performance of the designs in terms of two metrics: the percentage of simulations in which at least all of the active effects were identified, and the percentage of simulations in which the correct model was identified. These metrics are reported for every feasible combination of 2-5 ME and 0-2 2FI.

Identifying optimal nonregular designs based on empirical simulation results rather than optimality criteria related to the structure of the design matrix is a novel approach. The standard in the literature is for authors to introduce a design optimality criterion and provide a theoretical justification for how their criterion identifies designs with optimal model-fitting properties. The optimality criterion is then used to rank designs and identify which is optimal. Unfortunately designs often rank differently under various criteria and there are limited empirical studies which inform a practitioner about which are the best design-ranking criteria to use in selecting a design.

Many design-ranking criteria have been introduced and used to catalog and identify optimal nonregular designs. As an example, the optimal 16-run no-confounding designs were identified using the $E(s^2)$ metric, first suggested by Booth and Cox (1962) as a metric to rank supersaturated designs. Generalized aberration, or G-aberration, is also a common design-ranking criteria. Its introduction in Deng and Tang (1999) spurred a large variety of subsequent research into nonregular design evaluation.

Twenty-four popular and practical design-ranking criteria were used as the theoretical foundation for an empirical evaluation of OAs using Monte Carlo simulation. First, each OA in the catalog was given a ranking by each of the 24 criteria. Since many designs ranked lower than other designs according to all of the criteria, these dominated designs were removed from consideration to produce a Pareto efficient set of designs. Using several stages of simulation, an optimal m-factor 20-run no-confounding design for $m \in \{6, \ldots, 12\}$ factors was identified from its respective Pareto efficient design set. As a second contribution from this chapter, the empirical research indicated that the

G-aberration criterion may ultimately be the best way to identify optimal designs, particularly when using two-stage stepwise regression as an analysis method.

The remainder of this chapter is organized as follows. The background section discusses popular nonregular design optimality metrics and the concept of a Pareto efficient set. The methodology section discusses the details of the Monte Carlo simulation used to empirically determine the model-fitting capability of the OAs. This section also discusses the 24 design-ranking criteria used to reduce the complete non-isomorphic catalog of 20-run OAs to a Pareto efficient set of candidate designs. The section concludes with the process used to empirically assess the candidate designs and determine the optimal 20-run no-confounding designs. The optimal 20-run designs for 6-12 factors are presented in the results section along with the empirical data related to their model-fitting capability. Accompanying these results is a discussion of the high correlation between the best empirical results and a high G-aberration rank. This section also provides a summary of two different model-fitting approaches and a comparison of their effect on type I and type II variable-selection errors. The chapter concludes with a summarization of the results and suggestions for future research.

3.2 Background

This section presents some previous research concerning orthogonal arrays and the design-ranking criteria developed to create and evaluate them. Also included is a discussion of the concept of a Pareto efficient set and how it relates to the simultaneous optimization of multiple criteria. The previous research in design optimality criteria sets the stage for the empirical approach taken in this chapter to rank nonregular designs.

Empirical design analysis is particularly suited for small and completely-enumerated design catalogs such as the catalog of 20-run OAs.

The first to complete a catalog of non-isomorphic $OA(20; 2^m; 2)$ was Sun, Li et al. (2002). Subsequently, the non-isomorphism of designs in the catalog was verified by algorithms developed by Angelopoulos, Evangelaras et al. (2007), Evangelaras, Koukouvinos et al. (2007), Bulutoglu and Margot (2008) and Schoen, Eendebak et al. (2010). The complete catalogs of non-isomorphic designs exist for the $OA(16; 2^m; 2)$ and $OA(20; 2^m; 2)$. The complete catalogs of $OA(24; 2^m; 2)$ exist for the cases of $m = 6$ and $m = 7$ (Bulutoglu and Margot 2008). Unfortunately for $N > 24$ it has been computationally infeasible to obtain the complete catalog of non-isomorphic OAs of strength two.

During the past 15 years, articles in the experimental design literature have provided incomplete catalogs of non-regular experimental designs. These designs were ranked by a diverse set of design-ranking criteria. Each criterion relies on a theoretical basis for linking the highest ranked designs to optimal design performance in terms of model estimation. Unfortunately there are many sensible design criteria in the literature to choose from and designs which rank highest under one criterion do not necessarily rank highest under another.

3.2.1 Generalized Aberration

Deng and Tang (1999) proposed minimum G-aberration and Tang and Deng (1999) proposed minimum G_2-aberration to rank nonregular designs. These criteria use the J-characteristics of designs to compare them. Suppose there is an $n \times m$ design matrix $\mathbf{D} = (d_{ij})_{n \times m}$ with $d_{ij} = \pm 1$. Let $Z_m = \{1, \ldots, m\}$ be an index set of the columns

of D. For every subset s of Z_m, define $J_s(\mathbf{D}) = \sum_{i=1}^{n} \prod_{j \in s} d_{ij}$. The collection of J_s values is known as the J-characteristics of design $\mathbf{D}$. Let $n = 4t$ and f_{kj} be the frequency of k-column combinations (k is the cardinality of s) that give $|J_s| = 4(t + 1 - j)$ for $j = 1, \ldots, t$, where $|J_s|$ is the absolute value of a J-characteristic. The confounding frequency vector (CFV) of design $\mathbf{D}$ is defined to be

$$F(\mathbf{D}) = [(f_{11}, \ldots, f_{1t}), (f_{21}, \ldots, f_{2t}), (f_{31}, \ldots, f_{3t}), \ldots, (f_{m1}, \ldots, f_{mt})]$$

Note that $(f_{11}, \ldots, f_{1t})$ is simply the frequency of non-zero column sums of $\mathbf{D}$ in reverse order, which for a balanced two-level design are all zero since there are an equal number of ± 1.

Let $f_l(\mathbf{D_1})$ and $f_l(\mathbf{D_2})$ be the lth entries of $F(\mathbf{D_1})$ and $F(\mathbf{D_2})$, respectively. Let r be the smallest integer such that $f_r(\mathbf{D_1}) \neq f_r(\mathbf{D_2})$. If $f_r(\mathbf{D_1}) < f_r(\mathbf{D_2})$ then $\mathbf{D}_1$ has less G-aberration than $\mathbf{D_2}$. If no design has less G-aberration than $\mathbf{D_1}$, then $\mathbf{D_1}$ has minimum G-aberration (Tang and Deng 2003).

Minimum G_2-aberration designs are also derived from the J-characteristics. Let $A_k(\mathbf{D}) = N^{-2} \sum_{|s|=k} |J_k|^2(s; \mathbf{D})$. The generalized wordlength pattern is defined to be $(A_1(\mathbf{D}), \ldots, A_m(\mathbf{D}))$. For two designs $\mathbf{D_1}$ and $\mathbf{D_2}$, let r be the smallest integer such that $A_r(\mathbf{D_1}) \neq A_r(\mathbf{D_2})$. If $A_r(\mathbf{D_1}) < A_r(\mathbf{D_2})$ then $\mathbf{D_1}$ has less G_2-aberration than $\mathbf{D_2}$. If no design has less G_2-aberration than $\mathbf{D_1}$, then $\mathbf{D_1}$ has minimum G_2-aberration. Both G-aberration and G_2-aberration criteria reduce to minimum aberration for regular designs (Tang and Deng 2003). Minimum G-aberration and minimum G_2-aberration have been well researched and have been used to develop several design catalogues.

The J-characteristics of a design are indications of the non-orthogonality of effect columns in the design's model matrix. Therefore G-aberration and minimum

G_2-aberration are both related to the quality of the model coefficient estimates. Unfortunately for many designs, the G-aberration and G_2-aberration rankings do not agree.

3.2.2 Estimation Capacity

Another general way to evaluate designs is by model robustness, often measured by some variant of estimation capacity. Estimation capacity, as first introduced by Sun (1993), measures the number of estimable models containing all of the ME evaluated in a design and a given number of q 2FI's. Estimation capacity was used by Li and Nachtsheim (2000) and Li (2006) to create model robust designs for screening experiments.

Loeppky, Sitter et al. (2007) introduced a variant of estimation capacity known as projection estimation capacity (PEC). Instead of the potential model space involving all ME in the screening design, the model space for PEC involves all q ME in q-factor projections of the original design plus all of the associated 2FI. Using the completed catalog of non-isomorphic $\mathrm{OA}(16; 2^{\mathrm{m}}; 2)$ and $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$, Loeppky et al. were able to identify optimal 16-run and 20-run designs in terms of PEC.

The projection estimation capacity of experimental designs is important when screening all of the effects in a full ME and 2FI model. Schoen (2010) looked at OAs of strength three with full estimation capacity (the ability to estimate all terms in a full ME+2FI model) and discussed when such an OA performed better than an optimum design. A screening design for m factors will often be supersaturated for the number of terms in an m-factor full model. However, if an experimenter feels that due to factor sparsity only a few ME will be active then the experimenter will use the supersaturated

design and screen for effects in a full model for a subset of the original ME. Since it is not known in advance which ME are active, the projection estimation capacity can be evaluated to get an estimate of the design's model robustness.

3.2.3 The Pareto Front

It is common when ranking designs based on multiple design criteria that a design which is optimal for one criterion is not optimal for other criteria. According to Lu, Anderson-Cook et al. (2011), one criterion rarely encompasses all of the qualities a design should have to be effective. Consequently, it is recommended to consider the Pareto efficient set, or Pareto front, of designs. In general, a decision element A is included in a Pareto efficient set if there does not exist an element B which outranks element A in every optimality criterion. That is to say element A is not dominated by any other element. Examining the Pareto front of designs allows an experimenter to find a design which has favorable qualities across multiple criteria and will therefore have a greater chance of performing well in varying experimental contexts.

3.3 Methodology

The methodology used to identify and recommend 20-run no-confounding designs from the catalog of $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$ involved a three stage process. First every design in the catalog was evaluated and ranked according to 24 design metrics. Next a candidate set of optimal no-confounding designs was produced by determining the Pareto efficient set of designs based on the ranks of the 24 criteria. Finally, an empirical analysis of the model-fitting capability of the designs in the Pareto efficient set was conducted. The results of the empirical analysis ultimately determined which designs were recommended as the 20-run no-confounding designs.

The empirical results were also used as dependent variables in a supplementary regression analysis which used the design-ranking criteria as regressor variables. The goal of the analysis was to determine which design-ranking criteria best predicted the optimal design in terms of model-fitting capability. The information concerning design-ranking criteria will be valuable in studying OAs of larger dimension for which exhaustive catalogs do not exist, such as the OA $(28; 2^m; 2)$ catalog.

3.3.1 Developing the Monte Carlo Simulation

The empirical analysis was accomplished using a Monte Carlo simulation. The simulation was an uncommon aspect of the methodology used to identify the 20-run no-confounding designs. Marley and Woods (2010) used a Monte Carlo simulation to compare analysis strategies for supersaturated designs, but a journal article concerning Monte Carlo simulation as a method to compare designs has not been published.

Many details of the simulation were rigorously analyzed to ensure that the simulation results would accurately model the real-world application of OAs as screening designs. This subsection presents the details of the simulation development process. It covers three important aspects of the Monte Carlo simulation that needed to be determined: the analysis method and the associated p-value thresholds; the number of potential ME and 2FI in the simulated models; and the number of simulations. This discussion serves as a prelude to the subsequent presentation concerning the methodology used to identify the 20-run no-confounding designs.

The Monte Carlo simulation generated a polynomial response model with a randomly generated number of ME and 2FI, as well as random coefficients for the effects and random experimental error. A response vector was created based on the polynomial

model and the factor levels in a screening design. The simulation then automated the analysis of the randomly generated responses and fit a model.

The terms and coefficients in the estimated model were compared to the terms and coefficients in the randomly generated model and a binary variable recorded whether a type I or type II variable selection error occurred. The simulation was iterated five thousand times, each time generating an original random model. After a given number of iterations, the data was aggregated to report the percentage of models with at least the active effects, the percentage of models with type I variable selection errors and the percentage of models with type II variable selection errors. Designs with the minimum percentage of type II variable selection errors were identified as the optimal design among each of the m-column classes of OAs. See Figure 1 for a flow chart of the Monte Carlo simulation.

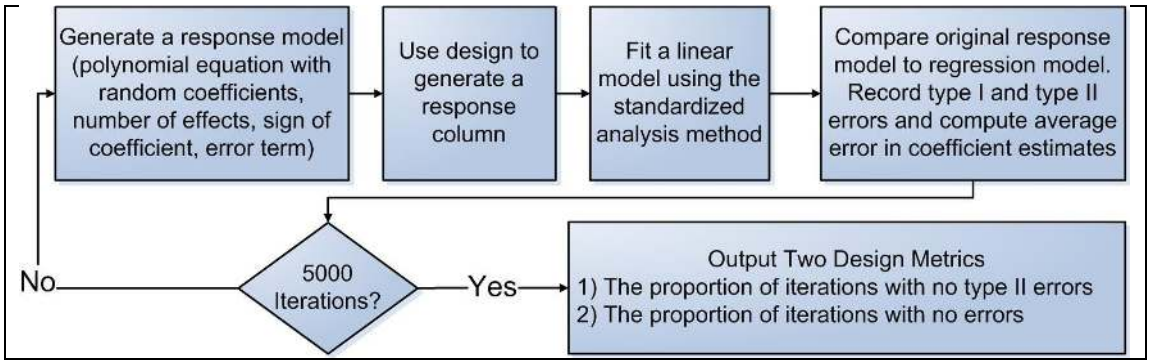

Figure 1. Monte Carlo Simulation Flow Chart

The number of simulations was based on the Normal approximation of the standard deviation for a binomial proportion $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. It was required that $n = 4444$ when $p = 0.5$ to achieve the desired standard deviation of $\sigma_{\hat{p}} = 0.75\%$. After weighing the simulation time against the requirements for the width of a 95% confidence

interval, it was determined that 5000 simulation runs would create sufficiently narrow confidence intervals to feasibly compare the various OAs.

3.3.1.1 Selecting the Analysis Method

Since the smallest ME+2FI full model under consideration was the 6-factor model with 22 terms, an analysis method suitable for supersaturated designs was required. Fifteen variants of forward mixed stepwise regression and forward stepwise regression were considered for the analysis method. The variants of stepwise regression involved different model selection methods using p-values and the Akaike's Information Criterion Corrected (AICc). AICc was a candidate model-selection criterion since it is particularly suited for situations where there are few active model terms. For the model selection approaches using p-values, several levels were considered for the threshold of the entering and exiting variables.

Doctoral research in Shinde (2012) suggested using a 2-stage forward stepwise analysis approach to both improve model accuracy and enable the estimation of larger models. In the first stage, the analysis is conducted considering only ME terms. In the second stage, only the active ME terms from stage one and all associated 2FI involving the active ME are considered. Both the forward mixed stepwise regression (p-values) and the forward stepwise regression (using AICc) were examined using a 1 and 2 stage process.

Since the analysis software did not perform bi-directional elimination when the AICc criterion was used to rank potential models, a modification of the forward selection procedure was considered. The modification involved eliminating variables in the final model which had a p-value above a certain threshold. Several "post-stepwise variable-

elimination p-value thresholds" were tested with the one-stage and two-stage stepwise procedures to determine if adding this step in the analysis would improve the model-fitting results.

To determine which of the 15 analysis techniques should be selected for the Monte Carlo simulation used to evaluate the OAs, an initial simulation was conducted to determine which analysis procedure resulted in the lowest percentage of type II errors and highest percentage of correct models. The OA chosen for the simulation was OA #74 in the $\mathrm{OA}(20; 2^6; 2)$ catalog. This design was selected due to preliminary analysis which indicated this design was in the Pareto efficient set. Ten thousand Monte Carlo simulations were run for each analysis technique.

The random models had two to five ME and zero to two 2FI. Li, Sudarsanam et al. (2006) conducted a meta-analysis of published experiments involving a DOE methodology and reported that for the 113 combined experimental data sets, 41% of the potential ME were active and 11% of the potential 2FI were active. According to these results, when considering six potential factors, it is expected that there will be 2-3 active ME and 1 active 2FI. The ranges for the number of ME and 2FI in the polynomial models were consistent with the model meta-analysis, as well as with the sparsity-of-effects principle.

Each ME coefficient random variable X_i ranged from $2\sigma^2 \leq X_i \leq 5.44\sigma^2$ and the 2FI coefficient random variable Y_i ranged from $2\sigma^2 \leq Y_i \leq 3.5\sigma^2$. The sign of the coefficient was determined by a binomial random variable with $p = 0.5$. Data was collected concerning the percentage of trials out of 10,000 in which at least the correct model was estimated (no type II errors occurred) and the percentage of trials out of

10,000 in which the correct model was estimated (no type I or type II errors occurred). Figure 2 shows the results of this preliminary experiment.

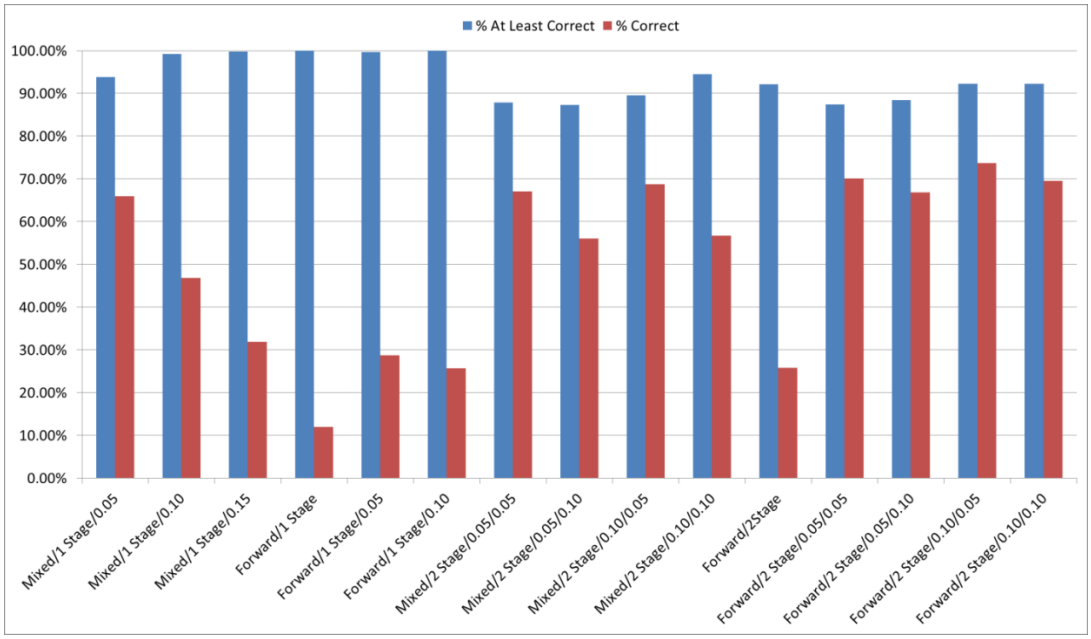

Figure 2. Simulation Results for Candidate Analysis Methods

Based on the results in Figure 2, the chosen analysis method was the 2-stage forward stepwise regression, with AICc as the model selection criteria and post-stepwise variable elimination p-value thresholds of 0.10 and 0.05 for stages 1 and 2 respectively. This choice of analysis procedure reflected the author's preference to risk a 7.78% type II error rate in order to have a 73.72% chance (the highest percentage out of all the analysis options) at producing a model with no type I or type II errors. This choice of analysis procedure allows some flexibility in comparing OAs on the basis of type II error percentages and is also consistent with the philosophy of using no-confounding designs under the constraints of a limited experimental budget. However an alternative analysis procedure preferred by some experimenters may be the 1-stage mixed stepwise regression using a p-value of $p = 0.10$ for the threshold to enter or remove variables.

This method reduces the type II error rate to a negligible 1% but the percentage of models without type I and type II errors falls to 47%.

3.3.1.2 The Number of ME and 2FI in Simulated Models

An important consideration for the Monte Carlo simulation was the appropriate maximum number of ME and 2FI effects in the randomly generated response models. For this research the models were built using the heredity constraint, i.e. 2FI were only allowed in the model if both of the ME were also in the model. Choosing integers p and q such that the simulation generated a random number of 2 to p ME and a random number of 0 to q 2FI required balancing poor estimation performance with a thorough design capability analysis. A 10,000-run simulation was run using OA #74 in the $\mathrm{OA}(20; 2^6; 2)$ catalog. A range of values for p and q were investigated with $p \in \{1, \dots, 5\}$ and $q \in \{0, \dots, 5\}$. For the ME and 2FI, the respective coefficient random variables X_i and Y_i ranged from $2 \leq X_i \leq 5.44$ and $2 \leq Y_i \leq 3.5$. The sign of the coefficient was determined by a Bernoulli random variable with $p = 0.5$. The analysis was conducted using 2-stage forward stepwise regression, with a post-stepwise variable-elimination p-value threshold of 0.10 for stage 1 and 0.05 for stage 2. Table 2 shows the results in terms of the percentage of trials that resulted in models without type II errors. This in an important performance metric for screening designs, as further experimentation will generally eliminate inactive effects that are selected during the screening process, but effects that are considered inactive in the screening phase are usually omitted in subsequent experimentation.

Table 2. Percentage of Type II Errors for p ME Effects and q 2FI Effects

% At Least Correct Model		Model Simulations = 5000					
	q (2FI)						Row Totals
p (ME)	0	1	2	3	4	5	
2	100.00%	100.00%					100.00%
3	100.00%	98.74%	89.03%	66.12%			88.71%
4	100.00%	96.29%	74.24%	48.28%	32.23%	22.12%	62.62%
5	100.00%	90.93%	58.16%	32.65%	17.57%	9.27%	52.06%
Column Totals	100.00%	97.67%	75.89%	51.71%	25.06%	15.90%	75.97%

Since the no-confounding designs are recommended as an option to experimenters with limited budgets who hope to avoid subsequent experimentation to de-alias fully confounded effects, a second important metric is the percentage of trials that resulted in completely correct models. Table 3 shows the percentage of simulations that resulted in models without type I or type II variable selection errors.

Table 3. Percentage of Correct Models for p ME Effects and q 2FI Effects

% Correct Model		Model Simulations = 5000					
	q (2FI)						Row Totals
p (ME)	0	1	2	3	4	5	
2	75.66%	98.74%					87.35%
3	74.22%	87.89%	85.32%	65.14%			78.22%
4	65.63%	74.94%	65.66%	43.60%	30.09%	21.18%	50.28%
5	62.50%	61.28%	47.28%	26.74%	15.10%	8.02%	37.29%
Column Totals	71.80%	86.53%	68.73%	48.29%	22.76%	14.81%	63.42%

In order to rigorously compare the OAs in the catalog, it was decided to randomly generate models with two to five ME and zero to two 2FI. This range of models not only covered the practical model size of two to five ME and 0 to one 2FI, but included larger models that might help differentiate the performance of the best OAs. As mentioned previously in this chapter, limiting the scope of analysis to models of at most seven effects was based on research by Li et al. (2006).

In summary, the Monte Carlo simulation repeatedly generated random response models and used an OA's factor level settings to generate a vector of responses. A realization of a standard normal random variable $Z \sim$ Normal(0,1) was added as a random noise term. There were several aspects of each response model that were randomly generated. The number of ME and 2FI were determined by discrete uniform random variables, $A \sim U(2,5)$ and $B \sim U(0,2)$, respectively. The coefficients of the ME were realizations of a random variable X which were generated via the equation $X_i = e^{0.69+V}$, with $V \sim$ Uniform(0,1). The result produced random coefficient values in the range $2 \leq X_i \leq 5.44$. The coefficients of the 2FI were realizations of a random variable Y_i which was generated via the equation $Y_i = e^{0.69+W}$, with $W \sim$ Uniform(0,0.56). The result produced random coefficient values in the range $2 \leq Y_i \leq 3.5$. Since the random error term was generated by a standard normal random variable, the coefficients can be viewed as multiples of $\sigma^2 = 1$. The sign of the coefficients was determined by a Bernoulli random variable $T \sim$ Bernoulli(0.5).

3.3.2 Reducing the Set of Candidate Designs

In order to identify the optimal OAs for the 20-run no-confounding designs, every non-isomorphic OA in the catalog of OA$(20; 2^m; 2)$ for m $\in \{6, \ldots, 12\}$ factors was evaluated. Table 4 provides the number of non-isomorphic OA for each value of m.

Table 4. Count of Non-isomorphic OAs

Number of columns	6	7	8	9	10	11	12
Count of Non-isomorphic OAs	75	474	1603	2477	2389	1914	1300

After computing the confounding frequency vectors, it was observed that $f_{31} = f_{41} = 0$ for every OA. Therefore, unlike 16-run OAs, there are no 20-run OAs

which produce ME+2FI model matrices that have fully confounded effects. In other words, every $OA(20; 2^m; 2)$ for $m \in \{6, \dots, 12\}$ factors is a no-confounding design with orthogonal ME. Consequently, no OAs could be removed from consideration based on the presence of fully-confounded effects. The simulation time required to perform 5000 simulated model analyses was approximately one hour. The time requirement for the simulation per design made it necessary to reduce the number of candidate designs before the simulation phase of the analysis.

A reduced design set was created for the $OA(20; 2^m; 2)$ for $m \in \{7, \dots, 12\}$ factors. There were only 75 non-isormorphic designs in the $OA(20; 2^6; 2)$ catalog so it was not reduced. The flow chart for the design reduction process is shown in Figure 3.

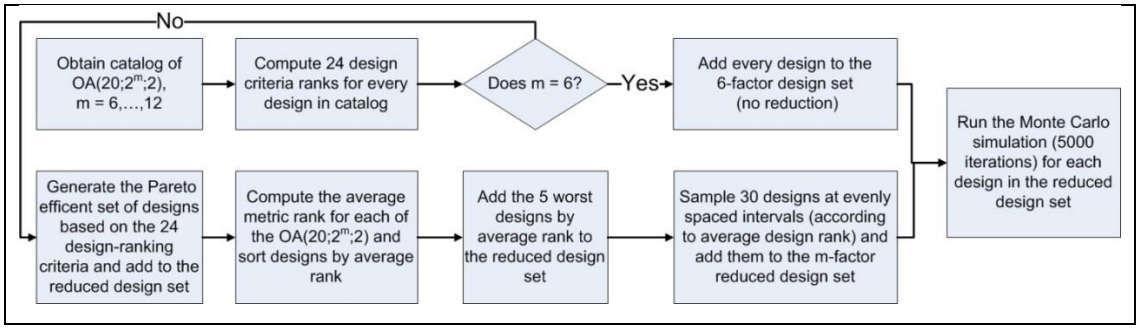

Figure 3. The Design Reduction Flow Chart

Twenty-four metric values were used to rank the OAs and create a Pareto efficient set of candidate optimal designs. The G-aberration and G_2-aberration metrics were calculated in the usual way but with an abridged CFV. The CFV of a design $\mathbf{D} \in OA(20; 2^m; 2)$ is defined to be

$$F(\mathbf{D}) = [(f_{11}, \dots, f_{15}), (f_{21}, \dots, f_{25}), (f_{31}, \dots, f_{35}), \dots, (f_{m1}, \dots, f_{m5})]$$

Now since every design $\mathbf{D} \in OA(20; 2^m; 2)$ is balanced and orthogonal, then $f_{11} = \cdots = f_{15} = f_{21} = \cdots = f_{25} = 0$. Also since it is often assumed that 3FI and higher

are insignificant, then $(f_{m1}, \ldots, f_{m5})$ can be disregarded for $m > 4$. Thus the abridged CFV used to develop the G-aberration, G_2-aberration, and several other design metrics is $F(\mathbf{D}) = [(f_{31}, \ldots, f_{35}), (f_{41}, \ldots, f_{45})]$. For the $OA(20; 2^m; 2)$, there was not a case where $f_{31} = f_{32} = \cdots = f_{45} = 0$. Therefore the rankings of G-aberration and G_2-aberration for the OAs in this chapter using the abridged CFV are completely consistent with the rest of the literature.

The $E(s^2)$ metric used previously in Jones and Montgomery (2010) to select the 16-run no-confounding designs from the $OA(16; 2^m; 2)$ was calculated in the usual way. The "sum of non-zero frequencies" metric is calculated from the CFV as $\sum_{m=3}^{4}\sum_{t=1}^{5} f_{mt}$. The "combined $F(\boldsymbol{D})$ G-aberration" metric is calculated by applying the G-aberration criteria to a combined CFV, $F(\mathbf{D})' = [(f_{31} + f_{41}, f_{32} + f_{42}, \ldots, f_{35} + f_{45})]$. "Combined $F(\boldsymbol{D})$ G-aberration" puts equal weight on the degree of non-orthogonality between 2FI and the degree of non-orthogonality between ME and 2FI.

The projection properties of the designs are evaluated with five types of metrics: the minimum number of unique factorial points out of all k-column projections for $k \in \{4,5,6\}$; the average number of unique factorial points in k-column projections for $k \in \{4,5,6\}$; the maximum number of unique factorial points out of all k-column projections for $k \in \{4,5,6\}$; the projection estimation capacity (PEC) of all 5-column projections; and the min, max and average variance inflation factors (VIF) for all k-column projections for $k \in \{3,4,5\}$.

The choice of k-column combinations for the metrics above were based on several properties of $OA(20; 2^m; 2)$. The 3-factor projections were not considered for most metrics because every three-column projection of an $OA(20; 2^m; 2)$ for $m \in$

$\{6, \ldots, 12\}$ factors contains a 2^3 full factorial. Hence the min, max, and average unique points for these projections always equal eight. As stated in Loeppky, Sitter et al. (2007), if $p_q(\boldsymbol{D})$ is the percentage of estimable models containing q main effects and all associated interactions for a design $\boldsymbol{D}$, then for the $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$ it is true that $p_1 = p_2 = p_3 = p_4 = 1$. Therefore only p_5 was examined. The VIFs for 6-column projections and p_6 were not evaluated because the full model for 6 factors has 22 terms, which is more than can be estimated with the 20-run designs.

The number of designs in the $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$ catalog for $\mathrm{m} \in \{7, \ldots, 12\}$ factors was reduced by identifying the Pareto efficient set of OAs. The Pareto front was defined by the 24 design-ranking criteria. The OAs on the Pareto front were those designs $\boldsymbol{D}_i$ for which there did not exist a design $\boldsymbol{D}_j$ which outranked design $\boldsymbol{D}_i$ in all 24 design ranking criteria. In some cases, such as G-aberration, the design criterion is not based on a single scalar value but rather on a vector of values (the CFV). Therefore criteria ranks were used as a surrogate for the actual criteria values so that each of the 24 design-ranking criteria could be represented by a scalar. The design ranks were then organized in a matrix with rows corresponding to an OA in the catalog and columns corresponding to the 24 criteria. The following algorithm was then used to identify the designs on the Pareto front.

Step 0: Create a matrix **A** with rows corresponding to every m-column OA and columns corresponding to the design criteria, so that a_{ij} is the ith OA's rank by criteria j

Step 1: Select the first non-evaluated design-criteria column j and sort in ascending order

Step 2: For each row $\boldsymbol{a}'_k, k > 1$,

a) Check if $a_{kl} \leq a_{1l}$ for any column $l \neq j$

b) If $a_{kl} \leq a_{1l}$ then retain the design row in the matrix

c) If $a_{kl} > a_{1l}$ for all l then delete the row from the matrix

Step 3: Return to Step 1 until all columns are evaluated

Step 4: The remaining matrix rows are the designs on the Pareto front

The Pareto front of designs was the largest portion of the reduced design set, but additional designs were added to the reduced design set so that the model fitting capability of the entire catalog of $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$ could be evaluated. First every design in the catalog of $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$ was assigned an average rank across all 24 design ranking criteria. Then the five worst OAs by average design rank across all 24 criteria were added to the reduced design set so that the empirical results of the best designs could be compared to the worst designs. Finally the OAs were sorted by average rank and an evenly spaced sample of 30 OAs were added to the reduced design list. This allowed for a more robust multiple regression analysis using the 24 design ranking criteria as independent variables and the simulation results as the dependent variables. The goal was to create a data sample that represented the complete spectrum of model estimation capability.

3.3.3 Identifying the No-confounding Designs

The optimal 20-run no-confounding design was determined by the empirical results of a Monte Carlo simulation. The output metric primarily used to assess the designs was the percentage of simulations that resulted in the estimation of at least every active effect, called the *% at least correct*. If inactive effects were also identified as active this was deemed acceptable in the function of a screening design. The percentage of simulations that resulted in a model with no type I or type II variable selection errors, called the *% correct*, was used as a secondary comparison metric.

The procedure to identify the optimal 20-run no-confounding designs was a multi-stage process involving an initial Monte Carlo simulation phase, a regression model analysis of the simulation output data, a model validation phase, and a secondary round of simulations on the final design candidate list. A flow chart of the optimal design identification process is shown in Figure 4.

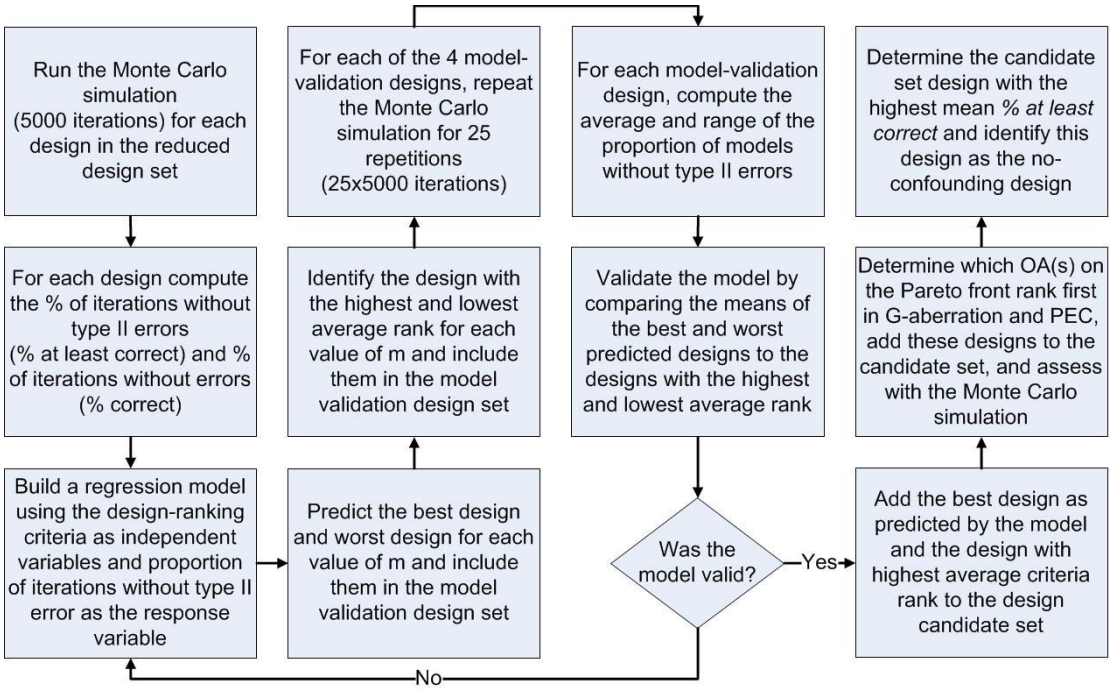

Figure 4. The No-confounding Design Identification Process

A Monte Carlo simulation involving the estimation of 5000 random models was run for each of the 75 OAs in the $\mathrm{OA}(20; 2^6; 2)$ catalog and for each design in the reduced design set for $m > 6$ factors. The model-fitting capability metrics were recorded for each design. The percentage of iterations without type II errors (the *% at least correct*) was used as the response variable to build a regression model which used the 24 design-ranking criteria and their interactions as independent variables.

The regression models were validated and improved using additional Monte Carlo simulation on reference OAs which had the highest and lowest average rank across all 24 design-ranking criteria. If the best and worst designs as predicted by the model were close in performance to the reference designs, the model was considered valid. Once a validated regression model was created, the best design of m-factors according to the model was placed in a final design candidate list.

In addition to the best design as predicted by the model, the final candidate list also included the design with the best average design-criteria rank, the design with the highest rank according to G-aberration, and the design with the highest rank according to PEC. Research with the 24-run no-confounding designs gave insight into the fact that having uncorrelated main effects was the primary indicator of the best performing designs. Unfortunately, there are no resolution IV 20-run designs. However, G-aberration ranks 20-run designs according to the minimum maximum ME correlation with 2FI. Thus designs which ranked first in G-aberration were included in the candidate set. The 24-run design research also indicated that considering PEC in conjunction with G-aberration was important. Therefore designs which ranked first in PEC were also included in the candidate set.

A final round of simulations was conducted for the short list of final design candidates. 60,000 simulations were run with random combinations of 2-5 ME and 0-2 2FI, providing at least 5000 iterations per combination of p ME and q 2FI. The data for each design was compared in order to determine the m-factor no-confounding designs.

3.4 Results

The Monte Carlo simulation proved to be an effective tool for the empirical analysis of the $\mathrm{OA}(20; 2^{m}; 2)$. The simulation data was not only useful in identifying the optimal 20-run no-confounding designs, but also in gaining insight into the design-ranking criteria and the analysis methods. Subsection 3.4.1 discusses the trade-offs involved, in terms of *% correct* and *% at least correct*, when using different stepwise analysis methods. Subsection 3.4.2 includes the regression analysis on the 24 design-ranking criteria. This subsection shows which criteria were the best predictors of the *% at least correct* in simple and multiple regression models. Finally the data used to identify the 20-run no-confounding designs is presented in subsection 3.4.3.

3.4.1 The Effects of the Model-Fitting Method on Type I and Type II Errors

In addition to using 2-stage forward stepwise regression with the AICc, the 1-stage mixed stepwise regression analysis method was also used to investigate the effect of different analysis methods across the entire catalog of $\mathrm{OA}(20; 2^{6}; 2)$. Figure 5 is a dot plot of the *% at least correct* after 5000 simulated models were analyzed by each OA.

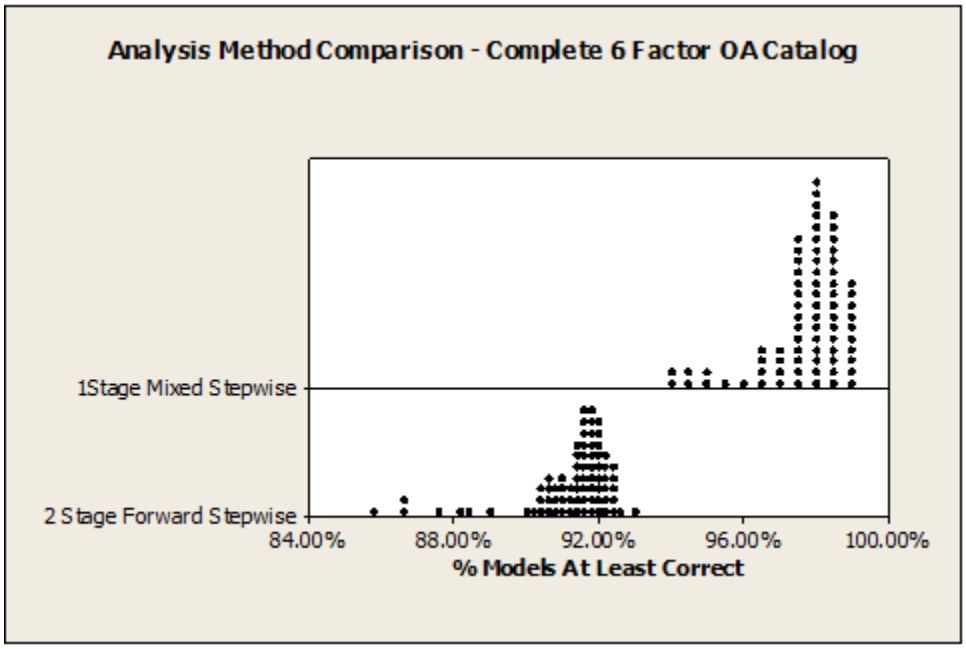

Figure 5. Analysis Method Comparison on OA(20;2^6;2) – *% at Least Correct*

It is evident that 1-stage mixed stepwise regression is the superior analysis method in terms of identifying all of the active model effects, although the average *% at least correct* using 2-stage forward stepwise regression is still close to 93%. However, there is an important trade-off made when using 1-stage mixed stepwise regression, i.e. the *% correct* drops considerably. This is illustrated in Figure 6, which is a dot plot of the *% correct* out of 5000 simulated models analyzed by each OA.

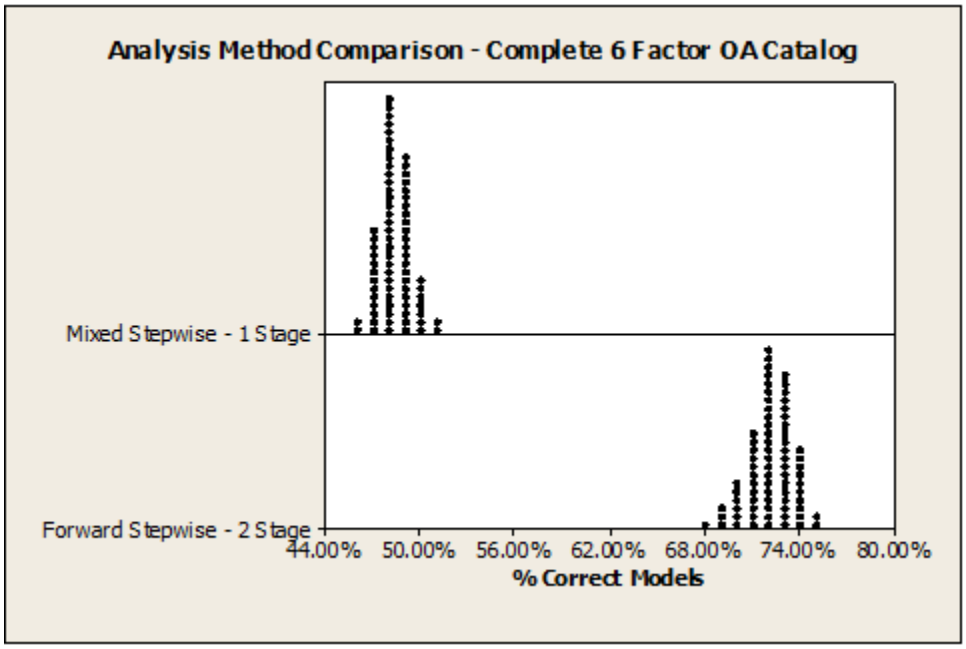

Figure 6. Analysis Method Comparison on OA(20;2^6;2) – *% Correct*

The figure shows that by using 2-stage forward stepwise regression, the *% correct* improves to a mean of 75% from a mean of 53% using 1-stage mixed stepwise regression.

3.4.2 The Regression Models and the Validation Process

To aid in identifying the no-confounding designs, simple and multiple regression models were built using the 24 design-ranking criteria as independent variables and the Monte Carlo simulation results as the dependent variable. Using the regression model for prediction of the optimal design was a methodology more robust to the simulation noise than simply picking the design corresponding to the best simulation value. Table 5 shows the results of the regression modeling process for all OAs of 6-12 factors.

Table 5. Regression Models

Factors	Model Type	Model Terms	R^2	R^2-Adjusted	Model Best OA	Model Worst OA
6	Simple Regression	Avg Unique Pts - 5 Col	0.704		74	9
	Multiple Regression	G-aberration Avg VIF – 4 Factor (G-aber)*(Avg VIF (4-fact)	0.8669	0.8613	74,75	7,9
7	Simple Regression	Avg Unique Pts - 6 Col	0.8731		472 (16 Tied)	44
	Multiple Regression	Avg Unique Pts - 6 Col PEC – 5 Col Proj Avg VIF – 3 Col (Avg Unique Pts – 6-Col)^2	0.9351	0.9307	337	44
8	Simple Regression	Avg VIF – 3 Col	0.345		1599 (6 Tied)	436
	Multiple Regression	Min Unique Pts - 6 Col Avg VIF – 3 Col	0.4122	0.3932	1599	150
9	Simple Regression	Avg VIF – 4 Col	0.2205		2476, 2477	1116
	Multiple Regression	N/A - No Improvement				
10	Simple Regression	Avg Unique Pts - 6 Col	0.112		2388	1213
	Multiple Regression	N/A - No Improvement				
11	Simple Regression	Avg VIF – 5 Col Proj	0.2654		1883	19
	Multiple Regression	N/A - No Improvement				
12	Simple Regression	Avg VIF – 5 Col	0.0659		1176	14
	Multiple Regression	G-aberration Min Unique Pts - 5 Col PEC – 5 Col	0.1459	0.1208	885	39

Column 3 in Table 5 lists the design-ranking criteria that are the independent variables in the best simple or multiple regression model. Columns 4 and 5 show the R^2 and R^2-adjusted values used to assess and choose between models. Columns 6 and 7 show the OA IDs of the best and worst OAs as predicted by the regression models. For example with the 6-factor designs, the best independent variable for the simple linear

regression model used to predict the mean *% at least correct* was the average unique points out of all 5-column projections of the OAs. The model had an R^2 value of 0.704 and predicted OA #74 would have the highest average *% at least correct*; OA# 9 was predicted to have the lowest average *% at least correct*.

Table 5 also shows that a multiple regression model improves the R^2 value to 0.8669 by using three independent variables: G-aberration, the average VIF for 4-factor full models, and an interaction term involving these variables. However, the prediction becomes more ambiguous as there are now two best models (OA#74 and #75) and two worst models (OA #7 and OA #9).

Figure 7 provides an example of the model validation process. The designs with the highest average design criteria rank (OA #72) and the lowest average design criteria rank (OA #7) were used as reference designs to compare with the best and worst predicted models. The analysis simulation of 5000 random models was repeated 25 times for each of the candidate and reference designs. The mean *% at least correct* metric was used to compare the designs. Figure 7 also includes the minimum and maximum *% at least correct* from the 25 repeated simulations of 5000 random models.

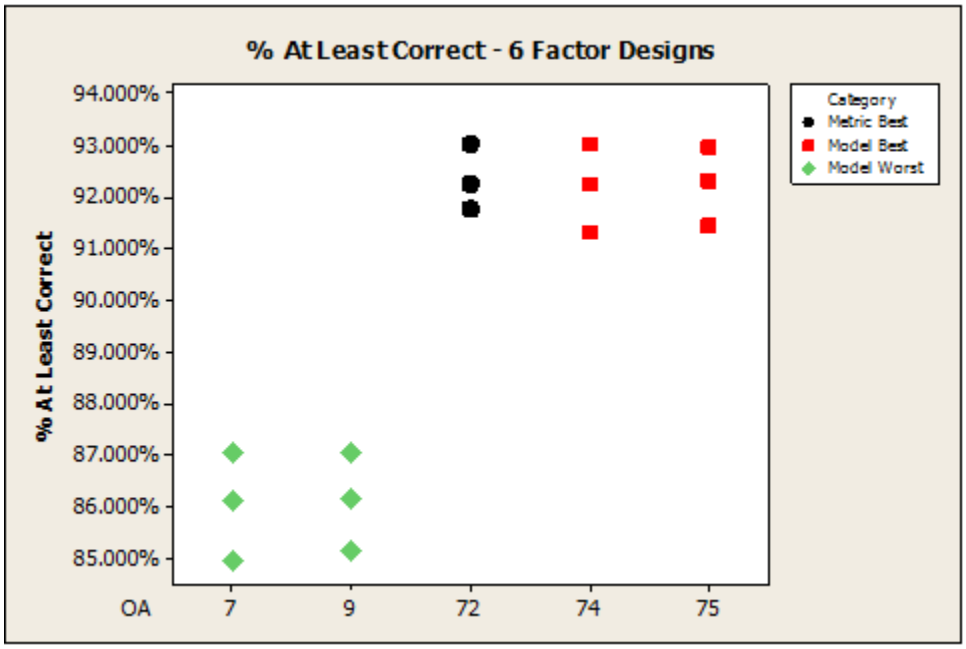

Figure 7. Candidate 6-Factor Optimal Design Comparison – *% at Least Correct*

The results show there is little empirical difference between OA #74 and OA #75. The data for OA #72 seems to validate that the model has selected the best candidate designs. The figure also shows a statistical difference between the mean *% at least correct* metric for the worst pair of OAs and the best triplet of OAs. However this difference is surprisingly small, as the mean *% at least correct* metric for the worst design is approximately 86%.

3.4.3 Selecting the 20-Run No-confounding Designs

A final design candidate list was created to compare the most promising designs from the OA catalog. The column headings in Table 6 are the categories used to identify the candidate designs and the rows correspond to the number of columns (factors) in the OAs.

Table 6. Final Candidate Designs

Factors	Best per Model OA ID	Best Avg Rank OA ID	G-aber #1 Rank OA ID	PEC #1 Rank OA ID
6	74,75	72	74	74
7	337	457	472	453
8	1599	1599	1599	1599
9	2476,2477	2232	2476,2477	2286
10	2388	2363	2389	2389
11	1883	1805	713	713
12	885	1299	426	426

Column 2 shows the designs that were predicted to be the best designs by the regression model. Column 3 shows the designs in each catalog that had the best average criteria rank. Column 4 shows the designs ranked first by G-aberration. Column 5 shows the designs ranked first by PEC.

Table 7 illustrates a result from the analysis of the G-aberration and PEC criteria which demonstrates the importance of considering multiple criteria in design selection.

Table 7. The Count of G-aberration and PEC Designs in the OA Catalog

Factors	G-aberration Count of #1 Ranks	PEC Count of #1 Ranks	Both Count of #1 Ranks
6	2	11	2
7	2	2	0
8	2	2	2
9	2	1	0
10	1	2	1
11	2	678	2
12	2	1	1

Column 2 shows the count of designs in the $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$ catalog that ranked first in G-aberration. Column 3 shows the count of designs in the $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$ catalog that ranked first in PEC. Finally, Column 4 shows the count of designs that ranked first in both categories. The rows correspond to $m = 6, \ldots, 12$.

By examining the data in Table 7, it is obvious that both categories are necessary to identify the best designs. Notice that 11 of the 75 designs in the 6-factor OA catalog

were ranked first in PEC. However, only two of these designs ranked first in G-aberration. The fact that two designs ranked first in both categories was a fortunate occurrence but is not guaranteed. As can be seen in the seven and nine factor cases, a design does not exist that ranks first in both the G-aberration and PEC criteria.

A final round of Monte Carlo simulation was conducted with each of the final candidate designs in Table 6. 60,000 simulations were run with random combinations of 2-5 ME and 0-2 2FI, providing at least 5000 iterations per combination of p ME and q 2FI. The *% at least correct* and the *% correct* were the metrics used to compare the designs. The *% at least correct* was the primary metric used in determining the optimal no-confounding design.

Figure 8 through Figure 21 show the simulation results, which are broken out by combinations of ME and 2FI. The notation (ME,2FI) = (4,1) means the data corresponds to simulated models with four ME and one 2FI. For each of the m-factor designs, the results for *% at least correct* are shown first, followed by the results for the *% correct*.

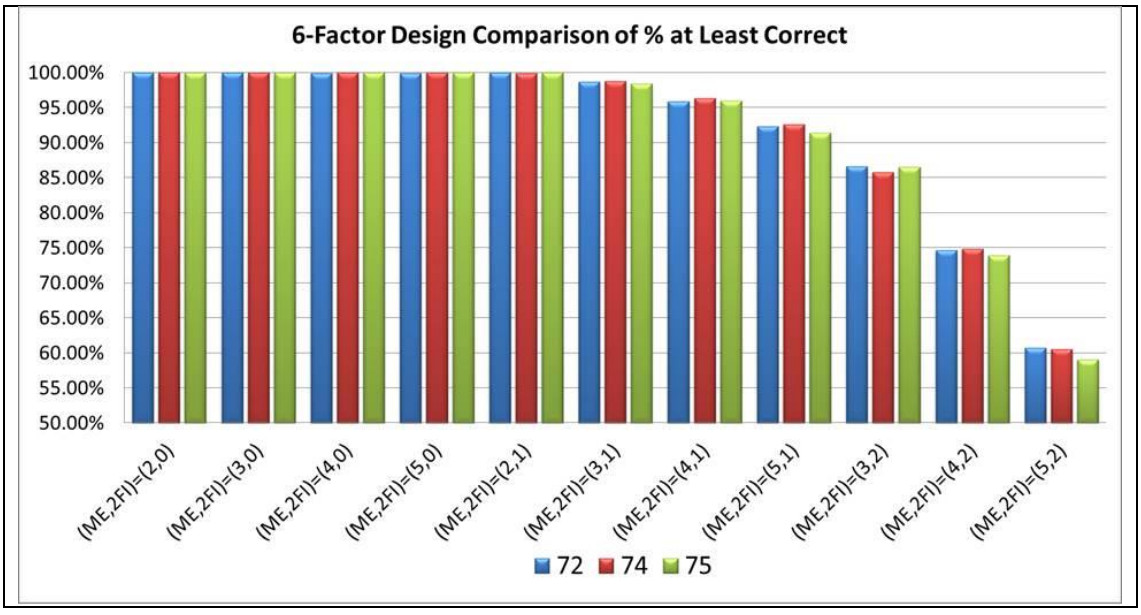

Figure 8. 6-Factor Design Comparison of *% at Least Correct* by p ME and q 2FI

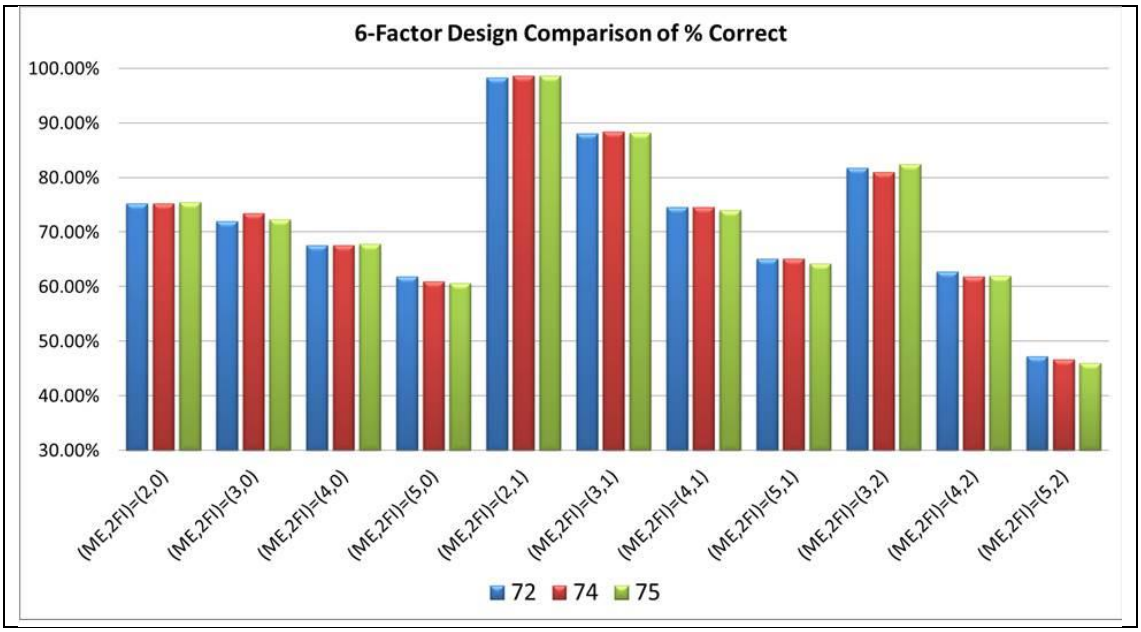

Figure 9. 6-Factor Design Comparison of *% Correct* by *p* ME and *q* 2FI

The differences in simulation results between the three 6-factor designs were very small and almost always within the range of noise in all but two combinations of ME and 2FI. On one of these combinations, the (ME,2FI) = (3,0) case, design 74 performed the best. Since OA #74 also ranked first in G-aberration and PEC, it was chosen as the best 6-factor no-confounding design. The design is shown in Appendix Table 75.

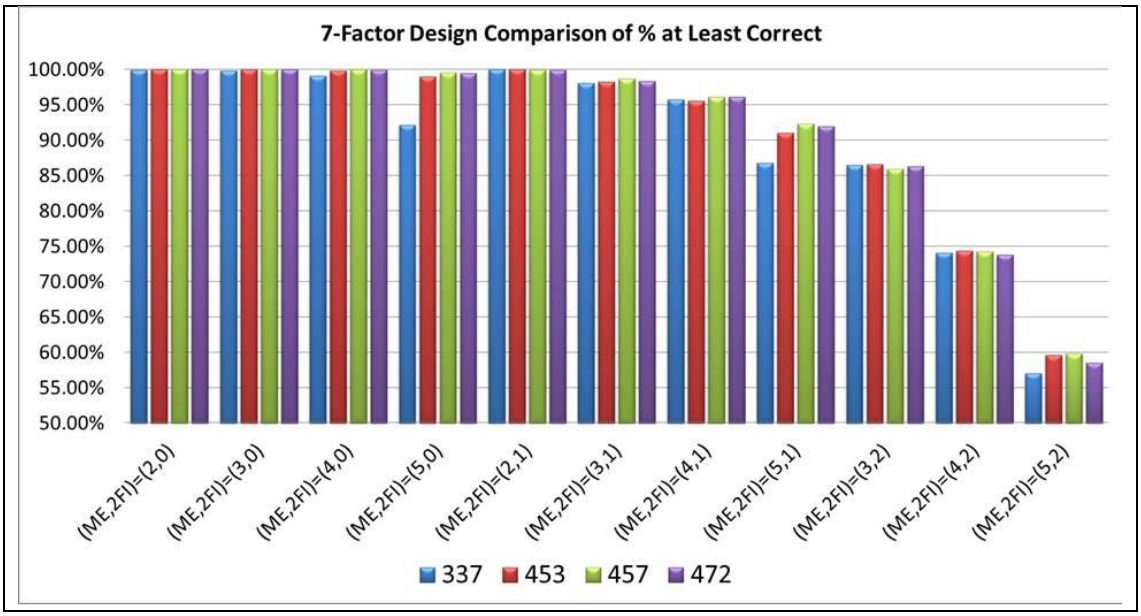

Figure 10. 7-Factor Design Comparison of *% at Least Correct* by *p* ME and *q* 2FI

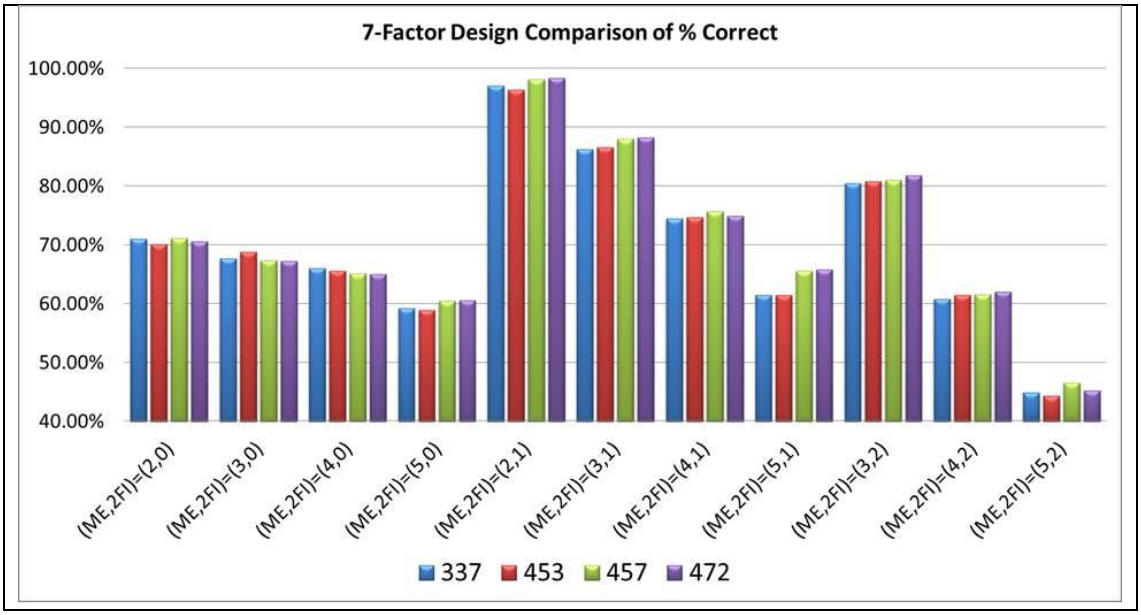

Figure 11. 7-Factor Design Comparison of *% Correct* by *p* ME and *q* 2FI

The data in Figure 10 and Figure 11 show the OA #457 is the best performing design overall. This design has the highest overall average for both *% at least correct* and *% correct*. In terms of the *% at least correct*, there are no instances where the better

performance of another design is statistically significant. The design is shown in Appendix Table 76.

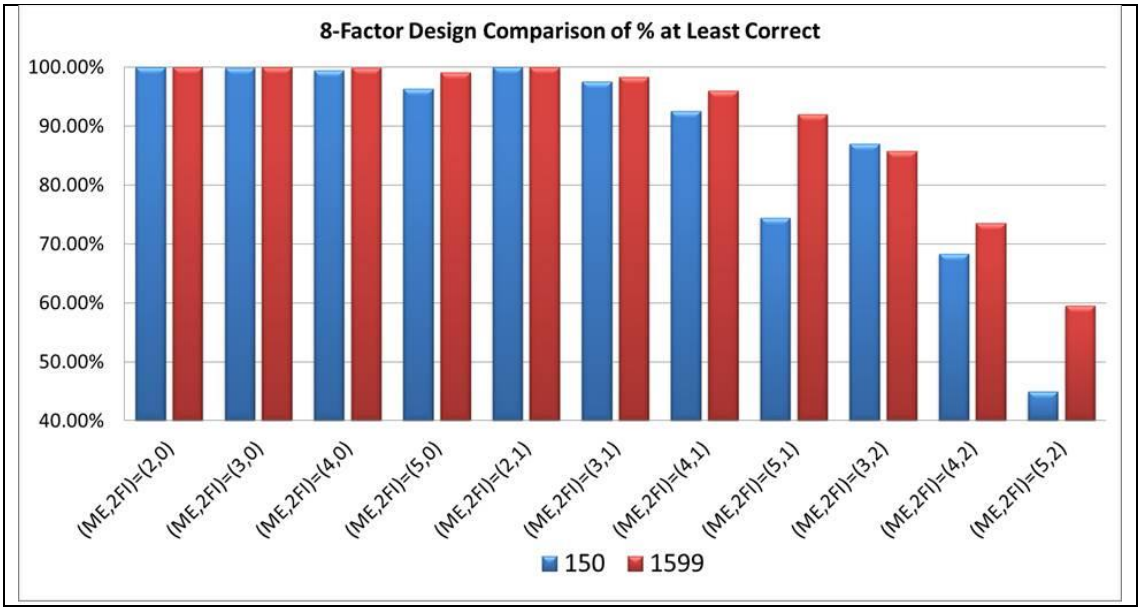

Figure 12. 8-Factor Design Comparison of *% at Least Correct* by *p* ME and *q* 2FI

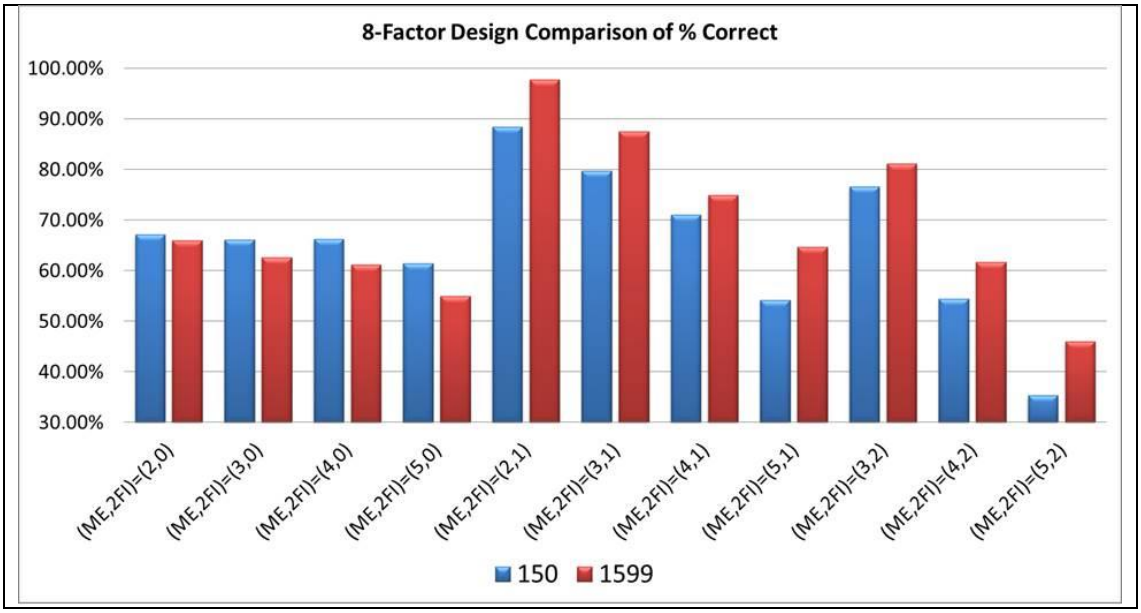

Figure 13. 8-Factor Design Comparison of *% Correct* by *p* ME and *q* 2FI

For the 8-factor designs, OA #1599 was the only design in the candidate list as it was first in all categories in Table 6. Therefore, the OA was compared to the worst OA as predicted by the model (OA#150) in Figure 12 and Figure 13. It was surprising to see

that in terms of *% correct*, OA #150 outperformed OA #1599 for ME-only models. However, in all categories OA#1599 performed significantly better, making this OA the best 8-factor no-confounding design. The design is shown in Appendix Table 77.

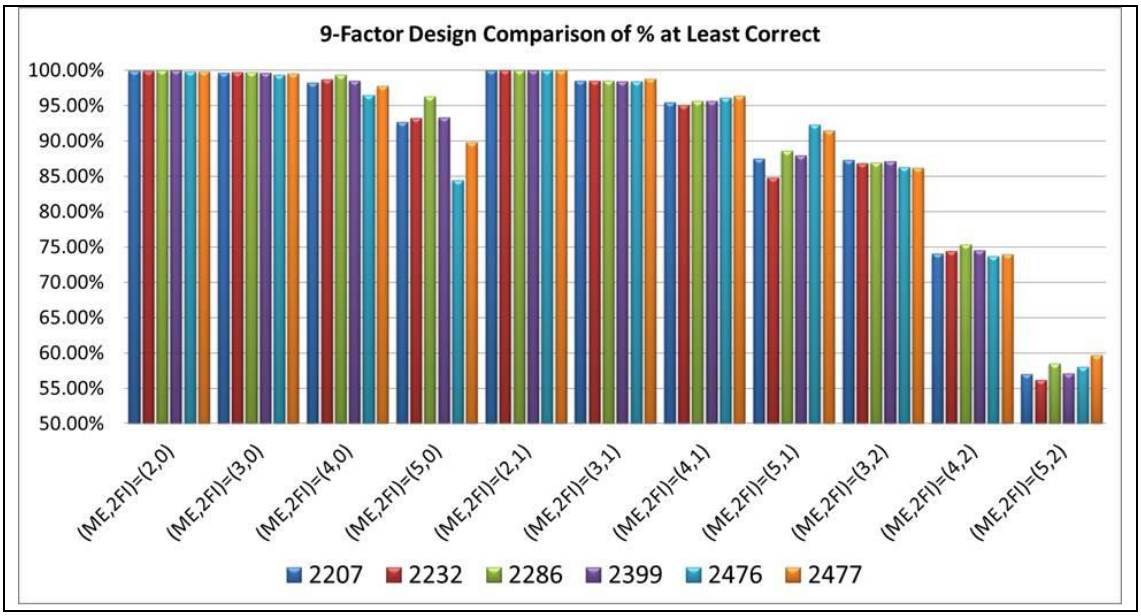

Figure 14. 9-Factor Design Comparison of *% at Least Correct* by *p* ME and *q* 2FI

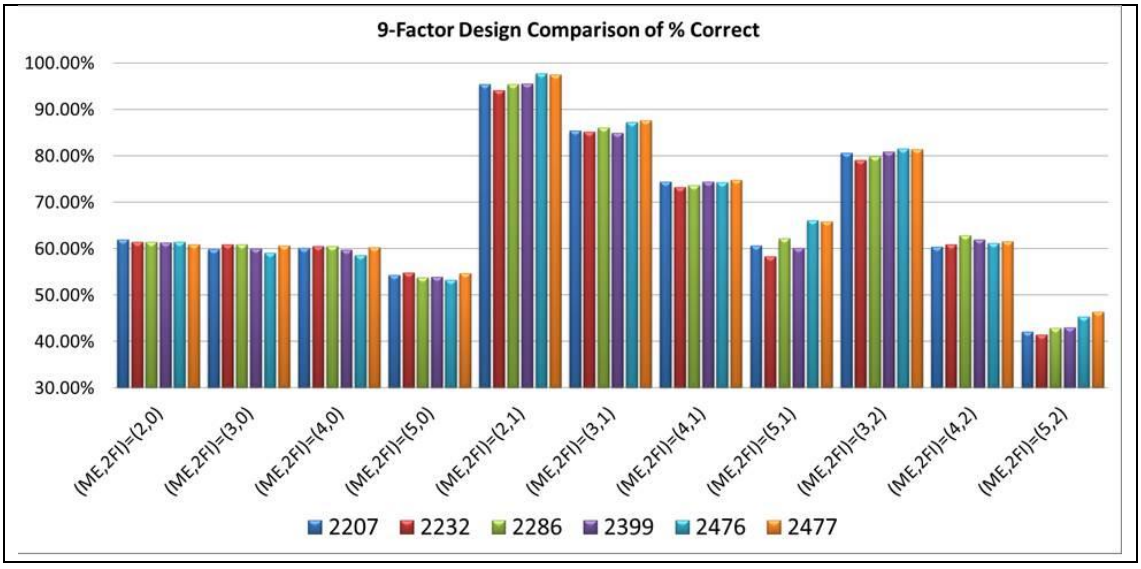

Figure 15. 9-Factor Design Comparison of *% Correct* by *p* ME and *q* 2FI

Two additional designs were added to the four designs in Table 6 for the 9-factor analysis. Even though OA #2207 and OA #2399 did not rank first in any category, they

were in the top 10 designs for G-aberration and PEC. This was in contrast to OA #2476 and OA #2477 which ranked first in G-aberration but ranked very poorly in terms of PEC. Neither OA #2207 or OA #2399 were chosen as the best design.

OA#2477 performed the best overall in terms of *% correct* and only compared poorly to other designs in terms of *% at least correct* for the cases where (ME,2FI) = (4,0) =(5,0). However, in these cases it performed significantly worse than the other designs, particularly OA#2286, which was the best performer overall for *% at least correct.* Ultimately OA#2286 was chosen as the best 9-factor no-confounding design since the *% at least correct metric* was deemed to be the most important metric. Also, this design ranked first in PEC and the OA performed better or comparable to OA#2477 when there was zero or one 2FI. The design is shown in Appendix Table 78.

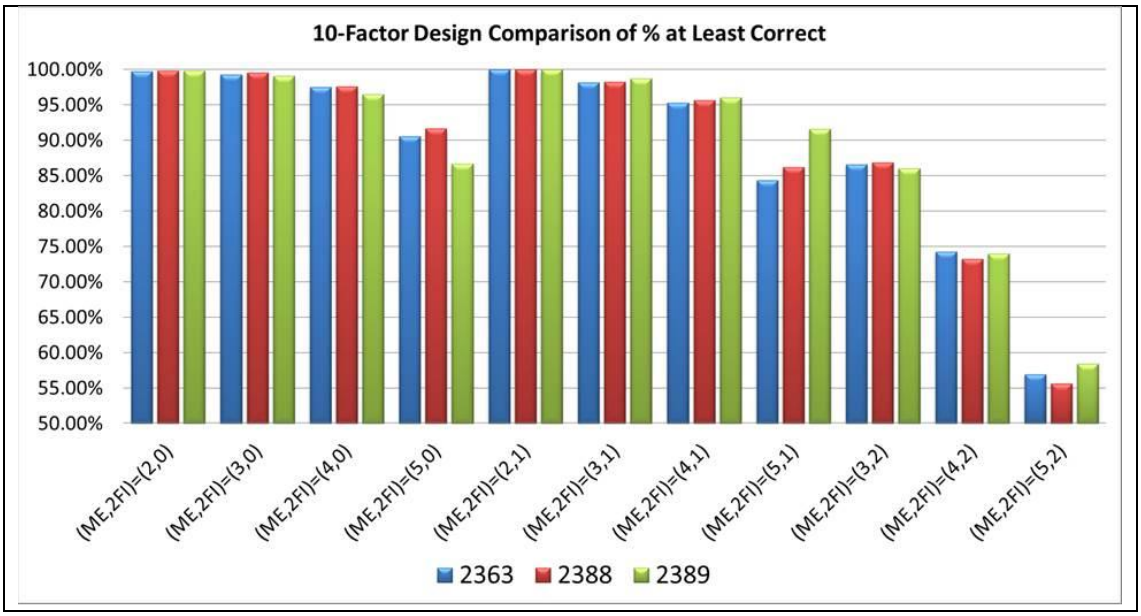

Figure 16. 10-Factor Design Comparison of *% at Least Correct* by *p* ME and *q* 2FI

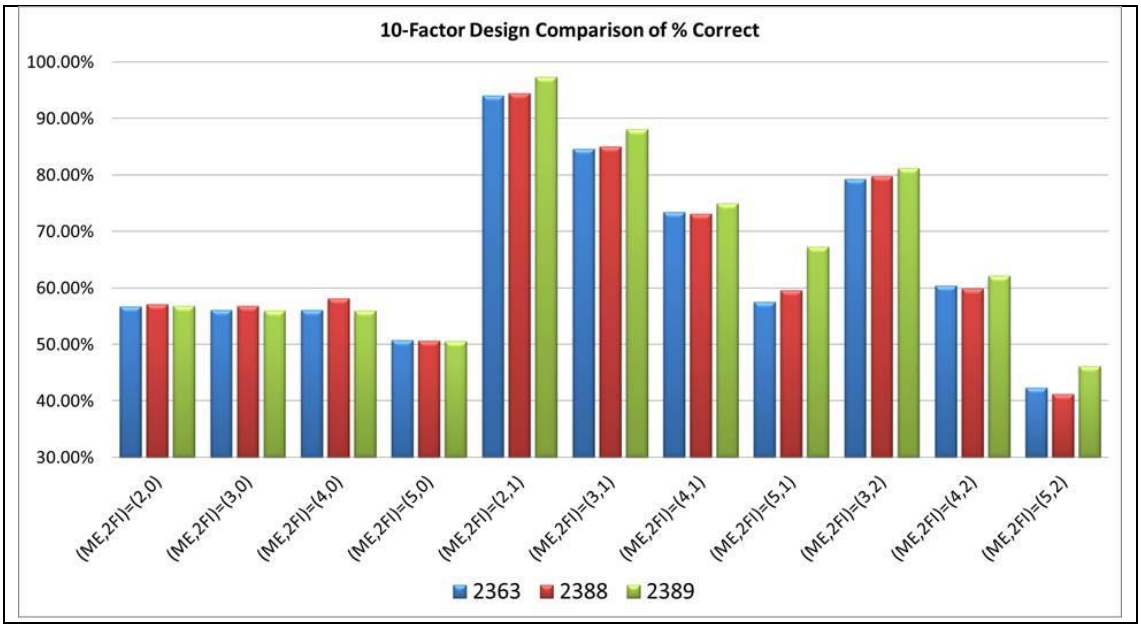

Figure 17. 10-Factor Design Comparison of *% Correct* by *p* ME and *q* 2FI

OA #2389 was an obvious choice for the best 10-factor no-confounding design. It ranked first in terms of G-aberration and PEC. Overall, it ranked highest in terms of *% at least correct* and ranked significantly higher in terms of *% correct*. The design is shown in Appendix Table 79.

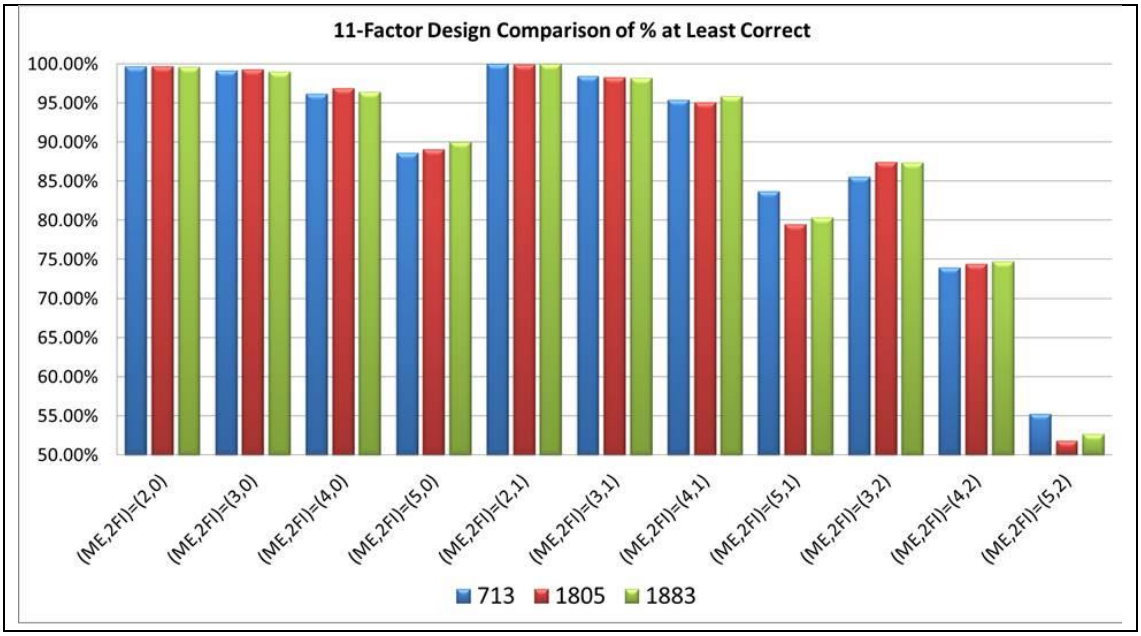

Figure 18. 11-Factor Design Comparison of *% at Least Correct* by *p* ME and *q* 2FI

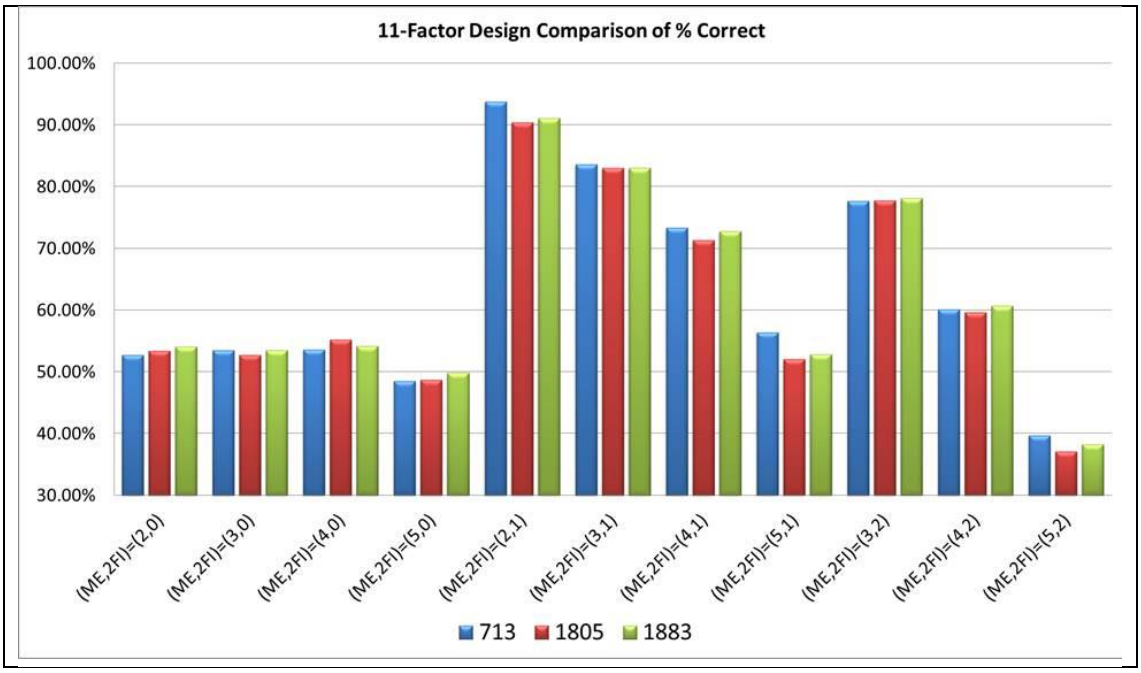

Figure 19. 11-Factor Design Comparison of *% Correct* by *p* ME and *q* 2FI

OA #713 was the choice for the best 11-factor no-confounding design. Overall, it ranked highest in terms of *% at least correct* and *% correct*, although the performance difference between the three designs was very close. In general, as the number of columns in the OAs increased, the difference in performance between the best and worst designs decreased. The tie-breaking factor was the fact that OA#713 ranked first in terms of both G-aberration and PEC. The design is shown in Appendix Table 80.

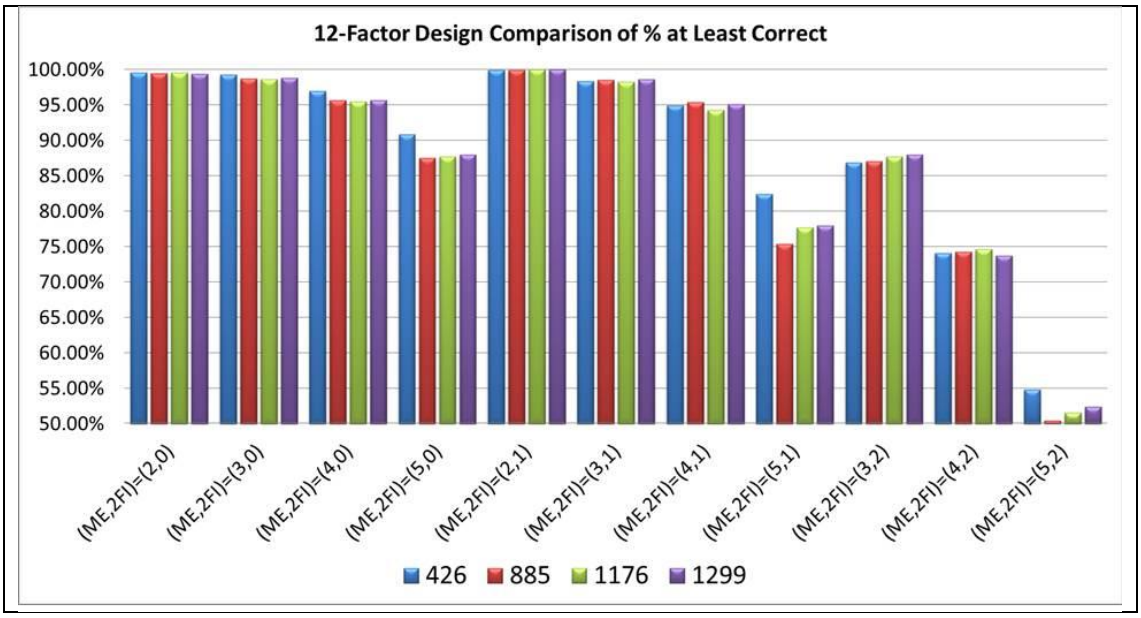

Figure 20. 12-Factor Design Comparison of *% at Least Correct* by *p* ME and *q* 2FI

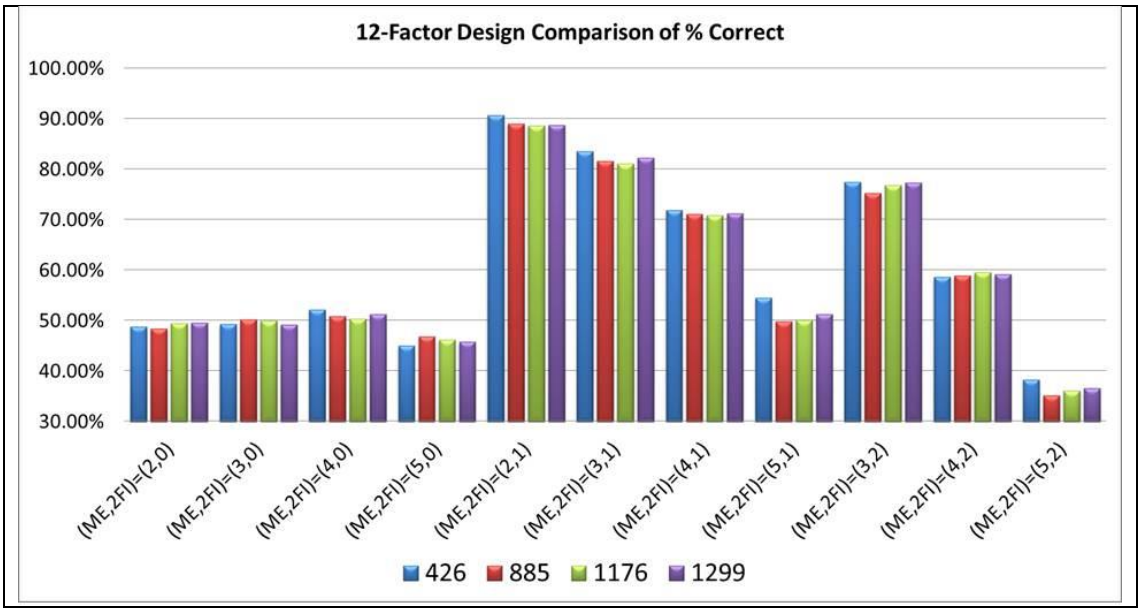

Figure 21. 12-Factor Design Comparison of *% Correct* by *p* ME and *q* 2FI

OA #426 was an obvious choice for the best 12-factor no-confounding design. It ranked first in terms of G-aberration and PEC. Overall, it ranked highest in terms of *% at least correct* and *% correct*. In the cases where (ME,2FI) = (5,0) and (ME,2FI) = (5,1) this design performed significantly better than the other candidate designs. The design is shown in Appendix Table 81.

Figure 22 and Figure 23 show the mean *% at least correct* and mean *% correct* respectively for each of the selected m-factor no-confounding designs.

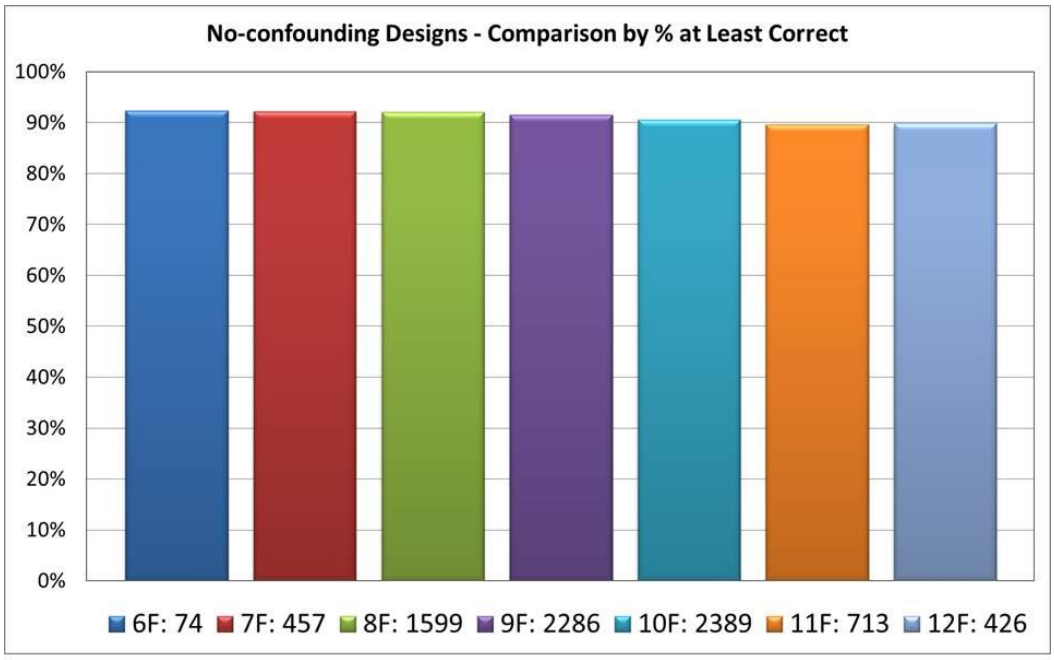

Figure 22. No-confounding Design Comparison - *% at Least Correct*

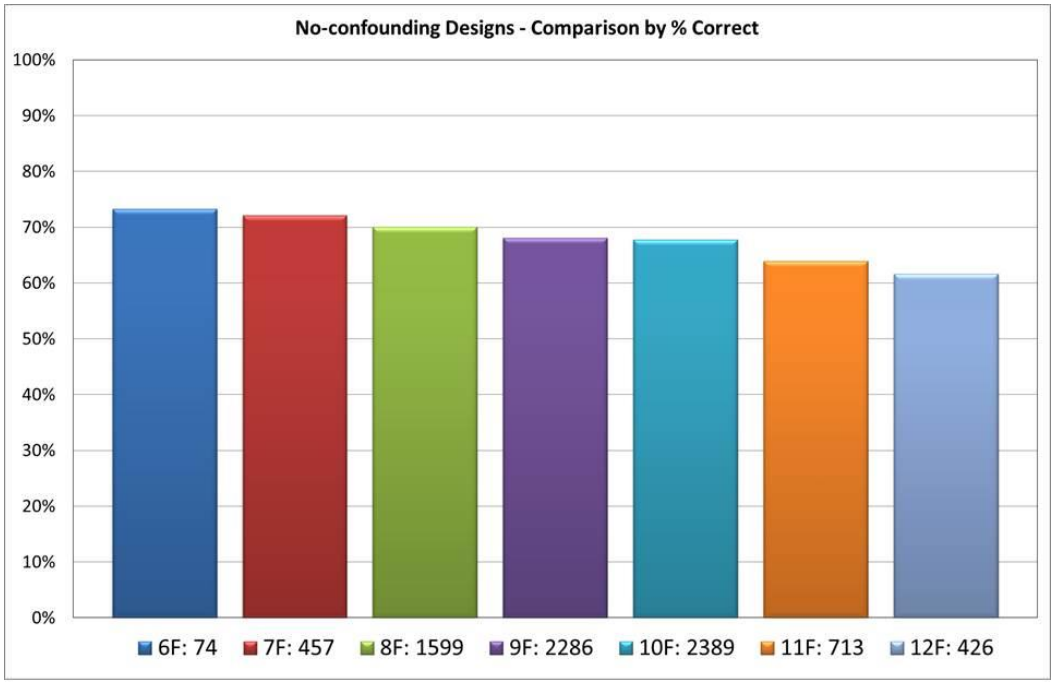

Figure 23. No-confounding Design Comparison - *% Correct*

Figure 22 shows that 20-run no-confounding designs perform very well as screening designs. The 6-factor design found at least the active effects 92% of the time for models of 2-5 ME and 0-2 2FI, while the 12-factor design found at least the active effects over 89% of the time. In examining the seven no-confounding designs together, there is not much difference in terms of the *% at least correct* metric (approximately 3 percentage points) between the 6-factor no-confounding design and the 12-factor no-confounding design. However, there is a significant difference between the 6-factor no-confounding design and the 12-factor no-confounding design in terms of the *% correct* metric (approximate 12 percentage points).

For each of the no-confounding designs, Table 8 - Table 21 show in tabular form the breakdown of the simulation metrics by the number of ME and 2FI. These tables give a better perspective on the outstanding performance of these designs in terms of the *% at least correct* metric, given sufficiently small models. This data was generated from pivot tables using data from 60,000 simulated design analyses for each of the m-factor no-confounding designs. For models with at most four ME and one 2FI, the 20-run no-confounding designs perform extremely well as screening designs, with a minimum *% at least correct* of 95%.

Table 8. *% at Least Correct* for *p* ME Effects and *q* 2FI Effects (6-Factor Design)

% At Least Correct Model		Model Simulations = 10000		
OA(20;2^6;2) #74	q (2FI)			Row Totals
p (ME)	0	1	2	
2	100.00%	99.96%		99.98%
3	100.00%	98.83%	85.80%	94.89%
4	100.00%	96.38%	74.87%	90.34%
5	99.98%	92.64%	60.54%	84.32%
Column Totals	100.00%	97.27%	73.69%	92.36%

Table 9. *% Correct* for *p* ME Effects and *q* 2FI Effects (6-Factor Design)

% Correct Model		Model Simulations = 10000		
OA(20;2^6;2) #74	q (2FI)			Row Totals
p (ME)	0	1	2	
2	75.25%	98.63%		86.91%
3	73.47%	88.39%	80.99%	80.91%
4	67.65%	74.61%	61.82%	68.04%
5	60.96%	65.16%	46.62%	57.56%
Column Totals	70.00%	83.48%	63.07%	73.31%

Table 10. *% at Least Correct* for *p* ME Effects and *q* 2FI Effects (7-Factor Design)

% At Least Correct Model		Model Simulations = 10000		
OA(20;2^7;2) #457	q (2FI)			Row Totals
p (ME)	0	1	2	
2	100.00%	99.95%		99.97%
3	100.00%	98.75%	85.96%	94.88%
4	100.00%	96.17%	74.29%	90.17%
5	99.55%	92.32%	59.85%	83.69%
Column Totals	99.90%	97.15%	73.29%	92.19%

Table 11. *% Correct* for *p* ME Effects and *q* 2FI Effects (7-Factor Design)

% Correct Model		Model Simulations = 10000		
OA(20;2^7;2) #457	q (2FI)			Row Totals
p (ME)	0	1	2	
2	71.14%	98.18%		84.60%
3	67.35%	88.01%	81.04%	78.82%
4	65.09%	75.70%	61.61%	67.46%
5	60.50%	65.55%	46.53%	57.44%
Column Totals	66.62%	83.70%	62.95%	72.08%

Table 12. *% at Least Correct* for *p* ME Effects and *q* 2FI Effects (8-Factor Design)

% At Least Correct Model		Model Simulations = 10000		
OA(20;2^8;2) #1599	q (2FI)			Row Totals
p (ME)	0	1	2	
2	100.00%	99.96%		99.98%
3	100.00%	98.33%	85.70%	94.62%
4	99.88%	96.00%	73.56%	89.86%
5	99.04%	91.94%	59.52%	83.54%
Column Totals	99.76%	96.91%	72.98%	92.01%

Table 13. *% Correct* for *p* ME Effects and *q* 2FI Effects (8-Factor Design)

% Correct Model		Model Simulations = 10000		
OA(20;2^8;2) #1599	q (2FI)			Row Totals
p (ME)	0	1	2	
2	65.98%	97.82%		81.69%
3	62.67%	87.53%	81.16%	77.11%
4	61.13%	75.00%	61.69%	66.06%
5	55.00%	64.67%	46.04%	55.25%
Column Totals	61.75%	82.98%	63.05%	70.03%

Table 14. *% at Least Correct* for *p* ME Effects and *q* 2FI Effects (9-Factor Design)

% At Least Correct Model		Model Simulations = 10000		
OA(20;2^9;2) #2286	q (2FI)			Row Totals
p (ME)	0	1	2	
2	99.97%	99.97%		99.97%
3	99.72%	98.52%	86.92%	95.07%
4	99.37%	95.59%	75.33%	90.06%
5	96.26%	88.62%	58.48%	81.06%
Column Totals	98.95%	96.12%	73.48%	91.49%

Table 15. *% Correct* for *p* ME Effects and *q* 2FI Effects (9-Factor Design)

% Correct Model		Model Simulations = 10000		
OA(20;2^9;2) #2286	q (2FI)			Row Totals
p (ME)	0	1	2	
2	61.45%	95.41%		78.36%
3	60.93%	86.04%	79.88%	75.65%
4	60.57%	73.66%	62.86%	65.63%
5	53.80%	62.21%	42.84%	52.93%
Column Totals	59.43%	81.04%	61.73%	68.08%

Table 16. *% at Least Correct* for *p* ME Effects and *q* 2FI Effects (10-Factor Design)

% At Least Correct Model		Model Simulations = 10000		
OA(20;2^{10};2) #2389	q (2FI)			Row Totals
p (ME)	0	1	2	
2	99.80%	99.97%		99.89%
3	99.06%	98.70%	86.05%	94.58%
4	96.51%	96.05%	73.98%	88.80%
5	86.64%	91.59%	58.42%	78.76%
Column Totals	95.97%	96.97%	72.74%	90.50%

Table 17. *% Correct* for *p* ME Effects and *q* 2FI Effects (10-Factor Design)

% Correct Model		Model Simulations = 10000		
OA(20;2^{10};2) #2389	q (2FI)			Row Totals
p (ME)	0	1	2	
2	56.84%	97.37%		76.94%
3	55.97%	88.08%	81.19%	75.20%
4	55.95%	74.91%	62.11%	64.34%
5	50.56%	67.24%	46.13%	54.55%
Column Totals	55.05%	83.65%	63.05%	67.75%

Table 18. *% at Least Correct* for *p* ME Effects and *q* 2FI Effects (11-Factor Design)

% At Least Correct Model		Model Simulations = 10000		
OA(20;2^{11};2) #713	q (2FI)			Row Totals
p (ME)	0	1	2	
2	99.67%	99.99%		99.83%
3	99.12%	98.40%	85.56%	94.35%
4	96.16%	95.44%	73.95%	88.59%
5	88.61%	83.69%	55.27%	75.87%
Column Totals	96.31%	95.00%	71.56%	89.66%

Table 19. *% Correct* for *p* ME Effects and *q* 2FI Effects (11-Factor Design)

% Correct Model		Model Simulations = 10000		
OA(20;2^{11};2) #713	q (2FI)			Row Totals
p (ME)	0	1	2	
2	52.73%	93.74%		73.31%
3	53.47%	83.62%	77.69%	71.73%
4	53.63%	73.32%	60.14%	62.49%
5	48.52%	56.40%	39.64%	48.22%
Column Totals	52.15%	78.62%	59.12%	63.93%

Table 20. *% at Least Correct* for *p* ME Effects and *q* 2FI Effects (12-Factor Design)

% At Least Correct Model		Model Simulations = 10000		
OA(20;2^{12};2) #426	q (2FI)			Row Totals
p (ME)	0	1	2	
2	99.58%	99.97%		99.78%
3	99.25%	98.39%	86.91%	94.89%
4	96.94%	94.92%	74.12%	88.62%
5	90.85%	82.46%	54.80%	75.88%
Column Totals	97.00%	94.70%	71.82%	89.84%

Table 21. *% Correct* for p ME Effects and q 2FI Effects (12-Factor Design)

% Correct Model		Model Simulations = 10000		
OA(20;2^{12};2) #426	q (2FI)			Row Totals
p (ME)	0	1	2	
2	48.81%	90.72%		69.61%
3	49.26%	83.52%	77.49%	70.01%
4	52.12%	71.95%	58.62%	60.84%
5	45.01%	54.53%	38.32%	45.83%
Column Totals	48.80%	77.10%	57.99%	61.62%

Finally, Table 22 provides information concerning the design ranking criteria values (or ranks as a substitute for non-scalar values) for each of the chosen m-factor no-confounding designs.

Table 22. Values of the 24 Design-Ranking Criteria for the No-confounding Designs

Factors	6	7	8	9	10	11	12
OA ID#	74	457	1599	2286	2389	713	426
G-aberration (Rank)	1	3	1	7	1	1	1
G_2-aberration (Rank)	1	3	1	7	1	1	1
Sum Nonzero Freq	35	70	126	210	330	495	715
Combined F(D) G-aberration (Rank)	1	12	1	456	2387	1908	1296
E(S^2)	8.000	11.937	13.257	14.836	18.424	17.343	17.439
Min Unique Pts – 4 Columns	15	12	12	12	12	12	12
Avg Unique Pts – 4 Columns	15.000	14.743	14.743	14.476	14.571	14.364	14.327
Max Unique Pts - 4 Columns	15	15	15	15	15	15	15
Min Unique Pts – 5 Columns	19	17	17	14	17	13	13
Avg Unique Pts - 5 Columns	19.333	18.714	18.714	18.381	18.286	18.177	18.182
Max Unique Pts - 5 Columns	20	20	20	20	20	20	20
Min Unique Pts – 6 Columns	20	19	19	18	18	18	14
Avg Unique Pts - 6 Columns	20.000	19.857	19.857	19.679	19.429	19.498	19.554
Max Unique Pts - 6 Columns	20	20	20	20	20	20	20
PEC	1.000	0.905	0.929	0.873	0.857	0.848	0.826
Min VIF -3 Columns	1.042	1.042	1.042	1.042	1.042	1.042	1.042
Avg VIF -3 Columns	1.042	1.042	1.042	1.054	1.042	1.057	1.061
Max VIF -3 Columns	1.042	1.042	1.042	1.563	1.042	1.563	1.563
Min VIF -4 Columns	1.161	1.161	1.161	1.161	1.161	1.161	1.161
Avg VIF -4 Columns	1.161	1.201	1.201	1.242	1.228	1.260	1.266
Max VIF -4 Columns	1.161	1.893	1.893	1.893	1.893	1.893	1.893
Min VIF -5 Columns	1.389	1.328	1.328	1.328	1.328	1.328	1.328
Avg VIF -5 Columns	1.446	1.812	1.831	1.975	2.112	2.101	2.043
Max VIF -5 Columns	1.500	5.035	5.035	5.035	5.035	5.035	5.035

3.5 Conclusion

The 20-run no-confounding designs perform very well as screening designs when the response model has up to five terms. For models of at most four ME and one 2FI, these designs will identify at least the active effects over 95% of the time, even when experimenting with as many as 12 factors. In the simulations with models of four ME and one 2FI, the 6-factor design detected at least the active main effects 96% of the time and there were no variable selection errors 75% of the time. In the best case scenario with two ME and one 2FI, the percentage at least correct and the percentage correct was 100% and 99% respectively. The 12-factor design detected at least the active main effects 95% of the time and there were no variable selection errors 72% of the time in the simulations with four ME and one 2FI. In the best case scenario with two ME and one 2FI, the percentage at least correct and the percentage correct was 100% and 91% respectively.

When there is more than one active 2FI in the response model, the analysis procedure has a difficult time differentiating between potential models. Consequently the type II error rate goes up significantly. Thus an experimenter who expects the final model to contain more than five terms should consider running more experiments.

The Monte Carlo analysis of the catalog of OAs was an effective technique in identifying the optimal no-confounding screening designs and characterizing their model estimation capabilities. The empirical analysis removed much of the confusion caused by the lack of a consistent design ranking for OAs across the commonly used design ranking criteria. Furthermore, the simulation results provided a detailed performance analysis of the designs across response models comprised of varying numbers of ME and 2FI.

The empirical analysis also showed that an experimenter can make trade-offs between type II errors and type I errors by selecting the appropriate analysis method. If an experimental budget significantly constrains additional experimentation beyond the screening design, there is a relatively high chance of successfully fitting the correct model using 2-stage stepwise regression and AICc. However, if the no-confounding design is truly used as a screening design in a multi-stage experiment, the type II error rate can be reduced by using a 1-stage mixed stepwise approach.

Finally, a high rank in terms of G-aberration is common among the selected no-confounding designs. The 6,8,10,11, and 12-factor no-confounding designs are ranked 1st, the 7-factor design is ranked 3rd, and the 9-factor design is ranked 7th. This implies that the minimization of the aliasing of ME with 2FI greatly improves model fitting results and should be a primary consideration when evaluating designs. However, since many designs are approximately equivalent in terms of G-aberration, it is important to consider multiple design-ranking criteria and determine the best designs from the Pareto efficient set of OAs.

There are many opportunities for future research involving no-confounding designs. Extending this work to 28-run designs is natural due to the many similarities in the combinatorial structure of 28-run designs and 20-run designs. Many insights gained from analyzing the optimal 20-run designs can be applied to search algorithms for finding good 28-run designs. In particular, optimal 28-run designs may be found by finding OAs with minimum G-aberration. It would also be valuable to determine if design performance could be better differentiated and characterized by using other model fitting techniques, such as the Dantzig selector. Designs that appear similar in performance

using 2-stage stepwise regression may in fact differ more significantly when fitting models with a Dantzig selector. Finally it may be beneficial to explore whether non-orthogonal designs offer some model-fitting advantages over orthogonal designs.

Chapter 4

NO-CONFOUNDING DESIGNS OF 24 RUNS

4.1 Introduction

When experimental budgets, time, or other resources are significantly constrained, it is often the case that sequential experimentation with Resolution IV 2^{k-p} fractional factorials is impractical or infeasible. If several two-factor interactions (2FI) are significant, a foldover experiment must be run to resolve the ambiguity of fully confounded 2FI effects. Jones and Montgomery (2010) introduced 16-run no-confounding designs as alternatives to Resolution IV 2^{k-p} fractional factorials. For the no-confounding designs of 6-10 factors, no pair or combination of main effects (ME) and 2FI effects are completely confounded. The absence of fully-confounded effects enables the unambiguous estimation of main effect (ME) and 2FI models without further experimentation – an advantage which can save significant experimental resources.

In this chapter, 24-run no-confounding designs are introduced as an additional option to resolution IV 2^{k-p} fractional factorials. Compared to the 16-run designs, the 24-run designs have more capability to accurately screen effects for models of six to seven terms. These designs also provide more accurate coefficient estimates for the cost of only eight additional runs.

The 6-12 factor no-confounding designs are identified through an evaluation, using Monte Carlo simulation, of published and algorithmically generated 24-run designs. In the current literature, 24-run nonregular designs have been ranked and cataloged by three popular design-ranking criteria: generalized aberration (G-aberration) introduced by Deng and Tang (1999); D-optimality; and projection estimation capacity

(PEC) introduced by Loeppky, Sitter et al. (2007). The originators of these design criteria offer substantial theoretical justification for their use as a tool to judge the model-fitting capability of designs. However, only PEC explicitly considers whether effects are fully confounded. The optimal 24-run designs according to minimum G-aberration, D-optimality (6-factor case only) and PEC are used as baseline designs to compare and evaluate the model-fitting performance of the algorithmically generated designs.

The algorithmically generated designs were created by two variants of a column exchange search algorithm. An interesting aspect of both column exchange algorithms was that they generated a Pareto efficient set, or Pareto front, of designs. This is in contrast to the more common method of generating a single design which optimizes a single design-ranking criterion, or possibly a pre-specified linear combination of multiple criteria. The Pareto efficient set of designs was generated based on three criteria: $E(s^2)$, projection estimation capacity (PEC) for 6-factor full models, and the average variance inflation factor (VIF) of 5-factor projections. One of the two column exchange algorithms was capable of efficiently exploring a design space that included unbalanced, non-orthogonal designs as well as balanced and orthogonal designs.

A second feature of the algorithms was that, unlike previously published column-exchange algorithms used to create 24-run designs, the exchanged columns were not restricted to Hadamard matrix columns. The algorithms were therefore able to explore designs which were not projections of Hadamard matrices. This was important as previous research has provided examples of optimal designs for certain criterion that are not Hadamard projections. See Loeppky, Sitter et al. (2007) for an example.

The designs on the Pareto efficient frontier of the multiple design-ranking criteria were evaluated via Monte Carlo simulation. The simulation created 5,000 randomly generated polynomial response models and automated the response variable analysis. Designs were evaluated based on two simulation metrics: the percentage of simulations which produced estimated models comprised of at least every active effect and the percentage of simulations that produced the correct model.

This chapter is organized as follows. Section 4.2 discusses background material on orthogonal arrays, existing research into nonregular design criteria, and the concept of the Pareto efficient set. Section 4.3 presents the methodology used to identify the 24-run no-confounding designs, including the details of the column-exchange search algorithm, the process of narrowing down and selecting designs from the Pareto efficient set, and the design evaluation process using Monte Carlo simulation.

Section 4.4 reports the results. It presents the analysis of the Monte Carlo simulation output for the baseline and algorithmically generated designs, and provides the recommendations for the optimal 24-run no-confounding designs of 6-12 factors. The results of the simulation conclusively show that, in terms of accurately identifying active effects, the nonregular minimum G-aberration designs identified in Ingram and Tang (2005) are superior to any other published or algorithmically generated designs. This is a bit surprising considering that these designs have a relatively high value of $E(s^2)$ compared to many of the other candidate designs. A drawback to these designs is that the full 6-factor model is not estimable in any 6-factor projection.

Fortunately, the column exchange algorithms were able to discover design options which can estimate the full 6-factor model for every 6-factor projection without a

significant decrease in model fitting capability. An additional contribution is that two of these designs – the seven and nine-factor designs - have the minimum number of maximum effect correlations of any published balanced and orthogonal-two-level design. Section 4.5 discusses important conclusions concerning the recommended 24-run no-confounding designs and suggests ideas for future research.

4.2 Background

No-confounding designs are nonregular designs, as opposed to regular designs such as the 2^{k-p} fractional factorial. Nonregular designs produce effect estimates that are partially confounded with other effects. More precisely, the alias matrix elements of nonregular designs take values other than ± 1 or 0. Sun and Wu (1993) introduced the term "nonregular design" when studying Hadamard matrices of order 16. A Hadamard matrix is an $n \times n$ square matrix $\boldsymbol{H}$ with ± 1 entries such that $\boldsymbol{H'H} = n\boldsymbol{I}_n$. Hadamard matrices are a special case of orthogonal arrays, which are discussed in the next section.

4.2.1 Orthogonal Arrays

Nonregular designs with pair-wise orthogonal columns are also known as orthogonal arrays. Orthogonal arrays are an important tool to facilitate factorial experimentation, not only in industrial contexts, but in a growing number of other disciplines as well. Orthogonal arrays were introduced by Rao (1947) and included Plackett-Burman designs as a special case. Formally, an OA is a $\mathrm{N} \times \mathrm{m}$ matrix whose ith column contains $\mathrm{s_i}$ levels. The OA has strength t if, within any t columns, every t-tuple of levels appears equally often. The general notation is $\mathrm{OA}(\mathrm{N}; \mathrm{s_1}, \ldots, \mathrm{s_m}; \mathrm{t})$. In the case of pure level arrays where each column has the same number of levels the notation is shortened to $\mathrm{OA}(\mathrm{N}; \mathrm{s^m}; \mathrm{t})$.

Orthogonal arrays include both regular and nonregular designs. Regular designs of resolution R are orthogonal arrays of strength $t = R - 1$. Cheng (1980) proved that for the main effects model, orthogonal arrays are universally optimal. An orthogonal array of strength t has projectivity t. Box and Tyssedal (1996) defined a design to be of projectivity t if the projection onto every subset of t factors contains at least a 2^t full factorial design. Hedayat, Sloane et al. (1999) give a detailed presentation of the theory and application of orthogonal arrays.

Sun, Li et al. (2002) developed an algorithm for sequentially constructing non-isomorphic regular and nonregular orthogonal arrays. Two OAs are said to be isomorphic if one OA can be obtained from the other by permuting rows, columns or relabeling factor levels. Their algorithm was used to obtain a complete catalog of $n \times k$ two-level orthogonal designs for $n \in \{12, 16, 20\}$ and $k \in \{2,3, \ldots, (n-1)\}$. The catalog has been used to identify generalized minimum aberration designs. Schoen and Mee (2012) classified all two-level orthogonal arrays of strength 3 with up to 48 runs. There are two non-isomorphic $\mathrm{OA}(24; 2^{\mathrm{m}}; 3)$ for $m = 6$, and one non-isomorphic $\mathrm{OA}(24; 2^{\mathrm{m}}; 3)$ for $m \in \{7, \ldots, 12\}$.

4.2.2 Nonregular Design Evaluation Criteria

An important advancement in nonregular design evaluation occurred when researchers were able to take the concept of minimum aberration, used to rank regular 2^{k-p} fractional factorials, and generalize it to nonregular designs. Experimenters identified regular 2^{k-p} fractional factorials that had the most favorable alias structure for ME and 2FI effects by using the minimum aberration criteria proposed by Fries and Hunter (1980). These designs were identified as having the minimum number of words

in the defining relation that were of minimum length. According to Montgomery (2012), a minimum aberration design of resolution R has the minimum number of ME aliased with $(R-1)$-factor interactions and the number of 2FI aliased with $(R-2)$-factor interactions.

4.2.3 Generalized Minimum Aberration

The first to generalize the concept of minimum aberration for nonregular fractional factorials was Deng and Tang (1999). Their paper introduced the concept of generalized resolution and minimum generalized aberration (minimum G-aberration). For regular designs, minimum G-aberration reduces to the traditional minimum aberration criterion.

In order to define minimum G-aberration, Deng and Tang first developed the idea of J-characteristics and the confounding frequency vector (CFV). Suppose there is an $n \times m$ design matrix $\boldsymbol{D} = (d_{ij})_{n\times m}$ with $d_{ij} = \pm 1$. Let $Z_m = \{1, \dots, m\}$ be an index set of the columns of $\boldsymbol{D}$. For every subset s of Z_m, define $J_s(\boldsymbol{D}) = \sum_{i=1}^{n} \prod_{j\in s} d_{ij}$. The collection of J_s values is known as the J-characteristics of design $\boldsymbol{D}$. Let $n = 4t$ and f_{kj} be the frequency of k column combinations (k is the cardinality of s) that give $|J_s| = 4(t + 1 - j)$ for $j = 1, \dots, t$, where $|J_s|$ is the absolute value of a J-characteristic. The CFV of design $\boldsymbol{D}$ is defined to be

$$F(\boldsymbol{D}) = [(f_{11}, \dots, f_{1t}), (f_{21}, \dots, f_{2t}), (f_{31}, \dots, f_{3t}), \dots, (f_{m1}, \dots, f_{mt})]$$

Note that $(f_{11}, \dots, f_{1t})$ is simply the frequency of non-zero column sums of $\boldsymbol{D}$ in reverse order, which for a balanced two-level design are all zero since all column sums are zero. Let $f_l(\boldsymbol{D_1})$ and $f_l(\boldsymbol{D_2})$ be the lth entries of $F(\boldsymbol{D_1})$ and $F(\boldsymbol{D_2})$, respectively. Let r be the smallest integer such that $f_r(\boldsymbol{D_1}) \neq f_r(\boldsymbol{D_2})$. If $f_r(\boldsymbol{D_1}) < f_r(\boldsymbol{D_2})$ then $\boldsymbol{D_1}$ has less G-

aberration than $\boldsymbol{D_2}$. If no design has less G-aberration than $\boldsymbol{D_1}$, then it has minimum G-aberration. Recently, Lin, Sitter et al. (2012) used G-aberration to catalog and rank two-level nonregular fractional factorial designs of 32 and 40 runs.

The CFV mentioned previously has been an important tool in nonregular design analysis. Deng and Tang (2002) used the CFV criterion to identify and rank non-isomorphic designs that are projections of Hadamard matrices. Their research demonstrated that the CFV was a powerful method for identifying non-isomorphic designs. Li, Deng et al. (2004) used CFVs to search for non-isomorphic OAs that were not projections of Hadamard matrices in order to determine if designs existed with lower G-aberration than found in Deng, Li et al. (2000). They discovered 15 designs outside Hadamard matrices but the best of these designs ranked 276th in terms of G-aberration.

There is an important link between the CFV and the $\mathbf{M} = \boldsymbol{X'X}$ matrix. Let $\boldsymbol{X}$ be a balanced and orthogonal $n \times m$ model matrix for a ME+2FI full model, and let n be a multiple of 4. In the CFV, if $f_{3t} > 0$, for $t = 1, \ldots, \frac{n}{4}$, then there are $6 * f_{3t}$ matrix elements x_{ij} in $\boldsymbol{M}$ corresponding to ME and 2FI correlations where $x_{ij} = 4(\frac{n}{4} + 1 - t)$. Also, if $f_{4t} > 0$, for $t = 1, \ldots, \frac{n}{4}$, then there are $6 * f_{4t}$ matrix elements x_{ij} in $\boldsymbol{X'X}$ corresponding to 2FI and 2FI interactions where $x_{ij} = 4(\frac{n}{4} + 1 - t)$. For example suppose there is a 24×6 design matrix $\boldsymbol{D}$ for which $f_{32} = 1$. Without loss of generality, let this correspond to $J_s(\boldsymbol{D}) = 8$ for $s = \{1,2,3\}$. Recall that s is a column index set and here corresponds to effect columns A,B and C. Let $\boldsymbol{X}$ be $\boldsymbol{D}$ expanded to a 6-factor full model matrix with no intercept column. Then for the $\boldsymbol{X'X}$ matrix it is true that $x_{1,12} =$

$x_{2,8} = x_{3,7} = x_{12,1} = x_{8,2} = x_{7,3} = 8$. Clearly working with CFVs is an efficient alternative to computing the frequency of nonzero elements in $\boldsymbol{X'X}$.

4.2.4 $E(s^2)$

Even before minimum aberration had been developed to compare regular fractional factorials, researchers discovered a way to compare supersaturated designs. A supersaturated design is a factorial design with n rows and k effect columns with $k > n$. Booth and Cox (1962) introduced $E(s^2)$ as a solution to the problem of evaluating supersaturated designs. Let $\mathbf{x}_i$ be the ith column of the $n \times k$ model matrix $\boldsymbol{X}$. Then

$$E(s^2) = \sum_{i<j} (\mathbf{x}_i \grave{} \mathbf{x}_j)^2 \Big/ \binom{k}{2}$$

Minimizing $E(s^2)$ is equivalent to minimizing the sum of squared off-diagonal elements of $\boldsymbol{X'X}$.

Lin (1993) was the first to rejuvenate interest in the metric. He constructed supersaturated designs of size $(n, m) = (2t, 4t - 2)$ using half fractions of Hadamard matrices. The $E(s^2)$ criterion was used to construct and evaluate supersaturated designs in Wu (1993), who used Hadamard matrices to develop supersaturated designs of 12 and 20 runs to investigate up to 66 and 124 factors respectively. Tang and Wu (1997) generalized the work in Lin (1993) by using Hadamard matrices to construct supersaturated designs. Jones and Montgomery (2010) used $E(s^2)$ to identify which projections of the Hadamard matrices of order 16 were optimal 16-run no-confounding designs.

4.2.5 Variance Inflation Factors

Montgomery (2012) mentions that a way to measure the multicollinearity of a model matrix is a diagnostic called the variance inflation factor (VIF). VIFs are calculated from the model matrix $\boldsymbol{X}$. First, unit length scaling is applied to the columns of $\boldsymbol{X}$ to create a new matrix $\boldsymbol{W}$ whose columns have mean zero and unit length. $\boldsymbol{W}'\boldsymbol{W}$ is in the form of a correlation matrix. The jth main diagonal element of $(\boldsymbol{W}'\boldsymbol{W})^{-1}$ is the VIF for the jth regression coefficient.

VIFs are commonly applied as a diagnostic tool in regression, where multicollinearity between regressor variables is common, but they are utilized much less in the context of DOE due to the emphasis on orthogonal designs. However, VIFs can be a useful tool for evaluating nonregular designs which may have orthogonal ME columns, but correlated columns in a model matrix for a ME+2FI full model. Thus far, VIFs have not been used in the literature to compare and rank nonregular designs.

4.2.6 Projection Estimation Capacity

Another general way to evaluate designs is by model robustness, often measured by some variant of estimation capacity. Estimation capacity was first introduced by Sun (1993) and measures the number of estimable models containing all of the ME evaluated in a design and a given number of q 2FI's. Estimation capacity was used by Li and Nachtsheim (2000) and Li (2006) to create model robust designs for screening experiments.

Loeppky, Sitter et al. (2007) introduced a variant of estimation capacity known as projection estimation capacity (PEC). Instead of the potential model space involving all ME in the screening design, the model space for PEC involves all q ME in q-factor

projections of the original design plus all of the associated 2FI. Some notation is introduced here that is used in the remainder of the chapter. If every q-factor projection of a design $\boldsymbol{D}$ is capable of estimating the full q-factor ME+2FI model, then for design $\boldsymbol{D}$ we have $p_q = 1$. As another example, if the full q-factor model is only estimable in 50% of the possible q-factor projections then for design $\boldsymbol{D}$ we have $p_q = 0.5$.

Using the completed catalog of non-isomorphic $\mathrm{OA}(16; 2^{\mathrm{m}}; 2)$ and $\mathrm{OA}(20; 2^{\mathrm{m}}; 2)$, Loeppky et al. were able to identify optimal 16-run and 20-run designs in terms of PEC. The authors also used PEC to create a ranked list of the top six projections of Hadamard matrices of orders 24 and 28.

4.2.7 Historical Research on Orthogonal Designs of 24 Runs

Evangelaras, Koukouvinos et al. (2007) evaluated and ranked the non-isomorphic $\mathrm{OA}(24; 2^6; 2)$ according to generalized minimum aberration. The CFV patterns for the top seven designs are shown in Table 23. To be consistent with the source literature, this section lists CFVs in the following format. Let $\mathrm{f}_{st}, t = 1, \ldots, 4$ be the frequency of s column combinations that give $|J_s| \in \{24,16,8,0\}$ for $s = 1, \ldots, 6$. Therefore we have $F(D) = [(f_{11}, \ldots, f_{14}), \ldots, (f_{61}, \ldots, f_{64})]$. As an example, if a design $\boldsymbol{D}$ has six columns and $(f_{41}, \ldots, f_{47}) = (0,0,9,6)_4$ then of the $\binom{6}{4} = 15$ possible element-wise products of four columns, nine have a column sum of 8 and six have a column sum of 0.

Table 23. The CFVs for Minimum G-aberration Designs (Evangelaras et al. 2007)

G-Aber_24 6.6	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,0,20)_3(0,0,15,0)_4(0,0,0,6)_5(0,0,0,1)_6]$
G-Aber_24 6.7	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,0,20)_3(0,0,15,0)_4(0,0,0,6)_5(0,1,0,0)_6]$
G-Aber_24 6.33	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,2,18)_3(0,0,9,6)_4(0,0,4,2)_5(0,0,0,1)_6]$
G-Aber_24 6.235	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,2,18)_3(0,0,11,4)_4(0,0,2,4)_5(0,0,0,1)_6]$
G-Aber_24 6.234	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,3,17)_3(0,0,8,7)_4(0,0,3,3)_5(0,0,1,0)_6]$
G-Aber_24 6.208	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,3,17)_3(0,0,10,5)_4(0,0,3,3)_5(0,0,1,0)_6]$
G-Aber_24 6.227	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,3,17)_3(0,1,8,6)_4(0,0,1,5)_5(0,0,1,0)_6]$

Evangelaras et al. recognized the problem of ranking designs on only one criterion. First they noted that designs 24.6.6, 24.6.7 and 24.6.235 cannot estimate the full model with the six main effects and each of the 15 two-factor interactions. After computing the D-efficiency of every OA, it was noted that the design 24.6.33 was ranked 41st by the D-criterion with a D-efficiency equal to 73%. The design 24.6.234 was ranked 62nd with D-efficiency 70.8%, the design 24.6.207 was ranked 84th with D-efficiency 69.6%, and the design 24.6.208 was ranked 303rd with D-efficiency 60.15%.

The authors provided an additional list of eight designs that were not in the list of designs with lowest generalized minimum aberration, but could estimate the full 6-factor ME+2FI model with the maximum D-efficiency observed (78.4%). The CFVs for these designs are listed in Table 24.

Table 24. The CFVs for D-optimal Designs (Evangelaras et al. 2007)

D-Opt_24 6.217 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4(0,0,2,4)_5(0,1,0,0)_6]$
D-Opt_24 6.218 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4(0,0,2,4)_5(0,1,0,0)_6]$
D-Opt_24 6.220 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4(0,1,1,4)_5(0,0,1,0)_6]$
D-Opt_24 6.221 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4(0,1,1,4)_5(0,0,1,0)_6]$
D-Opt_24 6.222 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4(0,1,2,3)_5(0,0,0,1)_6]$
D-Opt_24 6.224 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4(0,1,2,3)_5(0,0,0,1)_6]$
D-Opt_24 6.231 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4(0,1,2,3)_5(0,0,0,1)_6]$
D-Opt_24 6.232 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4(0,1,2,3)_5(0,0,0,1)_6]$

Ingram and Tang (2005) ranked the $\mathrm{OA}(24; 2^{\mathrm{m}}; 2)$ projections of Hadamard matrices for $m \in \{6, \ldots, 12\}$ according to minimum G-aberration. Table 25 shows the CFV patterns of the best design for each value of m.

Table 25. The CFVs for Min G-aberration Hadamard Designs (Ingram and Tang 2005)

G-Aber_24 H.1.6.1	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,0,20)_3(0,0,15,0)_4(0,0,0,6)_5(0,0,0,1)_6]$
G-Aber_24 H.1.7.1	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,0,35)_3(0,0,35,0)_4(0,0,0,21)_5(0,1,0,6)_6]$
G-Aber_24 H.1.8.1	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,0,56)_3(0,0,70,0)_4(0,0,0,56)_5(0,4,0,24)_6]$
G-Aber_24 H.1.9.1	$[(0,0,0,9)_1(0,0,0,36)_2(0,0,0,84)_3(0,0,126,0)_4(0,0,0,126)_5(0,12,0,72)_6]$
G-Aber_24 H.1.10.1	$[(0,0,0,10)_1(0,0,0,45)_2(0,0,0,120)_3(0,0,210,0)_4(0,0,0,252)_5(0,30,0,180)_6]$
G-Aber_24 H.1.11.1	$[(0,0,0,11)_1(0,0,0,55)_2(0,0,0,165)_3(0,0,330,0)_4(0,0,0,462)_5(0,66,0,396)_6]$
G-Aber_24 H.1.12.1	$[(0,0,0,12)_1(0,0,0,66)_2(0,0,0,220)_3(0,0,495,0)_4(0,0,0,792)_5(0,132,0,792)_6]$

These designs are all projections of Hadamard matrix H.24.1 as cataloged in Sloane (2013). There are 60 non-isomorphic Hadamard matrices of order 24. Note that H.24.1 is non-isomorphic to the 24 run Plackett-Burman design for 23 factors, which is H.24.60.

Searching through the 60 non-isomorphic Hadamard matrices cataloged in Sloane (2013), Loeppky, Sitter et al. (2007) ranked the top six PEC 24-run designs for $m \in \{6, \ldots, 12\}$ factors. Two search algorithms were used to generate these designs: a "bottom-up" algorithm and a "top-down" algorithm. Using the bottom-up algorithm, each of the top three PEC designs for $m \in \{6, \ldots, 12\}$ factors came from Hadamard matrix 58 (H.24.58). Using the top-down algorithm, each of the top three PEC designs for $m \in \{6, \ldots, 12\}$ factors came from Hadamard matrix 42 (H.24.42). Table 26 shows the CFV patterns for the top three PEC designs generated by top-down search and bottom-up search.

Table 26. The CFVs for PEC-Optimal Designs (Loeppky et al. 2007)

PEC_24 H.58.6.1 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4(0,1,2,3)_5(0,0,0,1)_6]$
PEC_24 H.58.6.2 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4(0,1,2,3)_5(0,0,0,1)_6]$
PEC_24 H.58.6.3 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4(0,1,1,4)_5(0,0,1,0)_6]$
PEC_24 H.42.6.1 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4(0,1,2,3)_5(0,0,0,1)_6]$
PEC_24 H.42.6.2 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4(0,1,2,3)_5(0,0,0,1)_6]$
PEC_24 H.42.6.3 $[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4(0,1,1,4)_5(0,0,1,0)_6]$
PEC_24 H.58.7.1 $[(0,0,0,7)_1(0,0,0,21)_2(0,0,8,27)_3(0,1,11,23)_4(0,2,4,15)_5(0,0,4,3)_6]$
PEC_24 H.58.7.2 $[(0,0,0,7)_1(0,0,0,21)_2(0,0,9,26)_3(0,2,9,24)_4(0,0,6,15)_5(0,1,2,4)_6]$
PEC_24 H.58.7.3 $[(0,0,0,7)_1(0,0,0,21)_2(0,0,9,26)_3(0,2,9,24)_4(0,1,2,18)_5(0,1,2,4)_6]$
PEC_24 H.42.7.1 $[(0,0,0,7)_1(0,0,0,21)_2(0,0,8,27)_3(0,0,15,20)_4(0,1,12,8)_5(0,0,0,7)_6]$
PEC_24 H.42.7.2 $[(0,0,0,7)_1(0,0,0,21)_2(0,0,10,25)_3(0,0,13,22)_4(0,1,10,10)_5(0,0,2,5)_6]$
PEC_24 H.42.7.3 $[(0,0,0,7)_1(0,0,0,21)_2(0,0,10,25)_3(0,0,13,22)_4(0,1,10,10)_5(0,0,2,5)_6]$
PEC_24 H.58.8.1 $[(0,0,0,8)_1(0,0,0,28)_2(0,0,13,43)_3(0,3,21,46)_4(0,5,14,37)_5(0,0,6,22)_6]$
PEC_24 H.58.8.2 $[(0,0,0,8)_1(0,0,0,28)_2(0,0,13,43)_3(0,4,21,45)_4(0,4,10,42)_5(0,0,10,18)_6]$
PEC_24 H.58.8.3 $[(0,0,0,8)_1(0,0,0,28)_2(0,1,13,42)_3(0,2,21,47)_4(0,4,14,38)_5(0,1,6,21)_6]$
PEC_24 H.42.8.1 $[(0,0,0,8)_1(0,0,0,28)_2(0,0,19,37)_3(0,0,27,43)_4(0,2,20,34)_5(0,0,12,16)_6]$
PEC_24 H.42.8.2 $[(0,0,0,8)_1(0,0,0,28)_2(0,0,17,39)_3(0,0,31,39)_4(0,1,24,31)_5(0,0,8,20)_6]$
PEC_24 H.42.8.3 $[(0,0,0,8)_1(0,0,0,28)_2(0,0,21,35)_3(0,1,23,46)_4(0,1,20,35)_5(0,0,12,16)_6]$
PEC_24 H.58.9.1 $[(0,0,0,9)_1(0,0,0,36)_2(0,0,0,84)_3(0,0,126,0)_4(0,0,0,126)_5(0,12,0,72)_6]$
PEC_24 H.58.9.2 $[(0,0,0,9)_1(0,0,0,36)_2(0,2,18,64)_3(0,3,44,79)_4(0,8,32,86)_5(0,2,16,66)_6]$
PEC_24 H.58.9.3 $[(0,0,0,9)_1(0,0,0,36)_2(0,1,21,62)_3(0,6,38,82)_4(0,7,27,92)_5(0,0,24,60)_6]$
PEC_24 H.42.9.1 $[(0,0,0,9)_1(0,0,0,36)_2(0,1,25,58)_3(0,2,46,78)_4(0,3,47,76)_5(0,1,24,59)_6]$
PEC_24 H.42.9.2 $[(0,0,0,9)_1(0,0,0,36)_2(0,1,24,59)_3(0,1,50,75)_4(0,1,58,67)_5(0,0,26,58)_6]$
PEC_24 H.42.9.3 $[(0,0,0,9)_1(0,0,0,36)_2(0,1,26,57)_3(0,2,46,78)_4(0,2,46,78)_5(0,1,26,57)_6]$
PEC_24 H.58.10.1 $[(0,0,0,10)_1(0,0,0,45)_2(0,0,0,120)_3(0,0,210,0)_4(0,0,0,252)_5(0,30,0,180)_6]$
PEC_24 H.58.10.2 $[(0,0,0,10)_1(0,0,0,45)_2(0,3,16,101)_3(0,0,126,84)_4(0,9,60,183)_5(0,12,0,198)_6]$
PEC_24 H.58.10.3 $[(0,0,0,10)_1(0,0,0,45)_2(0,3,18,99)_3(0,0,126,84)_4(0,6,66,180)_5(0,12,0,198)_6]$
PEC_24 H.42.10.1 $[(0,0,0,10)_1(0,0,0,45)_2(0,1,39,80)_3(0,4,75,131)_4(0,2,101,149)_5(0,0,71,139)_6]$
PEC_24 H.42.10.2 $[(0,0,0,10)_1(0,0,0,45)_2(0,1,40,79)_3(0,3,80,127)_4(0,2,100,150)_5(0,2,60,148)_6]$
PEC_24 H.42.10.3 $[(0,0,0,10)_1(0,0,0,45)_2(0,1,39,80)_3(0,3,79,128)_4(0,2,101,149)_5(0,1,67,142)_6$
PEC_24 H.58.11.1 $[(0,0,0,11)_1(0,0,0,55)_2(0,0,0,165)_3(0,0,330,0)_4(0,0,0,462)_5(0,66,0,396)_6]$
PEC_24 H.58.11.2 $[(0,0,0,11)_1(0,0,0,55)_2(0,4,20,141)_3(0,0,210,120)_4(0,14,100,348)_5(0,30,0,432)_6]$
PEC_24 H.58.11.3 $[(0,0,0,11)_1(0,0,0,55)_2(0,4,21,140)_3(0,0,210,120)_4(0,12,104,346)_5(0,30,0,432)_6]$
PEC_24 H.42.11.1 $[(0,0,0,11)_1(0,0,0,55)_2(0,1,56,108)_3(1,6,115,208)_4(0,4,174,284)_5(0,2,144,316)_6]$
PEC_24 H.42.11.2 $[(0,0,0,11)_1(0,0,0,55)_2(0,1,57,107)_3(1,4,119,206)_4(0,6,166,290)_5(0,3,148,311)_6]$
PEC_24 H.42.11.3 $[(0,0,0,11)_1(0,0,0,55)_2(0,1,60,104)_3(0,5,118,207)_4(0,6,164,292)_5(0,4,156,302)_6]$
PEC_24 H.58.12.1 $[(0,0,0,11)_1(0,0,0,55)_2(0,0,0,165)_3(0,0,330,0)_4(0,0,0,462)_5(0,66,0,396)_6]$
PEC_24 H.58.12.2 $[(0,0,0,11)_1(0,0,0,55)_2(0,4,20,141)_3(0,0,210,120)_4(0,14,100,348)_5(0,30,0,432)_6]$
PEC_24 H.58.12.3 $[(0,0,0,11)_1(0,0,0,55)_2(0,4,21,140)_3(0,0,210,120)_4(0,12,104,346)_5(0,30,0,432)_6]$
PEC_24 H.42.12.1 $[(0,0,0,11)_1(0,0,0,55)_2(0,1,56,108)_3(1,6,115,208)_4(0,4,174,284)_5(0,2,144,316)_6]$
PEC_24 H.42.12.2 $[(0,0,0,11)_1(0,0,0,55)_2(0,1,57,107)_3(1,4,119,206)_4(0,6,166,290)_5(0,3,148,311)_6]$
PEC_24 H.42.12.3 $[(0,0,0,11)_1(0,0,0,55)_2(0,1,60,104)_3(0,5,118,207)_4(0,6,164,292)_5(0,4,156,302)_6]$

Notice in Table 26 that the best design found using the bottom-up search for $m \in \{9, \ldots, 12\}$ has ME which are completely uncorrelated with 2FI. In addition, this design has no complete confounding of 2FI with other 2FI. This design is isomorphic to a design listed as the minimum G-aberration designs in Ingram and Tang (2005). Table

27, reproduced from Loeppky, Sitter et al. (2007), shows the corresponding PEC vectors for the designs in Table 26. See the subsection on PEC for an explanation of the p_q notation.

Table 27. The p_4, p_5 and p_6 Values for PEC-optimal Designs (Loeppky et al. 2007)

Design	(p_4,p_5,p_6)
PEC_24 H.58.6.1	(1,1,1.000)
PEC_24 H.58.6.2	(1,1,1.000)
PEC_24 H.58.6.3	(1,1,1.000)
PEC_24 H.42.6.1	(1,1,1.000)
PEC_24 H.42.6.2	(1,1,1.000)
PEC_24 H.42.6.3	(1,1,1.000)
PEC_24 H.58.7.1	(1,1,1.000)
PEC_24 H.58.7.2	(1,1,1.000)
PEC_24 H.58.7.3	(1,1,1.000)
PEC_24 H.42.7.1	(1,1,1.000)
PEC_24 H.42.7.2	(1,1,1.000)
PEC_24 H.42.7.3	(1,1,1.000)
PEC_24 H.58.8.1	(1,1,0.786)
PEC_24 H.58.8.2	(1,1,0.750)
PEC_24 H.58.8.3	(1,1,0.750)
PEC_24 H.42.8.1	(1,1,0.964)
PEC_24 H.42.8.2	(1,1,0.964)
PEC_24 H.42.8.3	(1,1,0.964)
PEC_24 H.58.9.1	(1,1.000,0.000)
PEC_24 H.58.9.2	(1,0.992,0.655)
PEC_24 H.58.9.3	(1,0.992,0.643)
PEC_24 H.42.9.1	(1,1.000,0.881)
PEC_24 H.42.9.2	(1,1.000,0.869)
PEC_24 H.42.9.3	(1,1.000,0.869)
PEC_24 H.58.10.1	(1,1.000,0.000)
PEC_24 H.58.10.2	(1,0.988,0.362)
PEC_24 H.58.10.3	(1,0.988,0.314)
PEC_24 H.42.10.1	(1,0.988,0.814)
PEC_24 H.42.10.2	(1,0.988,0.800)
PEC_24 H.42.10.3	(1,1.000,0.795)
PEC_24 H.58.11.1	(1.000,1.000,0.000)
PEC_24 H.58.11.2	(1.000,0.987,0.312)
PEC_24 H.58.11.3	(1.000,0.987,0.294)
PEC_24 H.42.11.1	(0.996,0.976,0.753)
PEC_24 H.42.11.2	(0.996,0.971,0.753)
PEC_24 H.42.11.3	(1.000,0.982,0.747)
PEC_24 H.58.12.1	(1.000,1.000,0.000)
PEC_24 H.58.12.2	(1.000,0.987,0.273)
PEC_24 H.58.12.3	(1.000,0.967,0.536)
PEC_24 H.42.12.1	(0.993,0.957,0.705)
PEC_24 H.42.12.2	(0.997,0.959,0.703)
PEC_24 H.42.12.3	(0.997,0.965,0.702)

Table 27 shows that for the minimum G-aberration designs, every 5-factor projection of these designs is capable of fitting the full 5-factor model, i.e. they have $p_5 = 1$. However, each of these designs has $p_6 = 0$, which means that a 6-factor projection does not exist that can fit the full model.

For $m \in \{6, \ldots, 10\}$ factors multiple designs have $p_5 = 1$. Therefore the authors ranked these designs on p_6. For the cases where $m = 6$ and $m = 7$ Loeppky et al. found designs where $p_6 = 1$. For the case where $m = 8$, the three projections of Hadamard H.42 all have $p_5 = 1$ and $p_6 = 0.964$. For the case where $m = 9$, the best projections of Hadamard H.42 had $p_6 = 0.881$. For the case where $m = 10$, the best projections of Hadamard H.42 had $p_6 = 0.795$. Notice that the minimum G-aberration designs were not PEC optimal for $m \in \{6, \ldots, 8\}$. This is due to the existence of designs for which $p_5 = 1$ and $p_6 > 0$.

Recently Smucker, del Castillo et al. (2012) created model robust two-level designs for 16, 20, 24 and 28 runs. They used a coordinate exchange algorithm with three steps. In Step 1, a randomly chosen initial design is constructed. In step 2, coordinate exchange is conducted to maximize the design's estimation capacity (EC), where the nature of the maximization is determined by a chosen model space. If EC = 1 after Step 2, Step 3 is conducted in which coordinate exchange is again used to maximize the minimum D-efficiency while maintaining EC = 1. Table 28 lists the CFVs for their published 6 and 8-factor designs. Since these designs are not necessarily balanced and orthogonal, the range of t is $t = 1, \ldots, 13$ and f_{st} is the frequency of s column combinations that give $|J_s| \in \{24,22,20,18,16,14,12,10,8,6,4,2,0\}$ for $s = 1, \ldots, 6$.

Table 28. The CFVs for CLF Designs (Smucker et al. 2012)

Factors	Design	CFV
6	CLF-EC	$[(0,0,0,0,0,0,0,0,0,2,2,0)_1(0,0,0,0,0,2,2,0,2,0,0,0)_2(0,2,2,0,2,0,0,0,0,0,0,0)_3$ $(2,0,0,0,0,0,0,0,0,1,3,5)_4(0,0,0,0,0,1,3,5,5,1,0,0)_5(0,1,3,5,5,1,0,0,0,0,0,0)_6]$
6	CLF-Maximin	$[(0,0,0,0,0,0,0,0,0,0,1,3)_1(0,0,0,0,0,0,1,3,2,0,0,0)_2(0,0,1,3,2,0,0,0,0,0,0,0)_3$ $(2,0,0,0,0,0,0,0,0,0,0,2)_4(0,0,0,0,0,0,0,2,9,4,0,0)_5(0,0,0,2,9,4,0,0,0,0,0,0)_6]$
8	CLF-EC	$[(0,0,0,0,0,0,0,0,0,2,2,3)_1(0,0,0,0,0,2,2,3,1,0,0,0)_2(0,2,2,3,1,0,0,0,0,0,0,0)_3$ $(1,0,0,0,0,0,0,0,0,0,3,4)_4(0,0,0,0,0,0,3,4,12,9,0,0)_5(0,0,3,4,12,9,0,0,0,0,0,0)_6]$
8	CLF-Maximin	$[(0,0,0,0,0,0,0,0,0,0,0,1)_1(0,0,0,0,0,0,0,1,7,0,0,0)_2(0,0,0,1,7,0,0,0,0,0,0,0)_3$ $(7,0,0,0,0,0,0,0,0,1,0,7)_4(0,0,0,0,0,1,0,7,7,13,0,0)_5(0,1,0,7,7,13,0,0,0,0,0,0)_6]$

From the CFVs for the CLF designs in Table 28, one can see nonzero values in $(f_{1,1}, \ldots, f_{1,13})$, indicating these designs are not balanced. Furthermore, nonzero values in $(f_{2,1}, \ldots, f_{2,13})$, particularly for $t < 9$, imply there are moderately high correlations between main effects. In addition, there are two-factor interactions which are fully confounded. Consequently these designs do not qualify as no-confounding designs and will not be considered further.

4.2.8 The Pareto Front

According to Lu, Anderson-Cook et al. (2011), one criterion rarely encompasses all of the qualities a design should have to be effective. Consequently, it is preferable to consider the Pareto front of designs generated by multiple design criteria. Examining the Pareto front allows an experimenter to find a design which has favorable qualities across multiple criteria and will therefore have a greater chance of performing well in varying experimental contexts. However, except for the designs published recently by Smucker, del Castillo et al. (2012), 24-run designs introduced in the literature have been evaluated on a single criteria.

In general when attempting to make an optimal decision based on multiple criteria, it is common that a choice which optimizes one criterion does not optimize another. One method of handling this problem is the approach used by Smucker et al. to

optimize one criterion and then optimize an additional criterion subject to the optimal level of the first criteria. A second approach is to create and evaluate a Pareto efficient set, or Pareto front, of choices based on the multiple criteria. A decision element A is included in a Pareto efficient set if there does not exist an element B which outranks element A in every criterion. That is to say element A is not dominated by any other element. Lu, Anderson-Cook et al. (2011) state that the optimal option on the Pareto front is often determined by a process of standardizing and weighting the optimality criteria.

4.3 Methodology

The methodology used to identify and recommend 24-run no-confounding designs involved a four stage process: an organization and assessment of nonregular designs in the current statistical literature; the creation and utilization of a column exchange search algorithm; the creation of a Pareto efficient set of candidate designs; and the evaluation of the Pareto efficient set using Monte Carlo simulation. Five categories of designs discussed in the background section of this chapter were organized from the literature. Collectively, these designs will be referred to in this chapter as the baseline designs, as they have already been established as optimal designs based on various criteria. These baseline designs serve as the benchmark for judging designs created by the column exchange algorithm.

The nomenclature in Table 29 is used to refer to the baseline designs from the literature and the algorithmically created designs. Although there are two categories of G-aberration designs, the categories are not merged in part to allow the reader to trace the design back to the literature source via the nomenclature, and in part because the

G-aberration designs in Evangelaras, Koukouvinos et al. (2007) only include 6-factor designs.

Table 29. Nomenclature for Baseline and Algorithmically-Generated Designs

Design Type (Source)	Nomenclature Format	Example
G-aberration (Ingram and Tang 2005)	G-Aber_24 H.*Source_Hadamard_ID*.Factors.*Literature ID*	G-Aber_24 H.1.6.1
PEC (Loeppky et al. 2007)	PEC_24 H.*Source_Hadamard_ID*.Factors.*Literature ID*	PEC_24 H.58.6.1
G-aberration (Evangelaras et al. 2007)	G-Aber_24 *Factors*.*Literature ID*	G-Aber_24 6.208
D-Optimal (Evangelaras et al. 2007)	D-Opt_24 *Factors*.*Literature ID*	D-Opt_24 6.217
Min Max Correlation	MinMax_24 *Factors*.Algorithm_*Result_ID*	MinMax_24 6.1
Column Exchange Algorithm	CEA_24 *Factors*.Algorithm_*Result_ID*	CEA_24 6.2025

4.3.1 The Baseline Designs

The first set of baseline designs are the minimum G-aberration designs of 24 runs and 6-12 factors from Ingram and Tang (2005). Recall that these designs have minimum G-aberration out of all projections of the 60 non-isomorphic Hadamard matrices of order 24.

Table 30 shows an abbreviated CFV for each of these designs as well as the number of nonzero correlations. The CFV is abbreviated to remove frequencies which are always null or involve irrelevant higher-order interactions. The abbreviated CFV has the following format. Let $f_{st}, t = 1, \dots, 4$ be the frequency of s column combinations that give $|J_s| = \{24,16,8,0\}$ for $s = 1, \dots, 4$. Then the abbreviated CFV is

$F(\boldsymbol{D}) = [(f_{11}, \dots, f_{14}), \dots, (f_{41}, \dots, f_{44})]$.

Table 30. CFVs for G-aberration Designs (Ingram and Tang 2005)

OA ID	CFV	Nonzero Column Sums
G-Aber_24 H.1.6.1	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,0,20)_3(0,0,15,0)_4]$	15
G-Aber_24 H.1.7.1	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,0,35)_3(0,0,35,0)_4]$	35
G-Aber_24 H.1.8.1	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,0,56)_3(0,0,70,0)_4]$	70
G-Aber_24 H.1.9.1	$[(0,0,0,9)_1(0,0,0,36)_2(0,0,0,84)_3(0,0,126,0)_4]$	126
G-Aber_24 H.1.10.1	$[(0,0,0,10)_1(0,0,0,45)_2(0,0,0,120)_3(0,0,210,0)_4]$	210
G-Aber_24 H.1.11.1	$[(0,0,0,11)_1(0,0,0,55)_2(0,0,0,165)_3(0,0,330,0)_4]$	330
G-Aber_24 H.1.12.1	$[(0,0,0,12)_1(0,0,0,66)_2(0,0,0,220)_3(0,0,495,0)_4]$	495

The PEC designs from Loeppky, Sitter et al. (2007) are the second set of designs. Recall that these designs were created by column exchange algorithms using Hadamard matrix columns. It turned out that the top six designs of m-factors came either from Hadamard matrix H.58 or H.42 as indexed in Sloane (2013). Table 31 shows the CFV and count of nonzero column sums for each of the PEC designs.

Table 31. CFVs and Correlations for PEC Designs (Loeppky et al. 2007)

OA ID	CFV	Nonzero Column Sums
PEC_24 H.58.6.1	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9
PEC_24 H.58.6.2	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9
PEC_24 H.58.6.3	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4]$	9
PEC_24 H.42.6.1	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9
PEC_24 H.42.6.2	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9
PEC_24 H.42.6.3	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4]$	9
PEC_24 H.58.7.1	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,8,27)_3(0,1,11,23)_4]$	20
PEC_24 H.58.7.2	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,9,26)_3(0,2,9,24)_4]$	20
PEC_24 H.58.7.3	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,9,26)_3(0,2,9,24)_4]$	20
PEC_24 H.42.7.1	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,8,27)_3(0,0,15,20)_4]$	23
PEC_24 H.42.7.2	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,10,25)_3(0,0,13,22)_4]$	23
PEC_24 H.42.7.3	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,10,25)_3(0,0,13,22)_4]$	23
PEC_24 H.58.8.1	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,13,43)_3(0,3,21,46)_4]$	37
PEC_24 H.58.8.2	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,13,43)_3(0,4,21,45)_4]$	38
PEC_24 H.58.8.3	$[(0,0,0,8)_1(0,0,0,28)_2(0,1,13,42)_3(0,2,21,47)_4]$	37
PEC_24 H.42.8.1	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,19,37)_3(0,0,27,43)_4]$	46
PEC_24 H.42.8.2	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,17,39)_3(0,0,31,39)_4]$	48
PEC_24 H.42.8.3	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,21,35)_3(0,1,23,46)_4]$	45
PEC_24 H.58.9.1	$[(0,0,0,9)_1(0,0,0,36)_2(0,0,0,84)_3(0,0,126,0)_4]$	126
PEC_24 H.58.9.2	$[(0,0,0,9)_1(0,0,0,36)_2(0,2,18,64)_3(0,3,44,79)_4]$	67
PEC_24 H.58.9.3	$[(0,0,0,9)_1(0,0,0,36)_2(0,1,21,62)_3(0,6,38,82)_4]$	66
PEC_24 H.42.9.1	$[(0,0,0,9)_1(0,0,0,36)_2(0,1,25,58)_3(0,2,46,78)_4]$	74
PEC_24 H.42.9.2	$[(0,0,0,9)_1(0,0,0,36)_2(0,1,24,59)_3(0,1,50,75)_4]$	76
PEC_24 H.42.9.3	$[(0,0,0,9)_1(0,0,0,36)_2(0,1,26,57)_3(0,2,46,78)_4]$	75
PEC_24 H.58.10.1	$[(0,0,0,10)_1(0,0,0,45)_2(0,0,0,120)_3(0,0,210,0)_4]$	210
PEC_24 H.58.10.2	$[(0,0,0,10)_1(0,0,0,45)_2(0,3,16,101)_3(0,0,126,84)_4]$	145
PEC_24 H.58.10.3	$[(0,0,0,10)_1(0,0,0,45)_2(0,3,18,99)_3(0,0,126,84)_4]$	147
PEC_24 H.42.10.1	$[(0,0,0,10)_1(0,0,0,45)_2(0,1,39,80)_3(0,4,75,131)_4]$	119
PEC_24 H.42.10.2	$[(0,0,0,10)_1(0,0,0,45)2(0,1,40,79)_3(0,3,80,127)_4]$	124
PEC_24 H.42.10.3	$[(0,0,0,10)_1(0,0,0,45)_2(0,1,39,80)_3(0,3,79,128)_4]$	122
PEC_24 H.58.11.1	$[(0,0,0,11)_1(0,0,0,55)_2(0,0,0,165)_3(0,0,330,0)_4]$	330
PEC_24 H.58.11.2	$[(0,0,0,11)_1(0,0,0,55)_2(0,4,20,141)_3(0,0,210,120)_4]$	234
PEC_24 H.58.11.3	$[(0,0,0,11)_1(0,0,0,55)_2(0,4,21,140)_3(0,0,210,120)_4]$	235
PEC_24 H.42.11.1	$[(0,0,0,11)_1(0,0,0,55)_2(0,1,56,108)_3(1,6,115,208)_4]$	179
PEC_24 H.42.11.2	$[(0,0,0,11)_1(0,0,0,55)_2(0,1,57,107)_3(1,4,119,206)_4]$	182
PEC_24 H.42.11.3	$[(0,0,0,11)_1(0,0,0,55)_2(0,1,60,104)_3(0,5,118,207)_4]$	184
PEC_24 H.58.12.1	$[(0,0,0,12)_1(0,0,0,66)_2(0,0,0,220)_3(0,0,495,0)_4]$	495
PEC_24 H.58.12.2	$[(0,0,0,12)_1(0,0,0,66)_2(0,5,25,190)_3(0,0,330,165)_4]$	360
PEC_24 H.58.12.3	$[(0,0,0,12)_1(0,0,0,66)_2(0,5,55,160)_3(0,25,140,330)_4]$	225
PEC_24 H.42.12.1	$[(0,0,0,12)_1(0,0,0,66)_2(0,2,77,141)_3(3,6,169,317)_4]$	257
PEC_24 H.42.12.2	$[(0,0,0,12)_1(0,0,0,66)_2(0,2,77,141)_3(1,6,181,307)_4]$	267
PEC_24 H.42.12.3	$[(0,0,0,12)_1(0,0,0,66)_2(0,3,73,144)_3(1,8,177,309)_4]$	262

The third and fourth sets of baseline designs were discussed in the background section of this chapter and come from the complete non-isomorphic catalog of

$OA(24; 2^6; 2)$. The five designs ranked highest by G-aberration and eight designs ranked highest by D-optimality were published in Evangelaras, Koukouvinos et al. (2007). The G-aberration designs are shown in Table 32. Note that two designs which appeared in Table 23 of the background section were removed. Design G-Aber_24 6.235 was removed because $p_6 < 1$ and G-Aber_24 6.227 was removed because there are several pairs of 2FI effects which have an excessively high correlation of 2/3.

Table 32. CFVs for 6-Factor Minimum G-aberration Designs

OA ID	CFV	Nonzero Column Sums
G-Aber_24 6.6	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,0,20)_3(0,0,15,0)_4]$	15
G-Aber_24 6.7	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,0,20)_3(0,0,15,0)_4]$	15
G-Aber_24 6.33	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,2,18)_3(0,0,9,6)_4]$	11
G-Aber_24 6.234	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,3,17)_3(0,0,8,7)_4]$	11
G-Aber_24 6.208	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,3,17)_3(0,0,10,5)_4]$	13

The CFVs and count of nonzero correlations for the eight designs which ranked highest according to D-optimality are shown in Table 33.

Table 33. CFVs for 6-Factor D-optimal Designs

OA ID	CFV	Nonzero Column Sums
D-Opt_24 6.217	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4]$	9
D-Opt_24 6.218	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4]$	9
D-Opt_24 6.220	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4]$	9
D-Opt_24 6.221	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,5,15)_3(0,0,4,11)_4]$	9
D-Opt_24 6.222	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4]$	9
D-Opt_24 6.224	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,6,14)_3(0,0,3,12)_4]$	9
D-Opt_24 6.231	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9
D-Opt_24 6.232	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9

A design subcategory was made from three of the PEC baseline designs. These designs have the property that all nonzero correlations between any combination of ME and 2FI effects are equal to 1/3. Recall that the only possible J-characteristics for 24-run designs with balanced orthogonal columns are 24, 16, 8 and 0, therefore 8/24 = 1/3 is the

minimum possible nonzero correlation in absolute value. Furthermore, these designs have the minimum number of correlations of 1/3 of all published designs for a given number of factors. A 24-run design of 6-12 factors with orthogonal 2FI does not exist (a resolution V design). Therefore over all known designs, these *MinMax Correlation* designs have a minimum frequency and minimum value for their maximum correlation. Table 34 shows the CFVs for the *MinMax Correlation* designs.

Table 34. CFVs for MinMax Correlation Designs

OA ID	PEC ID	CFV	Nonzero Column Sums
MinMax_24 6.1	PEC_24 H.58.6.1	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9
MinMax_24 7.1	PEC_24 H.42.7.1	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,8,27)_3(0,0,15,20)_4]$	23
MinMax_24 8.1	PEC_24 H.42.8.1	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,19,37)_3(0,0,27,43)_4]$	45

The *MinMax Correlation* designs for 9-12 factors are also the G-aberration designs from Ingram and Tang (2005). To avoid confusion, the G-aberration designs will not be double labeled as *MinMax Correlation* designs.

4.3.2 The Column Exchange Algorithms

Two column exchange search algorithms were developed to improve on the baseline designs from the literature. Completely enumerated catalogs of $\mathrm{OA}(24; 2^{\mathrm{m}}; 2)$ do not exist for $m > 7$ so a search algorithm was required that was capable of producing original OAs. The algorithm did not explicitly search the complete catalog of non-isomorphic $\mathrm{OA}(24; 2^{\mathrm{m}}; 2)$ for $m = 6$ and $m = 7$. This is because the $\mathrm{OA}(24; 2^7; 2)$ catalog provided in Bulutoglu and Margot (2008) contains a large number (57,389) of OAs and the $\mathrm{OA}(24; 2^6; 2)$ catalog has been thoroughly researched.

The search algorithms also provided the flexibility to create and assess non-orthogonal designs. Preliminary research had shown that by relaxing the requirement for orthogonal main effects, criteria such as $E(s^2)$ could be improved. Thus the search algorithm was designed to search both the space of balanced orthogonal designs, as well as unbalanced, non-orthogonal designs.

Two column-exchange algorithms were developed to search the design space. In the current literature, column exchange algorithms which generated 24-run OAs were based on projections of Hadamard matrices. Research of the complete catalog of 20-run non-isomorphic OAs has demonstrated the existence of OAs that are not Hadamard projections, yet have superior qualities to those which are projections. A goal of this research was to search the space of orthogonal and non-orthogonal arrays that were not projections of Hadamard matrices in order to determine if superior designs existed. Consequently, two column-exchange methods were developed that were not dependent on Hadamard matrix columns.

The first column-exchange algorithm, called the *Generating-Matrix Column Exchange Algorithm* was derived from research by Johnson and Jones (2010) on 16-run balanced orthogonal designs. Their paper demonstrated that columns in no-confounding designs of 6-8 factors could be constructed in a classical fashion with a 2^4 full factorial or replicated 2^3. That is to say that columns E, F, etc. could be constructed from a formula involving columns A-D. For eight of the 27 non-isomorphic designs with six factors and 16 runs, it is the case that the E and F main effects are only partially correlated with two-factor and higher interactions involving main effects A, B, C and D. These eight no-

confounding designs can all be written as a 2^4 factorial for columns A-D with columns E and F of the form

$$\boldsymbol{c} = \frac{1}{2}(\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} - \boldsymbol{x} * \boldsymbol{y} * \boldsymbol{z}) \quad (1)$$

Here $\boldsymbol{x}$,$\boldsymbol{y}$, and $\boldsymbol{z}$ are two-way or higher interaction columns of the A-D ME columns, for example AB, BC and CD. The $\boldsymbol{x} * \boldsymbol{y} * \boldsymbol{z}$ term represents the component-wise Hadamard product of the three columns, for example AB*BC*CD = AD. The authors also demonstrated that for the 55 designs with 7 factors in 16 runs, the subset of these designs with the no-confounding property also had a similar relationship among columns.

The observations in Johnson and Jones (2010) concerning the relationships between columns in equation (1) were utilized to create the *Generating-Matrix Column Exchange Algorithm*. For the 16-run designs, the interaction columns were created from the columns of a 2^4 full factorial. Since a 2^k full factorial option was not available for the 24-run case, the 5-factor 24-run resolution IV design with uncorrelated ME and 2FI was used as a substitute.

There is only one non-isomorphic OA with 24 runs and 5 factors of strength 3. In OA notation this is $OA(24; 2^5; 3)$. As this will be relevant to future discussion in this chapter, note that according to Schoen and Mee (2012) there is only one non-isomorphic $OA(24; 2^m; 3)$ for $m \in \{5,7,8,9,10,11,12\}$. There are two non-isomorphic $OA(24; 2^m; 3)$ for $m \in \{4,6\}$. An $OA(24; 2^m; 3)$ for $m > 12$ has not been found.

Two approaches were tested to determine how many columns should be held constant in the search algorithm. The first approach was to make columns A-E the 5-factor 24-run resolution IV design. The second approach was to only hold columns

A-C constant. An analysis of the baseline designs revealed that three columns of every design could be found such that if the rows were properly sorted, the first column would have twelve -1's followed by twelve +1's. The second column would have six -1's followed by six +1's, repeated twice. The third column would have three -1's followed by three +1's, with this pattern repeated four times. Note that these three columns are pair-wise orthogonal. Assessment of the algorithm performance revealed that holding three columns constant instead of five created more favorable diversity in the generated designs. Consequently, three columns were held constant rather than five.

For the *Generating-Matrix Column Exchange Algorithm*, the interaction columns used as arguments in equation (1) were generated from the columns of the 5-factor 24-run OA of strength 3. The power set of all possible component-wise column products was generated. From the power set, all 2FI,…,5FI columns were selected and placed in a 26-column matrix notated as $\boldsymbol{G}$. A candidate design column was generated by selecting three columns from $\boldsymbol{G}$ and plugging these columns into equation (1). The following are the steps of the *Generating-Matrix Column Exchange Algorithm* to construct a 24-run m-factor no-confounding design.

The *Generating-Matrix Column Exchange Algorithm*

1. Generate the matrix $\boldsymbol{G}$**,** the matrix of 2FI,…,5FI effects from the 5-factor 24-run resolution IV design
2. Generate a semi-random design $\boldsymbol{D}$ with the 3 constant columns and $m - 3$ random columns from the $OA(24; 2^{12}; 3)$
3. For column $i = 4$ to m

3.1. Select a new 3-column group from $\boldsymbol{G}$ and generate a candidate column $\boldsymbol{c}$ using equation (1)

3.2. Substitute $\boldsymbol{c}$ into column i of $\boldsymbol{D}$

3.3. Screen out undesirable designs

3.3.1. If $\boldsymbol{D}$ has fully confounded effects, break and return to 3.1

3.3.2. If p_5<1 for $\boldsymbol{D}$, break and return to 3.1

3.4. Compute three design metrics for $\boldsymbol{D}$

3.4.1. $E(s^2)$

3.4.2. PEC for 6-factor projections (p_6)

3.4.3. Average average VIF for the 5-factor full model over all 5-factor projections

3.5. Check if $\boldsymbol{D}$ belongs on the Pareto front for the three design metrics and add $\boldsymbol{D}$ if appropriate

3.6. Remove any dominated designs from the Pareto front

3.7. Repeat for all $\binom{26}{3}$ possible arguments for $\boldsymbol{x}$,$\boldsymbol{y}$, and $\boldsymbol{z}$ from $\boldsymbol{G}$

4. Go back to Step 3 and repeat for j cycles

5. Output the Pareto front of designs

The Pareto front of designs created by the *Generating-Matrix Column Exchange Algorithm* produced good designs relative to the baseline designs for $m \leq 10$. However, when $m > 10$, it required over 24 hours to make one cycle through the algorithm. This was due primarily to the computationally expensive design metrics, PEC and average VIF, which potentially had to be calculated $(\mathrm{m} - 3) * \binom{26}{3}$ times per cycle.

In order to create more diversity in the generated designs, a modification was made to the *Generating-Matrix Column Exchange Algorithm* where the arguments $\boldsymbol{x}$ and $\boldsymbol{y}$ in equation (1) were fixed for an algorithm cycle and only $\mathbf{z}$ was changed to the 26 possible columns in $\boldsymbol{G}$. For the next 26 cycles, $\boldsymbol{y}$ was changed to a new column of $\boldsymbol{G}$ and again $\mathbf{z}$ was changed to each of the 26 possible columns. Finally, after all possible values of $\boldsymbol{y}$ had been tried, $\boldsymbol{x}$ was changed. This process was repeated until all possible column combinations of $\boldsymbol{G}$ had been tried. This modified algorithm created more diversity in the generated designs.

The second algorithm, called the *Random Structured-Column Exchange Algorithm* was created as a computationally less expensive alternative to the *Generating-Matrix Column Exchange Algorithm*. Generating a completely random 2-level column was not efficient since there are $2^{24} \approx 1.678 \times 10^7$ of such columns. By considering only random balanced columns, the column search space is reduced to $\binom{24}{12} \approx 2.704 \times 10^6$ possible columns. In an effort to reduce the column search space further, the baseline designs were examined. It was discovered that for each of these designs, if a column was divided into four consecutive sub-columns of six, every block contained at least two +1's and at least two -1's. Furthermore, if the columns were divided into two sub-columns of twelve runs, each sub-column contained six +1's and six -1's. If this structure of ± 1 was enforced within a column, there would be 7.225×10^5 possible columns – a reduction in the column search space by 2 orders of magnitude from the completely random column approach (see Appendix B for the computations.)

The following are the steps in the *Random Structured-Column Exchange Algorithm* to construct a 24-run m-factor no-confounding design.

The *Random Structured-Column Exchange Algorithm.*

1. For column $i = 4$ to m
    1.1. While no improvement or for a max of 50 random columns do the following
        1.1.1. Generate a balanced, structured random column, $\boldsymbol{c}$
        1.1.2. Substitute $\boldsymbol{c}$ into column i of $\boldsymbol{D}$
        1.1.3. Screen out undesirable designs
            1.1.3.1. If $\boldsymbol{D}$ has fully confounded effects, break and return to 1.1
            1.1.3.2. If p_5<1 for $\boldsymbol{D}$, break and return to 1.1
        1.1.4. Compute three design metrics for $\boldsymbol{D}$
            1.1.4.1. Compute $E(s^2)$
            1.1.4.2. Compute PEC for 6-factor projections (p_6)
            1.1.4.3. Compute the average average VIF for the 5-factor full model over all 5-factor projections
        1.1.5. Check if $\boldsymbol{D}$ belongs on the Pareto front for the three design metrics and add $\boldsymbol{D}$ if appropriate
        1.1.6. Remove any dominated designs from the Pareto front
    1.2. Next i
2. Go back to step 1 and repeat for j cycles
3. Output the Pareto front of designs

Three metrics were chosen to create the Pareto front of designs: $E(s^2)$, p_6, and average average VIF for 5-factor full models. These design evaluation criteria were

defined in the background section. More criteria were not used due to the significant increase in computation time as a result of evaluating four or more metrics: The three evaluation criteria were chosen based on their ability to assess the designs both on their projection properties as well as their coefficient estimation capability.

The $E(s^2)$ was used to evaluate the 16-run no-confounding designs and was used to maintain some consistency with previous research. In addition it has proven to be an excellent metric for evaluating supersaturated designs, which is applicable to the case of 24-run designs when used to estimate full models of more than seven ME. An aspect of $E(s^2)$ that contrasts with G-aberration is that it does not prioritize the minimization of correlations of ME and 2FI over 2FI correlations, but rather gives both types of effect correlations equal weight.

PEC is an important quality in screening designs, as it is rare that all factors considered in the screening stage will ultimately be found to be active. Ideally $p_5 = p_6 = 1$ since 24-run designs have enough runs to estimate the full model for up to six factors. However, Evangelaras, Koukouvinos et al. (2007) noted that the designs in the complete catalog of $OA(24; 2^6; 2)$ which had minimum G-aberration had $p_6 < 1$. If an experimenter anticipates that only a subset of the considered factors will be active, then requiring that $p_6 = 1$ may be overly restrictive. Therefore, designs with a value of $p_5 < 1$ were rejected, but p_6 was used to define the efficient frontier of designs, in order to allow for designs which may perform well when there is a small number of active effects despite a relatively low value of p_6.

Since these designs have many partially correlated effects, there is a risk that the variance of some effect estimates is unacceptably high. The VIF is a method of

quantifying the variance of effect estimates. Unfortunately in the case of non-orthogonal designs the VIFs vary across effects. In order to assess a design in terms of VIFs, all possible 5-factor projections were examined and the VIFs for the 5-factor full model were computed. The VIFs were averaged across effects, and then averaged across every 5-factor projection. The result was termed the average average VIF.

4.3.3 Reducing the Size of the Pareto Efficient Set

Despite the fact that only three design metrics were used in the search algorithm, the Pareto front generated by the algorithm was populated with over 80 designs for $m > 8$. It was infeasible in terms of time to evaluate the entire list of designs on the Pareto front. Furthermore, many of these designs were very similar, differing in their metrics by a relatively small amount.

To reduce the quantity of designs on the Pareto front, the CFVs were generated for both the baseline designs and the Pareto front designs. The CFV had the form

$$\mathrm{F(D)} = [(\mathrm{f}_{11}, \ldots, \mathrm{f}_{17}), (\mathrm{f}_{21}, \ldots, \mathrm{f}_{27}), (\mathrm{f}_{31}, \ldots, \mathrm{f}_{37}), (\mathrm{f}_{41}, \ldots, \mathrm{f}_{47})]$$

Where $(\mathrm{f}_{11}, \ldots, \mathrm{f}_{17})$ corresponded to the frequencies of the following J-characteristics: $(24,20,16,12,8,4,0)$. The CFVs were summarized in six categories listed in Table 35.

Table 35. CFV Summary Categories

Category Expression	Description
$\sum_{t=1}^{4} f_{2t}$	Total ME correlations greater than 1/3
f_{25}	Total ME correlations equal to 1/3
$\sum_{i=1}^{4}\sum_{t=1}^{4} f_{it}$	Total correlations greater than 1/3 involving ME, ME and 2FI, and 2FI and 2FI
$\sum_{i=1}^{4} f_{i5}$	Total correlations equal to 1/3 involving ME, ME and 2FI, and 2FI and 2FI
$\sum_{i=1}^{4}\sum_{t=1}^{6} f_{it}$	Total nonzero correlations involving ME, ME and 2FI, and 2FI and 2FI
$\sum_{i=1}^{4}\sum_{t=1}^{5} f_{it}$	Total correlations greater than 1/6 involving ME, ME and 2FI, and 2FI and 2FI

The six CFV summary categories in Table 35 were computed for each design on the Pareto front. A second Pareto front based on this data was then generated. This reduced the total candidate design list for all factors to a manageable 267.

4.3.4 The Monte Carlo Simulation

A Monte Carlo simulation was used to narrow down the 267 designs. This was a rarely used approach in design comparison. In Marley and Woods (2010) a Monte Carlo simulation was used to compare analysis strategies for supersaturated designs, but a journal article concerning Monte Carlo simulation as a method to compare designs has not been published.

The simulation automated the analysis of the randomly generated responses and produced an estimated model. The terms and coefficients in the estimated model were compared to the terms and coefficients in the randomly generated model and a binary variable recorded whether a type I or type II variable selection error occurred. Data was

also recorded for the minimum, average, and maximum percent error in the coefficient estimates for all model terms. After a given number of simulations, the data was aggregated to report two simulation metrics: the percentage of models with at least the active effects, and the percentage of correct models (those models without type I or type II variable selection errors.

The number of simulations was determined using the Normal approximation of the standard deviation for a binomial proportion $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. When p = 0.5, if it is desired that $\sigma_{\hat{p}} = 0.75\%$ then it is required that $n = 4444$. After weighing the simulation time against the requirements for the width of a 95% confidence interval, it was determined that 5000 simulation runs would allow sufficiently small confidence intervals to compare the various OAs.

Since the ME+2FI full model for seven or more factors has more than 24 terms, an analysis method suitable for supersaturated designs was required. The method chosen was a modified two-stage forward stepwise analysis approach based on doctoral research in Shinde (2012). In the first stage, the analysis is conducted considering only ME terms. In the second stage, only the active ME terms from stage one and all associated 2FI involving the active ME are considered.

The Akaike's information criterion corrected (AICc) discussed in Hurvich and Tsai (1989) and Akaike (1974) was used as the model selection criterion. Since the analysis software did not perform bi-directional elimination when the AICc criterion was used to rank potential models, a modification of the forward selection procedure was used. The modification involved eliminating variables in the final model which had a p-

value above a certain threshold. The p-value thresholds of 0.10 in stage 1 and 0.05 in stage 2 were used. The added step of removing variables after the stepwise regression reduced the type I variable-selection error rate. This approach was identical to that used for the research in Chapter 3 of this dissertation.

An important consideration for the Monte Carlo simulation was the appropriate maximum number of ME and 2FI that the random response models might be composed of. Choosing integers p and q such that the simulation generated a random number of 2 to p ME and a random number of 0 to q 2FI required balancing poor estimation performance with a thorough design capability analysis. Li, Sudarsanam et al. (2006) conducted a meta-analysis of published experiments involving a DOE methodology and reported that for the 113 combined experimental data sets, 41% of the potential ME were active and 11% of the potential 2FI were active. According to these results, when considering six potential factors, it is expected that there are 2-3 active ME and 1 active 2FI.

For this simulation the number of ME was a random number in the range $2 \leq ME \leq 4$ and the number of 2FI was a random number in the range $1 \leq 2FI \leq 3$. These ranges for the number of ME and 2FI in the polynomial models were consistent with the model meta-analysis in Li, Sudarsanam et al. (2006), as well as with the sparsity-of-effects principle. For the ME and 2FI, the respective coefficient random variables X and Y ranged from $2 \leq \mathrm{X} \leq 5.44$ and $2 \leq \mathrm{Y} \leq 3.5$. The sign of the coefficient was determined by a Bernoulli random variable with p = 0.5.

After reviewing the results of the simulation for the 267 designs on the Pareto front, the list was reduced to 90 designs which ranked highest according to the simulation

metrics. At least 10 designs of m factors, $m \in \{6, \ldots, 12\}$, were chosen so that good designs which performed relatively poorly due to sampling variation would not be overlooked. Note that 5000 simulation runs were used to prevent large sample variance in the average simulation metrics.

To determine which of the 90 designs produced the best simulation results, a much larger simulation study was conducted. The simulation was repeated 20 times for each candidate design for a total of 20x5000 = 100,000 simulations. For each design, two design performance metrics were recorded: the average percentage of trials which produced correct models, and the average percentage of trials which produced at least the correct model. The binomial approximation was used to calculate the 95% confidence interval for the average simulation metrics over 100,000 simulations. The interval differed from the metric point estimate only in the 4th decimal place, allowing for precise differentiation between designs. For each number of factors m, the highest ranking design according to each of the two performance metrics was selected as the best algorithm design. These designs were then compared to the baseline designs to determine which should be recommended as the 24-run no-confounding designs.

4.4 Results

The results of the Monte Carlo simulation showed that the minimum G-aberration designs listed in Ingram and Tang (2005) are clearly the best 24-run no-confounding designs in terms of the average percentage of trials which produced correct models, and the average percentage of trials which produced at least the correct model. Even though the *MinMax Correlation* designs have fewer nonzero correlations between effects overall,

these designs have more ME correlated with 2FI, which according to the simulation this leads to inferior model-fitting performance.

The *Generating-Matrix Column Exchange Algorithm* and the *Random Structured-Column Exchange Algorithm* were able to discover two unpublished *MinMax Correlation* design for seven and nine factors respectively. Table 36 is an updated version of Table **34** with the CFV of the newly discovered *MinMax Correlation* designs (their original nomenclature was CEA_24_7_102 and CEA_24_9_2025). The CFVs in Table **36** have the form $F(\mathbf{D}) = [(f_{11}, \dots, f_{14}), (f_{21}, \dots, f_{24}), (f_{31}, \dots, f_{34}), (f_{41}, \dots, f_{44})]$, where $(f_{i1}, \dots, f_{i4})$ corresponds to the frequencies of the J-characteristics $(24,16,8,0)$.

Table 36. CFVs for Updated List of MinMax Correlation Designs

Original ID	MinMax ID	CFV	Nonzero Column Sums
PEC_24 58.6.1	MinMax_24 6.1	$[(0,0,0,6)_1(0,0,0,15)_2(0,0,4,16)_3(0,0,5,10)_4]$	9
CEA_24 7.102	MinMax_24 7.1	$[(0,0,0,7)_1(0,0,0,21)_2(0,0,6,29)_3(0,0,15,20)_4$	21
PEC_24 42.8.1	MinMax_24 8.1	$[(0,0,0,8)_1(0,0,0,28)_2(0,0,19,37)_3(0,0,27,43)_4]$	45
CEA_24 9.2025	MinMax_24 9.1	$[(0,0,0,9)_1(0,0,0,36)_2(0,0,10,74)_3(0,0,98,28)_4]$	108

The remainder of this section reports the simulation results in terms of two metrics. The first simulation metric is the percentage of models fit to the simulated responses which contained at least the active effects, referred to as the *% at least correct*. The second simulation metric is the percentage of models fit to the simulated responses which contained only the active effects, referred to as the *% correct.* The results include metrics for each baseline design and the best design created by the column exchange algorithms. In the event that an algorithmically created design is not ranked highest for both metrics, the results are reported for two designs. The results for the 6-factor designs include the additional 13 6-factor baseline designs from Evangelaras, Koukouvinos et al.

(2007). Figure 24 shows the simulation results for *% at least correct* for the 13 D-optimal and minimum G-aberration designs.

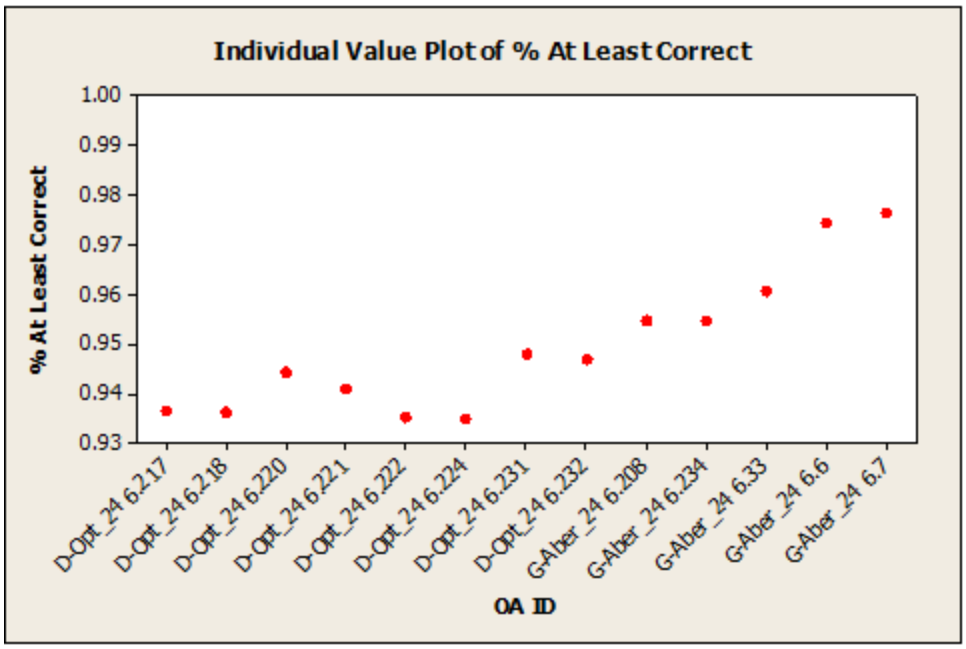

Figure 24. *% at Least Correct* for G-aberration and D-Optimal 6-Factor Designs

Figure 24 shows that the two non-isomorphic OAs of strength three, *G-Aber_24 6.6* and *G-Aber_24 6.7* outperform the D-optimal and other G-aberration designs in terms of *% at least correct.*

Figure 25 shows the simulation results for *% at least correct* for the G-aberration design in Ingram and Tang (2005) and the PEC designs in Loeppky, Sitter et al. (2007).

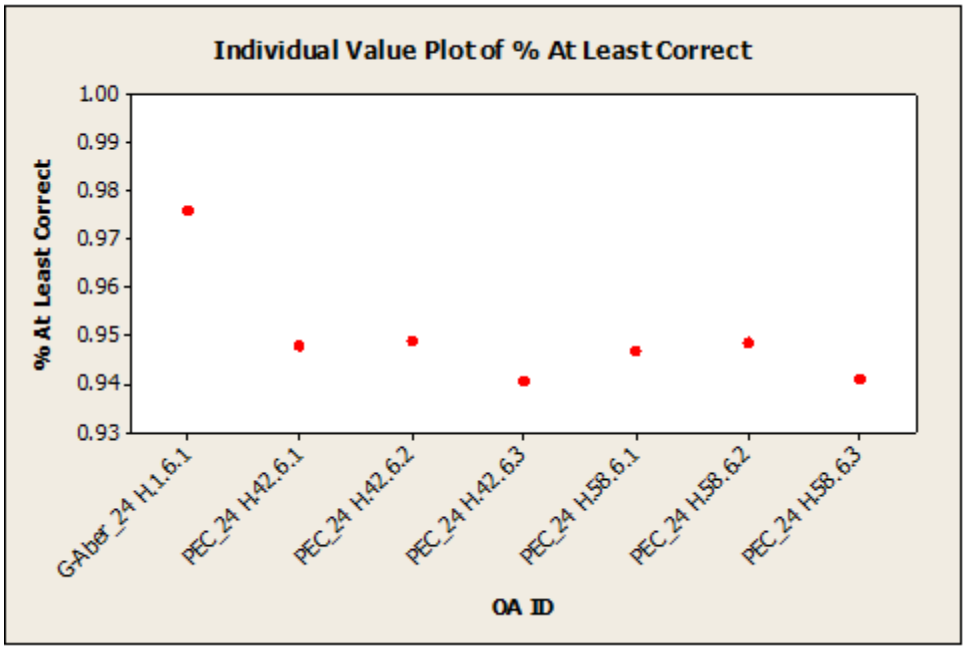

Figure 25. *% at Least Correct* for 6-Factor G-aberration and PEC Designs

Figure 25 shows that the minimum G-aberration design, *G-Aber_24 H.1.6.1* (which is isomorphic to *G-Aber_24 6.6*), performs significantly better than the PEC designs despite the fact that $p_6 = 1$ for the PEC designs and $p_6 = 0$ for the minimum G-aberration design.

Figure 26 compares the *% at least correct* for the best designs produced by the column exchange algorithm to the best performing baseline design in each category.

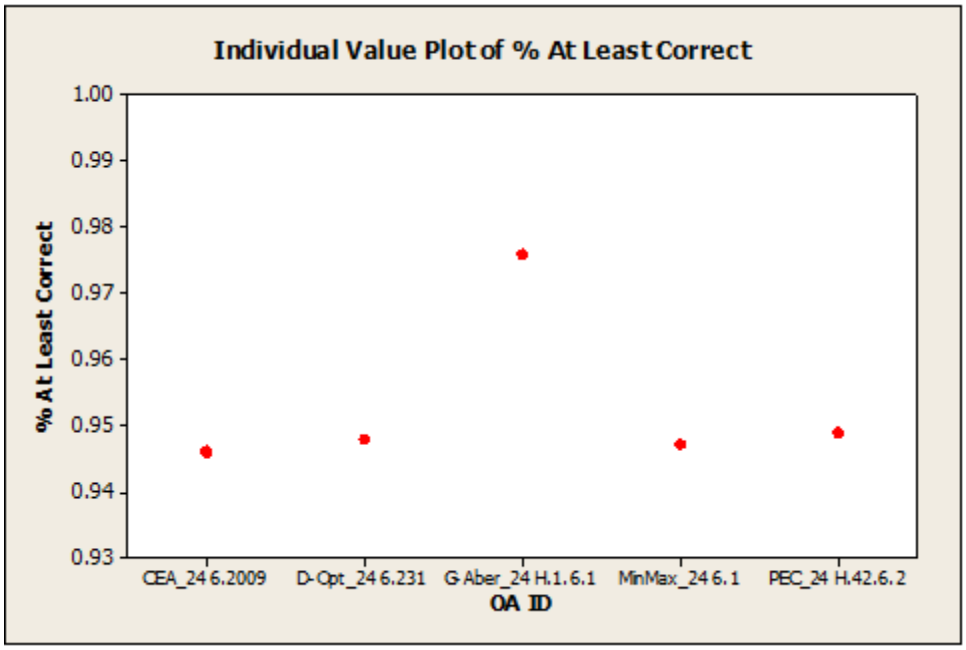

Figure 26. Comparison of *% at Least Correct* for 6-Factor Designs

The minimum G-aberration design, *G-Aber_24 H.1.6.1* (also identified as *G-Aber_24 6.6*), is the best performing design and is shown in Appendix Table 82. However, since $p_6 = 0$ for this design, a second design option is provided where $p_6 = 1$. The remaining four designs performed similarly in terms of the *% at least correct*, but CEA_24 6.2009, produced by the *Random Structured-Column Exchange Algorithm* performed slightly better in terms of *% correct.*

Table 37 lists the CVF and nonzero correlations for the best 6-factor design of each design type. The CFVs have the form

$F(D) = [(f_{11}, \dots, f_{17}), (f_{21}, \dots, f_{27}), (f_{31}, \dots, f_{37}), (f_{41}, \dots, f_{47})]$, where $(f_{i1}, \dots, f_{i7})$ corresponds to the frequencies of the J-characteristics $(24,20,16,12,8,4,0)$.

Table 37. CFVs and Nonzero Column Sums of Best 6-Factor Designs

OA ID	CFV	Nonzero Column Sums
CEA_24 6.2009	$[(0,0,0,0,0,0,6)_1(0,0,0,0,0,6,9)_2(0,0,0,0,0,10,10)_3(0,0,0,0,0,6,9)_4]$	22
D-Opt_24 6.231	$[(0,0,0,0,0,0,6)_1(0,0,0,0,0,0,15)_2(0,0,0,0,4,0,16)_3(0,0,0,0,5,0,10)_4]$	9
G-Aber_24 H.1.6.1	$[(0,0,0,0,0,0,6)_1(0,0,0,0,0,0,15)_2(0,0,0,0,0,0,20)_3(0,0,0,0,15,0,0)_4]$	15
MinMax_24 6.1	$[(0,0,0,0,0,0,6)_1(0,0,0,0,0,0,15)_2(0,0,0,0,4,0,16)_3(0,0,0,0,5,0,10)_4]$	9
PEC_24 H.42.6.2	$[(0,0,0,0,0,0,6)_1(0,0,0,0,0,0,15)_2(0,0,0,0,4,0,16)_3(0,0,0,0,5,0,10)_4]$	9

The table shows that CEA_24 6.2009 is not orthogonal for ME, but does not have any correlations above 1/6. Therefore, CEA_24 6.2009 is recommended as the 6-factor no-confounding design when enforcing $p_6 = 1$. CEA_24 6.2009 is shown in Appendix Table 89.

Figure 27 shows the *% at least correct* for the 7-factor G-aberration design in Ingram and Tang (2005) and the 7-factor PEC designs in Loeppky, Sitter et al. (2007).

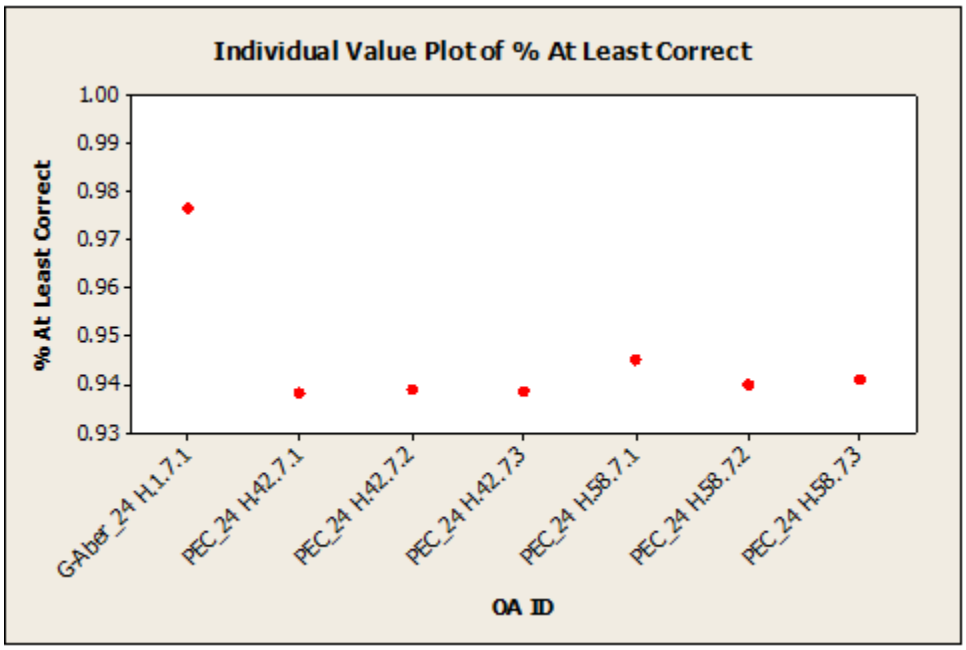

Figure 27. *% at Least Correct* for 7-factor G-aberration and PEC Baseline Designs

The minimum G-aberration design, *G-Aber_24 H.1.7.1* performs significantly better than the PEC designs.

Figure 28 compares the *% at least correct* for the best designs produced by the column exchange algorithm to the best performing baseline design in each category.

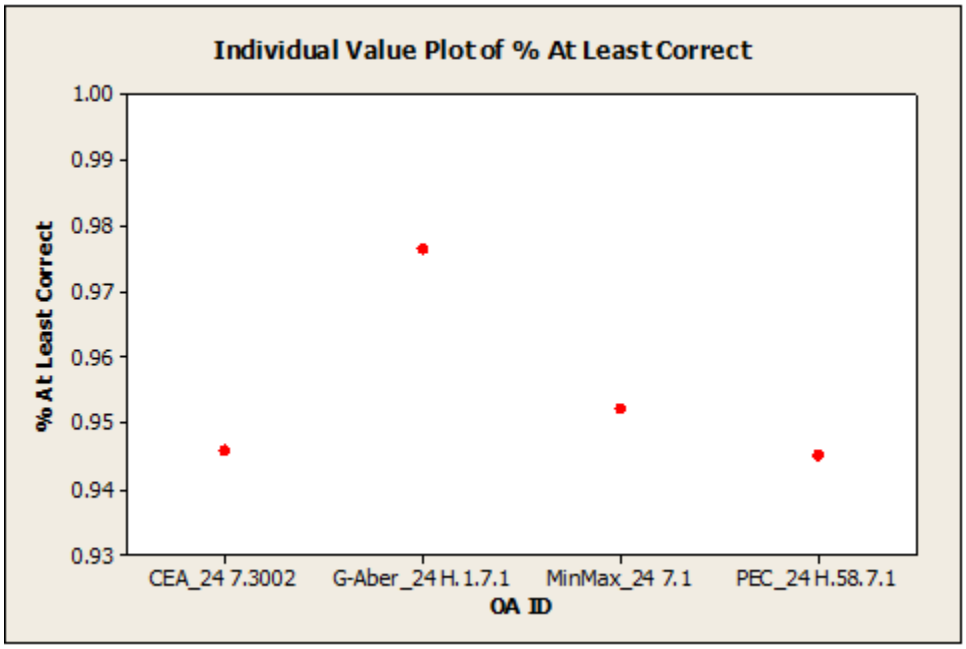

Figure 28. Comparison of *% at Least Correct* for 7-Factor Designs

The minimum G-aberration design, *G-Aber_24 H.1.7.1* is the best performing design and is shown in Appendix Table 83. However, since $p_6 = 0$ for this design, a second design option is provided where $p_6 = 1$. The design produced by the algorithm performed slightly better in terms of both simulation metrics. Table 38 lists the CVF and nonzero correlations for the best 7-factor designs for each design type.

Table 38. CFVs and Nonzero Column Sums of Best 7-Factor Designs

OA ID	CFV	Nonzero Column Sums
CEA_24 7.3002	$[(0,0,0,0,0,0,7)_1(0,0,0,0,0,0,2,19)_2(0,0,0,0,4,10,21)_3(0,0,0,1,7,17,10)_4]$	41
G-Aber_24 H.1.7.1	$[(0,0,0,0,0,0,7)_1(0,0,0,0,0,0,0,21)_2(0,0,0,0,0,0,0,35)_3(0,0,0,0,35,0,0)_4]$	35
MinMax_24 7.1	$[(0,0,0,0,0,0,7)_1(0,0,0,0,0,0,0,21)_2(0,0,0,0,6,0,29)_3(0,0,0,0,15,0,20)_4$	21
PEC_24 H.58.7.1	$[(0,0,0,0,0,0,7)_1(0,0,0,0,0,0,0,21)_2(0,0,0,0,8,0,27)_3(0,0,1,0,11,0,23)_4]$	20

The new *MinMax_24 7.1*, discovered by the *Generating-Matrix Column Exchange Algorithm*, is superior in both simulation metrics to *CEA_24 7.3002* and *PEC_24 H.58.7.1,* and in addition has orthogonal ME. *MinMax_24 7.1* is clearly the best design option among these designs when enforcing $p_6 = 1$. *MinMax_24 7.1* is shown in Appendix Table 90.

Figure 29 shows the *% at least correct* for the 8-factor G-aberration design in Ingram and Tang (2005) and the 8-factor PEC designs in Loeppky, Sitter et al. (2007).

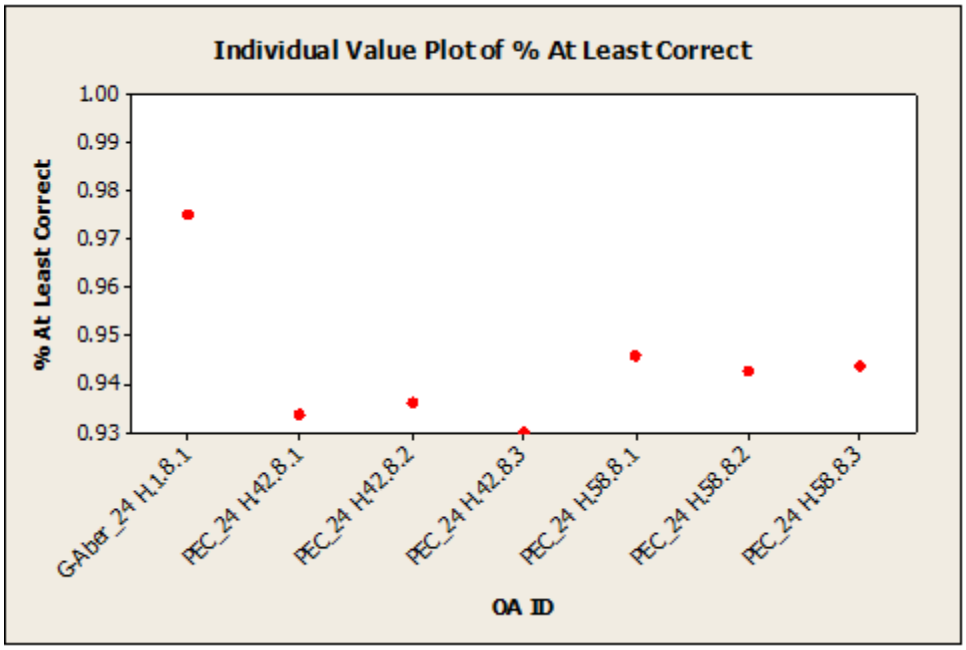

Figure 29. *% at Least Correct* for 8-factor G-aberration and PEC Baseline Designs

The minimum G-aberration design, *G-Aber_24 H.1.8.1*, performs significantly better than the PEC designs.

Figure 30 compares the *% at least correct* for the best designs produced by the column exchange algorithm to the best performing baseline design in each category.

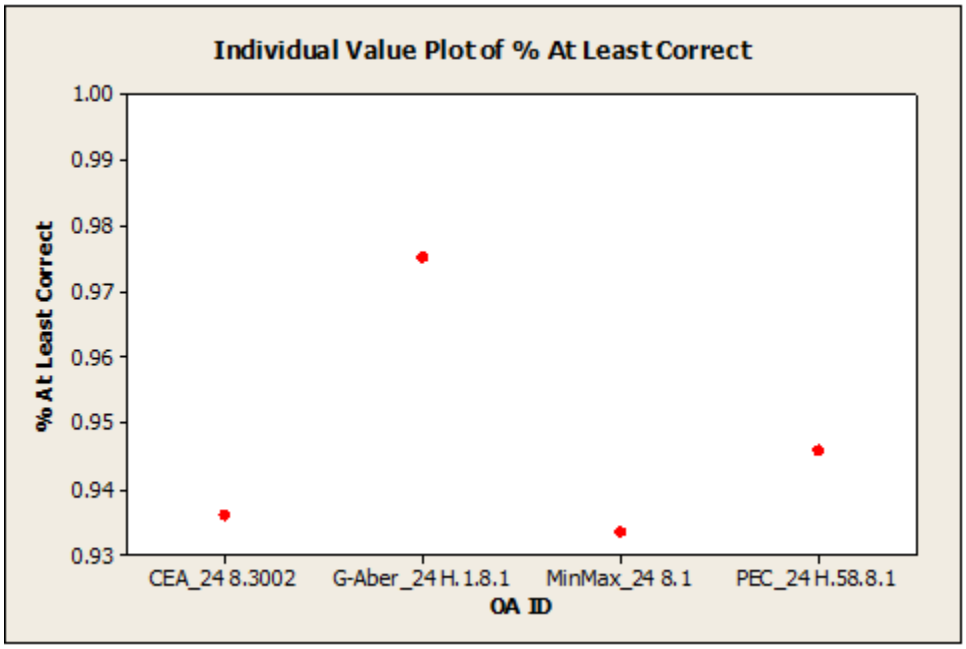

Figure 30. Comparison of *% at Least Correct* for 8-Factor Designs

The minimum G-aberration design, *G-Aber_24 H.1.8.1* is the best performing design and is shown in Appendix Table 84. However, since $p_6 = 0$ for this design, a second design option is provided where $p_6 = 1$. Design *CEA_24 8.3002*, produced by the *Random Structured-Column Exchange Algorithm*, performed slightly worse than the *PEC_24 H.58.8.1* design in terms of both simulation metrics. However, examining the CFVs for both designs in Table 39 reveals that for the *PEC_24 H.58.8.1* design, there are several pairs of 2FI correlated at 2/3. For *CEA_24 8.3002*, no pair of effects has a correlation above 1/2. Furthermore, $p_6 = 1$ for *CEA_24 8.3002* while $p_6 = 0.786$ for

PEC_24 H.58.8.1. Consequently *CEA_24 8.3002* is recommended as the best design option when enforcing $p_6 = 1$. *CEA_24 8.3002* is shown in Appendix Table 91.

Table 39. CFVs and Nonzero Column Sums of Best 8-Factor Designs

OA ID	CFV	Nonzero Column Sums
CEA_24 8.3002	$[(0,0,0,0,0,0,8)_1(0,0,0,0,0,4,24)_2(0,0,0,0,9,20,27)_3(0,0,0,5,11,31,23)_4]$	80
G-Aber_24 H.1.8.1	$[(0,0,0,0,0,0,8)_1(0,0,0,0,0,0,28)_2(0,0,0,0,0,0,56)_3(0,0,0,0,70,0,0)_4]$	70
MinMax_24 8.1	$[(0,0,0,0,0,0,8)_1(0,0,0,0,0,0,28)_2(0,0,0,0,19,0,37)_3(0,0,0,0,27,0,43)_4]$	46
PEC_24 H.58.8.1	$[(0,0,0,0,0,0,8)_1(0,0,0,0,0,0,28)_2(0,0,0,0,13,0,43)_3(0,0,3,0,21,0,46)_4]$	37

Figure 31 shows the *% at least correct* for the 9-factor G-aberration design in Ingram and Tang (2005) and the PEC designs in Loeppky, Sitter et al. (2007).

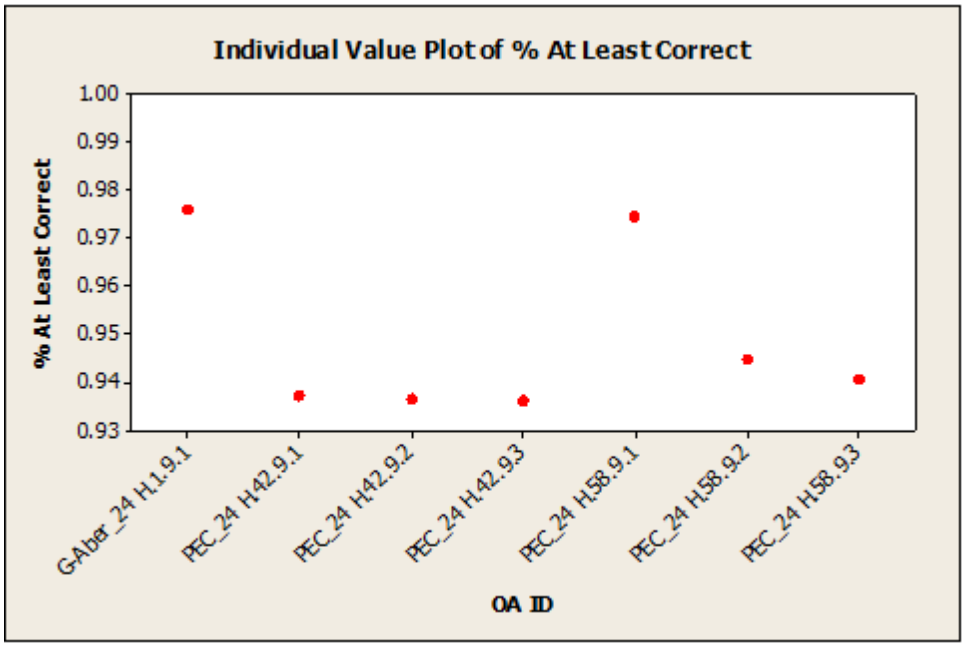

Figure 31. *% at Least Correct* for 9-Factor G-aberration and PEC Baseline Designs

The minimum G-aberration design, *G-Aber_24 H.1.9.1*, performs significantly better than the PEC designs with the exception of *PEC_24 58.9.1*, which is isomorphic to *G-Aber_24 H.1.9.1*.

It is important to point out here that the group of six PEC designs of 6-8 factors does not include the corresponding minimum G-aberration design. However for designs of 9-12 factors, the group of six PEC designs always includes the minimum G-aberration design. This is because the algorithm used by Loeppky, Sitter et al. (2007) discovered designs such as the *MinMax Correlation* designs which, similar to the minimum G-aberration designs, had $p_5 = 1$. However the PEC designs dominated the minimum G-aberration designs in terms of p_6, i.e. the minimum G-aberration design had $p_6 = 0$ while the six PEC designs had $p_6 = 1$. See Table 27 for the data. However, for designs of 9-12 factors, the only designs which the PEC algorithm could find with $p_5 = 1$ were the minimum G-aberration designs (with the exception of *PEC_24 H.42.10.3*). Therefore for each of the 9-12 factor designs, a minimum G-aberration design is represented among the top six PEC designs. The column exchange algorithms developed for the research in this chapter were able to find designs of 9-12 factors where $p_5 = p_6 = 1$.

Figure 32 compares the *% at least correct* for the best designs produced by the column exchange algorithm to the best performing baseline design in each category.

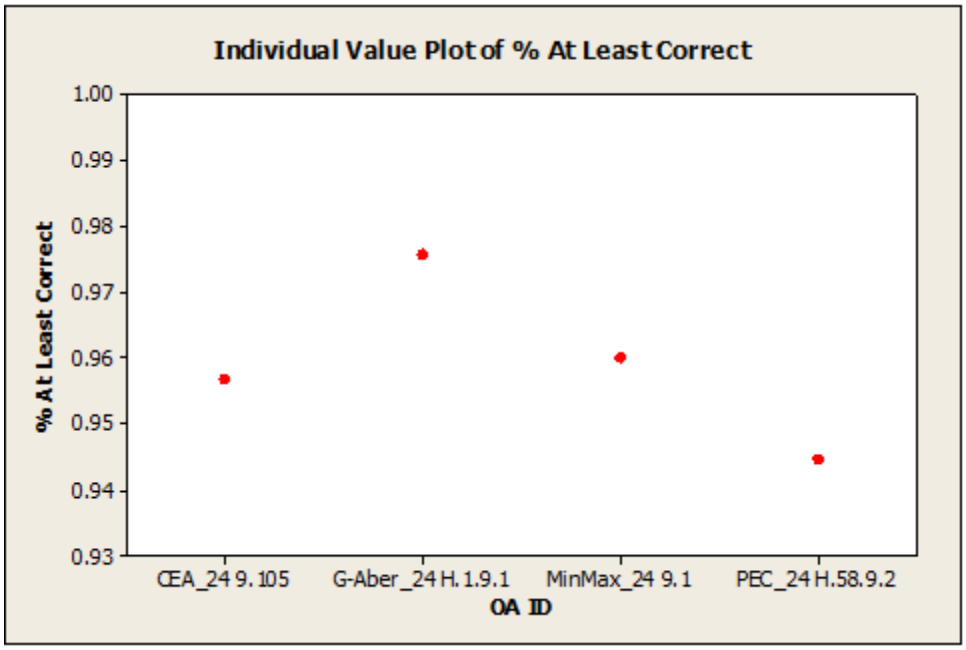

Figure 32. Comparison of *% at Least Correct* for 9-Factor Designs

The minimum G-aberration design, *G-Aber_24 H.1.9.1*, is the best performing design and is shown in Appendix Table 85. However, since $p_6 = 0$ for this design, a second design option is provided where $p_6 = 1$. Design *MinMax_24 9.1*, produced by the *Random Structured-Column Exchange Algorithm*, performed slightly better than *PEC_24 H.58.9.2* and *CEA_24 9.105* in terms of the *% at least correct* metric. It performed significantly better than *PEC_24 H.58.9.2* in terms of the *% correct* metric. In addition, $p_6 = 0.655$ for the PEC design. Table 40 shows that for *MinMax_24 9.1*, no pair of effects has a correlation above 1/3 and is therefore the recommended design option when enforcing $p_6 = 1$. *MinMax_24 9.1* is shown in Appendix Table 92.

Table 40. CFVs and Nonzero Column Sums of Best 9-Factor Designs

OA ID	CFV	Nonzero Column Sums
CEA_24 9.105	$[(0,0,0,0,0,0,9)_1(0,0,0,0,0,6,30)_2(0,0,0,0,6,12,66)_3(0,0,0,0,88,26,12)_4]$	138
G-Aber_24 H.1.9.1	$[(0,0,0,0,0,0,9)_1(0,0,0,0,0,0,36)_2(0,0,0,0,0,0,84)_3(0,0,0,0,126,0,0)_4]$	126
MinMax_24 9.1	$[(0,0,0,0,0,0,9)_1(0,0,0,0,0,0,36)_2(0,0,0,0,10,0,74)_3(0,0,0,0,98,0,28)_4]$	108
PEC_24 H.58.9.2	$[(0,0,0,0,0,0,9)_1(0,0,0,0,0,0,36)_2(0,0,2,0,18,0,64)_3(0,0,3,0,44,0,79)_4]$	125

Figure 33 shows the *% at least correct* for the 10-factor G-aberration design in Ingram and Tang (2005) and the 10-factor PEC designs in Loeppky, Sitter et al. (2007).

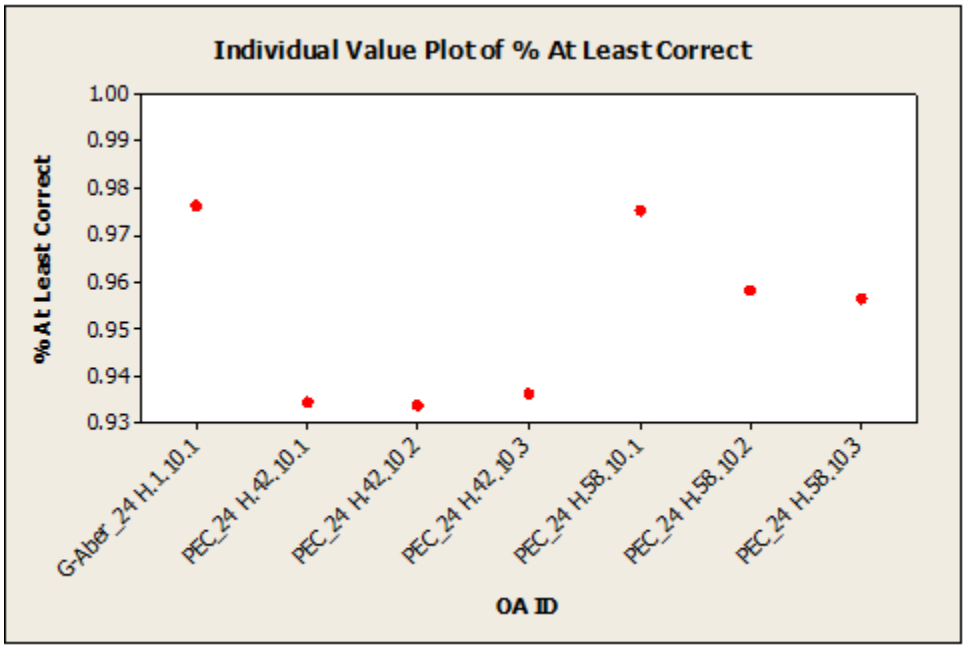

Figure 33. *% at Least Correct* for 10-Factor G-aberration and PEC Baseline Designs

The minimum G-aberration design, *G-Aber_24 H.1.10.1*, performs better than the PEC designs in terms of both simulation metrics with the exception of *PEC_24 58.10.1*, which is isomorphic to *G-Aber_24 H.1.10.1*.

Figure 34 compares the *% at least correct* for the best designs produced by the column exchange algorithm to the best performing baseline design in each category.

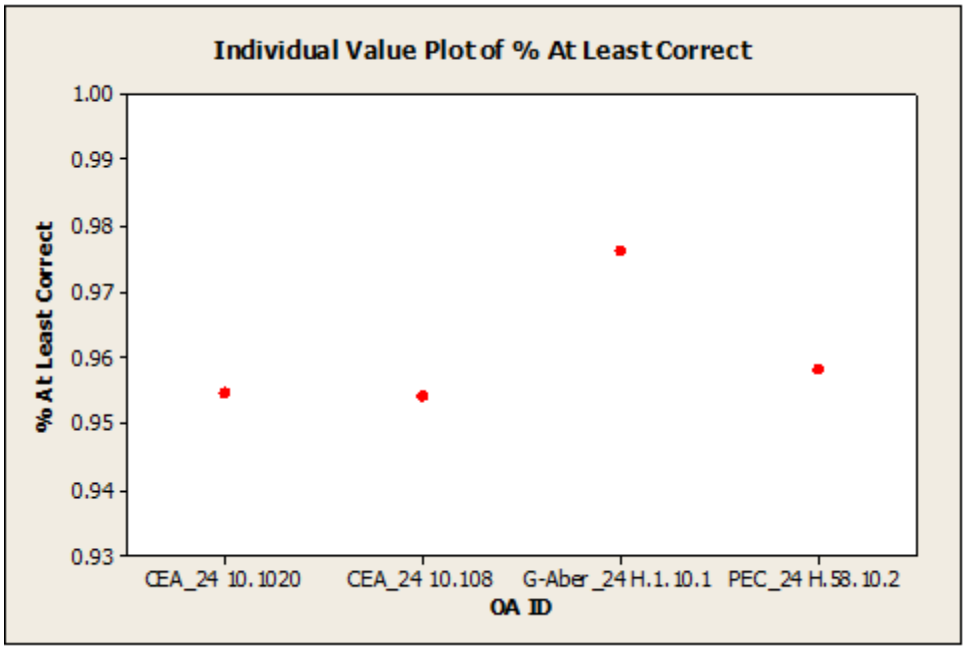

Figure 34. Comparison of *% at Least Correct* for 10-Factor Designs

The minimum G-aberration design, *G-Aber_24 H.1.10.1* is the best performing design and is shown in Appendix Table 86. However, since $p_6 = 0$ for this design, a second design option is provided where $p_6 = 1$. Design *CEA_24 10.108*, produced by the *Generating-Matrix Column Exchange Algorithm*, performed slightly better than *PEC_24 H.58.10.2* in terms of the *% correct* metric and only slightly worse in terms of the *% at least correct* metric. In addition, $p_6 = 0.362$ for the PEC design. Table 41 shows that for *CEA_24 10.108*, no pair of effects has a correlation above 1/2 and only a few 2FI effects are correlated above 1/3. Therefore *CEA_24 10.108* is the recommended design option when enforcing $p_6 = 1$. *CEA_24 10.108* is shown in Appendix Table 93.

Table 41. CFVs and Nonzero Column Sums of Best 10-Factor Designs

OA ID	CFV	Nonzero Column Sums
CEA_24 10.1020	$[(0,0,0,0,0,0,10)_1(0,0,0,0,0,3,42)_2(0,0,1,1,9,17,92)_3(0,0,0,0,140,46,24)_4]$	217
CEA_24 10.108	$[(0,0,0,0,0,0,9)_1(0,0,0,0,0,7,38)_2(0,0,0,0,10,14,96)_3(0,0,0,2,148,40,20)_4]$	221
G-Aber_24 H.1.10.1	$[(0,0,0,0,0,0,10)_1(0,0,0,0,0,0,45)_2(0,0,0,0,0,0,120)_3(0,0,0,0,210,0,0)_4]$	210
PEC_24 H.58.10.2	$[(0,0,0,0,0,0,10)_1(0,0,0,0,0,0,45)_2(0,0,3,0,16,0,101)_3(0,0,0,0,126,0,84)_4]$	145

Figure 35 shows the *% at least correct* for the 11-factor G-aberration design in Ingram and Tang (2005) and the 11-factor PEC designs in Loeppky, Sitter et al. (2007).

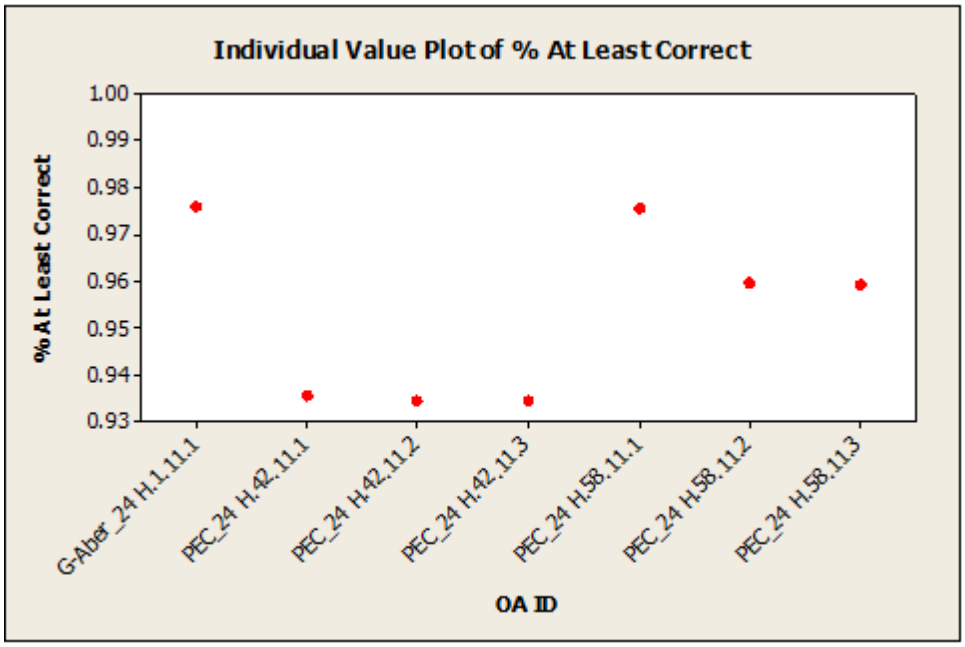

Figure 35. *% at Least Correct* for 11-Factor G-aberration and PEC Baseline Designs

The minimum G-aberration design, *G-Aber_24 H.1.11.1*, performs better than the PEC designs in terms of both simulation metrics with the exception of *PEC_24 58.11.1*, which is isomorphic to *G-Aber_24 H.1.11.1*.

Figure 36 compares the *% at least correct* for the best designs produced by the column exchange algorithm to the best performing baseline design in each category.

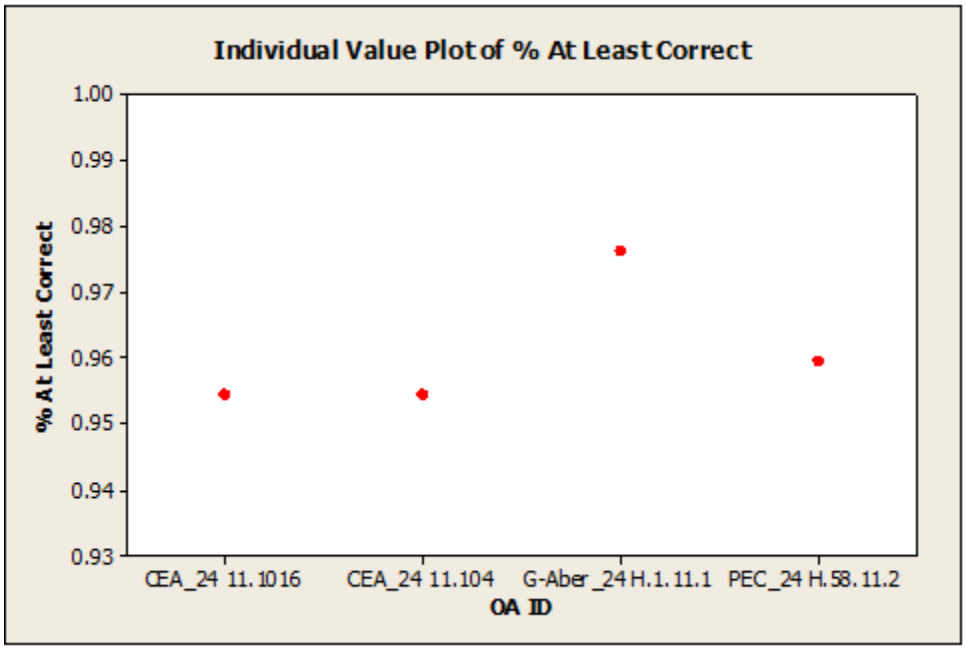

Figure 36. Comparison of *% at Least Correct* for 11-Factor Designs

The minimum G-aberration design, *G-Aber_24 H.1.11.1* is the best performing design and is shown in Appendix Table 87. However, since $p_6 = 0$ for this design, a second design option is provided where $p_6 = 1$. Design *CEA_24 11.1016*, produced by the *Random Structured-Column Exchange Algorithm*, performed slightly worse than *PEC_24 H.58.11.2* in terms of the *% at least correct* metric and significantly better in terms of the *% correct* metric. In addition, $p_6 = 0.312$ for the PEC design. Table 42 shows that for *CEA_24 11.1016*, no pair of effects has a correlation above 1/2 and only a limited number of ME and 2FI effects are correlated above 1/3. Therefore *CEA_24 11.1016* is the recommended design option when enforcing $p_6 = 1$. *CEA_24 11.1016* is shown in Appendix Table 94.

Table 42. CFVs and Nonzero Column Sums of Best 11-Factor Designs

OA ID	CFV	Nonzero Column Sums
CEA_24 11.1016	$[(0,0,0,0,0,0,11)_1(0,0,0,0,0,4,51)_2(0,0,0,4,10,20,131)_3(0,0,0,0,230,64,36)_4]$	332
CEA_24 11.104	$[(0,0,0,0,0,1,10)_1(0,0,0,0,0,4,51)_2(0,0,1,1,12,20,131)_3(0,0,0,0,230,64,36)_4]$	333
G-Aber_24 H.1.11.1	$[(0,0,0,0,0,0,11)_1(0,0,0,0,0,0,55)_2(0,0,0,0,0,0,165)_3(0,0,0,0,330,0,0)_4]$	330
PEC_24 H.58.11.2	$[(0,0,0,0,0,0,11)_1(0,0,0,0,0,0,55)_2(0,0,4,0,20,0,141)_3(0,0,0,0,210,0,120)_4]$	234

Figure 37 shows the *% at least correct* for the 12-factor G-aberration design in Ingram and Tang (2005) and the 12-factor PEC designs in Loeppky, Sitter et al. (2007).

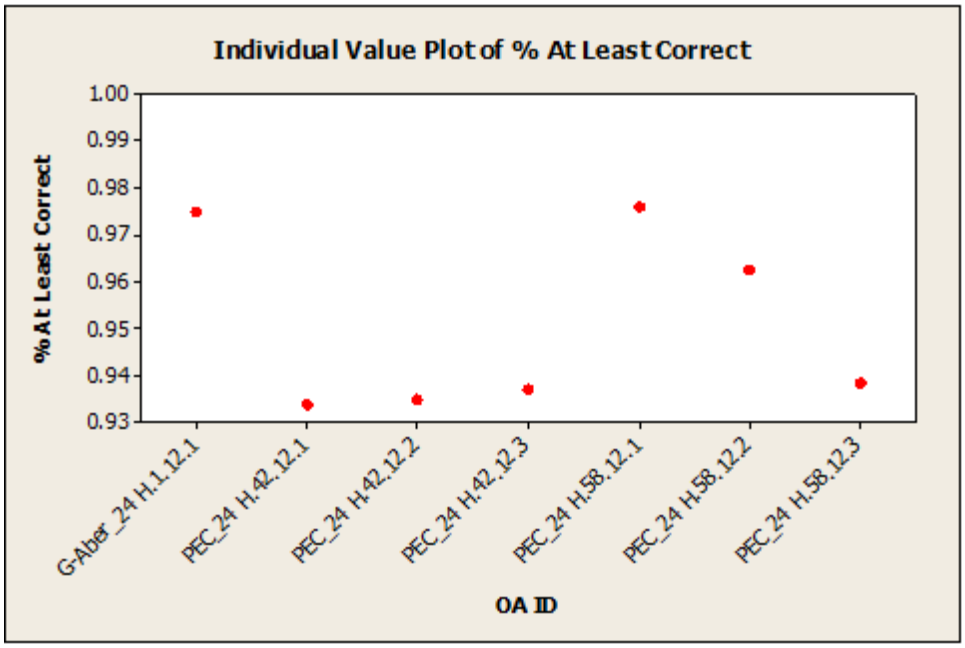

Figure 37. *% at Least Correct* for 12-Factor G-aberration and PEC Baseline Designs

The minimum G-aberration design, *G-Aber_24 H.1.12.1*, performs better than the PEC designs in terms of both simulation metrics, with the exception of *PEC_24 58.12.1*, which is isomorphic to *G-Aber_24 H.1.12.1*.

Figure 38 compares the *% at least correct* for the best designs produced by the column exchange algorithm to the best performing baseline design in each category.

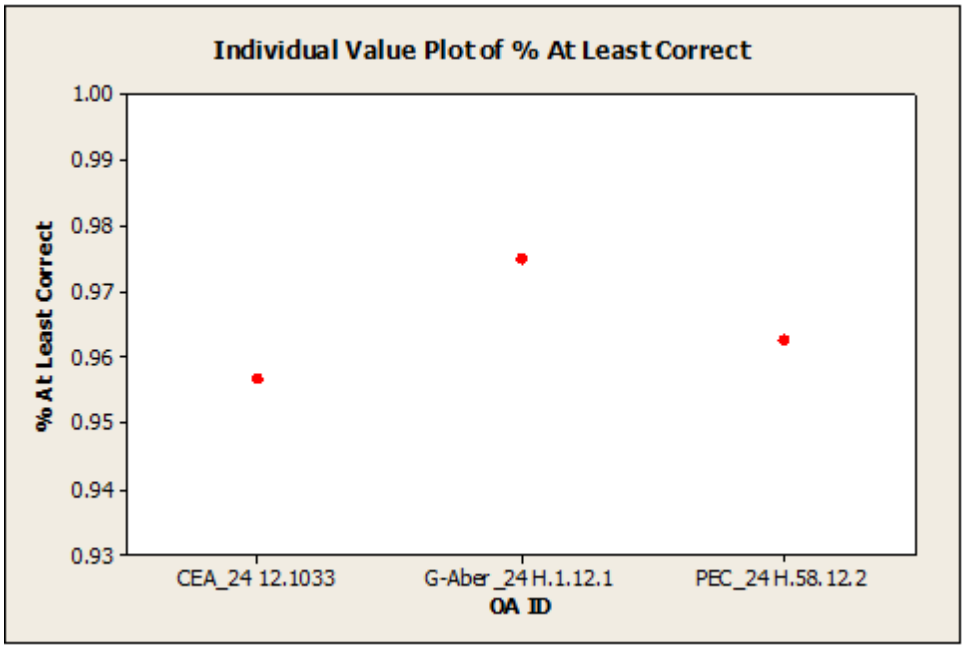

Figure 38. Comparison of *% at Least Correct* for 12-Factor Designs

The minimum G-aberration design, *G-Aber_24 H.1.12.1,* is the best performing design and is shown in Appendix Table 88. However, since $p_6 = 0$ for this design, a second design option is provided where $p_6 = 1$. Design *CEA_24 12.1033*, produced by the *Random Structured-Column Exchange Algorithm*, performed slightly worse than *PEC_24 H.58.12.2* in terms of the *% at least correct* metric and significantly better in terms of the *% correct* metric. In addition, $p_6 = 0.273$ for the PEC design. Table 43 shows that for *CEA_24 12.1033*, no pair of effects has a correlation greater than 1/2 and only a limited number of ME and 2FI effects have a correlation greater than 1/3. Therefore *CEA_24 12.1033* is the recommended design option when enforcing $p_6 = 1$. *CEA_24 12.1033* is shown in Appendix Table 95.

Table 43. CFVs and Nonzero Column Sums of Best 12-Factor Designs

OA ID	CFV	Nonzero Column Sums
CEA_24 12.1033	$[(0,0,0,0,0,0,12)_1(0,0,0,0,0,9,57)_2(0,0,0,2,14,16,188)_3(0,0,0,6,366,87,36)_4]$	500
G-Aber_24 H.1.12.1	$[(0,0,0,0,0,0,12)_1(0,0,0,0,0,0,66)_2(0,0,0,0,0,0,220)_3(0,0,0,0,495,0,0)_4]$	495
PEC_24 H.58.12.2	$[(0,0,0,0,0,0,12)_1(0,0,0,0,0,0,66)_2(0,0,5,0,25,0,190)_3(0,0,0,0,330,0,165)_4]$	360

The performance of the minimum G-aberration designs varies for different combinations of terms in the response models. A final round of simulations was conducted to produce data characterizing the performance across different response models. For each of the optimal no-confounding designs, 60,000 simulations were run with a random number of p ME where $2 \leq p \leq 6$ and a random number of q 2FI where $0 \leq q \leq 4$. Table 44 through Table 57 concern the minimum G-aberration designs and break down the *% at least correct* and *% correct* for each combination of p ME and q 2FI.

Table 44. Matrix of *% at Least Correct* for p ME and q 2FI - G-Aber_24 H.1.6.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	100.00%	99.48%	92.35%		97.99%
4	100.00%	100.00%	98.18%	86.88%	64.79%	89.76%
5	100.00%	100.00%	94.40%	72.95%	46.61%	82.78%
6	100.00%	99.78%	85.92%	48.37%	20.70%	70.55%
Grand Total	100.00%	99.97%	94.66%	75.46%	44.28%	88.14%

Table 45. Matrix of *% Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.6.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	73.16%	100.00%				86.69%
3	68.74%	90.69%	94.93%	92.35%		86.47%
4	63.27%	77.67%	78.38%	74.41%	59.07%	70.47%
5	54.47%	61.45%	62.53%	53.42%	37.93%	53.93%
6	52.96%	59.27%	54.81%	35.67%	15.95%	43.58%
Grand Total	65.11%	83.30%	73.62%	64.87%	37.90%	68.11%

Table 46. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.7.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	100.00%	99.43%	93.94%		98.35%
4	100.00%	100.00%	98.17%	86.40%	66.07%	90.14%
5	100.00%	100.00%	95.93%	74.17%	49.31%	84.20%
6	100.00%	99.79%	85.10%	49.36%	20.35%	70.42%
Grand Total	100.00%	99.97%	94.97%	77.02%	44.75%	88.64%

Table 47. Matrix of *% Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.7.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	67.55%	100.00%				83.61%
3	64.68%	89.47%	93.80%	93.94%		85.33%
4	57.90%	76.32%	76.63%	74.96%	59.80%	69.28%
5	49.15%	61.31%	63.11%	55.90%	39.74%	53.96%
6	51.98%	56.66%	53.04%	35.15%	15.93%	42.28%
Grand Total	60.60%	82.42%	73.02%	66.66%	38.06%	66.93%

Table 48. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.8.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	100.00%	99.46%	92.75%		98.02%
4	100.00%	100.00%	98.03%	85.68%	63.76%	89.54%
5	100.00%	99.96%	94.05%	72.12%	47.07%	83.15%
6	100.00%	99.76%	84.39%	49.17%	21.24%	71.08%
Grand Total	100.00%	99.96%	94.24%	76.05%	43.83%	88.33%

Table 49. Matrix of *% Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.8.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	62.28%	100.00%				81.30%
3	60.58%	89.48%	94.77%	92.75%		84.54%
4	53.98%	77.56%	78.57%	73.33%	58.46%	68.40%
5	45.08%	60.98%	62.84%	52.98%	38.19%	52.23%
6	48.14%	56.45%	54.56%	35.13%	16.94%	42.40%
Grand Total	56.10%	82.37%	73.74%	65.43%	37.70%	65.73%

Table 50. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.9.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	100.00%	99.54%	93.62%		98.32%
4	100.00%	99.96%	98.45%	86.22%	65.88%	89.89%
5	100.00%	100.00%	94.35%	71.39%	47.74%	82.72%
6	100.00%	99.87%	84.68%	46.37%	20.56%	70.31%
Grand Total	100.00%	99.98%	94.56%	75.31%	45.06%	88.24%

Table 51. Matrix of *% Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.9.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	56.77%	100.00%				78.46%
3	53.02%	89.94%	94.53%	93.59%		82.59%
4	50.89%	76.81%	80.56%	72.44%	59.95%	68.26%
5	42.46%	61.71%	63.53%	52.97%	38.14%	51.78%
6	45.59%	57.80%	52.28%	32.02%	16.06%	40.76%
Grand Total	51.43%	82.77%	74.04%	64.35%	38.39%	64.36%

Table 52. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.10.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	99.98%				99.99%
3	100.00%	100.00%	99.36%	93.90%		98.29%
4	100.00%	100.00%	97.78%	85.16%	65.37%	89.35%
5	100.00%	100.00%	93.95%	72.26%	47.55%	83.00%
6	100.00%	99.88%	85.37%	47.04%	21.40%	70.31%
Grand Total	100.00%	99.98%	94.44%	75.91%	44.77%	88.18%

Table 53. Matrix of *% Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.10.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	52.06%	99.98%				76.24%
3	45.94%	90.32%	94.42%	93.90%		80.75%
4	45.61%	76.99%	78.31%	72.00%	58.79%	66.24%
5	39.02%	62.82%	63.20%	53.55%	39.38%	51.60%
6	44.21%	57.89%	55.11%	33.22%	16.43%	41.15%
Grand Total	46.82%	82.93%	74.06%	65.25%	38.22%	63.20%

Table 54. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.11.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	100.00%	99.33%	93.91%		98.33%
4	100.00%	100.00%	97.75%	85.87%	65.60%	89.97%
5	100.00%	100.00%	94.79%	72.84%	47.33%	82.85%
6	100.00%	99.87%	84.21%	49.58%	20.21%	70.80%
Grand Total	100.00%	99.98%	94.28%	76.69%	44.06%	88.37%

Table 55. Matrix of *% Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.11.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	47.27%	99.98%				73.29%
3	43.02%	90.79%	94.09%	93.85%		80.46%
4	41.56%	77.15%	80.03%	74.15%	59.38%	66.67%
5	34.40%	62.33%	63.04%	54.44%	38.29%	50.41%
6	41.19%	58.09%	54.71%	35.56%	15.76%	41.05%
Grand Total	42.85%	83.09%	74.18%	66.19%	37.48%	62.33%

Table 56. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - G-Aber_24 H.1.12.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	99.98%				99.99%
3	100.00%	100.00%	99.60%	94.16%		98.39%
4	100.00%	100.00%	98.12%	84.83%	67.77%	89.98%
5	100.00%	99.87%	95.31%	73.04%	48.59%	83.36%
6	100.00%	99.92%	85.52%	47.99%	20.06%	71.24%
Grand Total	100.00%	99.96%	94.92%	76.41%	45.96%	88.60%

Table 57. Matrix of *% Correct* for p ME and q 2FI - G-Aber_24 H.1.12.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	42.79%	99.98%				71.28%
3	39.28%	90.68%	95.10%	94.13%		79.92%
4	38.39%	75.53%	79.68%	71.86%	61.37%	65.25%
5	32.45%	62.42%	64.48%	53.79%	39.07%	50.17%
6	39.83%	58.05%	54.71%	33.74%	15.52%	40.52%
Grand Total	39.47%	82.80%	74.80%	65.62%	39.11%	61.45%

Table 58 - Table 71 concern the alternative no-confounding designs and break down the *% at least correct* and *% correct* for each combination of p ME and q 2FI.

Table 58. Matrix of *% at Least Correct* for p ME and q 2FI - CEA_24 6_2009

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	99.98%				99.99%
3	100.00%	99.90%	97.18%	86.07%		95.69%
4	100.00%	99.41%	91.99%	73.16%	52.13%	83.26%
5	100.00%	98.67%	83.45%	58.37%	36.67%	75.60%
6	100.00%	96.47%	70.83%	41.74%	21.84%	66.39%
Grand Total	100.00%	99.16%	86.65%	66.43%	36.92%	84.22%

Table 59. Matrix of *% Correct* for p ME and q 2FI - CEA_24 6_2009

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	73.20%	99.21%				86.17%
3	69.21%	89.70%	92.07%	85.91%		84.25%
4	63.79%	76.98%	75.18%	62.10%	46.39%	64.87%
5	52.00%	60.25%	54.83%	43.68%	28.64%	47.93%
6	49.12%	51.97%	41.94%	27.09%	16.26%	37.36%
Grand Total	64.31%	81.31%	67.82%	57.04%	30.48%	64.18%

Table 60. Matrix of *% at Least Correct* for *p* ME and *q* 2FI – MinMax_24 7.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	99.64%	97.32%	89.28%		96.63%
4	100.00%	98.73%	93.20%	77.85%	58.01%	85.36%
5	100.00%	98.15%	87.66%	64.96%	43.22%	78.79%
6	100.00%	97.08%	76.82%	48.00%	27.26%	70.08%
Grand Total	100.00%	99.05%	89.26%	71.33%	43.00%	86.28%

Table 61. Matrix of *% Correct* for *p* ME and *q* 2FI - MinMax_24 7.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	69.57%	96.12%				83.10%
3	64.83%	86.18%	91.95%	88.81%		82.87%
4	57.87%	73.65%	74.84%	65.42%	52.79%	64.85%
5	48.32%	58.43%	57.74%	47.37%	34.70%	49.31%
6	47.92%	51.42%	47.15%	32.42%	20.50%	40.01%
Grand Total	60.66%	78.88%	69.39%	60.45%	36.16%	64.18%

Table 62. Matrix of *% at Least Correct* for *p* ME and *q* 2FI – CEA_24 8_3002

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	99.76%	95.36%	84.84%		95.05%
4	100.00%	98.95%	89.76%	72.29%	55.61%	83.34%
5	100.00%	97.27%	81.93%	58.02%	38.43%	74.89%
6	100.00%	95.90%	74.52%	43.13%	25.61%	67.62%
Grand Total	100.00%	98.79%	85.93%	65.62%	39.73%	84.13%

Table 63. Matrix of *% Correct* for *p* ME and *q* 2FI - CEA_24 8_3002

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	62.36%	94.34%				78.25%
3	59.94%	84.77%	88.29%	83.13%		78.93%
4	55.84%	74.03%	70.34%	60.60%	50.02%	62.16%
5	45.41%	59.31%	54.99%	43.21%	31.14%	46.78%
6	48.71%	52.71%	44.96%	29.17%	18.43%	38.70%
Grand Total	56.45%	78.22%	65.92%	55.50%	33.04%	60.89%

Table 64. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - MinMax_24 9.1

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	99.67%	98.06%	91.11%		97.18%
4	99.92%	99.50%	95.22%	81.63%	60.61%	87.28%
5	99.83%	98.71%	90.08%	68.77%	43.89%	80.43%
6	99.01%	97.92%	80.25%	50.50%	23.50%	70.26%
Grand Total	99.82%	99.37%	91.32%	74.40%	42.79%	87.10%

Table 65. Matrix of *% Correct* for *p* ME and *q* 2FI - MinMax_24 9.1

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	57.89%	95.48%				77.01%
3	53.52%	88.08%	91.77%	89.93%		80.78%
4	52.38%	73.99%	76.72%	70.17%	56.05%	65.79%
5	44.23%	59.81%	60.35%	49.63%	35.67%	50.12%
6	49.40%	56.01%	50.13%	34.48%	18.14%	41.65%
Grand Total	52.98%	79.89%	70.90%	63.20%	36.76%	63.15%

Table 66. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - CEA_24 10_108

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	99.69%	97.31%	88.90%		96.46%
4	99.92%	99.39%	93.74%	77.55%	58.95%	86.18%
5	99.92%	98.23%	87.60%	64.65%	42.79%	78.39%
6	99.41%	97.04%	76.08%	44.63%	22.22%	68.00%
Grand Total	99.89%	99.16%	89.22%	70.15%	41.34%	85.87%

Table 67. Matrix of *% Correct* for *p* ME and *q* 2FI - CEA_24 10_108

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	53.43%	96.53%				74.83%
3	49.66%	86.19%	91.19%	88.13%		78.95%
4	48.66%	74.27%	76.20%	65.56%	52.71%	63.58%
5	41.18%	61.14%	58.51%	47.10%	33.47%	48.23%
6	45.67%	54.48%	47.92%	30.85%	16.45%	39.15%
Grand Total	49.10%	79.97%	69.85%	59.67%	34.20%	61.00%

Table 68. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - CEA_24 11_1016

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	99.60%	97.21%	89.66%		96.59%
4	99.96%	99.17%	94.49%	78.55%	60.90%	86.48%
5	99.87%	97.96%	88.59%	66.29%	42.12%	78.89%
6	99.51%	96.45%	78.04%	45.27%	21.81%	68.42%
Grand Total	99.90%	98.97%	90.03%	71.45%	41.59%	86.08%

Table 69. Matrix of *% Correct* for *p* ME and *q* 2FI - CEA_24 11_1016

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	48.30%	93.93%				71.54%
3	45.30%	84.49%	89.75%	87.08%		76.53%
4	45.21%	72.44%	73.42%	65.98%	54.78%	62.31%
5	38.01%	57.31%	59.23%	48.50%	34.32%	47.46%
6	42.96%	51.31%	46.31%	31.65%	15.83%	37.66%
Grand Total	44.96%	77.20%	68.48%	60.46%	34.95%	59.11%

Table 70. Matrix of *% at Least Correct* for *p* ME and *q* 2FI - CEA_24 12_1033

% At Least Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	100.00%	100.00%				100.00%
3	100.00%	99.73%	98.19%	89.79%		96.89%
4	99.88%	99.46%	94.13%	79.71%	58.23%	86.22%
5	99.79%	98.66%	87.79%	66.48%	41.78%	78.90%
6	98.64%	96.69%	78.15%	44.45%	22.15%	68.23%
Grand Total	99.75%	99.19%	89.93%	71.38%	40.89%	86.03%

Table 71. Matrix of *% Correct* for *p* ME and *q* 2FI - CEA_24 12_1033

% Correct						
	Number of 2FI					
Number of ME	0	1	2	3	4	Grand Total
2	44.22%	95.99%				69.96%
3	41.08%	85.12%	90.70%	87.93%		76.43%
4	40.51%	73.70%	75.47%	66.75%	52.46%	61.67%
5	35.20%	59.15%	58.37%	50.53%	32.99%	47.24%
6	40.25%	53.65%	49.56%	30.99%	16.38%	38.28%
Grand Total	41.16%	79.04%	69.56%	60.94%	34.10%	58.71%

4.5 Conclusions

The results of the Monte Carlo simulation show that the 24-run no-confounding designs are outstanding screening designs. In terms of the fewest type I and type II model building errors, the minimum G-aberration designs with uncorrelated ME were clearly the best of the published and algorithmically generated OAs evaluated in this

research. Furthermore the simulation metrics for the *percent at least correct* were almost identical no matter how many design columns (investigated factors) there were. It appears that having ME uncorrelated with other ME and 2FI provides a significant advantage in reducing type I and type II model building errors, at least when 2-stage stepwise regression is the analysis method.

The *percent at least correct* for the minimum G-aberration designs was extremely high – even for models with as many as seven combined ME and 2FI terms. When there were between two and five ME and zero to one 2FI, the 24-run no-confounding designs detected all active effects over 99% of the time. When there were between two and five ME and zero to two 2FI, the worst case *percent at least correct* fell to only 94%. Furthermore, the worst case for the *percent correct* was 45%. These results show that no-confounding designs are an extremely cost-effective screening design option.

Even though it was not possible to exhaustively search the entire space of 24-run OAs, the data gives a strong indication that a design does not exist which performs better than a minimum G-aberration no-confounding design. The only downside to these designs was their inability to estimate the 6-factor full model in any 6-factor projection, i.e. $p_6 = 0$ for these designs. Therefore if an experimenter expects to fit a large model of more than five ME and two 2FI, other screening design options may be more preferable.

The algorithm was able to find alternative designs for all cases of 6-12 factors with $p_6 = 1$. Interestingly, the simulation results for these algorithmically created designs were inferior to the minimum G-aberration designs. The alternative designs however, were superior to the baseline PEC designs, both in performance and in the magnitude of p_6. The design search algorithms were able to find designs superior to the baseline

designs by investigating non-orthogonal designs and orthogonal arrays that were not projections of Hadamard matrices. No designs were discovered that provided a compelling reason to use unbalanced columns since effective balanced-column alternatives were always available.

The search algorithms discovered two unpublished *MinMax Correlation* designs of 7 and 9 factors. Both of these designs were recommended as alternative no-confounding designs. These designs are not OAs of strength three, but can estimate every term in the 6-factor full model unlike the minimum G-aberration designs.

There is some evidence that the *Random Structured-Column Exchange Algorithm* is preferable to the *Generating-Matrix Column Exchange Algorithm* as a search algorithm. Five of the seven no-confounding designs with $p_6 = 1$ were discovered by the *Random Structured-Column Exchange Algorithm*. The *Generating-Matrix Column Exchange Algorithm* cycled through the columns of a design $\boldsymbol{D}$ far less frequently than the *Random Structured-Column Exchange Algorithm.* When the number of columns in $\boldsymbol{D}$ was greater than nine, the *Generating-Matrix Column Exchange Algorithm* would only cycle through the columns of $\boldsymbol{D}$ between 1-2 cycles in a 24 hour period. In contrast, the *Random Structured-Column Exchange Algorithm* would spend only 5-10 minutes per column doing random column exchanges before advancing to the next column. Consequently, the diversity in the design space searched by the *Generating-Matrix Column Exchange Algorithm* was relatively small compared to the diversity of the space explored by the *Random Structured-Column Exchange Algorithm.* Ultimately, both algorithms proved to be effective search tools.

There are several avenues of future research in the area of no-confounding designs. A simulation using other model building techniques, such as those based on the Dantzig selector, may reveal designs which offer even better performance in terms of type I and type II model building errors. Designs that appear similar in performance using 2-stage stepwise regression may in fact differ more significantly when building models with a Dantzig selector.

The *Random Structured-Column Exchange Algorithm* only searched the balanced-column design space. If this algorithm was modified to allow for unbalanced columns, perhaps superior no-confounding designs could be discovered. Furthermore, considering other structures of the random sub-columns may perhaps lead to the discovery of better-performing designs. For example, the sub-column structure could be expanded to include a sub-column of one +1 or -1 and five elements of opposite sign.

Finally, it is expected that the results for the minimum G-aberration 24-run designs extend to the 32-run design case. That is to say a minimum G-aberration no-confounding design of 32 runs is likely optimal over all other OAs of similar dimensions. Monte Carlo simulation could be used to verify this. It would also be interesting to use the Monte Carlo simulation to compare the 2^{k-p} designs to a 32-run no-confounding designs of strength three. Just as the 16-run no-confounding designs provide alternatives to 16-run 2^{k-p} designs, perhaps no-confounding designs would be economically advantageous in the 32-run case as well.

Chapter 5

A SCREENING DESIGN SELECTION METHODOLOGY

5.1 Introduction

When experiments involve more than four continuous factors, it can be cost-prohibitive to use a full factorial experimental design. When the experimental budget or resources are significantly constrained, a more affordable alternative is to use a multi-stage experiment involving a screening design. Often an experimental budget must be established prior to the experiment. However, when experimenting in stages, it becomes more difficult to provide an initial estimate of the total cost of the experiment. This is because the total number of runs in each stage of experimentation - usually the key driver of experimental cost - depends on the number of significant main effects (ME), two-factor interactions (2FI) and quadratic effects (QE) required to adequately model the response variable.

The total number of experimental runs is heavily influenced by the choice of screening design. The screening design determines aspects of the data such as whether active effects are completely confounded with other effects and whether quadratic effects can be estimated in the screening phase. These aspects affect how many experimental runs are required in the later stages of experimentation. In this chapter, a methodology is provided to select a screening design from a set of alternatives such that the complete experiment has the lowest expected cost, as determined by the number of experimental runs and other cost factors.

Early work on design selection based on budgetary constraints includes Neuhardt, Bradley et al. (1973) and (1978) who examined the problem of selecting the lowest cost

2^{k-p} fractional factorial experiment based on the assignment of a cost c_i to each "cell" (factor level combination) in the full factorial experiment. Neuhardt, Bradley et al. (1978) discuss a dynamic programming approach to obtaining an optimal cost fractional factorial based on cell costs. See Mount-Campbell and Neuhardt (1980) and (1982) for discussions concerning cost-optimal 3^{n-r} fractional factorial designs and $3^{m-r}2^{n-s}$ fractional factorials respectively.

Pignatiello (1985) introduced a procedure to determine cost-optimal p^{n-r} fractional factorials (for prime p) that simultaneously considered cell costs and the ability to estimate a rank ordered list of specified main effects and two-factor interactions. Rafajlowicz (1989) developed an algorithm to find cost-efficient fractional factorials based on a specified information matrix.

Tack and Vandebroek (2001) were the first to simultaneously consider resource costs and run-transition costs as criteria for constructing cost-optimal designs. In Tack and Vandebroek (2002) they extend their cost-efficient and time trend-resistant optimal design approach to block designs. In Tack and Vandebroek (2004) the work is extended to situations where there is a budgetary constraint. Additional optimal design research was conducted by Park, Montgomery et al. (2006), who developed a genetic algorithm to create cost-constrained designs with good G-efficiency.

Our work extends the previous research on cost-constrained experimental design in three important ways. First, while the previous methods optimized the construction of a one-stage experimental design, our method minimizes the expected start-to-finish experimental cost of a multi-stage experiment involving a screening design, fold over runs, and axial runs. Second, instead of using an algorithm to select cost-optimal runs for

a fractional factorial design, our methodology is a process for design selection from a candidate list of cataloged screening designs. The methodology can be used with any traditional or modern experimental design, such as a definitive screening or no-confounding design. Finally, an expanded scope of three cost sources is considered: the cost of the experimental runs; the cost of building a model that omits significant model terms or includes superfluous model terms; and the cost of building a model with inaccurate estimates of the regression coefficients. The screening design with the lowest expected cost is reported for all feasible combinations of model probability and cost penalty values. This facilitates a sensitivity analysis that informs the experimenter of how robust the optimal screening design is to changes in the probability and cost penalty assumptions. Instead of making predictions or assumptions about specific probabilities or inaccurate model cost penalties, the experimenter has the easier task of predicting a region of probabilities and penalties.

5.2 Background

When experimenters investigate a set of continuous independent variables to determine how they affect a response variable, a common objective it to build a linear regression model for the response. This model has the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Where $\boldsymbol{y}$ is the $n \times 1$ vector of responses, $\boldsymbol{X}$ is the $n \times p$ model matrix, $\boldsymbol{\beta}$ is the $p \times 1$ vector of model coefficients, and $\boldsymbol{\varepsilon}$ is the $n \times 1$ vector of independent and normally distributed random errors with mean 0 and variance σ^2.

The screening design process is part of the first of three stages in a linear model building process known as response surface methodology (RSM). In stage one, a screening design is used to determine which of a set of candidate factors affect the response variable. An orthogonal design involving the active effects is then used to fit a first order model and compute a path of steepest ascent or descent. In stage two, single experiments are performed sequentially along the path to identify a region of the factor space which likely contains the optimal response. Finally, in stage three a response surface model is built to identify the optimum factor settings. See Myers, Montgomery et al. (2009) for a complete presentation of RSM. Cheng and Wu (2001) proposed a two-stage response surface analysis strategy applicable to situations where the first and third stages of the response surface methodology are conducted in the same experimental region. The focus of this chapter is also on experimental scenarios where the second stage of the response surface methodology is not required and there are very limited resources for experimentation in stage three.

5.2.1 Types of Screening Designs

Screening designs have been used in the first stage of the model building process since the introduction of the Plackett-Burman designs Plackett and Burman (1946). The traditional 2^{k-p} fractional introduced in Box and Hunter (1961) is the predominant choice for factor screening, owing to its compatibility with standard design analysis and the ease of design augmentation into a central composite design if higher order terms are active. Recently, designs such as the no-confounding design in Jones and Montgomery (2010) and the definitive screening designs in Jones and Nachtsheim (2011) have been presented in the literature.

5.2.1.1 Traditional Screening Designs

The 2^{k-p} fractional factorial design and the Plackett-Burman designs are traditional screening designs that are very commonly used for product and process design and for process improvement. 2^{k-p} fractional factorials, which use only a subset of the runs in a full factorial, rely on two important concepts: the sparsity of effects principle and sequential experimentation Montgomery (2012). Sparsity of effects is the idea that a system or process is likely to be determined by a relatively small number of main effects and low-order interactions. Sequential experimentation is the concept that the runs from separate fractional factorials can be combined to form a larger design to estimate the active effects and interactions in a process. A significant drawback of the 2^{k-p} is the relatively large number of experimental runs required for a sequential fold over experiment when two-factor interactions are completely confounded with other effects.

The Plackett-Burman design was introduced in 1946 as way to investigate up to $N-1$ factors with N experimental runs, where N is a multiple of four. All columns in the design matrix are pair-wise orthogonal, meaning main effects are completely de-aliased, but every main effect is partially aliased with every two-factor interaction not involving itself. Designs which produce effect estimates that are partially confounded with other effects are known as nonregular designs. Due to the complex alias structure of Plackett-Burman designs, experimenters have historically been cautioned about using them when several two-factor interactions are believed to be active. However, it has been demonstrated that the good projection properties of these designs can be exploited to allow for estimation of active main effects even when there are active two-factor interactions Tyssedal and Samset (1997).

5.2.1.2 Definitive Screening Designs

Definitive screening designs are a class of three-level nonregular screening designs that allow the estimation of all possible full quadratic models involving any three factors, as long as at least six factors are screened. To screen k factors, the design uses $2k + 1$ runs in a structure that includes one center run and runs called fold over pairs. The first run in each fold over pair has one factor at the zero level and the other $k - 1$ factors have levels of ± 1. The second run has the factor levels of the first run after multiplication by -1. For every factor under investigation, one fold over pair is added to the design.

For definitive screening designs, main effects are orthogonal to other main effects as well as quadratic effects. In addition, there is no complete confounding of any pair of second-order effects. The alias structure between two-factor interactions and other two-factor interactions, as well as between quadratic effects and two-factor interactions is complex. A full discussion of definitive screening designs can be found in Jones and Nachtsheim (2011).

5.2.1.3 No-confounding Designs

No confounding designs are a class of nonregular fractional factorial designs which have the defining feature that no main effect is completely confounded with any two-factor interaction. For many of these designs it is also true that no pair of two-factor interactions are completely confounded.

The 16-run designs for investigating six to ten factors are projections of the five orthogonal designs introduced in Hall (1961) for screening up to 15 factors in 16 runs. The no-confounding design for k factors is the optimal combination of k Hall-matrix

columns that minimizes $E(s^2)$. Let $\mathbf{x}_i$ be the ith column of the $n \times k$ model matrix $\boldsymbol{X}$. Then

$$E(s^2) = \sum_{i<j} (\mathbf{x}_i \grave{} \mathbf{x}_j)^2 \Big/ \binom{k}{2}$$

The $E(s^2)$ criterion was proposed by Booth and Cox (1962) as a metric to compare supersaturated designs. Minimizing $E(s^2)$ is equivalent to minimizing the sum of squared off-diagonal elements of the correlation matrix of the model matrix.

The lack of complete confounding of model effects presents a significant cost savings opportunity to an experimenter. If effect sparsity holds and there are no quadratic terms required in the model, analysis with a model building technique such as step-wise regression can produce an accurate model without additional experimentation beyond the screening design (Jones and Montgomery 2010). This can represent a significant savings in experimental costs.

5.2.2 Selecting Screening Designs

Nonregular two-level designs have the potential, when used in conjunction with model building techniques such as stepwise regression, to estimate the terms in a model with only main effects and two-factor interactions without additional experimentation beyond the original screening design. If a third factor level is added then it is possible to estimate terms in a quadratic model. When the screening design has the potential to be the terminal design, a potential largely influenced by effect sparsity, the number of runs in the screening design becomes an important cost consideration. Choosing the number of runs in a screening design involves a trade-off between experimental costs and the power to identify active effects. Small screening designs provide fewer degrees of

freedom to compute the mean square error (MSE) which results in a relatively larger variance for the estimate of σ^2. This affects the power to detect active factors, especially when the effect size of the active factor is less than $2\sigma^2$.

The number of runs in the design matrix also affects the variance of the estimates of the regression coefficients $\widehat{\boldsymbol{\beta}}$. Since $\text{Var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, where $\mathbf{X}$ is the design matrix in model form, then the variance of $\hat{\beta}_i$ is minimized when the diagonal elements of $\mathbf{X}'\mathbf{X}$ are as large as possible and $\boldsymbol{X}$ is orthogonal. When all factor settings are at the two levels of ± 1 then the magnitude of the diagonal elements of $\mathbf{X}'\mathbf{X}$ is equal to the number of runs in the design. Therefore if $\mathbf{X}$ is orthogonal then adding more runs with levels at ± 1 will always decrease the variance of $\hat{\beta}_i$ for all i - as long as orthogonality is preserved. It is up to the experimenter's judgment how to make trade-offs between saving experimental resources and running enough experiments to have good estimates of σ^2 and $\boldsymbol{\beta}$.

5.2.3 Screening Designs Analysis

An experimenter should carefully choose the analysis method used to develop a response model, as some analysis strategies fail to capitalize on the advantages of nonregular designs. Traditional design analysis methods and statistical tools such as ANOVA and normal probability plots are often used with factorial and regular 2^{k-p} fractional factorial designs. See Montgomery (2012) for a complete presentation of 2^{k-p} fractional factorial analysis. Hamada and Wu (1992) introduced an analysis strategy for nonregular designs that uses a forward stepwise regression procedure based on effect sparsity and effect heredity.

Forward stepwise regression is a model building technique that starts from a void model and sequentially adds variables in steps. At each step a variable is added from a candidate list that improves the model the most according to some criteria. When effect heredity is enforced, if a two-factor interaction or quadratic term is added to the model then all corresponding main effects are automatically added to the model.

Two-stage forward stepwise regression was suggested by Shinde (2012) as a method to reduce type I and type II model building errors in screening design analysis. First, forward stepwise regression is conducted considering only ME as candidate model terms. After the active ME are identified in the first stage, a second iteration of forward stepwise regression is conducted. In this second stage, the candidate model terms are based on the full model or full quadratic model for the active ME identified in stage one. There is an important caveat when using forward stepwise regression with two-level designs that include center runs. If all quadratic terms are included as potential model terms, the procedure will identify at most one quadratic term as active - even if the response model has more than one active quadratic effect. Obviously the quadratic terms are completely confounded due to only one unique run with a third factor (the center run) so the effect estimate is inaccurate. Thus the selection of a quadratic term in the second stage of the stepwise regression is a signal to augment the original screening design with runs involving a third level for each ME and re-estimate the second order terms.

The choice of model evaluation criteria used to select candidate variables and rank potential models is an important aspect of model building. For the example in this chapter a version of the Akaike information criterion (AICc) with a bias correction proposed by Hurvich and Tsai (1989) is used. The AIC, based on the concept of

information entropy, was introduced in Akaike (1974) as a measure that balances the reward for goodness of fit with a penalty for overfitting. The AICc utilizes a bias correction for small sample sizes making it a valuable statistical tool when building models from screening designs.

It is clear that an experimenter has many options to consider when selecting a screening design, but it is often unclear which option is the optimal choice. The primary difficulty in selecting a screening design is the lack of information regarding how many two-factor interaction or quadratic terms will ultimately be required in the regression model. The next section presents an expected cost methodology to select an optimal screening design under model uncertainty.

5.3 An Expected Cost Approach to Screening Design Selection

Selecting a screening design is a non-intuitive decision due to the competing requirements of the experiment; the experimenter wants to fit the most accurate model of the true system while consuming the minimum amount of testing resources. The difficulty of the decision comes from the fact that the advantages and disadvantages of a screening design relative to other designs are dependent on the nature of the response model, which is unknown. For example, 12-run Plackett-Burman designs are a suitable and economical choice when a small number of main effects and two-factor interactions are active but require augmentation when quadratic effects are active. On the other hand a definitive screening design, while having more initial runs than the 12-run Plackett-Burman, offers the prospect of estimating a full quadratic model in a small subset of factors without additional experimentation.

Our solution to the problem of selecting a screening design in the presence of model uncertainty is a methodology that involves the use of decision trees. In decision theory, decision trees are a tool that facilitates decision making when the consequence of the decision is subject to the chance outcome of a future event. The optimal decision is the course of action which leads to the optimal expected value of the decision consequence as measured by some metric. For more details on decision trees see Clemen and Reilly (2001).

The decision tree for the design selection methodology has the candidate screening designs as the decision nodes, categories of polynomial response functions as branches from the chance nodes, and the total end-to-end cost of experimentation as the metric evaluated at the end nodes. Polynomial response models are placed into mutually exclusive categories such as main effects models, main effects plus two-factor interaction models, and quadratic models. For each category of polynomial response model and for each candidate screening design, the experimenter assumes he will use a standard experimental process to estimate the regression model coefficients. For example, the experimenter could create the following standard analysis rule set to analyze a 2^{k-p} fractional factorial:

1. Add four center runs to a two-level screening design to enable a lack-of-fit test
2. Augment the screening design with a fold over to resolve any ambiguities associated with completely confounded effects.
3. Conduct a lack-of-fit test and, if required, augment the design again with two axial runs per active main effect in order to estimate the quadratic terms.

Note that the standard analysis method may have to be tailored to general classes of screening designs so that the most accurate analysis is performed. An experimenter may even outline a specific strategy for each screening design under consideration. The purpose is to make a realistic assessment of the total runs required across all experimental stages given a realization of the chance outcome of one of the polynomial response models. That is, for a given screening design and response function, the experimenter needs to assess the required number of screening runs, center runs, fold over runs and axial runs. The following sections present the cost model and the details of computing the expected cost for the end nodes of the decision tree. It should be emphasized that the methodology is a general construct for evaluating any candidate set of screening designs in the presence of uncertainty about a wide variety of possible response models.

5.3.1 The Experimental Cost Model

The experimental cost model in the screening design selection methodology has three components: the cost of experimental resources; the cost of utilizing a model with inactive effects or omitted active effects; and the cost of utilizing a model with inaccurate estimates of the regression model coefficients. The cost of experimental resources includes material, personnel and facility costs. There are several possible methodologies for assessing the resource costs in a multi-stage experimental environment. It is important to note here that much of the existing literature concerning cost-constrained designs uses the "cell" cost methodology as seen in Neuhardt, Bradley et al. (1978) and subsequent work. In this traditional methodology, the specific combination of factor levels drives the resource cost. However, suppose that the difference between the least expensive run and most expensive run was $500 and the average cost of an experiment

was \$5000. In this situation the number of runs in the experiment drives the resource cost. Consider a second example where the cost to set up a stage of an experiment is \$40,000 and the resources for each run average \$500. Here the number of stages in the experiment drives the resource cost. The use of the expected cost methodology requires an experimenter to use a cost modeling strategy that best matches the unique circumstances of their particular experiment.

The second source of design cost is the cost of producing a final model with incorrect model terms. If inactive effects are included in the model, a type I model building error has occurred. If active effects are not included in the model, a type II model building error has occurred. There are costs associated with including inactive effects in a model or excluding active effects. Examples of these costs include the costs of managing inactive factors, the cost of wasted resources when using a suboptimal model, and lost profits from producing a suboptimal product.

The third source of design cost in the methodology is the cost of producing a final model with inaccurate estimates of the regression model coefficients $\widehat{\boldsymbol{\beta}}$. The accuracy of an estimate $\hat{\beta}_j$ is determined both by the variance of random errors during the experiment, as well as the correlation between factor level settings of the independent variables. Increasing the number of experiments in the design and minimizing correlation of all model matrix columns mitigates against these sources of estimation inaccuracy. The costs associated with inaccurate estimates of $\widehat{\boldsymbol{\beta}}$ are similar to those for building an inaccurate model, i.e. the cost of wasted resources when using a suboptimal model and lost profits from producing an inferior product.

When an experimenter chooses a screening design, the final realization of the three sources of cost discussed above is determined in large part by the true nature of the system that generates the response data. For example, given that one of the active effects is a two-factor interaction, a foldover design is often required when screening with a 2^{k-p} fractional factorial of resolution III or IV. A full foldover experiment will double the cost of the total experiment. However if the response model involves only main effects, the original 2^{k-p} design is often sufficient to estimate the coefficients in the response function.

5.3.2 Computing the Total Experimental Runs

The expected cost screening design selection methodology has the flexibility to handle any experimental cost paradigm, whether it is based on run-to-run cost variation, experimental set-up costs or average run costs. To simplify the presentation of the methodology it is assumed that the variation in run-to-run cost due to the change in factor levels is small relative to the cost of the resources to conduct one run. This construct enables us to compute an average cost for an experimental run and convert the cost of an experiment from dollars to its number of runs. For example, instead of saying an experiment costs $32,000, the cost of the experiment is said to be 16 runs, where the average run cost is $2,000.

Given a response function, a candidate screening design and a standard experimental strategy, the number of experimental runs required is simply the sum of the number of runs in the screening design, the number of center runs, the number of fold over runs and the number of axial runs (or other appropriate type of design augmentation). Notice that by assuming a standard experimental strategy the screening

design determines the value of all the variables in this sum. Hence the number of total runs becomes deterministic once the screening design and the response function are fixed. For example refer to Table 72 which shows the number of experimental runs required given the use of a 2^{6-2} fractional factorial to investigate six factors and estimate terms in a full quadratic model with three main effects, one two-factor interaction and two quadratic terms.

Table 72. Run Total for 2^{6-2} Experiment

Experimental Run Type	Number of Runs
Screening	16
Center Runs	4
Fold Over	16
Axial (Assuming 3 active ME)	6
Total	42

Using the standard analysis procedure from the previous section, the number of runs in the complete experimental process is 42.

5.3.3 Computing the Cost of an Incorrect Model

The second experimental cost under consideration is the cost of having an incorrect model. This is defined as building a response functions with superfluous effects or omitting active effects in the model. Extra terms in a model result in the problem of overfitting the response variable and omitting terms results in a model whose coefficients are biased by absent active terms. The consequences of an inaccurate model include the opportunity costs of forgone revenue or cost savings that could have been realized had the experimenter built and used a better regression model. These costs are different from experimental run resource costs in that they are not paid out by the testing organization. In the expected cost methodology these costs function as a penalty assessed against

designs which produce relatively inferior models. A model that lacks active effects or includes inactive effects will be labeled as an incorrect model to distinguish it from an inaccurate model discussed in the next subsection.

The methodology requires that the cost penalty for an incorrect model is in units of experimental runs, so one example of a reference incorrect model cost is the number of runs required to repeat the experiment in the future to determine the correct model. This penalty could be the number of runs in the largest screening design in the set of candidate designs.

The establishment of a reference cost has many advantages. The experimenter has the flexibility to choose a number of runs above or below this value that best reflect the unique circumstances of the experiment. Choosing a number of runs below the baseline implies that there is little opportunity cost from using a relatively inferior model and that a subsequent experiment to improve the model would not be required. Selecting a number over the baseline implies that the opportunity costs associated with using an inaccurate model for a period of time are significant and should be added to the cost of performing an improved experiment in the future.

The baseline cost of having an incorrect model does not depend on the screening design or the truth model. However, the cost of an incorrect model can be multiplied by a term that represents the probability of an incorrect model given a combination of response function and screening design. Let Ω be the cost penalty for an incorrect model in units of runs and let γ_{ij} be the probability that given truth model i, an incorrect model will result from using screening design j. Then the cost penalty for screening design j given truth model i is $\Omega\,\gamma_{ij}$.

5.3.4 Computing the Cost of an Inaccurate Model

The third experimental cost under consideration is the cost of an inaccurate model, defined as the costs incurred as the result of using a model with inaccurate estimates of the regression model coefficients. As with the incorrect model cost, the inaccurate model cost represents an opportunity cost and must be in the units of experimental runs. Again, a reference inaccurate cost penalty will be used as a decision making tool that facilitates the conversion of accuracy cost penalties from units of dollars to units of runs.

The reference penalty makes all designs equivalent in terms of the average relative variance of the model coefficients and allows the experimenter to place a value on the lower average relative variance of the larger screening designs. The relative variance allows us to isolate how the structure of the model matrix $\boldsymbol{X}$ affects the variance of $\widehat{\boldsymbol{\beta}}$ independent of the effect of the random error variance. The vector of relative variances of the elements of $\widehat{\boldsymbol{\beta}}$ is the vector of diagonal elements of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Since the relative variances depend only on $\boldsymbol{X}$, it is a useful diagnostic tool to compare the ability of screening designs to provide accurate estimates of $\widehat{\boldsymbol{\beta}}$. The relative variance is not necessarily the same for every coefficient estimate in the regression model. Since one design comparison metric is more convenient, the relative variance for every model term is averaged to produce the average relative variance of a design.

The following is the procedure to compute the inaccurate model cost reference value for design $j \in \{1,2,\dots,d\}$ given response model category $i \in \{1,2,\dots,f\}$.

1. Create a specific response model i for every model category

2. Obtain the final model matrix $\boldsymbol{X}_{ij}$ which includes all augmentation necessary to estimate the p terms in model i starting with screening design j
3. Let RV_{ijk} be the relative variance of the kth term in the model with p terms constructed by using screening design j to analyze responses generated by response function i. Compute RV_{ijk} as the kth element of the diagonal of $(\boldsymbol{X}_{ij}{}'\boldsymbol{X}_{ij})^{-1}$
4. For each design j, compute the average relative variance given truth model i as

$$\overline{RV}_{ij} = \frac{1}{p}\sum_{k=1}^{p} RV_{ijk}$$

5. Determine the minimum average relative variance given truth model i as

$$\overline{RV}min_i = \min\{\overline{RV}_{i1}, \overline{RV}_{i2}, \ldots, \overline{RV}_{id}\}$$

6. For each truth model i, augment design j using the D-optimality criteria until $\overline{RV}^*_{ij} = \overline{RV}min_i$, where $\overline{RV}^*_{ij}$ is the average relative variance of the augmented design $\boldsymbol{X}^*_{ij}$
7. Let N_{ij} be the number of runs in $\boldsymbol{X}_{ij}$ and NA_{ij} be the number of runs in $\boldsymbol{X}^*_{ij}$. Compute the inaccurate model cost reference value as $\tau_{ij} = NA_{ij} - N_{ij}$

The τ_{ij} reference value is more nuanced than simply using the difference in runs between a design j and the design which produced $\overline{RV}min_i$. A smaller design may have a correlation structure that is superior to that of a larger design and so it is more efficient in terms of reduction in the relative variance per number of additional runs. In that case NA_{ij} would be smaller than the difference in runs between design j and the design whose average relative variance is equal to $\overline{RV}min_i$. Note that the actual experimental designs would not contain the augmentation runs used to calculate NA_{ij} but this process

establishes the reference value from which an experimenter can estimate the value of a smaller relative variance.

The experimenter can multiply τ_{ij} by a quantity Φ to decrease or increase the reference inaccurate cost penalty. For $0 \leq \Phi < 1$, this implies that the increased average relative variance of smaller designs has a smaller opportunity cost penalty than the cost of the additional runs required to achieve the minimum average relative variance. Conversely, setting $\Phi > 1$ implies that the opportunity cost of using a less accurate model exceeds the cost of the additional runs required to achieve the minimum average relative variance. The next subsection discusses how the three cost components are combined in an expected cost construct to identify a cost-optimal screening design from among a set of alternatives.

5.3.5 Computing the Expected Cost

The screening design selection expected cost methodology is as follows:

1. Create a candidate set of d screening designs
2. Create a set of f candidate response functions for each of the f model categories
3. Establish a standard analysis procedure for each screening design j
4. For each combination of design j and model i determine the following:

   n_{ij1} = The number of screening experiments in design j

   n_{ij2} = The number of center runs

   n_{ij3} = The number of fold over experiments

   n_{ij4} = The number of axial runs required to estimate quadratic effects

5. Let $\Omega\,\gamma_{ij}$ be the cost of an incorrect model and $\Phi\tau_{ij}$ be the cost of an inaccurate-model as calculated above. Compute the total cost (in units of runs) for design j given model i as

$$\mathrm{N_{ij}}=\left(\sum_{q=1}^{4}\mathrm{n_{ijq}}\right)+\Omega\,\gamma_{ij}+\Phi\tau_{ij}$$

6. Let P_i = Probability that model i is the true response function, $i = 1,2,\ldots,f$. Compute the expected cost in units of runs for design j as

$$E[N_j]=\sum_{i=1}^{f}(P_i * N_{ij})$$

7. Identify the cost-optimal screening experiment as design j where

$$E[N_j] = \min\{E[N_1], E[N_2],\ldots, E[N_d]\}$$

A significant advantage to this approach is that the experimenter does not have to be precise about the estimates of the parameters P_i, Ω and Φ. Rather the experimenter can compute the cost-optimal screening experiment for every factorial combination of a discretized set of levels of these parameters. Naturally, the combinations of P_i are restricted due to the fact that $\sum_{i=1}^{f}P_i = 1$. The output of this process will define regions of parameter settings where one screening design is cost-optimal. Now the experimenter can decide which region of parameter settings correspond to his particular situation and also determine the sensitivity of the design choice to those assumptions.

5.4 An Application of the Expected Cost Method

The expected cost methodology is now demonstrated by employing it in a scenario where six factors are under investigation and a screening design must be selected from five alternatives. Recall that to use the expected cost method the experimenter must create a candidate list of screening designs, create categories of

possible response function models and establish a standard analysis method. As previously mentioned, this is a general methodology which can be applied to any candidate set of screening designs and set of potential response models. The designs and response models chosen in this illustrative example are not intended to be taken as prescriptive recommendations but rather as a reference point for the reader to consider as they evaluate their own particular experimental situation. The following screening design candidate set was chosen to accomplish the additional contribution of a comparison of traditional screening designs to designs introduced in the recent literature. Evidence is provided that situations exist where it is advantageous to use the recently introduced definitive screening and no-confounding designs over the traditional 2^{k-p} and Plackett-Burman designs.

The five screening designs chosen for this example are in Table 73.

Table 73. The Candidate Screening Designs

Design	Runs
Plackett-Burman	12
Definitive Screening	17
No-confounding	16
No-confounding	24
2^{6-2} fractional factorial	16

Note that a D or I-optimal design was not included in the candidate list, although it is likely that many experimenters would choose to consider such a design. Optimal designs using the D and I criteria were investigated but since a full 6-factor quadratic model has 28 terms, a trade-off had to be made between model-robustness and the number of design runs. A candidate optimal design tended to be either too large to compete with the smaller designs, or lacked the model-robustness to compete with the larger designs. The analysis indicated that since an optimal design is created based on an

assumed response model, a D or I-optimal design (when compared to the model-robust designs in this example) only had a minimum expected cost over a small portion of the model probability space. Since D and I-optimal designs were in most cases dominated by the other designs in this example, they were omitted as incorporating them did not enhance the demonstration of the methodology and significantly increased the complexity of the results.

A 24-run no-confounding design was selected for this example that had a value of $E(s^2)$ equal to the best published value for any 24-run two-level design with six balanced and orthogonal columns. This was the alternative 24-run no-confounding design from Chapter 4 of this dissertation. It is important to note that the 13-run definitive screening design for six factors was not used due to the presence of relatively high correlation values in its extremely complex alias structure. The problem was mitigated by using a larger design: the 17-run definitive screening design for eight factors with two columns removed. The complex alias structure was less of a problem for the 12-run Plackett-Burman design for which the non-zero values in the alias matrix are only $\pm 1/3$. When center runs were added to the Plackett-Burman design, the degree of correlation between effects was not deleterious. Four center runs were added to all of the candidate designs except the 17-run definitive screening design, which already incorporated a third level for each factor.

The standard analysis procedure used for the Plackett-Burman and the no-confounding designs was as follows:

1. Use 2-stage forward stepwise regression with the AICc as the model evaluation criterion, considering all terms in a six-factor full quadratic model as potential terms

2. If any quadratic terms are significant then augment the design with two runs per active main effect using the I-optimality criteria
3. Re-run the 2nd stage of the forward stepwise regression to obtain the effect estimates

The analysis used for the definitive screening design was as follows:

1. Estimate the effects using 2-stage forward stepwise regression with AICc as the model evaluation criterion. Consider all terms in a six-factor full quadratic model as potential terms

The standard analysis procedure used for the regular 2^{6-2} factorial designs was as follows:

1. Use 2-stage forward stepwise regression with the AICc as the model evaluation criterion, considering all terms in a six-factor full quadratic model as potential terms
2. If confounded two-factor interactions are active, augment the design with 8-runs using the I-optimality criteria and re-run stage 2 of the forward stepwise regression
3. If any quadratic terms are significant then augment the design with two runs per active main effect using the I-optimality criterion, specifying a full quadratic model in the active ME
4. Re-run stage 2 of the forward stepwise regression and estimate the effects

When choosing the number of model categories, the experimenter must assess the tradeoffs between cost estimation accuracy and the interpretability of the results. Li, Sudarsanam et al. (2006) conducted a meta-analysis of published experiments involving a DOE methodology and reported that for the 113 combined experimental data sets, 41% of the potential ME were active and 11% of the potential 2FI were active. According to these results, when considering six potential factors, it is expected that there are 2-3

active ME and 1 active 2FI. Effect sparsity is another valid justification for assuming that there will not be a large number of terms in a model. If an experimenter expects a model of at most 7 effects, it seems reasonable to categorize models based on the types of terms in the model (ME, 2FI, QE) rather than by the presence or absence of specific effects. This categorization is even more reasonable if the alias structure of the screening design is similar within categories of effect types, as is the case with the Plackett-Burman, 2^{k-p}, no-confounding designs and definitive screening designs.

Three regression models were chosen for this example: a model with 2-4 possible ME; a model involving 2-4 ME and up to 3 2FI; and a model involving 2-4 ME, up to 3 2FI and up to 2 QE. The advantage of using three model categories is that the results are easily visualized on a triangle simplex coordinate system where the axes represent the model category probabilities. The disadvantage is that the results have less resolution, particularly with respect to the impact of incorrect and inaccurate model estimation penalties. For example, there may be a lower probability of a type II model-building error if a screening design is used to estimate a model with two ME and one 2FI than if that design was used to estimate a model with four ME and three 2FI.

The γ_{ij} term in the incorrect model penalty was computed via Monte Carlo simulation. 10,000 randomly generated response models were analyzed for each combination of screening design and response model category. The responses used for the analysis were generated by randomly producing a polynomial equation involving a random number of 2-4 ME, 0-3 2FI and 0-2 QE. The coefficients of the ME were realizations of a random variable X which was generated via the equation $X = e^{0.69+U}$, with U~ Uniform(0,1). The equation produced random ME coefficient values in the

range $2 \leq X \leq 5.44$. The coefficients of the 2FI were realizations of a random variable Y which was generated via the equation $Y = e^{0.69+V}$, with $V \sim$ Uniform(0,0.56). The equation produced random 2FI coefficient values in the range $2 \leq Y \leq 3.5$. The coefficients of the QE were realizations of a random variable Z which was generated via the equation $Z = e^{0.69+W}$, with $W \sim$ Uniform(0,1). The equation produced random QE coefficient values in the range $2 \leq Z \leq 5.44$. The sign of the coefficient was determined by a binomial random variable with p = 0.5. The random error term for each response was a standard normal random variable and hence the coefficients were specified to be multiples of the random error variance $\sigma^2 = 1$. At each of the 10,000 iterations, a response column was generated by plugging the factor level combinations for each design row into the randomly generated polynomial equation and adding the random error term to the computed response.

The inaccurate model penalty was computed as discussed in the previous section. The only assumption made was that an average of six additional runs would be required to estimate quadratic effects when the random responses came from the quadratic response model category. Table 74 summarizes the runs in each cost category for every design and response model combination.

Table 74. Experimental Cost (in Runs) by Design and Model

Design	Truth Model	Screening Runs	Center Runs	Fold over Runs	Augment Runs	Incorrect Model Probability γ_{ij}	Inaccurate Model Reference τ_{ij}	Total Runs
12-Run Plackett-Burman	ME	12	4	0	0	0.2354	12	28
	ME+2FI	12	4	0	0	0.3909	11	27
	Quadratic	12	4	0	6	0.6028	3	25
2^{6-2} Fractional Factorial	ME	16	4	0	0	0.3339	8	28
	ME+2FI	16	4	8	0	0.2812	0	28
	Quadratic	16	4	8	6	0.4035	0	34
17-Run Definitive Screening	ME	17	0	0	0	0.3621	11	28
	ME+2FI	17	0	0	0	0.2908	10	27
	Quadratic	17	0	0	0	0.4019	11	28
16-Run No-confounding Design	ME	16	4	0	0	0.3129	8	28
	ME+2FI	16	4	0	0	0.3106	7	27
	Quadratic	16	4	0	6	0.4729	1	27
24-Run No-confounding Design	ME	24	4	0	0	0.3541	0	28
	ME+2FI	24	4	0	0	0.2037	2	30
	Quadratic	24	4	0	6	0.3389	1	35

The final step in the expected cost method was to determine which candidate screening design had the lowest expected cost given the possible combinations of the parameters P_i, Ω and Φ. Recall that Ω is the cost penalty for an incorrect model and Φ is the multiple of the inaccurate model cost reference value that is used to calculate the inaccurate model penalty. P_i was defined as follows:

P_1 – The probability of a response model with only main effects

P_2 – The probability of a response model with main effects and two-factor interactions

P_3 – The probability of a full quadratic model response model

A simplex coordinate system for three components was used to plot the results for each of the settings of P_i, given fixed values of Ω and Φ. This coordinate system is often used for mixture experiments and other situations where the coordinate axes correspond

to three variables whose values always sum to one. It was practical to discretize the parameter values to an appropriate level of resolution in the following way: P_i ranged from 0 to 1 in increments of 0.02; Ω took values 0, 34 and 68; and Φ took values of 0.00, 1.00, and 2.00.

Figure 39 shows the screening design with the lowest expected cost for all possible combinations of the discretized values of P_i when $\Omega = 0$ and $\Phi = 0.00$. The data was interpolated for probability combinations between the discretized values so that the results could be presented as continuous regions in the figures. The legend for the figures is interpreted as follows: "DS" is the definitive screening design; "P-B" is the 12-run Plackett-Burman design; "NC(16)" is the 16-run no-confounding design; "NC(24)" is the 24-run no-confounding design; "2^{6-2}" is the 2^{6-2} fractional factorial design. Figure 40-Figure 47 show the screening design with the lowest expected cost for all possible combinations of the discretized values of P_i for the remaining combinations of the discretized values of Ω and Φ.

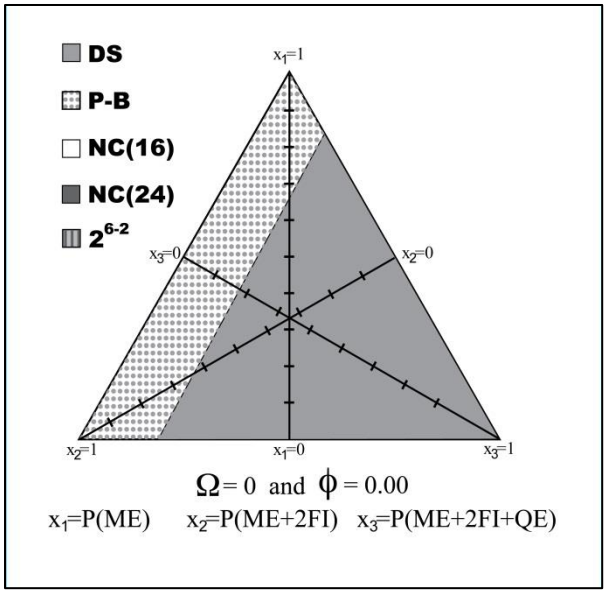

Figure 39. Results with $\boldsymbol{\Omega = 0}$ and $\boldsymbol{\Phi}$ = 0.00

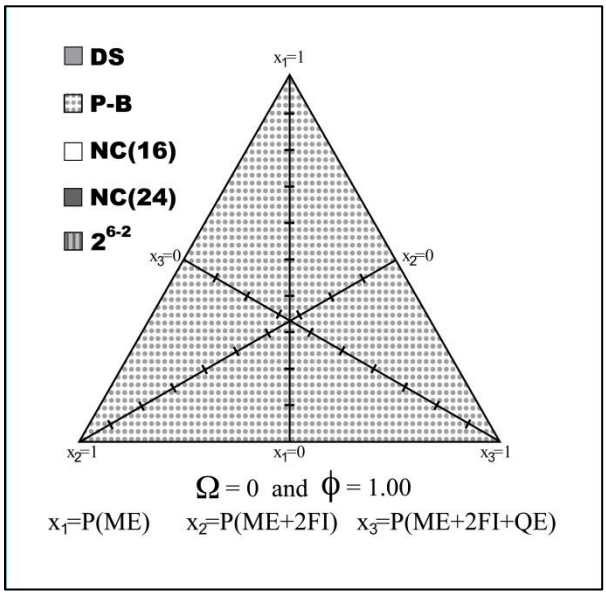

Figure 40. Results with $\boldsymbol{\Omega = 0}$ and $\boldsymbol{\Phi}$ = 1.00

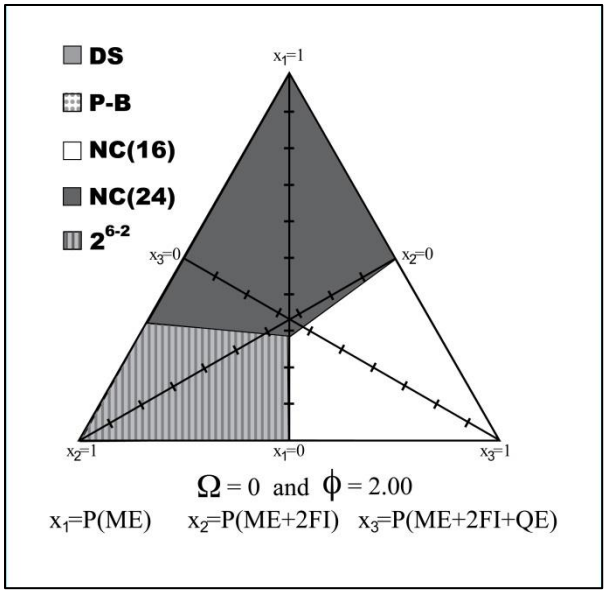

Figure 41. Results with $\boldsymbol{\Omega = 0}$ and $\boldsymbol{\Phi}$ = 2.00

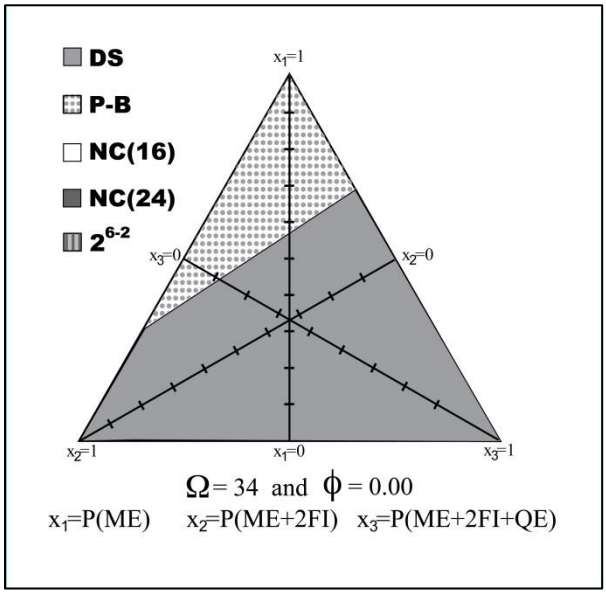

Figure 42. Results with $\boldsymbol{\Omega = 34}$ and $\boldsymbol{\Phi}$ = 0.00

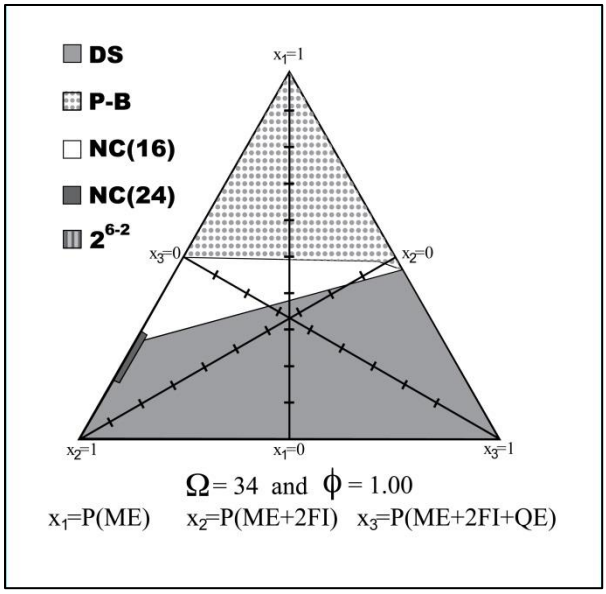

Figure 43. Results with $\boldsymbol{\Omega = 34}$ and $\boldsymbol{\Phi}$ = 1.00

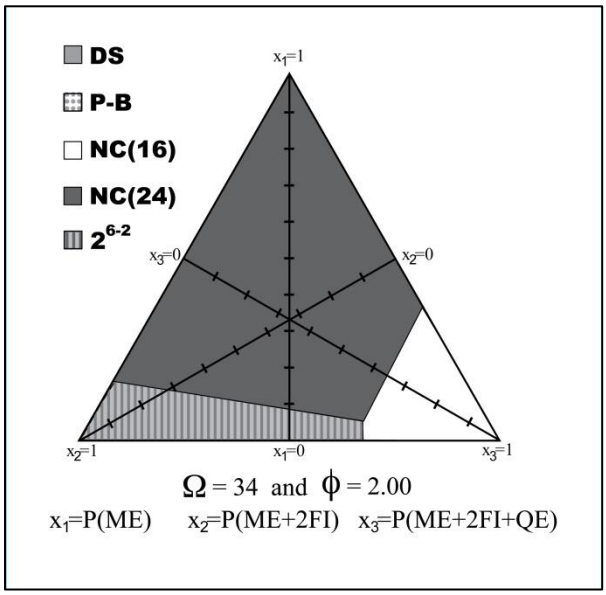

Figure 44. Results with $\boldsymbol{\Omega = 34}$ and $\boldsymbol{\Phi}$ = 2.00

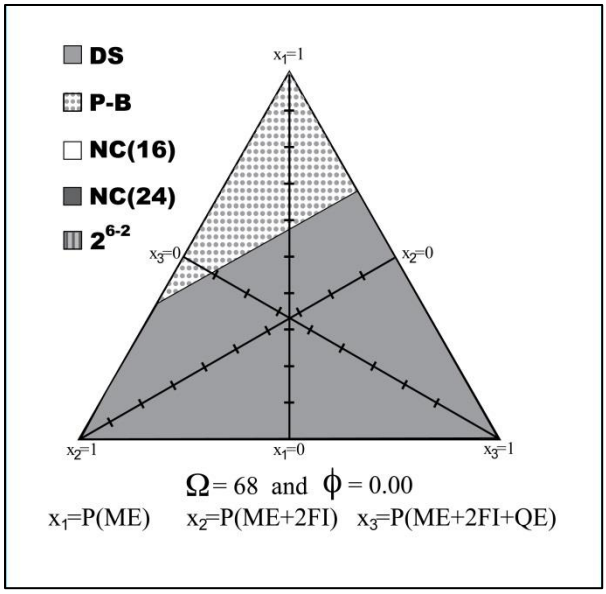

Figure 45. Results with $\boldsymbol{\Omega = 68}$ and $\boldsymbol{\Phi}$ = 0.00

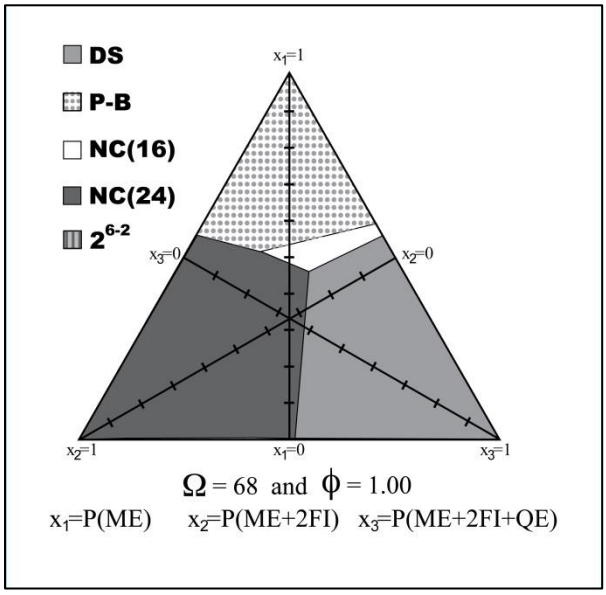

Figure 46. Results with $\boldsymbol{\Omega = 68}$ and $\boldsymbol{\Phi}$ = 1.00

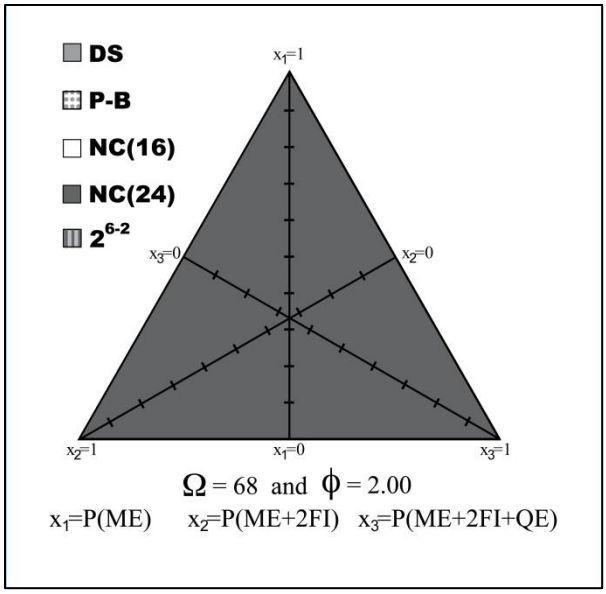

Figure 47. Results with $\boldsymbol{\Omega = 68}$ and $\boldsymbol{\Phi}$ = 2.00

The results in Figure 39 - Figure 47 provide valuable information to help an experimenter choose the most cost-effective design. Suppose an experimenter decides that the cost penalties for an incorrect or inaccurate model are negligible and there is little a priori knowledge about the type of model (ME, ME+2FI, quadratic) to be fit. By examining Figure 1 they can conclude that since the area corresponding to the definitive screening design is significantly greater than the Plackett-Burman design, the definitive screening design has the highest probability of having the lowest overall experimental cost. However, if there is very strong a priori evidence that there are no quadratic terms in the model then Figure 1 shows that a Plackett-Burman design is most likely to have the lowest cost.

Now suppose the experimenter decides that the inaccurate model cost is equivalent to the cost of running the extra experiments required to equal the lowest average relative variance of the coefficients among the alternative designs. Also it is decided the incorrect model penalty is equivalent to the cost of re-running the largest

alternative experiment. Then by referring to Figure 5 they can conclude that, under the assumption of equal probability for any model ($P_i = 1/3, i \in \{1,2,3\}$) or the assumption that the probability of a ME-only model is less than 30% ($P_1 < 30\%$), the definitive screening design is most likely to have the lowest cost. If the probability of a ME-only model is above 50% ($P_1 > 50\%$) then the Plackett-Burman design is most likely to have the lowest cost. If $30\% < P_1 < 50\%$ and if the probability of a ME and 2FI model is higher than the quadratic model, then the 16-run no-confounding design is most likely to have the lowest cost. For a very small set of circumstances, i.e. ($18\% < P_1 < 30\%$) and ($P_3 < 3\%$), the 24-run no-confounding design is expected to have the lowest cost.

5.5 Conclusion

This chapter has introduced an expected cost methodology to facilitate the selection of a screening design from a set of candidate experimental designs. The methodology gives an experimenter insight into which screening design will have the lowest expected cost given assumptions about the current experimental situation, i.e. the experimental resource costs and the opportunity cost of using an inferior regression model. Just as importantly, the methodology provides a sensitivity analysis for the optimal design choice vis-à-vis these assumptions. This methodology is a solution to a design selection problem that has become more challenging due to the large number of screening designs that have been introduced in the past decade.

The screening design selection example in this chapter demonstrates that one particular screening design will not have the lowest expected cost in all experimental circumstances. When the opportunity cost for an inaccurate model is relatively low, the small nonregular designs such as the Plackett-Burman and the definitive screening design

have the lowest expected cost. This is largely because an experimenter can often save considerable experimental resources by using a model building technique, such as stepwise regression, to obtain unambiguous effect estimates without augmenting the screening design.

Designs with fewer runs become less cost effective as the opportunity cost of using an inferior model increases. However, it is not simply the largest experiment, in terms of runs, which has the lowest expected cost. As was seen in Figure 41, the design with the lowest expected cost changed depending on which response model had a probability greater than 50%.

When an experimenter has the data to make reasonably accurate estimates of the probability distribution of potential response model categories, it is advantageous to increase the number of screening designs in the candidate set. Considering an additional screening design increases the granularity of the analysis. In the worst case nothing in the analysis changes, but in the best case the experimenter finds a design which, for his particular experimental situation, has a lower expected cost than the designs currently in the candidate set. For example, assume two screening designs are considered in the candidate set. Then the probability space (shown by the simplex coordinate system in the previous example) will be divided into either one or two regions, each representing a design with minimum expected cost. Now suppose a third design is considered. Either the probability space is divided in the previous way, or the new design has the minimum expected cost for some region in the probability space. The minimum expected cost for the new region will be lower than it was when only two screening designs were considered.

There are several ways future research could augment the concepts in this chapter. The empirical simulation study used to estimate γ_{ij}, the incorrect model penalty term, could be modified to study how alternative analysis methods affect the γ_{ij} estimates. Another avenue for research involves the development of a more complex experimental resource cost estimation model. The cost model used in this research was based on the number of experimental runs, but a more sophisticated model that simultaneously includes set-up costs and factor combination costs (cell costs) would give more resolution to the expected cost estimate. Finally, it might be beneficial to explore a simulation method to estimate the inaccurate model penalty rather than relying on the relative variance of the effect coefficients.

Chapter 6

CONCLUSION

The analysis of the catalog of $OA(20; 2^m; 2)$ and the $OA(24; 2^m; 2)$ using Monte Carlo simulation has been an effective technique in identifying the optimal no-confounding designs and characterizing their model-fitting capabilities. The empirical analysis removed much of the confusion caused by the conflicting ranks assigned to each design by the multitude of commonly used design ranking criteria. Furthermore, the simulation results provided a detailed performance analysis of the designs across response models comprised of varying numbers of ME and 2FI.

The research in this dissertation has shown that 20 run no-confounding screening designs are very effective when the response model is comprised of 1-4 ME and 0-1 2FI. These designs will identify at least the active effects over 95% of the time, even when experimenting with as many as 12 factors. In the simulations with models of four ME and one 2FI, the 6-factor design detected at least the active main effects 96% of the time and there were no variable selection errors 75% of the time. In the best case scenario with two ME and one 2FI, the percentage at least correct and the percentage correct was 100% and 99% respectively. The 12-factor design detected at least the active main effects 95% of the time and there were no variable selection errors 72% of the time in the simulations with four ME and one 2FI. In the best case scenario with two ME and one 2FI, the percentage at least correct and the percentage correct was 100% and 91% respectively.

When there was more than one active 2FI in the response model, the analysis procedure had a difficult time differentiating between potential models. Consequently

the type II error rate went up significantly. Thus an experimenter who expects the final model to contain more than five terms should consider running more experiments.

The empirical analysis of the 6-factor 20-run OAs also showed that an experimenter can make trade-offs between type II errors and type I errors by selecting the appropriate analysis method. If an experimental budget significantly constrains additional experimentation beyond the screening design, there is a relatively high chance of successfully fitting the correct model using 2-stage stepwise regression and AICc. However, if the no-confounding design is truly used as a screening design in a multi-stage experiment, the type II error rate can be reduced by using a 1-stage mixed stepwise approach.

Finally, a high rank in terms of G-aberration is common among the selected no-confounding designs. The 6,8,10,11, and 12-factor no-confounding designs are ranked 1st in G-aberration, the 7-factor design is ranked 3rd, and the 9-factor design is ranked 7th. A conclusion that can be drawn is that minimizing the aliasing of ME with 2FI greatly improves model fitting results and should be a primary consideration in designs. However, since many designs are approximately equivalent in terms of G-aberration, it is important to consider multiple factors and find designs which belong to the Pareto efficient set of OAs.

The empirical results from the analysis of the 24-run designs show that adding four additional runs significantly improves the model fitting capability over the 20-run designs. The simulation output indicated that no-confounding designs make outstanding screening designs. When there were between two and five ME and zero to two 2FI, the 24-run designs detected all active effects over 94% of the time. Furthermore, the worst

case for identifying the correct model was 45%. These results show that no-confounding designs are an extremely cost-effective screening design option.

In terms of the fewest type I and type II model building errors, the OAs of strength three with minimum G-aberration were the best no-confounding designs by a significant margin. Furthermore the simulation metrics for the *percent at least correct* were almost identical no matter how many design columns (investigated factors) there were. It appears that having ME uncorrelated with other ME and 2FI provides a significant advantage in reducing type I and type II model building errors, at least when 2-stage stepwise regression is the analysis method.

Even though it was not possible to exhaustively search the entire space of 24-run OAs, the data gives a strong indication there does not exist a design which performs better than the minimum G-aberration no-confounding designs of strength three. The only downside to these designs was their inability to estimate the 6-factor full model in any 6-factor projection, i.e. for these designs $p_6 = 0$. Therefore if an experimenter expects to fit a large model of more than five ME and two 2FI, other screening design options may be more effective.

The algorithm was able to find alternative designs for all cases of 6-12 factors with $p_6 = 1$. Interestingly, the simulation results for these algorithmically created designs were inferior to the minimum G-aberration designs. The alternative designs however, were superior to the baseline PEC designs, both in performance and in the magnitude of p_6.

Both column exchange algorithms were effective at identifying good OAs. The *Generating-Matrix Column Exchange Algorithm* was able to find designs superior to

most of the baseline designs by investigating non-orthogonal designs. No designs were discovered that provided a compelling reason to use unbalanced columns since effective balanced-column alternatives were always available.

The screening design selection methodology in Chapter 5 proved to be an important tool in choosing between designs. The example of the methodology demonstrated that one particular screening design option will not have the lowest expected cost in all experimental circumstances. When the opportunity cost for an inaccurate model is relatively low, the small nonregular designs such as the Plackett-Burman and the definitive screening design have the lowest expected cost. This is largely because an experimenter can often save considerable experimental resources by using an appropriate model building technique, such as stepwise regression, to obtain unambiguous effect estimates without augmenting the screening design. Designs with fewer runs become less cost effective as the opportunity cost of using an inferior model increases. However, it is not simply the largest experiment, in terms of runs, which has the lowest expected cost.

When an experimenter has the data to make reasonably accurate estimates of the probability distribution of potential response model categories, it is advantageous to increase the number of screening designs in the candidate set. Considering an additional screening design increases the granularity of the analysis. In the worst case nothing in the analysis changes, but in the best case the experimenter finds a design which, for that particular experimental environment, has a lower expected cost than the designs currently in the candidate set.

There are several ways the research in this dissertation could be extended. For designs with more than 12 factor columns, no-confounding designs of 28 and 32 runs may be required to achieve satisfactory model-fitting performance. Using simulations to characterize the performance of these larger designs would help experimenters weigh the risks of investigating more than 12 factors with only 32 runs. It would be very interesting to explore under what conditions a 32-run OA of strength three would be preferable to a 2^{k-p} fractional factorial of similar dimension.

For both the 20-run and 24-run designs, perhaps design performance could be better differentiated and characterized by using other model fitting techniques, such as those based on the Dantzig selector. Designs that appear similar in performance using 2-stage stepwise regression may in fact differ more significantly when fitting models with a Dantzig selector. It may also be that other analysis techniques could take advantage of aspects of non-orthogonal designs. Perhaps using a different analysis method, the non-orthogonal designs might perform better than the OAs of strength three.

The expected cost screening design selection methodology could be expanded and improved in many ways. The empirical simulation used to estimate the incorrect model penalty term could be modified to study how alternative analysis methods affect the penalty estimates. Another avenue for research involves the development of a more complex experimental resource-cost estimation model. The cost model used in this research was based on the number of experimental runs, but a more sophisticated model that simultaneously includes set-up costs and factor combination costs (cell costs) would give more resolution to the expected cost estimate. Finally, it might be beneficial to

explore a simulation method to estimate the inaccurate model penalty rather than relying on the relative variance of the effect coefficients.

REFERENCES

Ai, M.-Y. and R.-C. Zhang (2004). "Projection Justification of Generalized Minimum Aberration for Asymmetrical Fractional Factorial Designs." Metrika **60**: 279-285.

Akaike, H. (1974). "A new look at the statistical model identification." IEEE Transactions on Automatic Control **19**(6): 716-723.

Angelopoulos, P., H. Evangelaras, et al. (2007). "An effective step-down algorithm for the construction and the identification of nonisomorphic orthogonal arrays." Metrika **66**(2): 139-149.

Belcher-Novosad, S. and D. Ingram (2003). "Identifying Minimum G Aberration Designs from Hadamard Matrices of Order 28." Journal of the Arkansas Academy of Science **57**: 202-205.

Booth, K. H. V. and D. R. Cox (1962). "Some systematic supersaturated designs." Technometrics **4**: 489–495.

Box, G. and J. Tyssedal (1996). "Projective properties of certain orthogonal arrays." Biometrika **83**(4): 950-955.

Box, G. E. P. and J. S. Hunter (1961). "The 2^{k-p} Fractional Factorial Designs, Part I." Technometrics **3**: 311-352.

Box, G. E. P. and J. S. Hunter (1961). "The 2^{k-p} Fractional Factorial Designs, Part II." Technometrics **3**: 449–458.

Bulutoglu, D. A. and C. S. Cheng (2003). "Hidden projection properties of some nonregular fractional factorial designs and their applications." Annals of Statistics **31**(3): 1012-1026.

Bulutoglu, D. A. and F. Margot (2008). "Classification of orthogonal arrays by integer programming." Journal of Statistical Planning and Inference **138**(3): 654-666.

Butler, N. A. (2003). "Some Theory for Constructing Minimum Aberration Fractional Factorial Designs." Biometrika **90**(1): 233-238.

Butler, N. A. (2004). "Minimum G2-Aberration Properties of Two-Level Foldover Designs." Statistics and Probability Letters **67**: 121-132.

Butler, N. A. (2005). "Generalised Minimum Aberration Construction Results for Symmetrical Orthogonal Arrays." Biometrika **92**(2): 485-491.

Candes, E. and T. Tao (2007). "The Dantzig Selector: Statistical Estimation When p is Much Larger than n." The Annals of Statistics **35**(6): 2313-2351.

Cheng, C.-S. (1980). "Orthogonal Arrays with Variable Numbers of Symbols." The Annals of Statistics **8**(2): 447-453.

Cheng, C.-S. (1997). "E(s2)-optimal supersaturated designs." Statistica Sinica **7**: 929-940.

Cheng, C.-S. (2006). Projection Properties of Factorial Designs for Factor Screening. Screening. A. Dean and S. Lewis, Springer New York**:** 156-168.

Cheng, C. S. (1995). "Some projection properties of orthogonal arrays." Annals of Statistics **23**(4): 1223-1233.

Cheng, C. S. (1998). "Some hidden projection properties of orthogonal arrays with strength three." Biometrika **85**(2): 491-495.

Cheng, S. W. and C. F. J. Wu (2001). "Factor screening and response surface exploration." Statistica Sinica **11**(3): 553-580.

Clemen, R. T. and T. Reilly (2001). Making Hard Decisions with Decision Tools. Belmont, CA, Duxbury Press.

Deng, L.-Y., Y. Li, et al. (2000). "Catalogue of small runs nonregular designs from hadamard matrices with generalized minimum aberration." Communications in Statistics - Theory and Methods **29**(5-6): 1379-1395.

Deng, L. Y. and B. X. Tang (1999). "Generalized resolution and minimum aberration criteria for Plackett-Burman and other nonregular factorial designs." Statistica Sinica **9**(4): 1071-1082.

Deng, L. Y. and B. X. Tang (2002). "Design selection and classification for Hadamard matrices using generalized minimum aberration criteria." Technometrics **44**(2): 173-184.

Efron, B., T. Hastie, et al. (2004). "Least Angle Regression." The Annals of Statistics **32**(2): 451-495.

Evangelaras, H., C. Koukouvinos, et al. (2007). "Further Contributions to Nonisomorphic Two Level Orthogonal Arrays." Journal of Statistical Planning and Inference **137**: 2080-2086.

Fang, K.-T., A. Zhang, et al. (2007). "An Effective Algorithm for Generation of Factorial Designs with Generalized Minimum Aberration." Journal of Complexity **23**: 740-751.

Fries, A. and W. G. Hunter (1980). "Minimum Aberration 2k-p Designs." Technometrics **22**(4): 601-608.

Hall, M. J. (1961). Hadamard matrix of order 16. Jet Propulsion Laboratory Research Summary**:** 21‑26.

Hamada, M. and C. F. J. Wu (1992). "Analyzing Designed Experiments with Complex Aliasing." Journal of Quality Technology **24**: 30-37.

Hedayat, A., N. J. A. Sloane, et al. (1999). Orthogonal arrays: theory and applications, Springer Verlag.

Hurvich, C. M. and C.-L. Tsai (1989). "Regression and Time Series Model Selection in Small Samples." Biometrika **76**: 297-330.

Ingram, D. (2000). The Construction of Generalized Minimum Aberration Designs by Efficient Algorithm. Ph.D. Dissertation, University of Memphis.

Ingram, D. and B. Tang (2005). "Minimum G Aberration Design Construction and Design Tables for 24 Runs." Journal of quality technology **37**(2): 101-114.

Johnson, M. E. and B. Jones (2010). "Classical Design Structure of Orthogonal Designs with Six to Eight Factors and Sixteen Runs." Quality and Reliability Engineering International **27**: 61-70.

Johnson, M. E. and C. J. Nachtsheim (1983). "Some Guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces." Technometrics **25**(3): 271-277.

Jones, B. and D. C. Montgomery (2010). "Alternatives to Resolution IV Screening Designs in 16 Runs." International Journal of Experimental Design and Process Optimisation **1**(4): 285-295.

Jones, B. and C. Nachtsheim (2011). "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects." Journal of Quality Technology **43**(1): 1-15.

Li, W. (2006). Screening Designs for Model Selection. Screening. A. Dean and S. Lewis, Springer New York**:** 207-234.

Li, W. and C. Nachtsheim (2000). "Model-Robust Factorial Designs." Technometrics **42**: 345-352.

Li, W. and C. F. J. Wu (1997). "Columnwise-Pairwise Algorithms with Applications to the Construction of Supersaturated Designs." Technometrics **39**(2): 171-179.

Li, X. A., N. Sudarsanam, et al. (2006). "Regularities in data from factorial experiments." Complexity **11**(5): 32-45.

Li, Y. F., L. Y. Deng, et al. (2004). "Design catalog based on minimum G-aberration." Journal of Statistical Planning and Inference **124**(1): 219-230.

Lin, C. D., R. R. Sitter, et al. (2012). "Creating catalogues of two-level nonregular fractional factorial designs based on the criteria of generalized aberration." Journal of Statistical Planning and Inference **142**(2): 445-456.

Lin, D. K. J. (1993). "A New Class of Supersaturated Designs." Technometrics **35**(1): 28-31.

Lin, D. K. J. (1995). "Generating Systematic Supersaturated Designs." Technometrics **37**(2): 213-225.

Lin, D. K. J. and N. P. Draper (1992). "Projection Properties of Plackett and Burman Designs." Technometrics **34**: 423-428.

Liu, M. and R. Zhang (2000). "Construction of E(s2) Optimal Supersaturated Designs Using Cyclic BIBDs." Journal of Statistical Planning and Inference **91**: 139-150.

Loeppky, J. L., R. R. Sitter, et al. (2007). "Nonregular Designs With Desirable Projection Properties." Technometrics **49**: 454‑466.

Lu, L., C. M. Anderson-Cook, et al. (2011). "Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier." Technometrics **53**(4): 353-365.

Ma, C.-X. and K.-T. Fang (2001). "A Note on Generalized Aberration in Factorial Designs." Metrika **53**(85-93): 85.

Marley, C. J. and D. C. Woods (2010). "A Comparison of Design and Model Selection Methods for Supersaturated Experiments." Computational Statistics and Data Analysis **54**: 3158-3167.

Montgomery, D. C. (2012). Design and Analysis of Experiments. Hoboken, New Jersey, Wiley.

Mount-Campbell, C. A. and J. B. Neuhardt (1980). "Selecting Cost-Optimal Main-Effect Fractions of 3^n Factorials." AIIE Transactions **12**: 80-86.

Mount-Campbell, C. A. and J. B. Neuhardt (1982). "Selection of Cost-Optimal Fractional Factorials, 3m-r 2n-s Series." TIMS/Studies in the Management Sciences: Optimization in Statistics **19**: 221-231.

Myers, R. H., D. C. Montgomery, et al. (2009). Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Hoboken, NJ, Wiley.

Neuhardt, J. B., H. E. Bradley, et al. (1973). "Computational Results in Selecting Multi-Factor Experimental Arrangements." Journal of the American Statistical Association **68**: 608-611.

Neuhardt, J. B., H. E. Bradley, et al. (1978). "A Dynamic Program for Designing Factorial Experiments with Budget Constraints." AIIE Transactions **10**(2): 125-130.

Neuhardt, J. B., H. E. Bradley, et al. (1978). "Selection of Cost-Optimal 2^{k-p} Fractional Factorials." Communications in Statistics - Simulation and Computation **7**(4): 369-383.

Nguyen, N.-K. (1996). "An Algorithmic Approach to Constructing Supersaturated Designs." Technometrics **38**(1): 69-73.

Park, Y., D. C. Montgomery, et al. (2006). "Cost-constrained G-efficient Response Surface Designs for Cuboidal Regions." Quality and Reliability Engineering International **22**(2): 121-139.

Pignatiello, J. J. (1985). "A Minimum Cost Approach for Finding Fractional Factorials." IIE Transactions **17**(3): 212-218.

Plackett, R. L. and J. P. Burman (1946). "The Design of Optimum Multifactorial Experiments." Biometrika **33**(4): 305-325.

Rafajlowicz, E. (1989). "Minimum Cost Experimental Design with a Prescribed Information Matrix." Theory of Probability and its Applications **34**(2): 367-370.

Rao, C. R. (1947). "Factorial Experiments Derivable from Combinatorial Arrangements of Arrays." Supplement to the Journal of the Royal Statistical Society **9**(1): 128-139.

Schoen, E. D. (2010). "Optimum Designs Versus Orthogonal Arrays for Main Effects and Two-Factor Interactions." Journal of quality technology **42**(2): 197-208.

Schoen, E. D., P. T. Eendebak, et al. (2010). "Complete Enumeration of Pure-Level and Mixed-Level Orthogonal Arrays." Journal of Combinatorial Designs **18**(6): 488-488.

Schoen, E. D. and R. W. Mee (2012). "Two-level designs of strength 3 and up to 48 runs." Journal of the Royal Statistical Society Series C-Applied Statistics **61**: 163-174.

Shinde, S. (2012). Analysis Methods for Nonregular Fractional Factorial Designs. Ph.D. Dissertation, Arizona State University.

Sloane, N. J. A. (2013). "A Library of Hadamard Matrices." Retrieved 10 May, 2013, from http://www2.research.att.com/~njas/hadamard/.

Smucker, B. J., E. del Castillo, et al. (2012). "Model-robust two-level designs using coordinate exchange algorithms and a maximin criterion." Technometrics **54**(4): 367-375.

Stufken, J. and B. Tang (2007). "Complete Enumeration of Two-Level Orthogonal Arrays of Strength d with d+2 Constraints." The Annals of Statistics **35**(2): 793-814.

Sun, D. and C. Wu (1993). J. Statistical Properties of Hadamard Matrices of Order 16//Kuo W. Quality Through Engineering Design, Elsevier Science Publishers.

Sun, D. X. (1993). Estimation Capacity and Related Topics in Experimental Design, University of Waterloo.

Sun, D. X., W. Li, et al. (2002). An Algorithm for Sequentially Constructing Non-Isomorphic Orthogonal Designs and Its Applications. Technical Report SUNYSB-AMS-02-13, State University of New York at Stony Brook, Dept. of Applied Mathematics and Statistics.

Tack, L. and M. Vandebroek (2001). "(Dt,C)-optimal run orders." Journal of Statistical Planning and Inference **98**(1–2): 293-310.

Tack, L. and M. Vandebroek (2002). "Trend-resistant and cost-efficient block designs with fixed or random block effects." Journal of quality technology **34**(4): 422-436.

Tack, L. and M. Vandebroek (2004). "Budget constrained run orders in optimum design." Journal of Statistical Planning and Inference **124**(1): 231-249.

Tack, L. and M. Vandebroek (2004). "Trend-resistant and cost-efficient cross-over designs for mixed models." Computational Statistics & Data Analysis **46**(4): 721-746.

Tang, B. and L.-Y. Deng (1999). "Minimum G2-Aberration for Nonregular Fractional Factorial Designs." The Annals of Statistics **27**(6): 1914-1926.

Tang, B. and L.-Y. Deng (2003). "Construction of Generalized Minimum Aberration Designs of 3,4 and 5 Factors." Journal of Statistical Planning and Inference **113**: 335-340.

Tang, B. and C. F. J. Wu (1997). "A Method for Constructing Supersaturated Designs and its Es^2 Optimality." The Canadian Journal of Statistics **25**(2): 191-201.

Tang, B. X. (2001). "Theory of J-characteristics for fractional factorial designs and projection justification of minimum G(2)-aberration." Biometrika **88**(2): 401-407.

Tang, B. X. and L. Y. Deng (1999). "Minimum G(2)-aberration for nonregular fractional factorial designs." Annals of Statistics **27**(6): 1914-1926.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) **58**(1): 267-288.

Tibshirani, R. (2011). "Regression Shrinkage and Selection Via the Lasso: A Retrospective." Journal of the Royal Statistical Society, Ser B **73**(3): 273-282.

Tyssedal, J. and O. Samset (1997). Analysis of the 12-run Plackett-Burman design, The Norwegian University of Science and Technology.

Wang, J. C. and C. F. J. Wu (1995). "A Hidden Projection Property of Plackett-Burman and Related Designs." Statistica Sinica **5**: 235-250.

Wu, C. F. J. (1993). "Construction of Supersaturated Designs Through Partially Aliased Interactions." Biometrika **80**(3): 661-669.

Xu, H. (2003). "Minimum moment aberration for nonregular designs and supersaturated designs." Statistica Sinica **13**: 691-708.

Xu, H. and C. F. J. Wu (2001). "Generalized Minimum Aberration for Assymetrical Fractional Factorial Designs." The Annals of Statistics **29**(2): 549-560.

Xu, H. Q. and L. Y. Deng (2005). "Moment aberration projection for nonregular fractional factorial designs." Technometrics **47**(2): 121-131.

Zhang, Y., W. Li, et al. (2011). "Orthogonal Arrays Obtained by Generalized Difference Matrices with g Levels." Science China Mathematics **54**(1): 133-143.

Zhao, P. and B. Yu (2006). "On Model Selection Consistency of Lasso." Journal of Machine Learning Research **7**: 2541-2563.

APPENDIX A

20-RUN NO-CONFOUNDING DESIGNS

The 20-run No-confounding Design Matrices

Table 75. OA(20;2^6;2) #74

A	B	C	D	E	F
-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1
-1	-1	-1	1	1	-1
-1	-1	1	-1	1	-1
-1	-1	1	1	-1	1
-1	1	-1	-1	1	-1
-1	1	-1	1	-1	1
-1	1	1	-1	1	1
-1	1	1	1	-1	-1
-1	1	1	1	1	1
1	-1	-1	-1	1	1
1	-1	-1	1	1	-1
1	-1	1	-1	-1	1
1	-1	1	1	-1	-1
1	-1	1	1	1	1
1	1	-1	-1	-1	1
1	1	-1	1	-1	-1
1	1	-1	1	1	1
1	1	1	-1	-1	-1
1	1	1	-1	1	-1

Table 76. OA(20;2^7;2) #457

A	B	C	D	E	F	G
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	1
-1	-1	-1	-1	1	1	-1
-1	-1	1	1	-1	-1	-1
-1	-1	1	1	1	1	1
-1	1	-1	1	-1	1	1
-1	1	-1	1	1	-1	-1
-1	1	1	-1	-1	1	1
-1	1	1	-1	1	-1	1
-1	1	1	1	1	1	-1
1	-1	-1	1	-1	1	1
1	-1	-1	1	1	-1	1
1	-1	1	-1	-1	1	-1
1	-1	1	-1	1	1	1
1	-1	1	1	1	-1	-1
1	1	-1	-1	1	-1	1
1	1	-1	-1	1	1	-1
1	1	-1	1	-1	1	-1
1	1	1	-1	-1	-1	-1
1	1	1	1	-1	-1	1

Table 77. OA(20;2^8;2) #1599

A	B	C	D	E	F	G	H
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	1	1
-1	-1	-1	-1	1	1	-1	1
-1	-1	1	1	-1	-1	-1	-1
-1	-1	1	1	1	1	1	1
-1	1	-1	1	-1	1	1	-1
-1	1	-1	1	1	-1	-1	1
-1	1	1	-1	-1	1	1	1
-1	1	1	-1	1	-1	1	-1
-1	1	1	1	1	1	-1	-1
1	-1	-1	1	-1	1	1	-1
1	-1	-1	1	1	-1	1	-1
1	-1	1	-1	-1	1	-1	1
1	-1	1	-1	1	1	1	-1
1	-1	1	1	1	-1	-1	1
1	1	-1	-1	1	-1	1	1
1	1	-1	-1	1	1	-1	-1
1	1	-1	1	-1	1	-1	1
1	1	1	-1	-1	-1	-1	-1
1	1	1	1	-1	-1	1	1

Table 78. OA(20;2^9;2) #2286

A	B	C	D	E	F	G	H	I
-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	1	1	1
-1	-1	-1	-1	1	1	-1	-1	1
-1	-1	-1	1	-1	1	1	1	-1
-1	-1	1	1	1	-1	-1	1	-1
-1	1	-1	1	1	-1	1	-1	1
-1	1	1	-1	-1	1	-1	1	1
-1	1	1	-1	1	-1	1	-1	-1
-1	1	1	1	-1	1	-1	-1	-1
-1	1	1	1	1	1	1	1	1
1	-1	-1	1	1	1	1	-1	-1
1	-1	1	-1	-1	1	1	-1	1
1	-1	1	-1	1	1	-1	1	-1
1	-1	1	1	-1	-1	1	1	1
1	-1	1	1	1	-1	-1	-1	1
1	1	-1	-1	1	-1	-1	1	1
1	1	-1	-1	1	1	1	1	-1
1	1	-1	1	-1	-1	-1	1	-1
1	1	-1	1	-1	1	-1	-1	1
1	1	1	-1	-1	-1	1	-1	-1

Table 79. OA(20;2^{10};2) #2389

A	B	C	D	E	F	G	H	I	J
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1	1
-1	-1	-1	-1	1	1	1	1	1	-1
-1	-1	1	1	-1	-1	1	1	1	1
-1	-1	1	1	1	1	-1	-1	1	-1
-1	1	-1	1	-1	1	1	1	-1	-1
-1	1	-1	1	1	-1	-1	1	1	1
-1	1	1	-1	-1	1	1	-1	1	1
-1	1	1	-1	1	1	-1	1	-1	1
-1	1	1	1	1	-1	1	-1	-1	-1
1	-1	-1	1	-1	1	-1	1	1	1
1	-1	-1	1	1	1	1	-1	-1	1
1	-1	1	-1	-1	1	1	1	-1	-1
1	-1	1	-1	1	-1	1	-1	1	1
1	-1	1	1	1	-1	-1	1	-1	-1
1	1	-1	-1	1	-1	1	1	-1	1
1	1	-1	-1	1	1	-1	-1	1	-1
1	1	-1	1	-1	-1	1	-1	1	-1
1	1	1	-1	-1	-1	-1	1	1	-1
1	1	1	1	-1	1	-1	-1	-1	1

Table 80. OA(20;2^{11};2) #713

A	B	C	D	E	F	G	H	I	J	K
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	1	1	1	1
-1	-1	-1	-1	1	1	1	-1	-1	1	1
-1	-1	-1	1	-1	1	1	1	1	-1	-1
-1	-1	1	1	1	-1	-1	-1	1	-1	1
-1	1	-1	1	1	-1	-1	1	-1	1	-1
-1	1	1	-1	-1	1	1	-1	1	-1	1
-1	1	1	-1	-1	1	1	1	-1	1	-1
-1	1	1	1	1	-1	1	-1	-1	-1	-1
-1	1	1	1	1	1	-1	1	1	1	1
1	-1	-1	1	1	1	1	-1	-1	1	1
1	-1	1	-1	1	-1	1	1	1	1	-1
1	-1	1	-1	1	1	-1	1	-1	-1	-1
1	-1	1	1	-1	-1	1	1	-1	-1	1
1	-1	1	1	-1	1	-1	-1	1	1	-1
1	1	-1	-1	1	-1	1	1	1	-1	1
1	1	-1	-1	1	1	-1	-1	1	-1	-1
1	1	-1	1	-1	-1	1	-1	1	1	-1
1	1	-1	1	-1	1	-1	1	-1	-1	1
1	1	1	-1	-1	-1	-1	-1	-1	1	1

Table 81. OA(20;2^{12};2) #426

A	B	C	D	E	F	G	H	I	J	K	L
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1
-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1
-1	-1	-1	1	-1	1	1	-1	1	1	-1	-1
-1	-1	1	1	1	-1	1	1	-1	1	-1	1
-1	1	-1	1	1	-1	1	1	1	-1	1	-1
-1	1	1	-1	-1	1	-1	1	1	-1	-1	1
-1	1	1	-1	-1	1	1	1	-1	1	1	-1
-1	1	1	1	1	-1	-1	-1	1	1	1	1
-1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
1	-1	-1	1	1	1	-1	1	1	-1	-1	1
1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1
1	-1	1	-1	1	-1	1	-1	1	1	-1	-1
1	-1	1	1	-1	1	-1	-1	-1	1	1	1
1	-1	1	1	-1	1	1	1	1	-1	1	-1
1	1	-1	-1	1	1	-1	-1	1	1	1	-1
1	1	-1	-1	1	1	1	1	-1	1	-1	1
1	1	-1	1	-1	-1	-1	1	-1	1	-1	-1
1	1	-1	1	-1	-1	1	-1	-1	-1	1	1
1	1	1	-1	-1	-1	1	-1	1	-1	-1	1

APPENDIX B

24-RUN NO-CONFOUNDING DESIGNS

The 24-run No-confounding Design Matrices

Table 82. 6-Factor 24-run No-confounding Design (G-Aber_24 H.1.6.1)

A	B	C	D	E	F
1	1	1	1	1	1
1	1	1	1	-1	-1
1	-1	-1	-1	-1	-1
-1	1	1	-1	-1	-1
-1	1	-1	-1	1	1
-1	-1	-1	1	1	-1
1	1	-1	-1	1	-1
-1	-1	1	-1	-1	1
-1	-1	1	1	1	-1
1	-1	1	-1	1	1
-1	1	-1	1	-1	1
1	-1	-1	1	-1	1
-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	1	1
-1	1	1	1	1	1
1	-1	-1	1	1	1
1	-1	1	1	-1	-1
1	1	1	-1	-1	1
-1	-1	1	1	-1	1
1	1	-1	1	1	-1
1	1	-1	-1	-1	1
-1	1	-1	1	-1	-1
1	-1	1	-1	1	-1
-1	1	1	-1	1	-1

Table 83. 7-Factor 24-run No-confounding Design (G-Aber_24 H.1.7.1)

A	B	C	D	E	F	G
1	1	1	1	1	1	1
1	-1	-1	-1	-1	-1	-1
-1	1	1	1	-1	-1	-1
1	1	1	-1	1	-1	-1
-1	1	-1	-1	-1	1	1
-1	-1	1	-1	1	1	-1
-1	-1	-1	1	1	1	-1
1	-1	-1	1	1	-1	1
1	1	-1	1	-1	1	-1
1	-1	1	-1	-1	1	1
-1	-1	1	1	-1	-1	1
-1	1	-1	-1	1	-1	1
-1	-1	-1	-1	-1	-1	-1
-1	1	1	1	1	1	1
1	-1	-1	-1	1	1	1
-1	-1	-1	1	-1	1	1
1	-1	1	1	1	-1	-1
1	1	-1	1	-1	-1	1
1	1	1	-1	-1	-1	1
-1	1	1	-1	-1	1	-1
-1	-1	1	-1	1	-1	1
-1	1	-1	1	1	-1	-1
1	1	-1	-1	1	1	-1
1	-1	1	1	-1	1	-1

Table 84. 8-Factor 24-run No-confounding Design (G-Aber_24 H.1.8.1)

A	B	C	D	E	F	G	H
1	1	1	1	1	1	1	1
1	1	1	1	1	-1	-1	-1
1	1	1	-1	-1	-1	-1	-1
1	-1	-1	1	-1	1	-1	-1
1	1	-1	1	-1	-1	1	1
1	1	-1	-1	1	1	1	-1
1	-1	1	1	-1	1	1	-1
1	1	-1	-1	-1	1	-1	1
1	-1	-1	-1	1	-1	1	-1
1	-1	1	-1	-1	-1	1	1
1	-1	-1	1	1	-1	-1	1
1	-1	1	-1	1	1	-1	1
-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1	1	1
-1	-1	-1	1	1	1	1	1
-1	1	1	-1	1	-1	1	1
-1	-1	1	-1	1	1	-1	-1
-1	-1	1	1	-1	-1	-1	1
-1	1	-1	-1	1	-1	-1	1
-1	-1	1	1	1	-1	1	-1
-1	1	1	1	-1	1	-1	1
-1	1	-1	1	1	1	-1	-1
-1	1	1	-1	-1	1	1	-1
-1	1	-1	1	-1	-1	1	-1

Table 85. 9-Factor 24-run No-confounding Design (G-Aber_24 H.1.9.1)

A	B	C	D	E	F	G	H	I
1	1	1	1	1	1	1	1	1
1	1	1	1	1	-1	-1	-1	-1
1	1	1	-1	-1	1	1	-1	-1
1	-1	-1	1	1	1	1	-1	-1
1	1	-1	1	-1	1	-1	1	1
1	1	-1	-1	-1	-1	1	1	-1
1	-1	1	1	-1	-1	-1	1	-1
1	1	-1	-1	1	-1	-1	-1	1
1	-1	-1	-1	1	1	-1	1	-1
1	-1	1	-1	1	-1	1	1	1
1	-1	-1	1	-1	-1	1	-1	1
1	-1	1	-1	-1	1	-1	-1	1
-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1	1	1	1
-1	-1	-1	1	1	-1	-1	1	1
-1	1	1	-1	-1	-1	-1	1	1
-1	-1	1	-1	1	-1	1	-1	-1
-1	-1	1	1	1	1	-1	-1	1
-1	1	-1	-1	1	1	1	-1	1
-1	-1	1	1	-1	1	1	1	-1
-1	1	1	1	-1	-1	1	-1	1
-1	1	-1	1	-1	1	-1	-1	-1
-1	1	1	-1	1	1	-1	1	-1
-1	1	-1	1	1	-1	1	1	-1

Table 86. 10-Factor 24-run No-confounding Design (G-Aber_24 H.1.10.1)

A	B	C	D	E	F	G	H	I	J
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	-1	-1	-1	-1	-1
1	1	-1	-1	-1	1	1	-1	-1	-1
-1	-1	1	1	-1	1	-1	1	-1	-1
1	-1	1	-1	-1	1	-1	-1	1	1
1	-1	-1	-1	1	-1	-1	1	1	-1
-1	1	1	-1	-1	-1	1	1	1	-1
1	-1	-1	1	-1	-1	1	1	-1	1
-1	-1	-1	1	1	1	1	-1	1	-1
-1	1	-1	1	-1	-1	-1	-1	1	1
-1	-1	1	-1	1	-1	1	-1	-1	1
-1	1	-1	-1	1	1	-1	1	-1	1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1	1	1	1	1
-1	-1	1	1	1	-1	-1	1	1	1
1	1	-1	-1	1	-1	1	-1	1	1
-1	1	-1	1	1	-1	1	1	-1	-1
-1	1	1	1	-1	1	1	-1	-1	1
1	-1	-1	1	1	1	-1	-1	-1	1
-1	1	1	-1	1	1	-1	-1	1	-1
1	1	1	-1	-1	-1	-1	1	-1	1
1	-1	1	-1	1	1	1	1	-1	-1
1	1	-1	1	-1	1	-1	1	1	-1
1	-1	1	1	-1	-1	1	-1	1	-1

Table 87. 11-Factor 24-run No-confounding Design (G-Aber_24 H.1.11.1)

A	B	C	D	E	F	G	H	I	J	K
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	-1	-1	-1	-1	-1	-1
1	1	1	-1	-1	1	1	1	-1	-1	-1
1	-1	-1	1	-1	1	1	-1	1	-1	-1
1	1	-1	-1	-1	1	-1	-1	-1	1	1
1	1	-1	-1	1	-1	1	-1	1	1	-1
1	-1	1	-1	-1	-1	-1	1	1	1	-1
1	1	-1	1	-1	-1	-1	1	1	-1	1
1	-1	-1	1	1	1	-1	1	-1	1	-1
1	-1	1	1	-1	-1	1	-1	-1	1	1
1	-1	-1	-1	1	-1	1	1	-1	-1	1
1	-1	1	-1	1	1	-1	-1	1	-1	1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	1	1	1	1	1	1
-1	-1	-1	1	1	-1	-1	-1	1	1	1
-1	1	1	-1	1	-1	-1	1	-1	1	1
-1	-1	1	1	1	-1	1	1	1	-1	-1
-1	-1	1	1	-1	1	-1	1	-1	-1	1
-1	1	-1	1	1	1	1	-1	-1	-1	1
-1	-1	1	-1	1	1	1	-1	-1	1	-1
-1	1	1	-1	-1	-1	1	-1	1	-1	1
-1	1	-1	-1	1	1	-1	1	1	-1	-1
-1	1	1	1	-1	1	-1	-1	1	1	-1
-1	1	-1	1	-1	-1	1	1	-1	1	-1

Table 88. 12-Factor 24-run No-confounding Design (G-Aber_24 H.1.12.1)

A	B	C	D	E	F	G	H	I	J	K	L
1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
1	1	1	-1	-1	-1	1	1	1	-1	-1	-1
1	-1	-1	1	1	-1	1	1	-1	1	-1	-1
1	1	-1	1	-1	-1	1	-1	-1	-1	1	1
1	1	-1	-1	-1	1	-1	1	-1	1	1	-1
1	-1	1	1	-1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	1	-1	-1	-1	1	1	-1	1
1	-1	-1	-1	1	1	1	-1	1	-1	1	-1
1	-1	1	-1	1	-1	-1	1	-1	-1	1	1
1	-1	-1	1	-1	1	-1	1	1	-1	-1	1
1	-1	1	-1	-1	1	1	-1	-1	1	-1	1
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	1	1	1	1	1	1
-1	-1	-1	1	1	1	-1	-1	-1	1	1	1
-1	1	1	-1	-1	1	-1	-1	1	-1	1	1
-1	-1	1	-1	1	1	-1	1	1	1	-1	-1
-1	-1	1	1	1	-1	1	-1	1	-1	-1	1
-1	1	-1	-1	1	1	1	1	-1	-1	-1	1
-1	-1	1	1	-1	1	1	1	-1	-1	1	-1
-1	1	1	1	-1	-1	-1	1	-1	1	-1	1
-1	1	-1	1	-1	1	1	-1	1	1	-1	-1
-1	1	1	-1	1	-1	1	-1	-1	1	1	-1
-1	1	-1	1	1	-1	-1	1	1	-1	1	-1

Table 89. Alternate 6-Factor 24-run No-confounding Design (CEA_24 6.2009)

A	B	C	D	E	F
-1	-1	-1	1	1	-1
-1	-1	-1	-1	1	1
-1	-1	-1	1	-1	-1
-1	-1	1	1	1	1
-1	-1	1	-1	1	-1
-1	-1	1	-1	-1	-1
-1	1	-1	1	-1	1
-1	1	-1	-1	-1	-1
-1	1	-1	1	1	1
-1	1	1	-1	-1	1
-1	1	1	-1	1	1
-1	1	1	1	-1	-1
1	-1	-1	-1	-1	-1
1	-1	-1	-1	1	-1
1	-1	-1	1	1	1
1	-1	1	1	-1	-1
1	-1	1	-1	-1	1
1	-1	1	-1	1	1
1	1	-1	-1	-1	1
1	1	-1	1	-1	-1
1	1	-1	1	1	-1
1	1	1	1	1	1
1	1	1	1	-1	1
1	1	1	-1	1	-1

Table 90. Alternate 7-Factor 24-run No-confounding Design (MinMax_24 7.1)

A	B	C	D	E	F	G
-1	-1	-1	1	-1	-1	-1
-1	-1	-1	1	1	1	1
-1	-1	-1	-1	-1	-1	1
-1	-1	1	-1	-1	1	-1
-1	-1	1	1	1	1	-1
-1	-1	1	-1	1	-1	1
-1	1	-1	-1	1	-1	-1
-1	1	-1	1	1	1	-1
-1	1	-1	-1	-1	1	1
-1	1	1	1	-1	1	1
-1	1	1	1	1	-1	1
-1	1	1	-1	-1	-1	-1
1	-1	-1	-1	1	1	1
1	-1	-1	1	1	-1	-1
1	-1	-1	1	-1	1	-1
1	-1	1	-1	1	-1	-1
1	-1	1	1	-1	-1	1
1	-1	1	-1	-1	1	1
1	1	-1	-1	-1	1	-1
1	1	-1	1	-1	-1	1
1	1	-1	-1	1	-1	1
1	1	1	-1	1	1	-1
1	1	1	1	-1	-1	-1
1	1	1	1	1	1	1

Table 91. Alternate 8-Factor 24-run No-confounding Design (CEA_24 8.3002)

A	B	C	D	E	F	G	H
-1	-1	-1	1	1	1	-1	1
-1	-1	-1	1	1	-1	1	1
-1	-1	-1	-1	-1	1	-1	-1
-1	-1	1	-1	-1	1	1	1
-1	-1	1	-1	1	-1	-1	1
-1	-1	1	1	-1	-1	1	-1
-1	1	-1	-1	-1	-1	-1	-1
-1	1	-1	-1	-1	1	1	1
-1	1	-1	1	1	-1	-1	-1
-1	1	1	1	1	-1	1	1
-1	1	1	1	-1	1	-1	-1
-1	1	1	-1	1	1	1	-1
1	-1	-1	-1	1	1	1	1
1	-1	-1	-1	-1	-1	-1	1
1	-1	-1	1	-1	1	1	-1
1	-1	1	-1	1	-1	-1	-1
1	-1	1	1	1	1	1	-1
1	-1	1	1	-1	1	-1	1
1	1	-1	-1	1	1	-1	-1
1	1	-1	-1	1	-1	1	-1
1	1	-1	1	-1	-1	1	1
1	1	1	-1	-1	-1	1	1
1	1	1	1	-1	-1	-1	-1
1	1	1	1	1	1	-1	1

Table 92. Alternate 9-Factor 24-run No-confounding Design (MinMax_24 9.1)

A	B	C	D	E	F	G	H	I
-1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	1	-1	-1	1	1	1
-1	-1	-1	-1	1	1	-1	1	1
-1	-1	1	1	-1	1	-1	1	-1
-1	-1	1	-1	1	1	1	-1	-1
-1	-1	1	1	1	-1	1	-1	1
-1	1	-1	-1	-1	1	1	-1	-1
-1	1	-1	1	1	-1	1	1	-1
-1	1	-1	1	1	1	-1	-1	1
-1	1	1	1	-1	-1	-1	-1	1
-1	1	1	-1	-1	1	1	1	1
-1	1	1	-1	1	-1	-1	1	-1
1	-1	-1	-1	-1	1	1	-1	1
1	-1	-1	1	1	1	1	1	-1
1	-1	-1	1	1	-1	-1	-1	-1
1	-1	1	-1	-1	-1	1	1	-1
1	-1	1	1	-1	1	-1	-1	1
1	-1	1	-1	1	-1	-1	1	1
1	1	-1	1	-1	1	-1	1	-1
1	1	-1	-1	-1	-1	-1	1	1
1	1	-1	-1	1	-1	1	-1	1
1	1	1	1	-1	-1	1	-1	-1
1	1	1	-1	1	1	-1	-1	-1
1	1	1	1	1	1	1	1	1

Table 93. Alternate 10-Factor 24-run No-confounding Design (CEA_24 10.108)

A	B	C	D	E	F	G	H	I	J
-1	-1	-1	1	-1	-1	-1	-1	-1	-1
-1	-1	-1	1	1	-1	1	1	1	-1
-1	-1	-1	-1	-1	1	1	1	-1	1
-1	-1	1	1	1	1	1	-1	-1	1
-1	-1	1	-1	-1	-1	-1	-1	1	1
-1	-1	1	-1	1	1	-1	1	1	-1
-1	1	-1	1	1	1	-1	-1	1	1
-1	1	-1	-1	-1	1	1	-1	1	-1
-1	1	-1	-1	1	-1	-1	1	-1	1
-1	1	1	1	-1	1	-1	1	-1	-1
-1	1	1	1	-1	-1	1	1	1	1
-1	1	1	-1	1	-1	1	-1	-1	-1
1	-1	-1	1	-1	1	-1	1	1	1
1	-1	-1	-1	1	-1	1	-1	1	1
1	-1	-1	-1	1	1	-1	-1	-1	-1
1	-1	1	1	-1	1	1	-1	1	-1
1	-1	1	1	1	-1	-1	1	-1	1
1	-1	1	-1	-1	-1	1	1	-1	-1
1	1	-1	-1	-1	-1	1	-1	-1	1
1	1	-1	-1	1	1	1	1	-1	-1
1	1	-1	1	-1	-1	-1	1	1	-1
1	1	1	-1	1	-1	-1	-1	1	-1
1	1	1	1	-1	1	-1	-1	-1	1
1	1	1	1	1	1	1	1	1	1

Table 94. Alternate 11-Factor 24-run No-confounding Design (CEA_24 11.1016)

A	B	C	D	E	F	G	H	I	J	K
-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	1	1	-1	-1	1	1	1	1
-1	-1	-1	1	1	1	1	1	-1	-1	-1
-1	-1	1	-1	-1	1	-1	-1	1	-1	1
-1	-1	1	-1	-1	1	1	1	1	1	-1
-1	-1	1	-1	1	-1	1	-1	-1	1	1
-1	1	-1	-1	-1	1	-1	1	-1	1	1
-1	1	-1	-1	-1	-1	1	-1	1	1	-1
-1	1	-1	1	1	1	1	-1	1	-1	1
-1	1	1	-1	1	-1	-1	1	1	-1	-1
-1	1	1	1	1	1	-1	-1	-1	1	-1
-1	1	1	1	-1	-1	1	1	-1	-1	1
1	-1	-1	1	1	1	-1	-1	1	1	-1
1	-1	-1	-1	-1	1	1	-1	-1	1	1
1	-1	-1	1	-1	-1	1	1	1	-1	1
1	-1	1	-1	-1	-1	-1	1	-1	1	-1
1	-1	1	1	1	1	-1	1	-1	-1	1
1	-1	1	-1	1	-1	1	-1	1	-1	-1
1	1	-1	-1	-1	1	-1	1	1	-1	-1
1	1	-1	-1	1	-1	-1	-1	-1	-1	1
1	1	-1	1	1	-1	1	1	-1	1	-1
1	1	1	1	-1	-1	-1	-1	1	1	1
1	1	1	1	-1	1	1	-1	-1	-1	-1
1	1	1	-1	1	1	1	1	1	1	1

Table 95. Alternate 12-Factor 24-run No-confounding Design (CEA_24 12.1033)

A	B	C	D	E	F	G	H	I	J	K	L
-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	1	1	1	1	-1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1
-1	-1	1	-1	1	1	-1	-1	-1	1	1	1
-1	-1	1	1	1	-1	1	1	1	-1	-1	1
-1	-1	1	1	-1	1	1	-1	1	1	-1	-1
-1	1	-1	-1	-1	1	1	1	-1	1	-1	1
-1	1	-1	-1	1	-1	1	-1	1	1	1	-1
-1	1	-1	1	1	1	-1	-1	1	-1	-1	1
-1	1	1	1	1	-1	-1	1	-1	1	-1	-1
-1	1	1	1	-1	-1	1	-1	-1	-1	1	1
-1	1	1	-1	-1	1	-1	1	1	-1	1	-1
1	-1	-1	1	1	-1	1	-1	-1	1	-1	1
1	-1	-1	-1	-1	1	1	-1	1	-1	1	1
1	-1	-1	-1	1	1	-1	1	1	1	-1	-1
1	-1	1	-1	-1	-1	1	1	-1	1	1	-1
1	-1	1	1	-1	1	-1	1	-1	-1	-1	1
1	-1	1	1	1	-1	-1	-1	1	-1	1	-1
1	1	-1	1	1	-1	-1	1	-1	-1	1	1
1	1	-1	1	-1	1	-1	-1	-1	1	1	-1
1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1
1	1	1	-1	1	1	1	-1	-1	-1	-1	-1
1	1	1	-1	-1	-1	-1	-1	1	1	-1	1
1	1	1	1	1	1	1	1	1	1	1	1

Calculation of the Number of Possible Structured Column Combinations

Divide the 24 rows of a column c into two sub-columns of 12 rows and call them c_1 and c_2. The values in each of these sub-columns are independent of each other as long as both sub-columns contain 6 +1's, so without loss of generality, consider c1. c_1 can be divided into two sub-columns of six, c_{11} and c_{12}. These sub-columns can be of three types: Type I has three +1's; Type II has 2 +1's; Type III has 4 +1's. There are $\binom{6}{3} = 20$ possible Type I columns. There are $\binom{6}{2} = 15$ possible Type II columns and $\binom{6}{4} = 15$ possible Type III columns.

If c_{11} is Type I, then c_{12} is Type I and there are $20^2 = 400$ possible c_1 columns using Type I sub-columns. If c_{11} is not Type I, then it can be Type II or Type III, so there are 15+15 = 30 possibilities for c_{11}. There are only 15 possibilities for c_{12}, since it must be type II if c_{11} is type III or vice versa so that there are a total of 6 +1's in c_1. Therefore there are $30 * 15 = 450$ possible c_1 columns using Type II and Type III sub-columns. There are therefore $450 + 400 = 850$ possible combinations for c_1. Since c_1 is independent of c_2, there are $850^2 = 722{,}500$ possible columns for c.

BIOGRAPHICAL SKETCH

Captain Brian Benedict Stone was born in St. Louis, Missouri. He completed his secondary education at St. Louis University High School. Brian earned his Bachelor's degree in Mathematics at Truman State University in Kirksville, Missouri. Upon graduation, he was employed as a telecom provisioner at MCI WorldCom for three years. In May of 2003, Brian attended Air Force Officer Training School and was commissioned as a Second Lieutenant on August 8, 2003. After serving three years as a precision-guided munitions analyst, Lt Stone attended the Air Force Institute of Technology (AFIT) and earned a Master of Science degree in Operations Research. While at AFIT he was promoted to the rank of Captain. He then was assigned to the Pentagon where he served as a current operations analyst, manpower analyst, and executive officer to the Director of Studies and Analyses, Assessments and Lessons Learned. After serving almost three years at the Pentagon, Capt Stone enrolled in the Industrial Engineering PhD program at the Ira A. Fulton Schools of Engineering, Arizona State University. During his studies at ASU, he served as Vice President of the ASU INFORMS Student Chapter. In addition, he earned a certificate in statistics and completed the Six Sigma Black Belt program. Captain Stone will be assigned as an Assistant Professor of Operations Research at AFIT upon completion of his PhD program. His main research interests include design of experiments and response surface methodology.