Statistica Sinica **30** (2020), 1543-1559
doi:https://doi.org/10.5705/ss.202018.0120

ORDER-OF-ADDITION MODELING

Robert W. Mee

Nankai University and University of Tennessee

Abstract: The current literature on order-of-addition experiments generally relies on main effects models constructed from pair-wise ordering (PWO) factors. This study constructs models utilizing interactions of PWO factors to explain variations that are best accounted for by the ordering of sets of three or more components. Order-of-addition orthogonal arrays are optimal for fitting the main effects PWO model, but they differ in terms of their susceptibility to bias due to model misspecification. A measure computed from the alias matrix is proposed to identify robust PWO designs, and is illustrated for cases with four and five components. Applications with constraints on the ordering and order-of-addition experiments with additional mixture proportion and factorial factors are discussed briefly. Two drug sequence experiments based on data from private consultations are used to compare the usefulness of different order-of-addition models.

Key words and phrases: Alias matrix, component position model, interaction, mixture experiment, model misspecification, orthogonal array, pairwise order.

1. Introduction

The mixing order is critical when preparing tanks for spraying herbicides and pesticides. For example, Adama Agricultural Solutions Ltd. provide a 13-step guide specifying the order when adding up to 12 different categories of products, with water added in Steps 1, 10, and 12. They also give advice on pre-testing compatibility using a standard jar test prior to actual mixing, as well as limiting the number of products in a new mix. In the latter case, the more products one adds, the more likely it is that the active ingredients or adjuvants will be incompatible. Thus, even given well-established orders, some experimentation may be needed. In another agricultural application, Wagner (1995) investigated the order of mixing of feed rations and the time spent blending using two types of mixers. The order in which reagents are added also matters in a polymerase chain reaction (PCR), and the sequence in which drugs are administered can impact clinical outcomes (Ding et al. (2015)).

The literature on experimental designs for mixtures focuses primarily on the proportions or amounts of ingredients, rather than on the order of addition. Mod-

eling the effects of changing proportions can be achieved using linear, quadratic, and various cubic models (Scheffe (1958)). For modeling the effect of the mixing order, two classes of models have been proposed: pair-wise order models (Van Nostrand (1995)), and the recently proposed position effect model of Yang, Sun and Xu (2018). This study extends Van Nostrand's model for characterizing mixing order effects.

Order-of-addition models can be employed for any application where the order or sequence varies. When the product is a mixture of ingredients, it is natural to refer to the ingredients as *components*. When the sequence refers to the stages in an operation or process, one naturally refers to these as *steps*. We use the terms components and steps interchangeably. In the simplest cases, the m components are added (or steps are performed) sequentially. In other settings, some steps may be prepared in parallel and then combined, so that neither precedes the other. In this article, we discuss the design and modeling implications for both situations.

If there are m components added in sequence, then there exist $m!$ possible sequences, if no constraints are placed on the order. We consider this setting first, where all sequences are feasible. Van Nostrand (1995) proposed experimental designs that were a subset of the $m!$ sequences, and proposed fitting a model using pair-wise ordering (PWO) factors. Van Nostrand's model assumes that the response variable Y has expectation $E(Y)$, which depends solely on the sum of m-choose-2 PWO terms. In the next section, we examine how to detect higher-order effects, given data from the full set of $m!$ sequences. In particular, we propose models with terms that explain variation that depends on the order of triplets (or larger subsets) of components, and suggest analysis methods for full designs. In Section 3, we analyze data from a 40-run drug sequencing experiment involving five compounds. This example illustrates the need for higher-order models, as well as showing how to select a model based on only a fraction of the full 5! sequences. In Section 4, we consider the implications for fractional design choice based on models with higher-order effects, and propose a measure of the robustness to model misspecification based on an alias matrix. This robustness measure improves on Voelkel (2019) minimum moment aberration criterion. Section 5 discusses experiments in which the steps are only partially ordered. Section 6 describes more complicated experiments in which we vary the mixture proportions and/or the levels of other quantitative factors, in combination with order-of-addition experiments.

The contributions of this study are as follows. First, it extends the standard

PWO model for applications where Van Nostrand's model exhibits a lack of fit. By using selected interactions of the PWO factors, these new models are both flexible and interpretable. Second, we improve on Voelkel's criterion for ranking order-of-addition designs by deriving a measure based on the alias matrix. This provides a robustness-against-bias rationale for the ranking criterion. Third, we consider cases in which the sequence of some steps are fixed. Prior studies have largely ignored such applications, focusing instead on cases where all $m!$ permutations of the steps are deemed relevant. When this is not the case, the required design sizes are reduced and the corresponding models are simpler. Lastly, we present an analysis of two recent order-of-addition experiments.

2. Modeling Order Effects, Given All $m!$ Sequences

2.1. Van Nostrand's pairwise order-of-addition model

Suppose we number the components $1, 2, \ldots, m$, and denote feasible sequences by permutations of these numbers. Using this notation, Table 1 displays the 24 possible sequences for a four-component experiment, adding one component at a time. For any pair of components I and J, define the PWO factor $X_{I.J} = 1$ if I precedes J, -1 if J precedes I, and 0 if they are added simultaneously. Table 1 displays the six PWO factors. Van Nostrand (1995) model is

$$E(Y) = \beta_0 + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \beta_{i.j} X_{i.j}. \tag{2.1}$$

Voelkel (2019) refers to this as the *main effects PWO model*, because it contains no interactions; here, we refer to it as the *simple pairwise model*.

2.2. Lack of fit and the *triplets* order-of-addition model

There are six possible orderings for each set of three components. The main effects PWO model, having a constant and three PWO terms, would leave two degrees of freedom (df) for the lack of fit. Any two two-factor interactions of the PWO terms will account for this variation. If we had a full 2^3 factorial, there would be 3 df for two-factor interactions, and 1 df for the three-factor interaction. However, the 3! sequences result in a $3/4^{th}$ fraction of the 2^3 in the PWO factors, creating the following linear dependencies:

$$X_{I.J} - X_{I.K} + X_{J.K} = X_{I.J} X_{I.K} X_{J.K},$$

Table 1. Full sequence design for $m = 4$, with PWO factors.

Sequence	$X_{1.2}$	$X_{1.3}$	$X_{2.3}$	$X_{1.4}$	$X_{2.4}$	$X_{3.4}$	Y_1	Y_2*
1234	1	1	1	1	1	1	12	41.1
1243	1	1	1	1	1	-1	12	37.5
1324	1	1	-1	1	1	1	19.5	55.4
1342	1	1	-1	1	-1	1	17	56.5
1423	1	1	1	1	-1	-1	2	43.3
1432	1	1	-1	1	-1	-1	17	51.2
2134	-1	1	1	1	1	1	12	46.1
2143	-1	1	1	1	1	-1	12	27.8
2314	-1	-1	1	1	1	1	-3	39.5
2341	-1	-1	1	-1	1	1	2	46.4
2413	-1	1	1	-1	1	-1	32	34.4
2431	-1	-1	1	-1	1	-1	2	39.4
3124	1	-1	-1	1	1	1	4.5	53.5
3142	1	-1	-1	1	-1	1	2	51.2
3214	-1	-1	-1	1	1	1	4.5	50.8
3241	-1	-1	-1	-1	1	1	9.5	51.4
3412	1	-1	-1	-1	-1	1	7	52.9
3421	-1	-1	-1	-1	-1	1	7	53.4
4123	1	1	1	-1	-1	-1	22	39.1
4132	1	1	-1	-1	-1	-1	37	46.4
4213	-1	1	1	-1	-1	-1	22	37.2
4231	-1	-1	1	-1	-1	-1	-8	42.1
4312	1	-1	-1	-1	-1	-1	7	46.8
4321	-1	-1	-1	-1	-1	-1	7	41.8

* For Y_1, see Section 2.4. For Y_2, see Yang, Sun and Xu (2018).

Table 2. Full sequence design for $m = 3$, with PWO factors and interactions.

Sequence	$X_{1.2}$	$X_{1.3}$	$X_{2.3}$	$X_{1.2}X_{1.3}$	$X_{1.2}X_{2.3}$	$X_{1.3}X_{2.3}$	$X_{1.2}X_{1.3}X_{2.3}$
123	1	1	1	1	1	1	1
132	1	1	-1	1	-1	-1	-1
213	-1	1	1	-1	-1	1	-1
231	-1	-1	1	1	-1	-1	1
312	1	-1	-1	-1	-1	1	1
321	-1	-1	-1	1	1	1	-1

$$X_{I.J}X_{I.K} - X_{I.J}X_{J.K} + X_{I.K}X_{J.K} = 1.$$

These are easily verified by inspecting the six sequences; see Table 2. Thus, the three-factor interaction is in the column space of the PWO factors, and any two two-factor interactions make the third unnecessary.

We refer to the model with all $df_2 = m(m-1)/2$ PWO terms and any

full-rank set of $df_{3|2} = m(m-1)(m-2)/3$ two-factor interactions that share a common factor as the *triplets order-of-addition model*. For instance,

$$E(Y) = \beta_0 + \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\beta_{i.j}X_{i.j} + \sum_{i=1}^{m-2}\sum_{j=i+1}^{m-1}\sum_{k=j+1}^{m}[\beta_{i.j*i.k}X_{i.j}X_{i.k} + \beta_{i.j*j.k}X_{i.j}X_{j.k}]. \tag{2.2}$$

Note that two-factor interactions of PWO factors involve either three or four distinct components. The triplets model (2.2) only includes two-factor interactions that involve three distinct components, excluding interactions such as $X_{1.2}X_{3.4}$.

If the order-of-addition design consists entirely of pairs of opposite sequences, the resulting PWO treatment combinations will be mirror image pairs, and the design is a foldover design. As a result, each two-factor interaction that involves a shared component is orthogonal to the main effect PWO terms, but not orthogonal to the intercept.

Any design containing all $m!$ sequences, equally replicated, consists of mirror image pairs and, thus, enjoys the orthogonality property just described. For such designs, the variances of the estimators of the coefficients in the triplets order-of-addition model are simple functions of m and n. For $m = 3$, the estimators for the three PWO main effects and the two interactions all have the same variance, $1.5\sigma^2/n$, where n is the number of sequences (a multiple of $3! = 6$) and σ^2 is the common variance for Y at each given sequence. Voelkel (2019) reported that the main effects have variance inflation factors (VIF) of $3(m-1)/(m+1)$. Therefore, with $n = m!$ observations, these estimators have variance $3(m-1)\sigma^2/(m+1)!$. Furthermore, this remains true when fitting a triplet order-of-addition model, because the interactions are orthogonal to the main effects for the equally replicated $m!$ design. For $m > 3$, the estimators of the interaction terms in (2.2) have variance $3.75[(m-1)/(m+2)]\sigma^2/n$, as verified by inspection up to $m = 10$. Thus, for $m > 3$ and the full triplet model, the interaction estimators have variances that are slightly larger than those of the PWO main effect estimators. For reduced models with some terms eliminated, the variances are less; see Section 2.4.

2.3. Higher-order terms

We have already seen that two-factor interactions involve either three or four components. In the previous subsection, those involving just three components were included in model (2.2). Terms that depend on four components, such as $X_{1.2}X_{3.4}$, $X_{1.2}X_{1.3}X_{1.4}$, and $X_{1.2}X_{2.3}X_{1.4}$, would be considered next if the

triplets model showed a lack of fit. Note that three-factor interactions such as $X_{1.2}X_{1.3}X_{2.3}$ are ignored, because these lie within the column space of the main effects PWO model.

A saturated model for $m = 4$ has 23 df after the intercept. Because the triplets model has $df_2 + df_{3|2} = 14$ degrees of freedom, there are $23 - 14 = 9$ df to account for variation not attributable to the ordering of triplets of components. We have these 9 df for each set of four components, and thus define $df_{4|3} = 9\binom{m}{4}$. In general, the sequential degrees of freedom for each order may be obtained using a rencontres series (Riordan (1958, p.65)). Define $a_2 = 1$ and $a_r = r*a_{r-1}+(-1)^r$, such that $a_3 = 2$, $a_4 = 9$, $a_5 = 44$, and so on. These are the degrees of freedom for each additional order for combinations of r components. Thus, for $r = 2, \ldots, m$,

$$df_{r|r-1} = a_r \binom{m}{r}.$$

For instance, for $m = 5$, the degrees of freedom are partitioned as 10, 20, 45, and 44, which total 119. Interestingly, for any m, the last two counts always differ by ± 1.

2.4. Simple numerical example

The Y_1 column of Table 1 is obtained as follows. Suppose the four components have the following effects, starting with $Y_0 = 1$ before any components: 1: add 10; 2: subtract 5; 3: multiply by 4; 4: divide by 2. The response is maximized for the sequence 4132, resulting in $Y_1 = \{[(1/2) + 10] \times 4\} - 5 = 37$, and minimized for the sequence 4231, for which $Y_1 = \{[(1/2) - 5] \times 4\} + 10 = -8$.

The fitted parsimonious model requiring only four PWO main effects and two interactions is

$$\begin{aligned} Y_1 =& 11.375 + 11.250X_{1.3} - 6.250X_{1.4} - 3.750X_{1.3}X_{1.4} \\ & - 5.625X_{2.3} + 3.125X_{2.4} + 1.875X_{2.3}X_{2.4}. \end{aligned} \tag{2.3}$$

This model fits perfectly, exactly reproducing the values of Y_1 in Table 1. If a random error were added to the response, the true variances for the corresponding main effect estimators would be $1.4\sigma^2/24$, and for the interactions would be just $1.2\sigma^2/24$. If we were fitting the full triplets model, the constants in the numerators of these variances would be 1.8 and 1.875, respectively, as given by the formulae in Section 2.2.

Note how the three coefficients for subtraction are opposite in sign and half the magnitude of those for addition in (2.3). This makes sense because

the quantity subtracted (5) is half of the quantity added (10). Interpretation of the main effects is straightforward. The positive coefficient for $X_{1.3}$ ($X_{2.4}$) indicates the benefit of adding 10 before multiplying by 4 (subtracting 5 before dividing by 2). The negative coefficients for $X_{2.3}$ and $X_{1.4}$ indicate that subtracting-before-multiplication and addition-before-division both reduce the response. The $X_{1.3}X_{1.4}$ interaction shows that the addition-before-multiplication benefit is greater when $X_{1.4} = -1$ (i.e., when division precedes addition). Interpreting the $X_{2.3}X_{2.4}$ interaction, the subtraction-before-multiplication effect is greater in magnitude when subtraction follows division ($-5.625 - 1.875 = -7.5$) than when subtraction precedes division ($-5.625 + 1.875 = -3.75$). Equivalently, consider how the order for subtraction and multiplication ($X_{2.3}$) influences the effect associated with $X_{2.4}$. Conditional on subtracting before multiplying ($X_{2.3} = 1$), the coefficient of $X_{2.4}$ is $3.125 + 1.875 = 5$; conditional on subtracting after multiplying ($X_{2.3} = -1$), $X_{2.4}$'s effect is $3.125 - 1.875 = 1.25$. This discussion illustrates the interpretability of the triplet interaction terms.

2.5. Proposed analysis procedure

We now consider how to analyze a full $m!$ design, using the data in Table 3 of Yang, Sun and Xu (2018); see the last column of our Table 1. Table 3 provides the mean squares for the main effects PWO model, and for each additional order. Here, because the triplets do not add anything useful, we tentatively conclude that Van Nostrand's simple pairwise model should suffice, and proceed to identify which PWO factors to include. The main effects PWO model, with $R^2 = 83.25\%$, is

$$\hat{Y} = 45.22 + 1.21X_{1.2} + 0.04X_{1.3} - 3.61X_{2.3} - 0.40X_{1.4} - 1.19X_{2.4} + 3.98X_{3.4}.$$

By backward elimination, $X_{1.3}$, $X_{1.4}$, and $X_{1.2}$ are removed sequentially, with p-values 0.96, 0.66, and 0.19, respectively; this implies that it does not matter when drug 1 is administered. One more term ($X_{2.4}$) has a p-value of 0.078; removing this, the reduced model, with $R^2 = 78.2\%$, is $\hat{Y} = 45.22 - 4.70X_{2.3} + 3.07X_{3.4}$, with p-values < 0.001 for both remaining terms. Thus, to maximize the response, drug 3 should precede drugs 2 and 4. The eight sequences in which this is true all yielded a coded response above 50, while only one of the other 16 sequences achieved such response.

When the main effects PWO model holds, Yang et al.'s component-position effects plot (their Figure 1) should show mean responses that are approximately

Table 3. Basic ANOVA for Yang, Sun and Xu (2018).

Source	df	SS	MS
Pairs	6	1,051.274	175.21
+Triplets	8	117.236	14.65
+Quadruplets	9	94.303	10.48
Total	23	1,262.813	

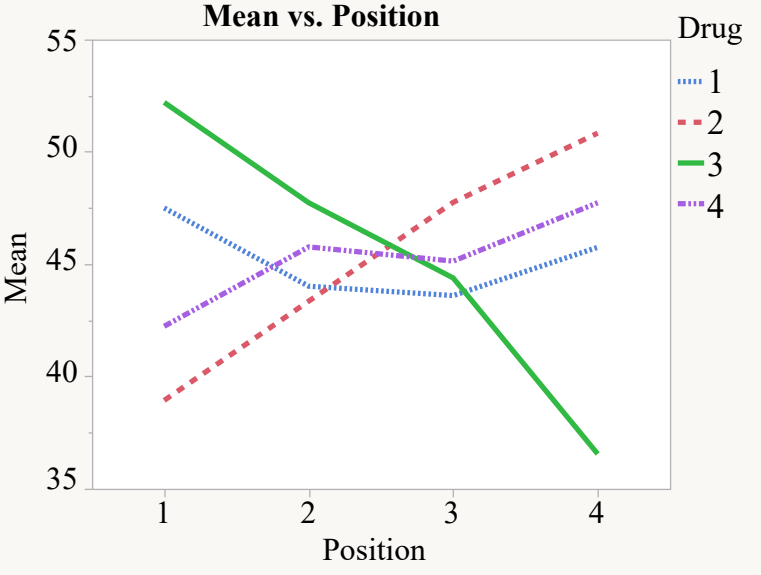

Figure 1. Component-position effects plot for 24-run $m = 4$ study.

linear when we have a full $m!$ design; see Figure 1. Only drug 1 appears to depart significantly from linear trend, but we have already seen that its position has no noticeable effect.

Yang, Sun and Xu (2018) fit their position-component model for Y_2; their model contains $(m-1)^2 = 9$ terms and achieves $R^2 = 84.43\%$. However, the Vuong (1989) test for non-nested models indicates that the simple PWO model (2.1) is significantly closer to the truth than the position-component model is, with p-value = 0.0152 using the Schwarz BIC adjustment for the three additional parameters. In the next section, we analyze a five-drug experiment where several triplet model terms are needed to explain the response.

3. Using Fractions of Full m! Designs: A Five-Drug Example

Table 4 presents a five-drug experiment conducted in two blocks of 20 runs each. Y denotes a coded, larger-the-better response. The runs have been sorted in ascending order within each block. Although the lack of similarity in the sequences with the largest responses suggests that a complicated model is likely required, we anticipate that beginning with either drug 3 or drug 5 will be preferred. Note too that the first block tended to yield better results, though we do not explore this further here. The PWO model, including a block main effect, has model sum of squares (SS) of 1,060.38 ($R^2 = 61.03\%$, RMSE = 4.92); the position-component model, with six additional parameters, produces a modest improvement ($R^2 = 72.62\%$, RMSE = 4.65). Vuong's test (with the BIC adjustment for the number of parameters) indicates a significant difference between these two models, favoring the PWO model (p-value = 0.0215) as closer to the truth. If we use forward selection with Bonferroni-adjusted p-values (Westfall, Young and Lin (1998)), after forcing in the block effect, we would add at most three PWO terms, because the adjusted p-value for the next term is 7×0.186; see Table 5. This simple model indicates that drug 3 should precede drug 4 ($b_{3.4} = 3.22$), and that drugs 1 and 2 should both follow drug 5 ($b_{1.5} = -1.79$, $b_{2.5} = -1.80$).

What about higher-order terms involving triplets? Partitioning the sum of squares as we did in Table 3 for the four-drug example, we find clear evidence for the need of triplet terms, though 1 df for triplets is lost as a result of the confounding with blocks (see Table 6). Using the remainder mean square as the denominator for the F-ratios, we are persuaded that the main effects PWO model cannot account for all systematic effects caused by drug ordering; that is, some interactions are needed (p-value = 0.0056).

The next step is to determine which triplets indicate useful interaction terms. One way to proceed with such an analysis is to use forward selection, using nominal effects for the 10 triplets as the eligible terms. Starting with a model containing the block effect and all PWO main effects, the results are shown in Table 7. Using Bonferroni adjusted p-values to reduce the chance of over-fitting the data, we have evidence for triplets 235 and 135, because their adjusted p-values are less than 0.20. Note that triplet 345 had the second smallest unadjusted p-value in each step: 0.0193, 0.1331, and 0.0828 in steps 1, 2, and 3, respectively. No other triplet had an unadjusted p-value of less than 0.125 in any of the steps.

Forward selection with 30 individual interactions yields results similar to

Table 4. Five-drug, 40-run experiment.

Block	Sequence	Y	Block	Sequence	Y
1	42135	4.93	2	42351	5.53
1	21345	13.63	2	21453	7.72
1	41253	15.57	2	12435	10.96
1	43512	18.47	2	24315	12.09
1	54123	19.50	2	41532	13.84
1	12543	20.23	2	14523	16.25
1	24531	21.47	2	15342	16.37
1	15234	21.59	2	43125	17.97
1	13425	23.55	2	53412	19.71
1	14352	23.61	2	54231	20.35
1	23154	23.85	2	25134	20.40
1	45321	25.23	2	13254	22.06
1	53241	25.62	2	32514	22.35
1	32451	26.08	2	31245	23.37
1	51432	26.75	2	45213	23.40
1	25413	28.38	2	52143	24.31
1	34215	29.43	2	23541	24.65
1	35142	30.52	2	34152	25.99
1	31524	31.27	2	35421	26.30
1	52314	31.96	2	51324	26.49

Table 5. Five-drug example: Forward selection for PWO terms.

Step	Parameter	p-value	No. Eligible	Adj. p-value	RSquare
1	$X_{3.4}$	0.0008	10	0.0084	0.3337
2	$X_{2.5}$	0.0065	9	0.0588	0.4590
3	$X_{1.5}$	0.0324	8	0.2589	0.5263

Table 6. Five-drug example: ANOVA.

Source	df	Sequential SS	Mean Sq.	F	p-value
Block	1	166.22	166.22	28.714	0.0005
2 | Block	10	894.16	89.416	15.446	0.0002
3 | 2, Block	19	624.96	32.893	5.682	0.0056
Remainder	9	52.10	5.789		
Total	39	1737.44			

those shown in Table 7. The first four interactions to be added are $X_{2.3}X_{3.5}$, $X_{1.3}X_{1.5}$, $X_{1.2}X_{1.5}$, and $X_{3.4}X_{3.5}$, with unadjusted p-values 0.0035, 0.0054, 0.0206, and 0.0276, respectively; the next step had unadjusted p-value > 0.05. The model with just two interactions still exhibits a possible lack of fit, based on a test using MSE = 5.789 from Table 6 as the error mean square. This leads us to consider

Table 7. Five-drug example: Forward selection for adding triplets to simple PWO model.

Step	Triplet	Adjusted p-value	RSquare
1	{235}	$10 * 0.0133 = 0.133$	0.721
2	{135}	$9 * 0.0215 = 0.193$	0.797
3	{125}	$8 * 0.0754 = 0.603$	0.840

two models, one with just two interactions, and a fuller model with four interactions. Table 8 displays the fuller model with $R^2 = 86.4\%$, 24 df for error, and RMSE = 3.14. Without the last two interactions, $R^2 = 79.1\%$ and RMSE = 3.73, with $b_{2.3*3.5} = -2.06$ and $b_{1.3*1.5} = -1.95$. Note how much larger in magnitude $b_{2.3*3.5}$ is in this model than in the Table 8 model.

The primary use of these models is to identify which sequences are likely to yield the highest average response. The best observed response was for sequence 52314, but our models do not commend this sequence. Sequence 52314 was ranked 15th (29th) out of 120 sequences based on the model with two (four) interactions. Instead, the models with two and four interactions both rank the sequences 31524 and 35214 as best and second best; sequence 31524 was the second-best observed response, while 35214 does not appear in the Table 4 design. The main effects PWO model with the batch effect does not agree with either of the models with interactions, ranking 31524 (35214) as 12th (3rd) best. Thus, the interactions improve the fit and change which sequences have the highest predicted responses.

4. Robustness of Order-of-Addition Orthogonal Arrays

Voelkel (2019) introduced the concept of order-of-addition orthogonal arrays, as follows: a design is an *order-of-addition orthogonal array* (OofA-OA) of strength t if, for every subset of size t in the PWO factors, the frequencies are proportional to those of the full $m!$ design. Voelkel showed that strength-2 (3) OofA-OAs only exist for run sizes that are integer multiples of 12 (24). Thus, for $m = 4$, the only OofA-OA's (with $n < 24$) are half-fractions. It can be shown that any 12-run OofA-OA is isomorphic to one of Voelkel (2019) two 12-run fractions for $m = 4$.

Although every strength-2 OofA-OA is fully efficient for the PWO model, they generally have correlations between the main effects and the two-factor interactions. Any design that consists of foldover pairs, including the full $m!$ design, does not have these correlations, but is not necessarily fully efficient for the PWO model. (The 40-run design in Table 4 does not consist of foldover pairs, and each

Table 8. Five-drug example: Fitted model with four interactions ($R^2 = 86.4\%$, MSE = 9.87).

Term	Estimate	Std Error	t Ratio	Prob> $\|t\|$	VIF
Intercept	27.54	1.71	16.10	<0.0001	
batch	-4.78	1.07	-4.47	0.0002	1.16
$X_{1.2}$	0.00	0.76	0.00	0.9980	2.34
$X_{1.3}$	-0.95	0.74	-1.27	0.2154	2.24
$X_{1.4}$	0.55	0.76	0.72	0.4778	2.33
$X_{1.5}$	-1.41	0.75	-1.87	0.0734	2.29
$X_{2.3}$	-1.46	0.75	-1.93	0.0653	2.30
$X_{2.4}$	0.57	0.77	0.74	0.4664	2.40
$X_{2.5}$	-1.57	0.75	-2.10	0.0469	2.27
$X_{3.4}$	1.74	0.74	2.34	0.0277	2.24
$X_{3.5}$	-0.10	0.75	-0.14	0.8896	2.27
$X_{4.5}$	-0.76	0.75	-1.01	0.3225	2.28
$X_{2.3} * X_{3.5}$	-0.93	0.63	-1.48	0.1507	1.45
$X_{1.3} * X_{1.5}$	-2.35	0.57	-4.09	0.0004	1.22
$X_{1.2} * X_{1.5}$	1.71	0.65	2.63	0.0145	1.43
$X_{3.4} * X_{3.5}$	1.35	0.57	2.35	0.0276	1.22

effect estimate is correlated with every other estimate.)

One common design diagnostic is the alias matrix. For main effect models, the alias matrix is commonly constructed by taking all two-factor interactions as the omitted effects. Let D_1 be the $n \times df_2$ matrix of PWO factors, and let $[D_2\ X_2]$ be the matrix of all two-factor interactions, partitioned into those that share a common component and those that do not. That is, D_2 has $m(m-1)(m-2)/2$ columns ($>$ rank(D_2)), whereas X_2 has $m!/[8(m-4)!]$ columns. Define the alias matrix as

$$A = (D_1'D_1)^{-1}D_1'[D_2,\ X_2] = [(D_1'D_1)^{-1}D_1'D_2, (D_1'D_1)^{-1}D_1'X_2] = [A_3, A_4]. \tag{4.1}$$

If A were zero, as for the full $m!$ design, then omitting the two-factor interactions would not bias the main effect estimates, and their inclusion would not change those estimates or increase their VIFs. We prefer the coefficients of A to be small. One scalar measure of size is the sum of the square of all elements of A:

$$\text{trace}(A'A) = \text{trace}(A_3'A_3) + \text{trace}(A_4'A_4). \tag{4.2}$$

If the columns of D_1 were orthogonal to one another, then (4.1) would be a matrix of the correlations between all PWO effects and their two-factor interactions. Furthermore, (4.2) would be the sum of the squares of these correlations. For

regular 2^{k-p} fractions, (4.2) is simply three times the number of length-3 words in the defining relation, whereas for nonregular fractions of strength 2, (4.2) is $3B_3$, where B_3 is the leading term in the generalized word length pattern (gwlp). Thus, one motivation for the G_2-aberration criterion (Tang and Deng (1999)), which is based on sequentially minimizing the elements of the gwlp, is to minimize the bias caused by omitted interactions. Voelkel (2019) Sim_3 is equivalent to the ranking based on B_3; here, the criterion (4.2) is preferred because it takes into account the lack of orthogonality for the columns of D_1, and because it enables us to distinguish between two-factor interactions associated with triplets versus those with quadruplets of components.

For full $m!$ order-of-addition designs, and any fraction that consists of foldover pairs, (4.1) is a matrix of zeros. However, where there does not exist an $(n/2) \times k$ matrix E such that $D_1 = [E'; -E']'$, generally $\text{trace}(A'A) > 0$ and $B_3 > 0$. We opt to assess the potential bias for strength-2 OofA-OA and other small fraction designs using $\text{trace}(A_3'A_3)$ and $\text{trace}(A_4'A_4)$.

Voelkel (2019) 12-run Design 1 for $m = 4$, with $\text{trace}(A'A) = 12.96 = 12.48 + 0.48$, is such that the three two-factor interactions involving all four components, if present, would cause little bias to the PWO estimates. However, Design 2, with $\text{trace}(A'A) = 15.84 = 9.6 + 6.24$, is less susceptible to bias caused by omitting triplet effects. Thus, which design is more robust depends on whether active two-factor interactions are associated with triplets or quadruplets.

We also compute (4.2) for three strength-2 OofA-OA designs from Voelkel (2019) with $m = 5$ and $n = 24$ orderings. By this measure, his Design 65, with $\text{trace}(A'A) = 24.\bar{5} = 14.\bar{3} + 10.\bar{2}$, is dominated by both Design 5 [$\text{trace}(A'A) = 20.5 = 14 + 6.5$] and Design 33 [$\text{trace}(A'A) = 18.61 = 9.5 + 9.1$]. Design 33 is preferred over Design 5, both for triplet model robustness ($9.5 < 14$) and with respect to all two-factor interactions ($18.61 < 20.5$). The ranking of these designs based on $\text{trace}(A'A)$ coincides with the ranking in Voelkel's Table 4, based on his minimum moment aberration criterion (Sim_3), as well as his average ($\chi^2_{ave,3}$) and fraction with perfect ($F0_3$) balance for sets of three PWO factors.

Augmenting any of these 24-run OofA-OA by folding over would produce an alias matrix $A = 0$. However, none of the foldovers of Designs 5, 33, or 65 produce a design that supports an estimation of the triplets model (2.2). For instance, the foldover of Design 65 has 48 distinct treatment combinations, but $\text{rank}(D_2) = 19$, which is less than 21, as required. Yet, there do exist OofA-OA$(n = 24, m = 5, t = 2)$ for which the $n = 48$ foldover has $\text{rank}(D_2) = 21$. One such design is given in Table 9. Its D-efficiency for the triplets model, relative to

Table 9. OofA-OA(24,5,2) for which foldover supports estimation of the triplet model.

12354	31425	43251
12453	31542	45213
13524	32145	51243
14352	32541	51342
21543	34512	52314
23415	41253	53241
24351	41532	54231
25134	42135	54312

the full 5! design, is only 83.76%, and its largest VIF is 9.47. In addition, although foldover designs are fully efficient for main effects and the alias matrix (4.1) is null, they have at most $n/2 - 1$ degrees of freedom for two-factor interactions. As a result, if two-factor interactions involving four components are active, but omitted from the model, these will bias estimates for the two-factor interactions involving three components.

We can improve on the D-efficiency of the Table 9, $n = 48$ triplets design by increasing the run size. SAS's Proc Optex readily generates D-optimal designs of size 31 or 36 for the triplets model ($n = 31$ is the minimum size for this model). Two advantages of $n = 36$ are that this size may result in an OofA-OA, and its foldover is potentially able to estimate all two-factor interactions, even those involving four components. Repeated searches using Proc Optex yielded triplet designs of size 31 and 36 with D-efficiencies of 79.4% and 87.2%, relative to the full $5! = 120$-run design. Note that the 36-run has a higher D-efficiency than the Table 9 design with 48 runs. Folding over these 31- and 36-run designs yields D-efficiencies of 92.1% and 95.4%, respectively, for the resulting 62- and 72-run designs. The online Supplementary Material lists each of these designs.

For triplet designs, the alias matrix should be redefined to measure the potential bias caused by omitting quadruplet terms, corresponding to the columns of X_2 and D_3, where D_3 is the matrix of three-factor interactions that involve exactly four components. There are $16m!/[4!(m-4)!]$ columns in D_3. Ignoring the block effect, the Table 4 design is very efficient for estimating the triplet model (2.2). Its D-efficiency, relative to that of the full 5! design, is $0.6004/0.6613 = 90.8\%$. An alias matrix for this design would entail 30 rows, one for each effect in the full triplet model, and 95 columns: 15 for X_2, and 80 for D_3.

5. Applications with Constrained Orderings

Consider an assay that involves four components, where the first two can be combined in either order, and the last two can be combined in either order. However, the second pair always follows the first pair. In such a case, rather than having 4! orderings to consider, there are only four possible orderings: 1234, 2134, 1243, and 2143. To generalize this, suppose that the components appear in g groups, with $m_1, m_2, \ldots, m_g$ components, respectively. The number of possible sequences, assuming no constraints within a group, is $m_1!\ m_2! \cdots m_g! < (m_1 + m_2 + \cdots + m_g)!$. Depending on the application, the groups may be performed sequentially, or in parallel and then combined. Let I be a component from group 1 and J a component from group 2. If group 1 always precedes group 2, then $X_{I.J} = +1$ for all runs. However, if the groups are done in parallel and then combined, $X_{I.J} = 0$ for all runs. As long as the relationship between the groups is fixed, this does not impact which terms are estimable, because $X_{I.J}$ is constant in either case.

The degrees of freedom for the main effects PWO model with g groups is $df_2 = \sum_{i=1}^{g} 0.5 m_i(m_i - 1)$. For instance, if $m_1 = m_2 = 2$ and $m_3 = 3$, there are only 24 sequences rather than 7! = 5,040 for the unconstrained case. Given these constraints, the PWO model has just 5 df, whereas the triplets model would add just 2 df more to account for the lone triplet in Group 3. Thus, such restrictions greatly reduce the number of orderings to be considered, and will often make it reasonable to perform all of the reduced set of feasible sequences.

In many applications, such as in mixing pesticides, the sequence may be largely specified. In this case, experiments are used only to vary the order within small groups of steps. The advantage of such restrictions is that sequences known to be incompatible or inferior are avoided. The resulting, smaller experiments are focused on answering a more limited set of questions involving order within groups.

6. Applications with Additional Factors

Consider an assay where, in addition to different orderings of the components, we vary the quantities of components and the incubation times following each addition. Varying the quantities makes this a mixture proportion experiment, whereas varying the incubation times includes additional quantitative factors. If the order of addition were fixed, such an experiment would be a mixture design with additional quantitative factors; see Piepel and Cornell (1994). To the best

of our knowledge, the mixture experiment literature has not considered order-of-addition aspects, but this is a natural extension.

Let O denote the set of feasible addition orderings expressed in terms of the PWO factors, let P denote the set of component mixture proportions to be considered, and let F be the treatment combinations for the other numerical factors. Then, an experiment involving order of addition, component proportions, and other factors might conceivably sample from the set of conditions formed by the product array $O \times P \times F$. Taking this as a candidate set of points, one might use standard software such as SAS' Proc Optex to select an efficient design for a given specified model. See Voelkel (2019, Sec.6) for design construction examples from $O \times F$.

7. Conclusion

This study extends the modeling of order-of-addition experiments beyond the simplistic model that assumes that the response depends additively on pairwise effects. The triplets model provides an extension to the standard PWO model for applications where Van Nostrand's model exhibits a lack of fit. If necessary, interactions involving four components (terms in a quadruplets model) can be added, although our examples do not demonstrate a need for higher-order terms than two factor interactions sharing a common component. The alias matrix-based measure (4.2) improves on one of Voelkel (2019) criteria for ranking order-of-addition designs. Section 5 showed how the standard PWO model and the triplets model are simplified for cases where the steps can be partitioned into groups, with the order of groups fixed, and only the order within groups varying. Most of this article has focused on proposing more versatile models and comparing the fit of these models with other models in the literature for actual order-of-addition experiments. Given that these new models are useful, more work is needed related to design construction, the results of which will be reported elsewhere.

Supplementary Material

The online Supplementary Material contains the D-efficient designs mentioned in Section 4.

Acknowledgement

This work was supported in part by the Tianjin Regional Thousand Talents program, by the National Natural Science Foundation of China (Grant No. 11431006), and the Tianjin "131" Talents Program.

References

Ding, X., Matsuo, K., Xu, L., Yang, J. and Zheng, L. (2015). Optimized combinations of bortezomib, camptothecin, and doxorubicin show increased efficacy and reduced toxicity in treating oral cancer. *Anti-Cancer Drugs* **26**, 547–554.

Piepel, G. F. and Cornell, J. A. (1994). Mixture experiment approaches – examples, discussion, and recommendations. *J. of Quality Techn.* **26**, 177–196.

Riordan, J. (1958). *An Introduction to Combinatorial Analysis*. Wiley, New York.

Scheffe, H. (1958). Experiments with mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **20**, 340-366.

Tang, B. and Deng, L. Y. (1999). Minimum G_2-aberration for nonregular fractional factorial design. *Ann. Statist.* **27**, 1914–1926.

Van Nostrand, R. C. (1995). Design of experiments where the order of addition is important. In *ASA Proceedings, the Phys. and Eng. Sci. Section*, 155–160. Alexandria, VA: American Statistical Association.

Voelkel, J. G. (2019). The design of order-of-addition experiments. *J. of Quality Techn.* **51**, 230-241.

Vuong, D. K. J. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307–333.

Wagner, J. J. (1995). Sequencing of feed ingredients for ration mixing. CATTLE 95-14. *SD Beef Report. SDAES, Brookings*, 52–54.

Westfall, P. H., Young, S. S. and Lin, Q. H. (1998). Forward selection error control in the analysis of supersaturated designs. *Statistica Sinica* **8**, 101–117.

Yang, J-F., Sun, F. and Xu, H. (2018). A component-position model, analysis and design for order-of-addition experiments. Preprint.

School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tianjin 300071, China.

Department of Business Analytics and Statistics, University of Tennessee, Knoxville, TN 37996-0525, U.S.A.

E-mail: rmee@utk.edu

(Received July 2017; accepted September 2018)