Practical comparison of traditional and definitive screening designs in chemical process development

Rebecca Olsen, John Lawson*,
Nathaniel Rohbock and Brian Woodfield

Department of Chemistry and Statistics,
Brigham Young University,
Provo, UT 84602, USA
Email: olsen29@gmail.com
Email: lawson@byu.edu
Email: nathaniel.rohbock@gmail.com
Email: brian_woodfield@chem.byu.edu
*Corresponding author

Abstract: Traditional Screening Designs, such as resolution III 2^{k-p} fractional factorials are used routinely in the initial stages of process development. These designs are used to determine which process variables have the largest effect on process outcomes. Once a screening design is complete and the data are analysed, follow-up experiments are normally required in order to develop useful prediction equations involving the important variables and to identify the optimal process operating conditions. Recently developed definitive screening designs allow researchers to identify important variables and optimum process conditions after one set of experiments, eliminating the need for follow-up experiments. This leads to the question: What is now the role of traditional and definitive screening designs in process optimisation? We share our insights gained from using both of these designs in developing a process to produce catalyst support material to shed light on these questions.

Keywords: catalyst support; definitive screening design; resolution III; screening design.

Reference to this paper should be made as follows: Olsen, R., Lawson, J., Rohbock, N. and Woodfield, B. (2016) 'Practical comparison of traditional and definitive screening designs in chemical process development', *Int. J. Experimental Design and Process Optimisation*, Vol. 5, Nos. 1/2, pp.1–22.

Biographical notes: Rebecca Olsen is currently working at Brigham Young University as an Adjunct Faculty in the Chemistry and Biochemistry Department. She holds a BSc in Chemistry from Brigham Young University and a PhD in Physical Chemistry from Brigham Young University. Her research interests include nanosynthesis, nanocharacterisation, oxides, catalysis, process optimisation, and design of experiments.

John Lawson is a Professor in the Department of Statistics at Brigham Young University (BYU) in Provo, Utah. He holds a BS from BYU and an MS from Rutgers University in Statistics, and a PhD in Applied Statistics from the Polytechnic Institute of New York. His interests are in experimental design, process improvement, and statistical methods for improvement of quality and reliability.

Nathaniel Rohbok received Bachelor's from the University of Utah in Mathematics, and Master's in Statistics from BYU. After starting his career working for BioStat Solutions Inc. as a Statistical Analyst, a desire to work in quality improvement, design of experiments, and statistical process control led him to his current position as a Quality Engineer (ASQ CQE) at Myriad Genetics Laboratories in SLC, Utah.

Brian Woodfield is a Professor of Chemistry and Biochemistry at Brigham Young University (BYU) in Provo, Utah. He holds a BS and MS in Chemistry from BYU and a PhD in Physical Chemistry from the University of Californiaat Berkeley. His research interests are in the fields of low-temperature physics, chemical thermodynamics, nanomaterials, and catalysts and catalyst supports.

1 Introduction

Good manufacturing processes should rely on a sound understanding of the impact of various process inputs upon the efficiency of process operation and the critical characteristics of the resulting product. For example, the US Food and Drug Administration's Current Good Manufacturing Practices (CGMPs) initiative for the 21st century emphasises this approach (US Food and Drug Administration, 2004, 2011). This is normally done in a two-step process where an initial screening design is used to identify important process inputs, and further follow-up experiments are conducted with the important factors (according to a response surface design) to define the optimal operating conditions (Abu-Absi et al., 2010).

Box (1999) previously elucidated the role of experimental designs and response surface methods in a process of sequential learning to innovate or design a manufacturing process. This process usually starts with a two-level screening design followed by a factorial design with the important factors in order to estimate two-factor interactions. The method of steepest ascent can be used, if needed, to move to a new location in the factor space. Finally, follow-up experiments can augment the two-level factorial design and expand it to a central composite design that will provide data to estimate the full quadratic model and identify optimum operating conditions. Normally, it is recommended that no more than 25% of the research budget be allocated to the initial screening experiments since it may be discovered that inappropriate factor levels were chosen or that one or more factors, later recognised as important, were not included in the design. For this reason, very efficient two-level resolution III fractional factorial or Plackett-Burman designs have been traditionally recommended for initial screening experiments.

Recently developed three-level definitive screening designs (Jones and Nachtsheim, 2011, 2013; Xiao et al. 2012) provide main effects' estimates that are unbiased by quadratic effects and linear by linear interaction effects. In addition, no two linear-by-linear interactions are completely confounded. These designs are very efficient. They require only one run more than twice the number of factors in the design, and when there

are at least six factors in the design, they permit efficient estimation of a full quadratic response surface in any subset of three factors or less. For this reason, developers of these designs suggest that their use may eliminate the need for follow-up experiments and that one design can be used for both screening and response surface exploration. Since there is no defining relation, the alias pattern for a definitive screening design, which includes partial confounding among two-factor interactions and among two-factor interactions and quadratic terms, cannot be listed simply like the alias pattern for a regular fractional factorial design. It can be best visualised by looking at the colourmap of the correlations computed from the design matrix that includes main effects, quadratic effects, and two-factor interactions (Jones and Nachtsheim, 2011). These designs are only described briefly (if at all) in the recent reference books on experimental design.

Data from definitive screening designs cannot be analysed using a saturated set of orthogonal contrasts (the normal practice with regular fractional factorial designs). The interaction effects are partially confounded with quadratic effects and other interaction effects. The data can be analysed using regression subset procedures similar to those used to detect interactions with Plackett-Burman designs or orthogonal arrays (Hamada and Wu, 1992; Chipman, Hamada and Wu, 1997; Lin, 1999; Lawson, 2002). Jones and Nachtsheim (2011) suggested a simple procedure using a forward stepwise regression that enforces effect heredity (Hamada and Wu, 1992). They recommend all terms in the quadratic response surface model be used as candidate terms. To enforce effect heredity, they modify the forward stepwise regression in the following way. If an interaction or quadratic effect is the next term to enter the model, then the main effect(s) that define that quadratic or interaction term should also be added to the model (if they are not already in the model). This modified forward stepwise procedure is available using the *combine* option in the forward stepwise regression in JMP or with the fhstep function in the R package daewr (Lawson, 2015a). Some may argue that the stopping rules used in forward stepwise regression are a little more subjective than graphically examining a set of orthogonal contrast effects as is normally done when analysing the data from a regular fractional factorial design. Therefore, those who are accustomed to using and analysing data from regular fractional factorials may feel uncomfortable with the new screening designs.

Can adequate screening and response optimisation really be accomplished after conducting one set of experiments according to a definitive screening design? If so, what is the role of traditional resolution III two-level screening designs and sequential follow-up experimentation as recommended by Box and others? The answer to these questions will depend upon the unknown but underlying model that generates the data. By choice of a simulation model, either the traditional sequential approach to optimisation, or the use of a definitive screening experiment could be made to appear more efficient in a simulation study. Rather than trying to give conclusive answers to these questions, we attempt to shed some light on the quandary by sharing our experience in using both approaches in characterising the process of synthesising Al-modified anatase TiO_2 catalyst supports (Olsen et al., 2014).

TiO_2 is an excellent support material for metal and metal oxide catalysts in oxidative synthesis and pollution control reactions, such as CO oxidation at low temperatures, low-

temperature synthesis of hydrogen peroxide, complete oxidation of volatile organic compounds, etc. Each application of TiO_2, as a catalyst support, requires different anatase or anatase/rutile properties such as, pore volume and pore diameter, while maintaining a large surface area. Therefore, prediction equations had to be determined that could identify process input values that would be appropriate for various applications.

The study began utilising a traditional 2^{k-p} fractional factorial design to identify important process inputs because it could be easily augmented to form a central composite design in the factors found to be important. This would allow a quadratic model to be fit characterising the impact of process inputs upon the important anatase or anatase/rutile properties of TiO_2. However, after completion of the experiments and analysis of the data, it was discovered that many additional follow-up experiments could be required (using the traditional approach) before a quadratic model could be fit to the data. Since the number of follow-up experiments required to form a central composite in the important variables could be even greater than the number of experiments required by a definitive screening design of Jones and Nachtsheim (2011), it was decided to start over using this new design rather than completing the necessary follow-up experiments to the 2^{k-p} design. This allowed us to make a practical comparison of traditional screening and sequential follow-up with the definitive screening design. Based on the results of this comparison, we give our insights regarding the traditional approach and the newer definitive screening designs in process development and process improvement.

The remainder of this article is organised as follows. Section 2 describes the TiO_2 synthesis process and the traditional 2_{III}^{10-6} fractional factorial design used in the initial screening experiment. Section 3 describes the models obtained from the analysis of the fractional factorial design. Section 4 describes alternative approaches to follow-up experiments that could be used for sequential augmentation of the initial design. Section 5 shows the definitive screening design, and Section 6 describes results obtained from analysing the data. Section 7 compares the information obtained from the initial fractional factorial to the information obtained from the definitive screening design. Section 8 compares the information obtained from the definitive screening design to what may have been obtained from the initial fractional factorial design plus sequential follow-up experiments. In the final section, we discuss the insights we have gained in attempting to use the traditional sequential screening-optimisation, and definitive screening design on the same process.

2 Synthesis process and fractional 2^{k-p} design

The process of synthesising TiO_2 in the lab is a 4-step process as illustrated in Figure 1. Steps 3 and 4 could either be Rinse followed by Calcine or Calcine followed by Rinse. The starting materials were $TiCl_4$, $Al(NO)_3{\cdot}9H2O$, and NH_4HCO_3 (ABC). At each step of this process, there were at least two factors that could be varied, and at the end of the process, the resulting material was analysed to determine the surface area, pore volume and pore diameter.

Figure 1 Process for synthesising TiO_2 in the lab

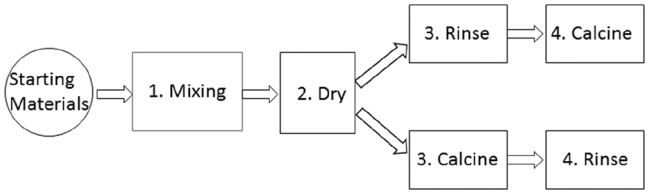


Table 1 shows the factors that could be varied at each step of the synthesis process and the two alternative levels for each factor. At the mixing step, the levels of mixing order are defined as follows: Ti/Al means that $TiCl_4$ and $Al(NO)_3{\cdot}9H_2O$ were mixed together in a mortar and pestle for 1 min after which NH_4HCO_3 (ABC) was added and mixed for an additional minute; Al/ABC means that $Al(NO)_3{\cdot}9H_2O$ and NH_4HCO_3 (ABC) were mixed together in a mortar and pestle for 1 min after which $TiCl_4$ was added and mixed for an additional minute. Next, the amount of distilled H_2O indicated by the level of factor C was added at the rate indicated by factor B, and the slurry was mixed an additional 5 min to form a stabilised anatase precursor. Factor D, the order of rinsing, is illustrated by the two possible paths in Figure 1. Using the upper alternative path, the material was first rinsed with distilled water using a vacuum filtration system and then immediately calcined. Using the lower alternative path in Figure 1, the material was first calcined and then rinsed with distilled water using a vacuum filtration system. In the drying step, both the time and temperature of drying were varied. In the calcination step, temperature, time and ramp rate were varied. Finally, Factor K (the mole % Al) is determined by the amount of the starting materials used in the mixing step.

Table 1 Factors and levels for the fractional factorial design

Factor		*Low level (–)*	*High level (+)*
Mixing step			
A	Mixing order	Ti/Al	Al/ABC
B	Speed of H_2O addition	Slow	Fast
C	Amount H_2O (ml)	7	22
D	Order of rinsing	Rinse then calcine	Calcine then rinse
Drying step			
E	Drying time (h)	3	24
F	Drying temperature (°C)	25	100
Calcination step			
G	Calcination ramp rate (°C/min)	2	20
H	Calcination temperature (°C)	400	700
J	Calcination time (h)	2	20
K	Mole % Al	5	22

Using the conventional wisdom that the screening design should take no more that 25% of the total resources, we decided to start with a 16-run 2_{III}^{10-6} fractional factorial design. Using the factor labels in Table 1, the design generators were E=ABCD, F=BCD, G=ACD, H=CD, J=ABD and K=ABC. The alias structure up to two-factor interactions is shown in Table 2.

Table 2 Alias structure for estimable effects

	Confounded	
Factor	*Label*	*Two-factor interactions*
Mixing order	A	EF + GH
Speed of H_2O addition	B	FH + EG
Amount of H_2O	C	DH + EJ
Order of rinsing	D	EK + CH
Drying time	E	DK + CJ + BG + AF
Drying temperature	F	BH + AE
Calcination ramp	G	BE + AH
Calcination temperature	H	BF + JK + CD + AG
Calcination time	J	HK + CE
Mol % Al	K	HJ + DE
	AB	CK + DJ + EH + FG
	AC	BK + DG + FJ
	AD	BJ + CG + FK
	AJ	BD + CF + GK
	AK	BC + DF + GJ

Although this is a resolution III design (with two-factor interactions confounded with main effects and other two-factor interactions), it was hoped that by running these 16 experiments the factors that had the largest effects could be identified, and follow-up experiments could focus on the important factors. Table 3 shows the design and resulting responses (properties of the resulting catalyst support material) in standard order. The actual experiments were performed in a random order to prevent biases from uncontrollable variables such as the reaction temperature.

Table 3 Fractional factorial design and results

	Factors										*Responses*		
Run	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*K*	*Surf. area (m^2/g)*	*Pore vol. (cm^3/g)*	*Pore Dia. (nm)*
1	–	–	–	–	+	–	–	+	–	–	50	0.11	6.4
2	+	–	–	–	–	–	+	+	+	+	67	0.12	5.7
3	–	+	–	–	–	+	–	+	+	+	104	0.21	6.1
4	+	+	–	–	+	+	+	+	–	–	61	0.16	7.6
5	–	–	+	–	–	+	+	–	–	+	375	0.33	3.5

Table 3 Fractional factorial design and results (continued)

	Factors										*Responses*		
Run	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*K*	*Surf. area (m^2/g)*	*Pore vol. (cm^3/g)*	*Pore Dia. (nm)*
6	+	–	+	–	+	+	–	–	+	–	332	0.35	3.5
7	–	+	+	–	+	–	+	–	+	–	203	0.2	3.7
8	+	+	+	–	–	–	–	–	–	+	177	0.31	6.5
9	–	–	–	+	–	+	+	–	+	–	123	0.36	10.1
10	+	–	–	+	+	+	–	–	–	+	369	0.31	3.6
11	–	+	–	+	+	–	+	–	–	+	169	0.52	15.6
12	+	+	–	+	–	–	–	–	+	–	150	0.49	12.3
13	–	–	+	+	+	–	–	+	+	+	124	0.41	12.1
14	+	–	+	+	–	–	+	+	–	–	95	0.46	17.3
15	–	+	+	+	–	+	–	+	–	–	308	0.84	14.9
16	+	+	+	+	+	+	+	+	+	+	94	0.41	15.3

3 Results of analysis of the fractional 2^{k-p} design

Figures 2, 3 and 4 show Pareto charts of the orthogonal estimated effects (Lenth, 2015) upon the three responses (or catalyst support properties that require different values for different catalyst applications). These figures were produced by the JMP Analyze/Modeling/Screening menu using the method of Lenth (1989). The bars that protrude beyond the solid vertical error lines on the left or right side of the table, or whose *P*-values are less than 0.10, were judged to be significant. Models for each response were refit including only the largest three effects in Figures 2, 3 and 4, and any insignificant terms were then dropped to reach the final models shown in Table 4.

Figure 2 Effects on surface area

Term	Estimate	Relative Std Error	Lenth t-Ratio	P-Value
H	-62.1875	0.25	-2.30	0.0702
F	45.6875	0.25	1.69	0.1526
C	38.4375	0.25	1.42	0.2152
A*C	-32.0625	0.25	-1.18	0.2898
A*B	-30.8125	0.25	-1.14	0.3070
G	-26.6875	0.25	-0.99	0.3699
J	-25.4375	0.25	-0.94	0.3909
A*J	18.0625	0.25	0.67	0.5345
B	-16.8125	0.25	-0.62	0.5621
K	9.8125	0.25	0.36	0.7320
A	-6.9375	0.25	-0.26	0.8081
A*D	4.9375	0.25	0.18	0.8626
D	3.9375	0.25	0.15	0.8901
A*K	-1.1875	0.25	-0.04	0.9667
E	0.1875	0.25	0.01	0.9947

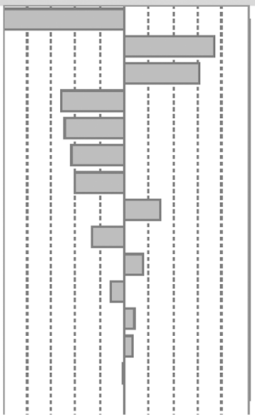


Figure 3 Effects on pore volume

Term	Estimate	Relative Std Error	Lenth t-Ratio	P-Value
D	0.125625	0.25	2.98	0.0309*
C	0.064375	0.25	1.53	0.1876
A*J	0.046875	0.25	1.11	0.3171
B	0.043125	0.25	1.02	0.3536
E	-0.040625	0.25	-0.96	0.3798
A*D	-0.034375	0.25	-0.81	0.4522
J	-0.030625	0.25	-0.73	0.5004
G	-0.029375	0.25	-0.70	0.5173
A*B	-0.026875	0.25	-0.64	0.5521
A	-0.023125	0.25	-0.55	0.6072
F	0.021875	0.25	0.52	0.6262
K	-0.021875	0.25	-0.52	0.6262
A*K	-0.016875	0.25	-0.40	0.7057
H	-0.009375	0.25	-0.22	0.8329
A*C	-0.008125	0.25	-0.19	0.8549

Figure 4 Effects on pore diameter

Term	Estimate	Relative Std Error	Lenth t-Ratio	P-Value
D	3.6375	0.25	3.96	0.0108*
H	1.6625	0.25	1.81	0.1302
B	1.2375	0.25	1.35	0.2358
A*C	1.0875	0.25	1.18	0.2897
F	-0.9375	0.25	-1.02	0.3543
G	0.8375	0.25	0.91	0.4038
A*K	-0.7375	0.25	-0.80	0.4586
A*J	0.6375	0.25	0.69	0.5187
C	0.5875	0.25	0.64	0.5507
E	-0.5375	0.25	-0.59	0.5839
A*D	-0.4875	0.25	-0.53	0.6184
K	-0.4625	0.25	-0.50	0.6361
J	-0.4125	0.25	-0.45	0.6722
A*B	0.2125	0.25	0.23	0.8263
A	-0.0375	0.25	-0.04	0.9690

Table 4 Final models for fractional factorial

Term	*Estimate*	*STD error*	*t Ratio*	*Prob > |t|*
Effects for surface area				
Intercept	175.0625	20.77128	8.43	<0.0001*
H	–62.1875	20.77128	–2.99	0.0104*
F	45.6875	20.77128	2.20	0.0465*
Adjusted R^2 = 0.44				
Effects for pore volume				
Intercept	0.349375	0.029538	11.83	<0.0001*
D	0.125625	0.029538	4.25	0.0009*
C	0.064375	0.029538	2.18	0.0483*
Adjusted R^2 = 0.58				
Effects for pore diameter				
Intercept	9.0125	0.70966	12.70	<0.0001*
D	3.6375	0.70966	5.13	0.0002*
H	1.6625	0.70966	2.34	0.0357*
Adjusted R^2 = 0.58				

Factor H (Calcination Temperature) and D (Order of Rinsing) appear to have the largest effects, and only factor A (Mixing Order) appears to have a negligible effect for all three responses. Only two factors were significant in the final models for each response, but they were not the same two factors in each model. Although all the terms in the simple final models were significant at the 0.05 level, the adjusted R^2 was less than 0.60 for all three models. Furthermore, the range of predicted pore volumes (within the experimental region), while maintaining a large surface area, was not wide enough for many catalytic applications. Therefore, better models were sought for the data.

4 Alternative approaches to follow-up experiments

One way to make the models better would be to include interaction terms or expand the models to the full quadratic models (including quadratic effects and linear by linear interaction terms) in the factors found important. However, this would require additional follow-up experiments. There are different alternative approaches for determining a list of follow-up experiments depending on the assumptions made.

One approach would be based on the assumption that the four factors in all the final simple models in Table 4 (i.e., C - Amount of H_2O, D - order of rinsing, F - drying temperature, and H - calcination temperature) were the only ones important, and that the other six factors could be ignored. Ignoring all but C, D, F and H in the original fractional factorial design, the main effect C is still completely confounded with the DH interaction, the main effect for D is still completely confounded with the CH interaction, and the

main effect for H is still confounded with the CD interaction. These three main effects could be made completely orthogonal to the two-factor interactions they are confounded with by augmenting the design with an additional block of 16 experiments with the signs on factor C reversed (Box, Hunter and Hunter, 2005). Running these 16 experiments would result in a full 2^3 factorial design in factors C, F and H, at each level of Factor D, with half the runs replicated. This would allow unconfounded estimates of the main effects for C, F and H and all two-factor interactions among these three factors. Next, a third block of experiments could be run that included axial and centre points for factors C, F and H at each level of Factor D. This would allow fitting a full quadratic model in factors C, F and H at each level of D (Order of Rinsing). This approach would require approximately 34 follow-up experiments.

However, if the experimenter was uncomfortable in assuming that all factors except C, D, F and H had negligible effects after the initial 16 experiments, another approach could be taken. First, the original 16-run fractional factorial could be augmented with a mirror image design (signs reversed on all 10 factors Lawson, 2015b). This would allow all 10 main effects to be estimated independently of the strings of the two-factor interactions they are confounded with. In addition, the confounded strings of two-factor interactions beginning with AB, AC, AD, AJ and AK (shown in Table 2) could be estimated. More data would be available to determine which of the 10 factors and two-factor interactions were important. If the significant contrasts in the combined 16-run fractional factorial plus mirror image design could be interpreted (by the effect heredity principle (Hamada and Wu, 1992) to represent a few main effects and interactions among them, then the combined design could be augmented with axial and centre points (at each level of Factor D) to allow fitting quadratic models in the important factors at each level of D. This would require about 32–34 follow-up experiments.

However, if no clear interpretation of the significant effects were possible using the effect heredity principle, additional experiments would be necessary to separate confounded strings of interactions before adding centre points and axial points. For example, if after the analysis of the original and mirror image designs, main effects B, C, D, F and H were found to be significant, in addition to the two confounded strings of two-factor interactions AJ+BD+CF+GK and BF+JK+CD+AG, then the effect heredity principle would not help to determine whether the first string of two-factor interactions represented BD or CF (since the four main effects B, C, D and F were all significant). Additionally, the effect heredity principle would not help to determine whether the second string of interactions represented BF or CD, again since the main effects B, C, D and F were significant. Therefore, additional experiments would be required to allow separate estimates of BD and CF, and BF and CD, before the design could be augmented with centre points and axial points. One way this could be done is to complete another 32 experiments folded on factor B, or factor C or factor F (Montgomery and Runger 1996). This would allow independent estimation of BD, CF, BF and CD. A smaller set of follow-up experiments that would allow separate, but not independent, estimates of these four interactions could be obtained by selecting a subset of additional runs to maximise the determinant of $X'X$, where X is the model matrix (i.e., containing B, C, D, F, H, BD, CF, BF and CD) for the combined set of runs (Dykstra, 1971). After adding centre and

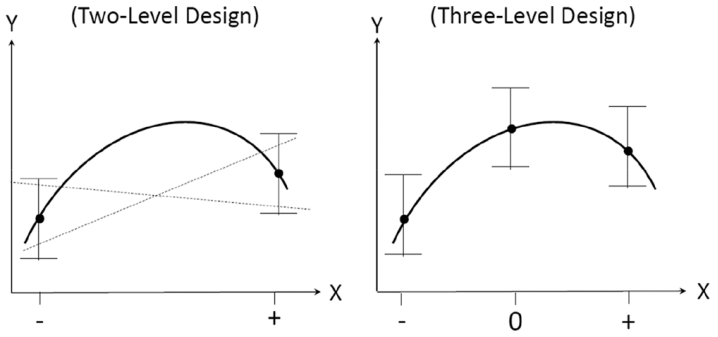
axial points to allow fitting a quadratic model (at each level of D - Order of Rinsing), as many as 40–68 follow-up experiments would be required using this approach.

Another consideration in planning follow-up experiments was the fact that the range of predicted pore diameters (with high surface area) derived from the final simple models fit to the data from the screening experiment was not wide enough. That may have been due to the fact that quadratic effects were not in the final models or that range of important factor levels in the original experiment were too narrow. Therefore, it was also desirable to expand the range on the suspected important factors in follow-up experiments. This could be done using the two approaches above by expanding the range of the factors found to be important in the axial portion of the design. However, if all factors but C, D, F and H were held constant in the follow-up experiments, the range could only be expanded for factors C, F and H. If all factors were varied in the follow-up experiments, it is still possible that the range on some important factors would not be increased. If the range of a factor in the initial fractional factorial portion of the design (including follow-up experiments) was too narrow, its effect might not be detected due to the experimental error. As a result, axial points would not be added for this factor.

Due to the uncertainties associated with augmenting the original 16 experiments, we decided to try a different approach. Jones and Nachtsheim (2011, 2013) had recently published papers describing a new class of screening designs called *definitive screening designs*. These designs incorporate three levels on each quantitative factor (allowing for possible quadratic effects) and are efficient for screening a large number of factors. Only $2k+1$ experiments are required to study k quantitative factors and only 22–24 experiments are required to study 8 quantitative factors and 2 two-level factors. Since all the factors, except A and D in Table 1, are quantitative, it was decided to plan a separate definitive screening design in factors B, C, E, F, G, H, J and K at each level of the discrete factor D (Order of Rinsing). Factor A (Mixing Order) was held constant since it seemed to have the least effect on any of the responses in the initial screening experiment. This would require only $2(8) + 1 = 17$ runs in each of the two designs, or a total of 34 additional experiments. This would be no more than required by either of the two more traditional approaches to augmenting the initial 16 experiments, and the range of settings on the factors felt to be important by the experimenters could be widened.

When designing a two-level screening design, like a regular fractional factorial design, an experimenter should be bold in choosing factor levels that are separated widely so that the effect of the factor can be detected above the level of the random experimental error. However, if the relationship between the response and a factor is non-linear and can be approximated by a quadratic function, choosing a range of factor levels that is too wide can in some cases reduce the power of detecting the linear effect above the experimental error, as shown in the left side of Figure 5. If, on the other hand, a three-level design (like the definitive screening design) is used, the probability of this happening is greatly reduced since both linear and quadratic terms can be fit. Therefore, when using a definitive screening design, an experimenter has more to gain and less to lose by choosing a wide range of factor levels.

Figure 5 Effect of curvature on effect



5 Definitive screening design

A 17-run definitive screening design was run using the DRC (dry→rinse→calcine) procedure shown in the top branch of Figure 1. The design and resulting data are shown in Table 5. The actual factor levels are shown in this table and the runs are in a random order. It can be seen that there is a centre value for every factor in this table and that ranges on Factors: B (Speed of H_2O addition), C (Amount of H_2O), G (Calcination Ramp Rate) and K (Mole % Al) were increased over what they were in the initial 16-run fractional factorial design. A second 17-run definitive screening design was run using the DCR (dry→calcine→rinse) path shown in the bottom branch of Figure 1. This design and the resulting data are shown in Table 6.

Table 5 Definitive screening design for rinsing order DRC (randomised order)

Run	*Factor B speed H_2O addition*	*Factor C amount of H_2O*	*Factor E drying time*	*Factor F drying temp*	*Factor G Calc ramp rate*	*Factor H Calc temp*	*Factor J Calc time*	*Factor K Mol % Al*	*Surface area (m^2/g)*	*Pore volume (cm^3/g)*	*Pore diameter (nm)*
1	2	20	13.5	62.5	12	550	11	0.15	245.86	0.31	5.35
2	3	5	24	62.5	22	700	2	0.05	63.08	0.15	11.4
3	2	35	3	25	22	400	2	0.05	219.41	0.23	4.35
4	2	5	24	100	2	700	20	0.25	193.84	0.25	5.4
5	3	20	24	25	2	400	2	0.25	362.5	0.282	3.79
6	3	35	3	100	12	700	2	0.25	130.16	0.26	11.4
7	1	35	24	25	22	700	11	0.25	127.95	0.25	9.54
8	1	35	24	100	2	550	2	0.05	142.16	0.23	5.9
9	3	35	24	100	22	400	20	0.15	333.97	0.39	3.919
10	3	5	3	25	22	550	20	0.25	112.26	0.23	10.09
11	1	5	24	25	12	400	20	0.05	240.29	0.22	3.56
12	1	5	3	25	2	700	2	0.15	83.09	0.13	8.1
13	1	35	3	62.5	2	400	20	0.25	394.18	0.37	4.29
14	3	5	3	100	2	400	11	0.05	208.94	0.21	4.53
15	3	35	13.5	25	2	700	20	0.05	58.44	0.16	14.05
16	1	5	13.5	100	22	400	2	0.25	449.81	0.36	4.15
17	1	20	3	100	22	700	20	0.05	59.99	0.16	12.91

Table 6 Definitive screening design for rinsing order DCR (randomised order)

Run	Factor B speed H_2O addition	Factor C amount of H_2O	Factor E drying time	Factor F drying temp	Factor G Calc ramp rate	Factor H Calc temp	Factor J Calc time	Factor K Mol % Al	Surface area (m^2/g)	Pore volume (cm^3/g)	Pore diameter (nm)
1	3	5	3	100	2	400	11	0.05	126.6849	0.381962	10
2	1	5	13.5	100	22	400	2	0.25	189.27	0.332965	6.62
3	3	5	3	25	22	550	20	0.25	101.8725	0.339303	11.16
4	1	20	3	100	22	700	20	0.05	82.1226	0.362516	14.64
5	1	35	24	25	22	700	11	0.25	132.5421	0.344219	8.48
6	1	35	24	100	2	550	2	0.05	137.2066	0.390948	9.33
7	2	35	3	25	22	400	2	0.05	139.2732	0.410541	10.3
8	3	35	24	100	22	400	20	0.15	137.5256	0.437526	18.61
9	1	5	24	25	12	400	20	0.05	139.9975	0.446502	11.02
10	3	5	24	62.5	22	700	2	0.05	80.574	0.338265	15.34
11	1	5	3	25	2	700	2	0.15	126.3628	0.403241	11.75
12	3	35	13.5	25	2	700	20	0.05	97.0894	0.407181	13.29
13	2	5	24	100	2	700	20	0.25	147.9825	0.38116	8.97
14	2	20	13.5	62.5	12	550	11	0.15	118.9704	0.337983	12.47
15	1	35	3	62.5	2	400	20	0.25	192.152	0.437772	7.76
16	3	20	24	25	2	400	2	0.25	182.199	0.557723	16.86
17	3	35	3	100	12	700	2	0.25	136.8539	0.516929	14.01

In definitive screening designs, linear main effects are not confounded with two-factor interactions or quadratic effects, and linear-by-linear two-factor interactions are only partially confounded with quadratic effects. Therefore, the confusion caused by the confounding of main effects and interactions in the initial 16-run 2_{III}^{10-6} fractional factorial design could be avoided with the definitive screening design. A full quadratic model cannot be fit to the data from a definitive screening design because there would be more terms in the model $\left(8 \text{ linear} + 8 \text{ quadratic} + \binom{8}{2} \text{interactions}\right)$ than there are runs in the design (17).

We used Jones and Nachtsheim (2011)'s modified forward stepwise regression (described in the introduction) to analyse the data in Tables 5 and 6. The *combine* option in the JMP forward stepwise regression procedure was used to do this. When combine option entered insignificant terms into the model, the backward elimination procedure was used to trim them from the model. This procedure was used to select models for predicting surface area, pore diameter and pore volume from the results of the two definitive screening designs.

6 Results of analysis of data from the definitive screening designs

The coded and scaled factors used as candidate terms in the model for the stepwise regression were:

$X_B = (\text{speed of } H_2O \text{ Addition-2})/1$,

$X_C = (\text{Amount } H_2O - 20)/15$,

$X_E = (\text{Drying Time -13.5})/10.5$,

$X_F = (\text{Drying Temp. - 62.5})/37.5$,

$X_G = (\text{Calc. Ramp -12})/10$,

$X_H = (\text{Calc. Temp. -550})/150$,

$X_J = (\text{Calc. Time -11})/9$,

$X_K = (\text{Mole \% AL} - .15)/.10$.

We found the procedure for fitting the equations to be fairly straightforward and repeatable. Forward steps were completed until the terms entering the equation no longer appeared significant. Due to the *combine* option, this resulted in models that contained all main effects involved in interactions and quadratic terms in the model. Backward elimination steps were then completed to sequentially remove terms that had the largest *P*-value and were not significant at the 0.05 level. A non-significant main effect was not removed if there was a significant interaction or quadratic term in the model that involved this main effect. This preserved the effect heredity in all cases. This procedure only took between 2 and 5 steps of the forward regression followed by zero to two steps of backward elimination. The coefficients for the resulting models are shown in Table 7. By examination of the R^2 values and diagnostics plots, it could be seen that these equations fit the data from the definitive screening designs much better than the linear models (Table 7) fit the data from the fractional design.

Table 7 Results from definitive screening design

DRC regression coefficients for surface area					DCR regression coefficients for surface area				
Term	*Estimate*	*STD error*	*t Ratio*	*Prob>|t|*	*Term*	*Estimate*	*STD error*	*t Ratio*	*Prob>|t|*
Intercept	201.52529	7.294279	27.63	<.0001*	Intercept	133.45171	2.584694	51.63	<.0001*
X_B	−16.29429	8.037904	−2.03	0.0733	X_B	−9.775307	2.848194	−3.43	0.0064*
X_E	18.268571	8.037904	2.27	0.0491*	X_G	−10.46409	2.848194	−3.67	0.0043*
X_F	22.495	8.037904	2.80	0.0208*	X_H	−21.68392	2.848194	−7.61	<.0001*
X_H	−106.6107	8.037904	−13.26	<.0001*	X_K	19.994557	2.848194	7.02	<.0001*
X_K	55.599286	8.037904	6.92	<.0001*	X_J	−6.642671	2.848194	−2.33	0.0419*
X_BX_E	27.920833	8.681929	3.22	0.0106*	X_GX_J	−6.587667	3.076401	−2.14	0.0579
X_FX_K	23.179167	8.681929	2.67	0.0256*	$R^2 = 0.934$				
$R^2 = 0.966$									
DRC regression coefficients for pore volume					DCR regression coefficients for pore volume				
Term	*Estimate*	*STD error*	*t Ratio*	*Prob>|t|*	*Term*	*Estimate*	*STD error*	*t Ratio*	*Prob>|t|*
Intercept	0.2465882	0.007944	31.04	<.0001*	Intercept	0.4015727	0.007776	51.65	<.0001*
X_C	0.0242857	0.008754	2.77	0.0168*	X_B	0.0186233	0.008568	2.17	0.0615
X_F	0.0255714	0.008754	2.92	0.0128*	X_C	0.0229799	0.008568	2.68	0.0278*
X_H	−0.050143	0.008754	−5.73	<.0001*	X_G	−0.028189	0.008568	−3.29	0.0110*
X_K	0.0458571	0.008754	5.24	0.0002*	X_H	−0.017963	0.008568	−2.10	0.0693
$R^2 = 0.864$					X_J	−0.009904	0.008568	−1.16	0.2811
					X_BX_C	0.0323291	0.00939	3.44	0.0088*
					X_BX_J	−0.038037	0.009523	−3.99	0.0040*
					X_GX_J	0.0216206	0.00939	2.30	0.0503
					$R^2 = 0.873$				

Table 7 Results from definitive screening design (continued)

DRC regression coefficients for pore diameter					DCR regression coefficients for pore diameter				
Term	*Estimate*	*STD error*	*t Ratio*	*Prob>|t|*	*Term*	*Estimate*	*STD error*	*t Ratio*	*Prob>|t|*
Intercept	5.0333333	0.83611	6.02	<.0001*	Intercept	11.800588	0.358826	32.89	<.0001*
X_B	0.7663571	0.387044	1.98	0.0733	X_B	2.1192857	0.395407	5.36	0.0002*
X_E	–0.868643	0.387044	–2.24	0.0464*	X_E	0.6421429	0.395407	1.62	0.1327
X_G	0.7356429	0.387044	1.90	0.0839	X_H	0.3792857	0.395407	0.96	0.3581
X_H	3.1579286	0.387044	8.16	<.0001*	X_BX_E	1.4245714	0.433147	3.29	0.0072*
X^2_B	2.6544524	0.921349	2.88	0.0149*	X_EX_H	–1.932571	0.433147	–4.46	0.0010*
$R^2 = 0.888$					$R^2 = 0.862$				

7 Comparison of models from 2^{k-p} and definitive screening designs

The two definitive screening designs required a total of 34 experiments (17+17), while the 2_{III}^{10-6} fractional factorial design took only 16 experiments. There were only four factors found to have significant effects on the three responses in the fractional factorial designs, and the fitted models were not accurate enough to accurately predict the input factor settings necessary to produce a wide enough range of pore volume needed for different applications. The remainder of this section discusses the additional factors and interactions found in the prediction models fit to the data from the definitive screening designs and how they improved the predictions.

Table 8 indicates the terms in the models fit to both the fractional factorial design and the definitive screening designs. The left column shows the abbreviation for factors and interaction terms in the models. The columns indicate to which design and response the model was fit. For example, the FF column indicates the model fit to the original fractional factorial design, and the DSD (DRC) column indicates the model fit to the definitive screening design used on the DRC (dry→rinse→calcine) process, etc. The + or − signs in the body of the table indicate the signs of the coefficients in the models. Signs followed by a superscript NS were not significant at the 0.05 level but were retained in the model to preserve effect heredity. It can be seen that all the effects (and the direction of effects) found significant in the original fractional factorial design are again confirmed by being found significant in the definitive screening designs. In addition, many other effects were found to be significant in the definitive screening designs and increased the predictive value of the models.

Table 8 Comparison of models

	Surface area			*Pore volume*			*Pore diameter*		
Model term	*FF*	*DSD (DRC)*	*DSD (DCR)*	*FF*	*DSD (DRC)*	*DSD (DCR)*	*FF*	*DSD (DRC)*	*DSD (DCR)*
Calc. temp (H)	–	–	–		–	$-^{NS}$	+	+	$+^{NS}$
Dry temp (F)	+	+			+				
Amt. H_2O (C)				+	+	+			
Speed H_2O add (B)		$-^{NS}$	–			$+^{NS}$		$+^{NS}$	+
Calc. ramp (G)			–			–		$+^{NS}$	
Mole % Al (K)		+	+		+				
Calc. time (J)			–			$-^{NS}$			
Dry time (E)		+						–	$+^{NS}$
(B^2)								+	
(BC)						+			
(BJ)						–			
(GJ)			–			+			
(BE)		+							+
(EH)									–
(FK)		+							
s	83	30	11	.12	.0033	.0032	2.84	1.44	1.48

It can be seen in the table that factor B (Speed of H_2O addition) did not have a significant effect on any of the three responses in the models fit to the data from the fractional factorial design (even though it had the third largest effect on pore diameter). However, in the definitive screening designs, it did have significant effects. For the DRC (dry→rinse→calcine) process, it had a negative effect on the surface area and a positive and quadratic effect on pore diameter. For the DCR (dry→calcine→rinse) process, it had a somewhat negative effect on surface area that depended on the level of factor E (Drying time). It had a positive effect on pore volume that depended on the level of factor C (Amount of H_2O), and it had a positive effect on pore diameter that depended on the level of factor E (Drying time).

It can be seen that factor G (Calcination ramp rate) was not found to have any significant effects on any of the three responses in the fractional factorial design. However, it also had significant effects in the definitive screening designs. In the DCR (dry→calcine→rinse) process, it had a significant negative effect on both pore volume and pore diameter that depended on the level of factor J - Calcination time.

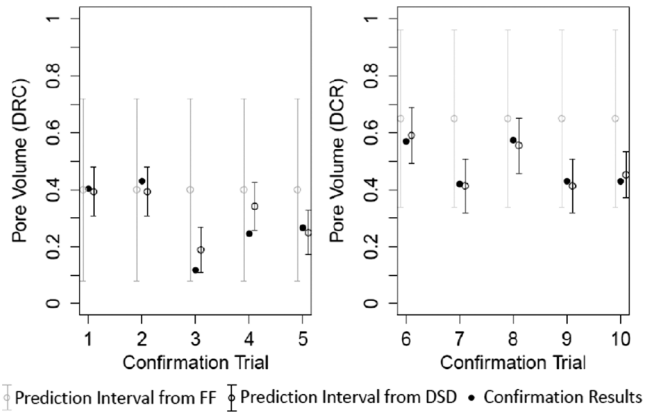
Factor K (Mole % Al) did not have a significant effect on any response in the fractional factorial design, but it did have significant effects on surface area and pore volume in the definitive screening designs. In the DRC (dry→rinse→calcine) process, it had a positive effect on surface area that depended on the level of factor F (Drying Temp.), and it had a positive effect on pore volume. It also had a positive effect on surface area in the DCR (dry→calcine→rinse) process.

In the definitive screening design conducted on the DCR (dry→calcine→rinse) process, factor J (Calcination time) had a negative effect on surface area that depended on the level of factor G (Calcination ramp rate), and it had a somewhat negative effect on pore volume that depended on the levels of both factors B (Speed of H_2O addition) and G. Again, factor J did not have any significant effect on the responses in the fractional factorial design.

Finally, although factor E (Drying time) had no significant effects on the responses in the fractional factorial design, it did in the definitive screening designs. For experiments conducted with the DRC (dry→rinse→calcine) process, it had a positive effect on surface area that depended on the level of factor B (Speed of H_2O addition), and it had a negative effect on pore volume. For experiments conducted in the DCR (dry→calcine→rinse) process, it had a somewhat positive effect on pore diameter that depended on the level of both factors B and H (Calcination Temp.).

Including these additional factors and interactions to the models fit to the definitive screening designs improved the accuracy of predicting pore volume by reducing the root mean square error (s) (shown at the bottom of Table 8) by more than 70%. Confirmation trials were shown in Olsen et al. (2014) to verify the accuracy of the prediction equations. A comparison of the 95% predicted intervals for pore volume on these trials between the equation fit to the fractional factorial design and the equations fit to the definitive screening designs are shown in Figure 6. In addition, the results of the confirmation trials are also shown in the figure.

Figure 6 Predicted pore volume from confirmation experiments



The graph on the left side of the figure shows predictions intervals and the result of confirmation experiments for the DRC (dry→rinse→calcine) process. The grey lines show the prediction intervals obtained from the model fit to the fractional factorial design. These intervals were wide and the predicted values were constant because factor C was constant in these confirmation trials, and it was the only factor beside rinse order found to affect pore volume in the fractional factorial design. The black lines show the prediction intervals from the model fit to the definitive screening design. The predicted pore volumes from this model are lower for trials 3, 4 and 5 because the level of factor H (Calcination Temp) was increased and the level of factor F (Drying Temp.) was decreased during these trials. These two factors had significant effects on pore volume in the model for the DRC process.

The graph on the right side of the figure shows prediction intervals and the result of confirmation experiments for the DCR (dry→calcine→rinse) process. The grey lines show the prediction intervals obtained from the model fit to the fractional factorial design. Again, these intervals were wide and the predicted values were constant because factor C was constant in these confirmation trials. The black lines show the prediction intervals from the model fit to the definitive screening design, and it can be seen that this model did a much better job of predicting the results of these confirmation trials.

Although the results of the confirmation trial fall within nearly all the prediction intervals obtained from the models fit to the data from both the fractional factorial design and the definitive screening designs, the range of predicted pore volume (0.19–0.59) is much greater from the models fit to the definitive screening designs than the range of predicted values (0.40–0.65) from the model fit to fractional factorial design. Therefore, the models fit to the definitive screening designs are much more useful for identifying process conditions useful for catalyst applications requiring large or small pore volume. The width of the prediction intervals from the models fit to the definitive screening designs are also much narrower that those obtained from the model fit to the fractional factorial design.

8 Comparison of results from definitive screening designs to potential results from traditional follow-up experiments to the 2^{k-p} design

The two definitive screening designs required a total of 34 experiments, which was no more than would have been required using either of the two more traditional approaches for running follow-up experiments described in Section 4.

From the models fit to the definitive screening designs, seven of the original 10 factors were found to have significant effects on one or more of the three responses, and several interaction effects and one quadratic effect were found to be significant. If the assumption had been made that only factors C, D, F and H were important enough to vary in follow-up experiments to the original fractional factorial design, three important factors and all of the interactions and the quadratic term would have been missed. The ranges on factors B, C, G and K were increased in the definitive screening designs, and this could be another reason that factors B, G, K and interactions with these factors were found to be significant. Follow-up experiments using the mirror image design on all factors may not have detected these effects using the original range of factor levels.

9 Discussion and insights

Section 7 compared the equations and predictions made from the fractional factorial design to the equations and predictions made from the two definitive screening designs. With roughly twice the number of experiments required to investigate one less factor, and expanded ranges on four factors, it was clear that the equations fit to the results of the definitive screening designs were more useful for the intended application and were more accurate in predicting the results of the 10 confirmation experiments. With more data, the results from the definitive screening designs not only confirmed the major findings of the fractional factorial designs but identified additional important factors and allowed for more accurate prediction equations to be fit. Of course, the accuracy of the equations fit to the data from the fractional factorial designs could have been improved by conducting follow-up experiments. However, for the situation we studied, the number of required follow-up experiments could have been as many or more than those required by the definitive screening designs.

When there are multiple responses and the set of factors that affect each of these responses is different, the job of determining how to augment a resolution III fractional factorial screening design becomes much more difficult. Most textbooks on experimental design only illustrate traditional approaches to augmenting a fractional factorial design when there is only one response of interest. We considered two possible approaches to augmenting the fractional factorial design, and the first approach (which would have required fewer follow-up experiments) would not have been effective. On the other hand, using the definitive screening design for follow-up experiments worked quite well. Therefore, in situations like this with multiple responses and different sets of factors found to affect each response in an initial screening experiment, we would recommend the use of a definitive screening experiment in place of traditional follow-up experiments to a resolution III design.

In screening designs, it is important to vary the continuous factors over a wide range so that their effects can be identified above the level of the experimental error. However, if there are curvilinear relationships or diminishing returns as factor levels are increased,

a wider range of factor settings in a two-level design may actually decrease the power for detecting an effect. Therefore, in the case where curvilinear relationships are suspected from the outset, the definitive screening design may be a better choice than a resolution III two-level design for initial screening.

We believe traditional resolution III two-level screening experiments and sequential follow-up experiments are still valuable, especially when used in early stages of investigations or where many of the factors are binary and can only be turned on or off. Resolution III designs are extremely efficient in terms of the number of runs required and will waste the fewest resources when experimenters are learning how to vary factor levels and measure appropriate responses. Additionally, if there is only one response (or a consistent set of important factors for all responses), there may only be a minimal number of follow-up experiments required to augment a resolution III design to allow fitting an adequate optimisation model.

On the other hand, if the research is at a stage where a majority of factors can be varied over a continuous range or quadratic relationships are suspected, we would recommend using the definitive screening designs over the traditional resolution IV or V two-level screening designs. Resolution V designs require more runs than definitive screening designs and they are rarely needed because all two-factor interactions are usually not important. There is inherently more confounding among two-factor interactions in resolution IV designs, and unlike definitive screening designs, they are inadequate for estimating quadratic effects.

The central composite or Box-Behnken designs traditionally used when collecting data to fit curvilinear models require many more runs than a definitive screening design. For example, a definitive screening design with nine factors requires only 19 experiments, while a central composite or Box-Behnken design would require 130–156 experiments. This is because these traditional response surface designs allow estimation of all coefficients in the general quadratic model, while in the definitive screening designs two-factor interactions and quadratic effects are partially confounded. If more than six factors are under study, the extra experiments needed for a traditional response surface design are rarely needed because all two-factor interactions and quadratic effects are usually not important. Therefore, with six or more factors, we would recommend starting with a definitive screening design rather than a traditional response surface design. If after analysis of the data from a definitive screening design, a simple and adequate model cannot be determined (as in the example presented in this paper), the design can be augmented to maximise the determinant of $\boldsymbol{X'X}$ as discussed in Section 4 of this paper, or better yet to maximise the I-optimality as described by Jones and Goos (2012).

Definitive screening designs cannot be analysed in the same way as traditional fractional factorial designs by looking at a graphical display (i.e., normal plot or Lenth plot) of an independent set of orthogonal effect contrasts. This may make those used to the traditional designs uncomfortable. However, we found the identification of a useful model involving main effects, two-factor interactions, and quadratic terms to be fairly straightforward using Jones and Nachtsheim (2011)'s modified forward stepwise regression.

Based on our experience, we cannot say that traditional resolution III fractional factorial screening experiments are now obsolete. Nor can we say that definitive screening designs should completely replace the traditional approach for finding a model good enough to identify optimum operating conditions or to make useful predictions. We

do find the definitive screening designs to be a welcome addition. We found that they were useful as follow-up experiments in a complicated situation with multiple responses. We also believe that they can provide an alternative one-step screening and optimisation opportunity in some situations.

References

Abu-Absi, S.F., Yang, L., Thompson, P., Jang, C., Kandula, S., Schilling, B. and Shukla, A.A. (2010) 'Defining process design space for monoclonal antibody cell culture', *Biotechnology and Bioengineering*, Vol. 106, No. 6, pp.894–905.

Box, G.E.P. (1999) 'Statistics as a catalyst to learning by scientific method part I', *Journal of Quality Technology*, Vol. 31, No. 1, pp.16–29.

Box, G.E.P., Hunter, J.S. and Hunter, W.G. (2005) *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed., John Wiley and Sons, Hoboken, NJ.

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

Dykstra, O. (1971) 'The augmentation of experimental data to maximize $|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}|$', *Technometrics*, Vol. 13, pp.682–688.

Hamada, M. and Wu, C.F.J. (1992) 'Analysis of designed experiments with complex aliasing', *Journal of Quality Technology*, Vol. 24, No. 2, pp.130–137.

Jones, B. and Goos, P. (2012) 'I-optimal versus D-optimal split-plot response surface designs', *Journal of Quality Technology*, Vol. 44, No. 2, pp.85–101.

Jones, B. and Nachtsheim, C.J. (2011) 'A class of three-level designs for definitive screening in the presence of second-order effects', *Journal of Quality Technology*, Vol. No. 43, No. 1, pp.1–15.

Jones, B. and Nachtsheim, C.J. (2013) 'Definitive screening designs with added two-level categorical factors', *Journal of Quality Technology*, Vol. 45, No. 2, pp.121–129.

Lawson, J. (2002) 'Regression analysis of experiments with complex confounding patterns guided by the alias matrix', *Computational Statistics and Data Analysis*, Vol. 39, pp.227–241.

Lawson, J. (2015a) daewr: Design and analysis of experiments with, R package version 1.1-5.

Lawson, J. (2015b) *Design and Analysis of Experiments with R*, Chapter 6, CRC Press, Boca Raton, FL, pp.248–249.

Lenth, R.V. (1989) 'Quick and easy analysis of unreplicated factorials', *Technometrics*, Vol. No. 31, pp.469–473.

Lenth, R.V. (2015) 'The case against normal plots of effects', *Journal of Quality Technology*, Vol. 47, pp.91–97.

Lin, D.K.J. (1999) 'Spotlight interaction effects in main effect plans: a supersaturated design approach', *Quality Engineering*, Vol. 11, pp.133–139.

Montgomery, D.C. and Runger, G.C. (1996) 'Foldovers of 2k-p resolution IV experimental designs', *Journal of Quality Technology*, Vol. 28, No. 4, pp.446–450.

Olsen, R.E., Bartholomew, H., Enfield, D.B., Lawson, J.S., Rohbok, N., Scott, S.B. and Woodfield, B.F. (2014) 'Optimizing the synthesis and properties of Al-modified anatase catalyst supports by statistical experimental design', *Journal of Porous Materials*, Vol. 21, No. 5, pp.827–837.

US Food and Drug Administration. (2004) Pharmaceutical cGMPs for the 21st century; a risk based approach.

US Food and Drug Administration. (2011) Guidelines for industry: process validation: general principles and practices (revision 1).

Xiao, L., Lin, D.K.J. and Bai, F. (2012) 'Constructing definitive screening designs using conference matrices', *Journal of Quality Technology*, Vol. 44, No. 1, pp.2–8.