Statistica Sinica Preprint No: SS-2020-0178	
Title	A Position-Based Approach for Design and Analysis of Order-of-Addition Experiments
Manuscript ID	SS-2020-0178
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0178
Complete List of Authors	Zack Stokes and Hongquan Xu
Corresponding Author	Hongquan Xu
E-mail	hqxu@stat.ucla.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

A Position-Based Approach for Design and Analysis of Order-of-Addition Experiments

Zack Stokes and Hongquan Xu

University of California, Los Angeles

Abstract: In many physical and computer experiments, the order in which the steps of a process are performed may have a substantial impact on the measured response. Often, the goal in these situations is to uncover the order that optimizes the response according to some metric. However, the brute force approach of performing all permutations quickly becomes impractical as the number of components in the process increases. Instead, we seek to develop order-of-addition experiments that choose an economically viable subset of permutations to test. The statistical literature on this topic is sparse, and many researchers rely on ad-hoc methods to study the effect of process order. In this work, we present a series of novel developments, including a modeling framework that exploits certain structures of the data, a method for constructing optimal designs under this proposed framework, and an evaluation of the performance and robustness of the constructed designs. We use data from a drug combination therapy problem to highlight the benefits of our approach.

Key words and phrases: Experimental design, drug combination experiment, generalized minimum aberration, Latin square, optimal design, orthogonal array.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

1. Introduction

In many experiments, the order in which a process is executed or components are added can have a substantial impact on the response. Researchers must therefore consider this effect when designing their experiments, or they run the risk of producing sub-optimal conclusions. However, the combinatorial explosion that occurs in experiments with more than a few components quickly renders running a trial for every permutation impractical. The solution to this problem is to design order-of-addition experiments in which the goal is to choose an appropriate subset of all possible permutations, such that the study objective can be appropriately met, while satisfying the computational and financial constraints.

Order-of-addition experiments are popular for studying physical and simulated phenomena in areas such as medical science, pharmaceutical science, bio-chemistry, nutritional science, food science, and mechanics and engineering; see Lin and Peng (2019) for a review of these applications. We encounter order-of-addition experiments in both past and present drug combination projects. Combination chemotherapy has become common in cancer treatment, viral infection eradication, and super bacteria inhibition (Ding et al. (2013), Jaynes et al. (2013), Ding et al. (2015), Silva et al. (2016), Xiao et al. (2019)). However, a major limitation of the current

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

techniques is that drugs are added simultaneously and the drug sequence is not considered. The drug sequence often plays a major role in deciding the endpoint efficacy because the early addition of certain drugs could prepare the biological system to better accept or defend the later drugs. Pre-clinical and clinical studies indicate that the drug sequence is important to improving the effect of the treatment (MacBeath and Yaffe (2012), Wang et al. (2020)).

Nevertheless, references for the design and modeling of such experiments are rather primitive. Traditional factorial designs and orthogonal arrays cannot be used for order-of-addition experiments, because each run must be a permutation of the components, and existing methods fall short when working with complex, real data. In this work, Section 2 begins with an overview of the current order-of-addition literature. Next, Section 3 proposes new models and presents the results of applying them to the drug-sequencing problem discussed above in the context of treating lymphoma. Section 4 introduces a novel construction method, and covers general optimality results for a class of models. Section 5 evaluates the performance and properties of the designs from our algorithm and compares them with those of existing designs. This includes a study that demonstrates that the proposed designs are robust under algorithm tuning and model misspecifi-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

cation. Section 6 concludes the paper with a summary and discussion. All proofs are given in Section S1 of the Supplementary Material.

2. Background

We call each material or drug to be added in an experiment a component. If the experiment involves m components, denoted by $0, 1, \ldots, m-1$, then there are $m!$ possible permutations. Let $\mathbf{F}_m$ be the full design with $m!$ distinct rows and m columns, where each row is a permutation of m components. Performing all possible permutations quickly becomes unfeasible, even for experiments with five or more components. To save time and cost, it is necessary to choose a subset of the runs to perform. A natural question then arises of which subset to choose and how to model the response.

There have been a few recent studies on the design and analysis of order-of-addition experiments, as formulated above. Van Nostrand (1995) and Voelkel (2019) studied order-of-addition experiments by creating a set of pseudo-factors $\{I_{ij},\ 0 \leq i < j \leq m-1\}$, such that each corresponds to the pairwise ordering of the components. For example, in the case of $m = 4$ components, the six pairwise ordering factors are $I_{01}, I_{02}, I_{03}, I_{12}, I_{13}$, and I_{23}. Each factor I_{ij} has two levels, 1 and -1, indicating whether or not, respectively, component i is added before component j. Furthermore, they

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

considered the following pairwise ordering (PWO) model:

$$y = \beta_0 + \sum_{i<j} \beta_{ij} I_{ij} + \varepsilon, \tag{2.1}$$

with random error $\varepsilon \sim N(0, \sigma^2)$. Voelkel (2019) constructed optimal designs for this PWO model and employed the D-criterion to assess their properties and make comparisons. Peng et al. (2019) showed that the full design $\mathbf{F}_m$ is optimal for the PWO model under any concave and signed permutation invariant criterion. The authors also constructed a class of fractional designs that are optimal under these same conditions. However, their designs often have an excessive number of runs, and may be less useful in practice. Zhao et al. (2020) constructed minimally supported designs for the PWO model containing only one point per parameter. Mee (2020) extended the PWO model to include interactions of the pairwise ordering factors, which we briefly consider in Section 3. Lin and Peng (2019) provided a good summary of PWO models.

Yang et al. (2020) took a different approach to the problem by measuring the absolute position effects instead of the relative position effects. They framed the order-of-addition experiment with n runs and m components as a design matrix $\boldsymbol{A} = (a_{ij})$, where a_{ij} is the component added

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

in the jth position of the ith run. They constructed an indicator $z_{kj}^{(i)}$ for each component-position pair (k, j), such that $z_{kj}^{(i)}$ is one if $a_{ij} = k$, and zero otherwise. Because exactly one component is used at each position, we have $\sum_{k=0}^{m-1} z_{kj}^{(i)} = 1$ for any i and j. Thus, $m-1$ contrasts are needed to represent the effects of m components for each position. Because each run is a permutation of m distinct components, we also have $\sum_{j=1}^{m} z_{kj}^{(i)} = 1$ for any i and k. As a result, we can only include $m-1$ positions in the model. Given these constraints, an appropriate regression model, called the component-position (CP) model, is

$$y = \gamma_0 + \sum_{k=1}^{m-1} \sum_{j=1}^{m-1} z_{kj}\gamma_{kj} + \varepsilon, \tag{2.2}$$

where y is the response, γ_0 is the intercept, z_{kj} is an indicator for the component-position pair (k, j), as described above, γ_{kj} is the parameter representing the effect of component k being added at the jth position, and ε is an independent normal random error. Yang et al. (2020) also proposed the following class of designs for the CP model (2.2).

Definition 1. An $n \times m$ matrix with entries from $\{0, 1, \ldots, m-1\}$ is called a *component orthogonal array* (COA) of n runs and m factors if each row is a permutation of $\{0, 1, \ldots, m-1\}$ and, for any subarray of two columns,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

each level combination (i, j), with $i \neq j$ and $i, j = 0, 1, \ldots, m-1$, appears equally often. Such an array is denoted by COA(n, m).

By Definition 1, every level combination (i, j), with $i \neq j$ and $i, j = 0, 1, \ldots, m-1$, must appear equally often in every two-column sub-array of a COA(n, m). Thus, all designs must have $n = \lambda m(m-1)$, where λ is an integer. Indeed, COAs are orthogonal arrays of type I, as defined by Rao (1961). Hedayat et al. (1999) provided a comprehensive introduction to orthogonal arrays.

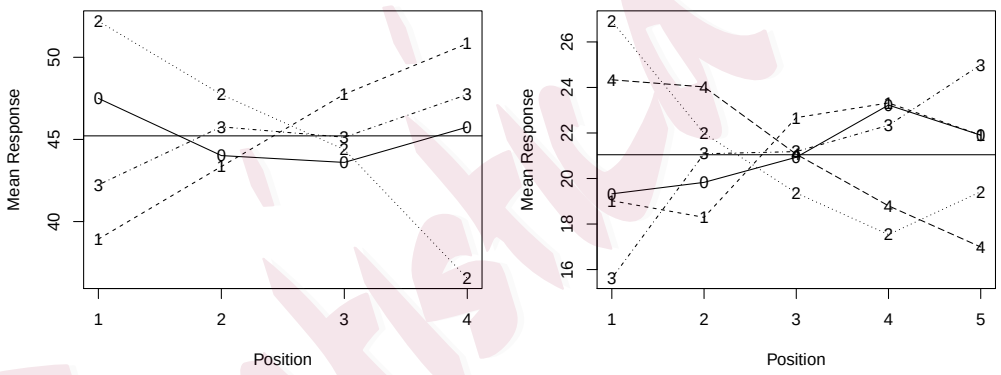

Figure 1: Component-position effects plots for four-drug (left) and five-drug (right) order-of-addition experiments.

Several of these recent works have focused on the aforementioned problem of choosing an optimal sequence for drug administration. Figure 1 shows the component-position effects plots for four-drug (left) and five-drug (right) order-of-addition experiments from Yang et al. (2020) and Mee

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

(2020). These experiments considered four and five chemotherapeutics, respectively, for treating lymphoma that have received FDA approval for clinical testing (Wang et al., 2020). Each drug was tested at a fixed dosage estimated from a preliminary dose-response study. For each sequence, a drug was administered every four hours in the four-drug study, and every three hours in the five-drug study. In each plot, the horizontal axis denotes the position at which a drug is added, and the vertical axis denotes the mean response, in this case, a measure of cancer cell inhibition 24 hours after the first drug was administered. Each point denotes the mean response of all runs in which the labeled drug is at the fixed position. For each drug, the m dots corresponding to m different positions are connected to visualize the trend as that drug is shifted to a later position in the sequence. The solid horizontal line, used as a reference, represents the average response of all observations. Both plots show that the effect of a drug on tumor inhibition depends on its position. The four-drug plot suggests that the component effects have a nearly linear relationship with the position. In this case, the authors found that both the PWO model (2.1) and the CP model (2.2) fit the data well, with predictive R^2 of 0.67 and 0.54, respectively. On the other hand, the five-drug plot suggests that the relationship between the component effects and the position is nonlinear. Neither model fits the data

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

well, with predictive R^2 of 0.20 and 0.09, respectively.

The two existing models do not fit the five-drug data well because they lack interaction terms. It is common in practice to find that a few two-factor interactions are also significant, in addition to the main effects. For this reason, Mee (2020) proposed a triplets order-of-addition model that expands the PWO model to include two-factor interactions involving exactly three distinct components. The triplets model has many more parameters than both existing models, so it requires a much larger run size to estimate. This is a major shortcoming. We hope to improve this body of literature by proposing new models and designs that can handle increasingly complex situations, such as the five-drug example, without requiring an excessive number of runs.

3. Flexible Position Models

Before presenting our proposed models, we give some notation and definitions. Given an $n \times m$ component matrix $\boldsymbol{A} = (a_{ij})$, where each row is a permutation of the components $0, 1, \ldots, m-1$, we define a new $n \times m$ matrix $\boldsymbol{B} = (b_{ik})$ as follows: $b_{ik} = j$ if $a_{ij} = k - 1$, for $k = 1, \ldots, m$. Note that a_{ij} is the component used at the jth position of the ith run, while b_{ik} is the position of component $k - 1$ in the ith run. For example, the left

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

side of row 14 in Table 1 indicates that the four components should appear in the order (2, 0, 3, 1), while the right-hand side equivalently states that positions (2, 4, 1, 3) should be assigned to components 0, 1, 2, and 3, respectively. Each row of $\boldsymbol{B}$ is a permutation of the m positions $1, \ldots, m$. To maintain the previous notation, we refer to $\boldsymbol{B}$ as the position matrix.

To compare our new models against the existing ones, we use the previously discussed four- and five-drug data presented in Yang et al. (2020) and Mee (2020), respectively. The data from these two experiments are given in Table 1 (matrices $\boldsymbol{A}$ and $\boldsymbol{B}$) and Table 2 (matrix $\boldsymbol{A}$), respectively. Recall from Section 2 that the existing methods are not sufficient for efficiently estimating the interaction effects between the drugs. We propose the following new, broader class of linear models based on the position matrix $\boldsymbol{B} = (b_{ik})$ that overcomes this weakness:

$$y = \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon, \tag{3.1}$$

where $\boldsymbol{x}$ is a row of the position matrix $\boldsymbol{B}$, $\boldsymbol{f}(\boldsymbol{x})$ is a vector of some basis functions, $\boldsymbol{\beta}$ is a vector of unknown coefficients, and $\varepsilon \sim N(0, \sigma^2)$ consists of independent normal errors. Using $\boldsymbol{B}$, we can represent the two existing models as special cases of this model. Specifically, the PWO model uses a set

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

Table 1: Design and data for a four-drug order-of-addition experiment.

	Components				Positions				
Run	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	y
1	0	1	2	3	1	2	3	4	41.1
2*	0	1	3	2	1	2	4	3	37.5
3*	0	2	1	3	1	3	2	4	55.4
4	0	2	3	1	1	4	2	3	56.5
5	0	3	1	2	1	3	4	2	43.3
6*	0	3	2	1	1	4	3	2	51.2
7*	1	0	2	3	2	1	3	4	46.1
8	1	0	3	2	2	1	4	3	27.8
9	1	2	0	3	3	1	2	4	39.5
10*	1	2	3	0	4	1	2	3	46.4
11*	1	3	0	2	3	1	4	2	34.4
12	1	3	2	0	4	1	3	2	39.4
13	2	0	1	3	2	3	1	4	53.5
14*	2	0	3	1	2	4	1	3	51.2
15*	2	1	0	3	3	2	1	4	50.8
16	2	1	3	0	4	2	1	3	51.4
17	2	3	0	1	3	4	1	2	52.9
18*	2	3	1	0	4	3	1	2	53.4
19*	3	0	1	2	2	3	4	1	39.1
20	3	0	2	1	2	4	3	1	46.4
21	3	1	0	2	3	2	4	1	37.2
22*	3	1	2	0	4	2	3	1	42.1
23*	3	2	0	1	3	4	2	1	46.8
24	3	2	1	0	4	3	2	1	41.8

Note: The 12 ($*$) runs were used in Example 1 to fit the models and compare the quality of the out-of-sample predictions.

of basis functions that return the sign of $b_k - b_l$ for each pair of components $k-1$ and $l-1$ when $\boldsymbol{x} = (b_1, \ldots, b_m)$. The CP model similarly includes one indicator function for every component-position pair (k, j). However, these methods do not take full advantage of the benefit provided by this

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

new position-based perspective.

Table 2: Design and data for a five-drug order-of-addition experiment.

Run	Components					y	Run	Components					y
1	3	1	0	2	4	4.93	21	3	1	2	4	0	5.53
2	1	0	2	3	4	13.63	22	1	0	3	4	2	7.72
3	3	0	1	4	2	15.57	23	0	1	3	2	4	10.96
4	3	2	4	0	1	18.47	24	1	3	2	0	4	12.09
5	4	3	0	1	2	19.5	25	3	0	4	2	1	13.84
6	0	1	4	3	2	20.23	26	0	3	4	1	2	16.25
7	1	3	4	2	0	21.47	27	0	4	2	3	1	16.37
8	0	4	1	2	3	21.59	28	3	2	0	1	4	17.97
9	0	2	3	1	4	23.55	29	4	2	3	0	1	19.71
10	0	3	2	4	1	23.61	30	4	3	1	2	0	20.35
11	1	2	0	4	3	23.85	31	1	4	0	2	3	20.4
12	3	4	2	1	0	25.23	32	0	2	1	4	3	22.06
13	4	2	1	3	0	25.62	33	2	1	4	0	3	22.35
14	2	1	3	4	0	26.08	34	2	0	1	3	4	23.37
15	4	0	3	2	1	26.75	35	3	4	1	0	2	23.4
16	1	4	3	0	2	28.38	36	4	1	0	3	2	24.31
17	2	3	1	0	4	29.43	37	1	2	4	3	0	24.65
18	2	4	0	3	1	30.52	38	2	3	0	4	1	25.99
19	2	0	4	1	3	31.27	39	2	4	3	1	0	26.3
20	4	1	2	0	3	31.96	40	4	0	2	1	3	26.49

Because positions have a natural order, we can study their effects using polynomial functions (e.g., Wu and Hamada (2009)). Such a model was proposed by Anderson-Cook and Lu (2019), but no framework or details were given. We define the orthogonal polynomials of degree 1 and 2 over

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

the set of positions as

$$p_1(x) = c_1\left(x - \frac{m+1}{2}\right) \text{ and } p_2(x) = c_2\left[\left(x - \frac{m+1}{2}\right)^2 - \left(\frac{m^2-1}{12}\right)\right],$$

respectively, where c_1 and c_2 are scalars that ensure that the length of each contrast vector is $\sqrt{m}$. For example, when $m = 4$, $c_1 = 2/\sqrt{5}$ and $c_2 = 2$, and $(p_1(x), p_2(x)) = (-1.5c_1, 1), (-0.5c_1, -1), (0.5c_1, -1)$, and $(1.5c_1, 1)$, for $x = 1$, 2, 3, and 4, respectively. When $m = 5$, $c_1 = \sqrt{1/2}$ and $c_2 = \sqrt{5/14}$, and $(p_1(x), p_2(x)) = (-2c_1, 2c_2), (-c_1, -c_2), (0, -2c_2)$, $(c_1, -c_2)$, and $(2c_1, 2c_2)$, for $x = 1$, 2, 3, 4, and 5, respectively.

The orthogonal polynomials have the following constraints:

$$(a)\ \sum_{x=1}^{m} p_j(x) = 0, \quad (b)\ \sum_{x=1}^{m} p_j^2(x) = m, \tag{3.2}$$

for $j = 1, 2$. These constraints complicate the modeling and the study of the design optimality for order-of-addition experiments, because each row of the position matrix $\boldsymbol{B}$ is a permutation of $\{1, \ldots, m\}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

Using these polynomials, we consider three specific models:

$$y = \beta_0 + \sum_{k=1}^{m-1} p_1(b_k)\beta_k + \varepsilon, \tag{3.3}$$

$$y = \beta_0 + \sum_{k=1}^{m-1} p_1(b_k)\beta_k + \sum_{k=1}^{m-1} p_2(b_k)\beta_{kk} + \varepsilon, \tag{3.4}$$

$$y = \beta_0 + \sum_{k=1}^{m-1} p_1(b_k)\beta_k + \sum_{k=1}^{m-2} p_2(b_k)\beta_{kk} + \sum_{1\le k<l\le m-1} p_1(b_k)p_1(b_l)\beta_{kl} + \varepsilon, \tag{3.5}$$

where y is the response, $b_1, \ldots, b_m$ are the positions of the m components, β_0 is the intercept, β_k, β_{kk}, and β_{kl} are unknown parameters, and $\varepsilon \sim N(0, \sigma^2)$ is a random error. We can interpret the main effect parameters as the expected change in the response after moving the specified component one position later in the sequence. Because each row of the position matrix is a permutation of $\{1, \ldots, m\}$ and the orthogonal polynomials obey the constraints in (3.2), we must remove one component effect from models (3.3) and (3.4) in order to make the models estimable. Furthermore, model (3.5) only includes β_{kk}, for $k = 1, 2, \ldots, m-2$, and removes any interaction terms involving component $m-1$. We can similarly craft more complicated models with higher-order terms, if needed. For convenience, we refer to models (3.3), (3.4), and (3.5) as the first-order, quadratic, and second-order

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

Table 3: Number of parameters of models for $m = 3$–10.

	m							
Model	3	4	5	6	7	8	9	10
PWO Model	4	7	11	16	22	29	37	46
CP Model	5	10	17	26	37	50	65	82
First-order Model	3	4	5	6	7	8	9	10
Quadratic Model	5	7	9	11	13	15	17	19
Second-order Model	5	9	14	20	27	35	44	54

position models, with $m - 1$, $2m - 1$, and $(m - 1)(m + 2)/2$ parameters, respectively.

Table 3 shows the number of parameters of five models for $m = 3, \ldots, 10$, including the PWO model and the CP model. The first-order and quadratic position models have fewer parameters than the others when $m > 4$. The second-order position model has a few more parameters than the PWO model, but has fewer parameters than the CP model as m increases. The new position models are both parsimonious and flexible. We demonstrate these traits using both drug sequencing experiments, each of which has two objectives: fitting an accurate model, and locating the optimal drug sequence.

Example 1. Consider the four-drug order-of-addition experiment in Table 1. We first fit the five models to the full data. The PWO and CP models have predictive R^2 of 0.67 and 0.54, respectively. The first-order, quadratic,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

and second-order position models have predictive R^2 of 0.69, 0.66, and 0.65, respectively. The root mean squared errors (RMSEs) for the PWO and CP models are 2.97 and 2.86, respectively, and 3.34, 3.00, and 2.67, respectively, for the position models. From this, we see that all five models have a similar goodness of fit. The first-order model with four parameters is the simplest and achieves the best predictive R^2 value.

To further compare the predictive accuracy of the models, we train each on the COA(12,4) given by the runs with $*$ in Table 1, and predict across all 24 sequences. The PWO and CP models have predicted versus observed correlations of 0.90 and 0.87, respectively, while the position models have correlations 0.87, 0.88, and 0.89, respectively. All models achieve comparable prediction accuracy, but the first-order model is able to do so with fewer parameters and has a better predictive R^2 when considering the full data set. Thus, for the simpler data set in which the relationship appears linear (Figure 1), our succinct models fit well and produce accurate predictions.

In order to interpret the position models, we first simplify each model (fit to all 24 runs) using forward and backward stepwise variable selection with respect to the AIC. We start from a constant model, and instead of removing the last effect, as in (3.3)-(3.5), we allow for the choice of any

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

effect. The resulting models are

$$\hat{y} = 45.22 + 2.03B - 5.55C - 1.81A, \quad (3.6)$$

$$\hat{y} = 45.22 - 1.81A + 2.03B - 5.55C + 1.41A^2, \quad (3.7)$$

$$\hat{y} = 44.68 - 1.81A + 2.03B - 5.55C + 0.98A^2 - 1.62AB, \quad (3.8)$$

where each drug has been replaced with a letter to make the conclusions clearer (e.g., A and A^2 represent the linear and quadratic effects, respectively, of drug 0). The predictive R^2 values for these three models are 0.69, 0.72, and 0.74, and the RMSEs are 3.34, 3.03, and 2.76, respectively. Further examination reveals that A, A^2, and A^2 and AB are not significant at the 5% level in models (3.6)-(3.8), respectively. After removing the nonsignificant terms, we have the reduced model $\hat{y} = 45.22 + 2.93B - 4.65C$.

In this model, the negative coefficient of drug C can be interpreted as the response being maximized when it comes earlier in the sequence, and the positive coefficient of drug B signifies that the response increases when it is placed later. These interpretations reflect the linear trends we see in the four-drug component-position effects plot in Figure 1. Mee (2020) and Yang et al. (2020) also performed a stepwise regression to simplify the PWO and CP models, respectively. Their simplified PWO and CP models

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

are comparable to models (3.6)-(3.8) in terms of their predictive R^2.

Example 2. Consider the five-drug order-of-addition experiment in Table 2. The experiment was conducted in batches. The first 20 runs were used in a batch, and the second 20 runs were used in another batch. After fitting each model to all 40 runs, including a block variable representing the batch effect, the PWO and CP models have predictive R^2 value of 0.20 and 0.09, respectively, and the first-order, quadratic, and second-order position models have predictive R^2 values of 0.44, 0.41, and 0.52, respectively. The RMSEs for the PWO and CP models are 4.11 and 3.45, respectively, and 4.18, 3.80, and 2.85, respectively, for the position models. The position models show a greater ability to capture the nonlinear trends present in Figure 1. The second-order model not only produces the overall best fit, but also generalizes well.

In order to improve interpretability and keep the final model concise, variable selection is used to choose the most appropriate to include effects from the second-order model. Starting with a constant model, forward and backward stepwise regressions are used to produce a model with a small AIC. Because the choice of which effects to remove from the position models was arbitrary, we allow for the selection of any linear, quadratic, or two-factor interaction effects, as in Example 1. We also allow for the

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

selection of a block effect that represents the two batches. With this in mind, the resulting model has a total of eight terms, has a predictive R^2 of 0.68 (larger than any competitor), and is given by

$$\hat{y} = 23.13 - 4.08\Delta + 3.19A + 3.45B + 4.49D + 1.05C^2 + 1.82BE - 1.64CE. \tag{3.9}$$

In this model, the block variable is given by Δ, and each drug is again represented by a letter to facilitate substantive conclusions. We see that the quadratic effect of drug C and two interactions involving drug E are included in the final model. Owing to the constraints in (3.2), we have $A + B + C + D + E = 0$. Therefore, if we replace A with $-B - C - D - E$, then we get an equivalent model that follows the effect hierarchy principle (Wu and Hamada, 2009).

A direct interpretation of these significant effects is complicated by the inclusion of C^2, BE, and CE, so we consider the top 10 predicted sequences: CEBAD, CAEBD, CEABD, CBEAD, CADEB, CEBDA, EBADC, CAEDB, CABED, and EBACD. While most of these sequences are not in the Table 2 design, the sequence CAEBD has the second-highest predicted response and the second-highest observed response.

In order to overcome the shortcomings of the PWO model when fitting

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

to the data, Mee (2020) considered expanded pairwise models that include interactions of the pairwise factors I_{jk}. The model that includes only factor interactions that involve exactly three components is dubbed the triplets model, and is given by

$$y = \beta_0 + \sum_{j<k} \beta_{jk} I_{jk} + \sum_{j=1}^{m-2} \sum_{k=j+1}^{m-1} \sum_{l=k+1}^{m} [\beta_{jk\star jl} I_{jk} I_{jl} + \beta_{jk\star kl} I_{jk} I_{kl}] + \varepsilon. \quad (3.10)$$

This model contains too many parameters to be useful in many cases; however, this also gives it additional flexibility that may produce a better fit. Also using a forward stepwise regression, Mee (2020) found two models that include some of the additional interaction terms. Our stepwise model (3.9) with df=32, predictive $R^2 = 0.68$, and RMSE = 3.32 is competitive with both of these triplets models (df = 24, predictive $R^2 = 0.60$, RMSE = 3.14, and df = 26, predictive $R^2 = 0.51$, RMSE = 3.73), and becomes more appealing when considering the use of fewer parameters. Furthermore, the top two predicted sequences from both of these models are CAEBD and CEBAD, aligning with the top two predicted sequences from the position model. Our model is also better than the PWO and CP models with interactions reported by Yang et al. (2020) in terms of various measures, including the predictive R^2 and RMSE. This further substantiates our claim that the

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

second-order model is able to achieve an intuitive and cost-effective fit on complex order-of-addition data, which until now has not been possible.

Note that while our models fit well to the real data in these examples, they assume that the absolute rather than the relative component positions are most predictive of the response. While the substantive conclusions are similar between the two model types, it is important to recognize that, in practice, the details of the application should be considered when assuming a model. For example, the absolute position assumption may be more valid in the drug administration problem, in which early exposure to a drug may produce better results. On the other hand, the relative position assumption makes more sense in cases where the components are known to react with each other, as in the experiments considered by Voelkel and Gallagher (2019). We study the robustness of our models to this assumption in Section 5.

4. Design Construction and Optimality

The positive results shown in the previous examples inspire us to study the properties of these new models and design construction. Because we do not know in advance which model is the best in a practical situation, we would like to have a class of designs that can perform well with different models

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

and various run sizes. To achieve this goal, we propose a novel construction method which, for many values of m, can quickly generate efficient designs of any run size.

For a prime or a prime power m, let $GF(m) = \{\omega_0, \omega_1, \ldots, \omega_{m-1}\}$ be a Galois field of order m, with ω_0 being the zero element (Barker, 1986). When m is a prime, $GF(m) = \{0, 1, \ldots, m-1\}$ is a ring of integers modulo m. The following algorithm constructs an $n \times m$ design for any $n \leq m!$.

Algorithm 1.

Step 1. For $k = 1, \ldots, m-1$, define an $m \times m$ matrix $\boldsymbol{L}_k$ such that its (i, j)th element is $\omega_i + \omega_k * \omega_j$, for $i, j = 0, \ldots, m-1$, where the addition and multiplication are defined on $GF(m)$.

Step 2. Construct an $(m^2 - m) \times m$ matrix $\boldsymbol{C}_1$ by row-wise concatenating $\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{m-1}$.

Step 3. Keep the first two columns of $\boldsymbol{C}_1$ fixed and permute the last $m-2$ columns of $\boldsymbol{C}_1$ in a systematic way. There are $(m-2)!$ permutations, admitting a total of $(m-2)!$ permuted matrices, denoted as $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_{(m-2)!}$.

Step 4. Construct an $m! \times m$ matrix $\mathbf{F}_m$ by row-wise concatenating $\boldsymbol{C}_1, \ldots, \boldsymbol{C}_{(m-2)!}$ and replacing ω_i with number i, for $i = 0, \ldots, m-1$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

Step 5. Let $\mathbf{F}_{n,m}$ be the $n \times m$ design formed by the first n rows of $\mathbf{F}_m$.

Step 6. Permute the columns of $\mathbf{F}_{n,m}$ to improve its performance under a chosen criterion.

Each $\boldsymbol{L}_k$ in Step 1 is an $m \times m$ Latin square, and the $(m-1)$ Latin squares ($\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{m-1}$) are mutually orthogonal. (Two Latin squares are orthogonal if, when they are superimposed, each pair (i, j) appears exactly once for any $i, j = 0, \ldots, m-1$.) Mutually orthogonal Latin squares are traditionally used to construct balanced incomplete block designs and orthogonal arrays. We use them for a different purpose.

The design $\boldsymbol{C}_1$ constructed in Step 2, as well as any $\boldsymbol{C}_i$ in Step 3, is a COA$(m^2 - m, m)$. Any pair of $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ in Step 3 do not share any common permutations. The $m! \times m$ matrix $\mathbf{F}_m$ constructed in Step 4 consists of all $m!$ permutations of m components. Step 5 simply chooses the first n rows of $\mathbf{F}_m$ as a candidate design, which often has good properties already. Specifically, Anderson-Cook and Lu (2019) outline the benefits of constructing designs from Latin squares and choosing a run size that is a multiple of m. Step 6, to be discussed later, can be used to further improve the design according to a specific criterion.

We consider the $m = 4$ case in which the full design $\mathbf{F}_m$ consists of 24 permutations in the order given in Table 4. The first four permutations

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

Table 4: The full 24-run design $\mathbf{F}_4$, as generated by Algorithm 1.

Run			a_1	a_2	a_3	a_4	Run		a_1	a_2	a_3	a_4
1			0	1	2	3	13		0	1	3	2
2		$\boldsymbol{L}_1$	1	0	3	2	14		1	0	2	3
3			2	3	0	1	15		2	3	1	0
4			3	2	1	0	16		3	2	0	1
5			0	2	3	1	17		0	2	1	3
6	$\boldsymbol{C}_1$	$\boldsymbol{L}_2$	1	3	2	0	18	$\boldsymbol{C}_2$	1	3	0	2
7			2	0	1	3	19		2	0	3	1
8			3	1	0	2	20		3	1	2	0
9			0	3	1	2	21		0	3	2	1
10		$\boldsymbol{L}_3$	1	2	0	3	22		1	2	3	0
11			2	1	3	0	23		2	1	0	3
12			3	0	2	1	24		3	0	1	2

form a 4×4 Latin square $\boldsymbol{L}_1$, the next four permutations form another Latin square $\boldsymbol{L}_2$, and so on. The first 12 permutations form a COA(12,4). The last 12 permutations are obtained from the first 12 by permuting the last two columns.

When $n = m(m-1)$, $\mathbf{F}_{n,m}$ is equivalent to the COA$(m(m-1), m)$ constructed by Yang et al. (2020). However, their construction does not provide designs with other run sizes. When m is not a prime power, a Galois field of order m does not exist. In these cases, an exchange algorithm may produce good designs. In future work, we will explore construction methods for this situation.

One criterion that we can use to assess the goodness of the designs produced by Algorithm 1 is the generalized wordlength pattern (GWLP), which

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

can be computed using the GWLP function in the R package DoE.base (Groemping et al., 2014). The GWLP $(W_1, \ldots, W_m)$ measures the aliasing of factorial effects, where $W_i \geq 0$ measures the overall aliasing of i-factor interactions on the general mean under the standard ANOVA model. An important property of the GWLP is that it characterizes the orthogonality or strength of a design. Xu and Wu (2001) showed that a design is an orthogonal array of strength t if and only if $W_1 = \cdots = W_t = 0$. Applying this result, we have $W_1 = 0$ if and only if the design is level balanced; that is, each level appears the same number of times in each column. Among level balanced designs, designs with small W_2 are preferred. The generalized minimum aberration criterion (Xu and Wu, 2001) favors designs that sequentially minimize $W_1, W_2, \ldots$ The generalized minimum aberration criterion includes the minimum aberration criterion (Fries and Hunter, 1980), the minimum G_2-aberration criterion (Tang and Deng, 1999), and many optimality criteria as special cases (Xu, 2003; Xu et al., 2009). Generalized minimum aberration designs are model robust in the sense that they minimize contamination of higher-order effects on the estimation of lower-order effects (Xu and Wu, 2001).

To assess the orthogonality of different types of designs, we use the GWLP of the component matrix $\boldsymbol{A}$ directly. Note that the component

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

matrix $\boldsymbol{A}$ and the position matrix $\boldsymbol{B}$ have the same GWLP, provided that every component appears in every position at least once. With this in mind, the designs $\mathbf{F}_{n,m}$ produced by Algorithm 1 have several desirable properties.

Theorem 1. *The design $\mathbf{F}_{n,m}$ has the following properties:*

(i) For any $n = qm + r$ with integers $q > 0$ and $0 \leq r < m$, $\mathbf{F}_{n,m}$ has $W_1 = mr(m-r)/n^2$, which is the minimum among all possible designs with n runs, m columns, and m levels.

(ii) If n is a multiple of m, then $\mathbf{F}_{n,m}$ has $W_1 = 0$.

(iii) If $m \leq n \leq m(m-1)$, then $\mathbf{F}_{n,m}$ has generalized minimum aberration among all possible designs with n runs, m columns, and m levels.

(iv) If n is a multiple of $m(m-1)$, then $\mathbf{F}_{n,m}$ is a COA(n, m).

Theorem 1 (i) and (ii) imply that our designs always have the desirable property of being level balanced or nearly balanced for each column. Theorem 1 (iii) and (iv) indicate that our designs tend to minimize the correlation between columns and reduce aliasing among first-order and second-order effects. Finally, having shown in Theorem 1 (iv) that $\mathbf{F}_{n,m}$ is a COA when n is a multiple of $m(m-1)$, Theorem 2 shows that these designs are D-optimal under the first-order and quadratic models.

Cheng et al. (2002) and Mandal and Mukerjee (2005) showed that generalized minimum aberration designs have high efficiency under model uncer-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

tainty for factorial experiments. Thus, Theorem 1 implies that the designs constructed from our algorithm have high efficiency under various models for order-of-addition experiments. Evidence of this property is presented in the next section.

We can also assess the constructed designs using the popular D- and A-optimality criteria. For an n-run design $\boldsymbol{\xi} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, let $\boldsymbol{X} = (\boldsymbol{f}(\boldsymbol{x}_1), \boldsymbol{f}(\boldsymbol{x}_2), \ldots, \boldsymbol{f}(\boldsymbol{x}_n))^{\mathrm{T}}$ be the model matrix of the linear model (3.1), and let $\boldsymbol{M}(\boldsymbol{\xi}) = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}/n$ be the per-run information matrix. A D-optimal design maximizes $|\boldsymbol{M}(\boldsymbol{\xi})|$, while an A-optimal design minimizes $tr(\boldsymbol{M}(\boldsymbol{\xi})^{-1})$. The D-optimality criterion seeks to minimize the volume of the confidence ellipsoid around the parameter estimates, and the A-optimality criterion minimizes the sum of the variances of the parameter estimates. The full design $\mathbf{F}_m$ with all $m!$ permutations is D-optimal for both the PWO model and the CP model (Peng et al., 2019; Yang et al., 2020). Additionally, Peng et al. (2019) showed that this design is also A-optimal for the PWO model. We can therefore compare the quality of any proposed design to this optimal one. For convenience, we define the D- and A-efficiency of $\boldsymbol{\xi}$ under model (3.1) relative to $\mathbf{F}_m$ respectively as

$$D(\boldsymbol{\xi}) = \{|\boldsymbol{M}(\boldsymbol{\xi})|/|\boldsymbol{M}(\mathbf{F}_m)|\}^{1/p}, \ A(\boldsymbol{\xi}) = \{tr(\boldsymbol{M}(\boldsymbol{\xi}))/tr(\boldsymbol{M}(\mathbf{F}_m))\}, \quad (4.1)$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

where p is the number of columns of the model matrix $\boldsymbol{X}$.

We need to determine whether the full design $\mathbf{F}_m$ is indeed optimal under the three position models. In this process, we rely on the checking condition for optimality provided by the equivalence theorem (Silvey, 1980). The equivalence theorem for models of the form (3.1) states that a design $\boldsymbol{\xi}^*$ is D- or A-optimal for a model with regression function $\boldsymbol{f}(\boldsymbol{x})$ over compact space Ω if and only if

$$D: \qquad \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{\xi}^*)^{-1}\boldsymbol{f}(\boldsymbol{x}) - p \leq 0 \ \ \forall \boldsymbol{x} \in \Omega, \tag{4.2}$$

$$A: \quad \boldsymbol{f}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{\xi}^*)^{-2}\boldsymbol{f}(\boldsymbol{x}) - tr(\boldsymbol{M}(\boldsymbol{\xi}^*)^{-1}) \leq 0 \ \ \forall \boldsymbol{x} \in \Omega, \tag{4.3}$$

with equality obtained at the design points $\boldsymbol{x} \in \boldsymbol{\xi}^*$. In the case of the three position models, each $\boldsymbol{x}$ is a permutation, and Ω is the space of all permutations of $\{1, \ldots, m\}$. We have the following important results regarding the full design and COAs.

Theorem 2. *The full design* $\mathbf{F}_m$ *is D-optimal under the first-order and quadratic position models, as is every* COA(n, m).

Theorem 3. *The full design* $\mathbf{F}_m$ *is D-optimal for the second-order position model.*

Remark 1. Note that the result of Theorem 1 from Peng et al. (2019),

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

which shows the optimality of $\mathbf{F}_m$ under the PWO model for any concave and signed permutation invariant criteria, does not apply to the three position models. As a counterexample, we consider the A-optimality criterion. If the result of their theorem held for the position models, then we would be able to confirm A-optimality of the full design numerically. However, as shown below in Remark 2, this is not the case. Furthermore, the information matrices for the first-order and quadratic models are block diagonal (see Section S1), yet the closed form of the information matrix for the second-order model is too complex to work with directly. However, the proof of Theorem 3 is general and can be applied to the PWO and CP models, as well as to other models, such as a third-order model that includes all estimable terms.

Remark 2. We leave a detailed investigation of A-optimal designs for these models to future work. Through preliminary investigation, we conclude that the design $\mathbf{F}_m$ does not satisfy the checking condition (4.3), and is thus not A-optimal, for any of the position models (see Section S2). Similarly, I-optimal designs under the position models can ensure that we are able to predict the best ordering, but they are beyond the scope of this study and warrant further examination. The D-optimality remains the most popular design criterion, so we focus on it for the remainder of this work.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

The D-efficiency of a design varies with respect to column permutations; that is, permuting the columns of a design may lead to different D-efficiencies. For this reason, we can permute the columns in Step 6 to maximize the D-efficiency for a specific model. We consider other opportunities for improved efficiency using level and $\boldsymbol{C}_i$ permutations in Section 5. In contrast, the GWLP is invariant with respect to column permutations, and instead studies the combinatorial properties of the design, such as balance and orthogonality.

5. Efficiency and Robustness of Designs

We compare our designs with those from Voelkel (2019) with $m = 4, 5, 7$ and various run sizes. We also compare them with the design given in Table 2 from Mee (2020) with $(n, m) = (40, 5)$. In order to present a fair comparison, we consider one design $\mathbf{F}_{n,m}$, which is generated by taking the first n rows of the full design $\mathbf{F}_m$ in Step 5, and another design $\mathbf{F}^*_{n,m}$, which permutes the columns of $\mathbf{F}_{n,m}$ in Step 6 to maximize the geometric mean efficiency of the models of interest. Table 5 compares the D-efficiencies of these designs under the five models and the first two terms (W_1 and W_2) of the GWLP. Designs $\mathbf{F}^*_{n,m}$ for which there is no improvement over $\mathbf{F}_{n,m}$ are omitted. While Voelkel (2019) used many other criteria to compare

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

his order-of-addition designs, many of these are derivatives of the two we consider here, and are thus not necessary to include. Because our algorithm is able to generate designs with variable run sizes, we also include the efficiencies of designs with various n between $m(m-1)$ and $m!$.

Voelkel's designs are constructed for the PWO model, and thus perform well under this model. However, they exhibit poor performance under the CP model and have large W_1 or W_2 values. In contrast, our designs are robust and perform well under all models (with the exception of the PWO model in certain situations) and always have small W_1 and W_2 values. Recall $W_1 = 0$ if and only if a design is level balanced for each column. Voelkel's designs are not level balanced, except for one case (Voelkel.12a), while our designs are all level balanced or nearly balanced. Mee's design performs comparably to $\mathbf{F}_{40,5}$ for all models except the PWO model, which it outperforms. When allowing for column permutations, Mee's design has similar properties to those of $\mathbf{F}^*_{40,5}$. In general, the $\mathbf{F}^*_{n,m}$ designs in Table 5, which maximize the geometric mean efficiency, may not necessarily be optimal for any of the five models, but they are model robust and have high D-efficiencies for all models; see Section S3 of the Supplementary Material.

Figure 2 shows the maximal D-efficiency obtained using Algorithm 1 for each model across a range of run sizes, n, by using a brute force search

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

Table 5: Comparison of D-efficiency and GWLP across designs and models.

			D-efficiency					GWLP	
n	m	Design	D_{PWO}	D_{CP}	D_{FO}	D_{PQ}	D_{SO}	W_1	W_2
12	4	Voelkel.12a	1	0.758	1	0.955	0.953	0	4.667
		Voelkel.12b	1	0	1	0.916	0.877	0.361	2.514
		$\mathbf{F}_{12,4}$	0.909	1	1	1	1	0	2
16	4	$\mathbf{F}_{16,4}$	0.917	0.950	0.977	0.963	0.953	0	3
20	4	$\mathbf{F}_{20,4}$	0.954	0.957	0.983	0.970	0.961	0	2
20	5	Voelkel.20a	0.903	0.588	0.997	0.945	0.861	0.35	10.35
		Voelkel.20b	0.97	0.623	0.993	0.931	0.854	0.625	8
		$\mathbf{F}_{20,5}$	0	1	1	1	0.959	0	2.5
		$\mathbf{F}^*_{20,5}$	0.898	1	1	1	0.950	0	2.5
24	5	Voelkel.24a	1	0.094	1	0.969	0.933	0.573	8.281
		Voelkel.24b	1	0	1	0.967	0.94	0.639	7.635
		Voelkel.24c	1	0.668	1	0.969	0.944	0.226	8.368
		$\mathbf{F}_{24,5}$	0.545	0.961	0.99	0.982	0.949	0.035	3.75
		$\mathbf{F}^*_{24,5}$	0.926	0.961	0.996	0.981	0.950	0.035	3.75
40	5	Mee.40	0.969	1	1	1	0.994	0	2.5
		$\mathbf{F}_{40,5}$	0.889	1	1	1	0.999	0	2.5
		$\mathbf{F}^*_{40,5}$	0.969	1	1	1	0.995	0	2.5
60	5	$\mathbf{F}_{60,5}$	0.977	1	1	1	0.986	0	2.5
		$\mathbf{F}^*_{60,5}$	0.977	1	1	1	0.999	0	2.5
24	7	Voelkel.24	0.990	-	1	0	-	2.125	20.688
		$\mathbf{F}_{24,7}$	0	-	0.989	0.686	-	0.146	21
		$\mathbf{F}^*_{24,7}$	0.742	-	0.989	0.873	-	0.146	21
36	7	Voelkel.36	0.970	-	1	0.911	0.764	0.789	23.722
		$\mathbf{F}_{36,7}$	0	-	1	0.923	0.809	0.032	7.389
		$\mathbf{F}^*_{36,7}$	0.881	-	0.987	0.979	0.839	0.032	7.389
48	7	Voelkel.48	0.986	0.587	1	0.950	0.872	0.924	17.099
		$\mathbf{F}_{48,7}$	0	0.967	0.993	0.985	0.876	0.018	5.688
		$\mathbf{F}^*_{48,7}$	0.943	0.967	0.995	0.989	0.798	0.018	5.688

Note: D-efficiency: the larger, the better; GWLP: the smaller, the better. D_X is the D-efficiency under model X (PWO, CP, first-order, pure quadratic, and second-order, respectively). In some cases, the chosen run size does not permit estimation of some models (with D-efficiencies marked by "-"). Designs $\mathbf{F}_{n,m}$ are obtained via Algorithm 1 without Step 6, while $\mathbf{F}^*_{n,m}$ are obtained with column permutations in Step 6 to maximize the geometric mean efficiency of the estimable models.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

over all column permutations in Step 6. From these plots, we find that the algorithm is able to produce highly efficient designs for many values of n. Specifically, we find that with a proper selection of a column permutation, our designs achieve high D-efficiency ($> 85\%$) for every model.

In addition to column permutations, the assignment of values to $\omega_0, \omega_1, \ldots, \omega_{m-1}$ in Step 1 and the order of permutations of the last $m-2$ columns to create the $\boldsymbol{C}_i$ in Step 3 can both be manipulated. To understand the effects of these permutations, we repeat the algorithm many times, with each iteration using a different combination of level, $\boldsymbol{C}_i$, and column permutations. This detailed study has demonstrated that for small values of m, the effect of the choice of permutations on efficiency is large under the PWO model, and small for the other models. Upon studying each choice of permutation in turn, we find that improvements to the D-efficiency of the best column-permuted design are small when allowing for level and $\boldsymbol{C}_i$ permutations. This justifies the inclusion of column permutations in Step 6 of Algorithm 1. The full results of this study can be found in Section S3.

Having shown that Algorithm 1 can, in general, produce designs that are optimal or near-optimal for many models when accounting for choices in the algorithm, we now examine the robustness of our designs to model misspecification. For example, we see from Table 5 that Voelkel's designs

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

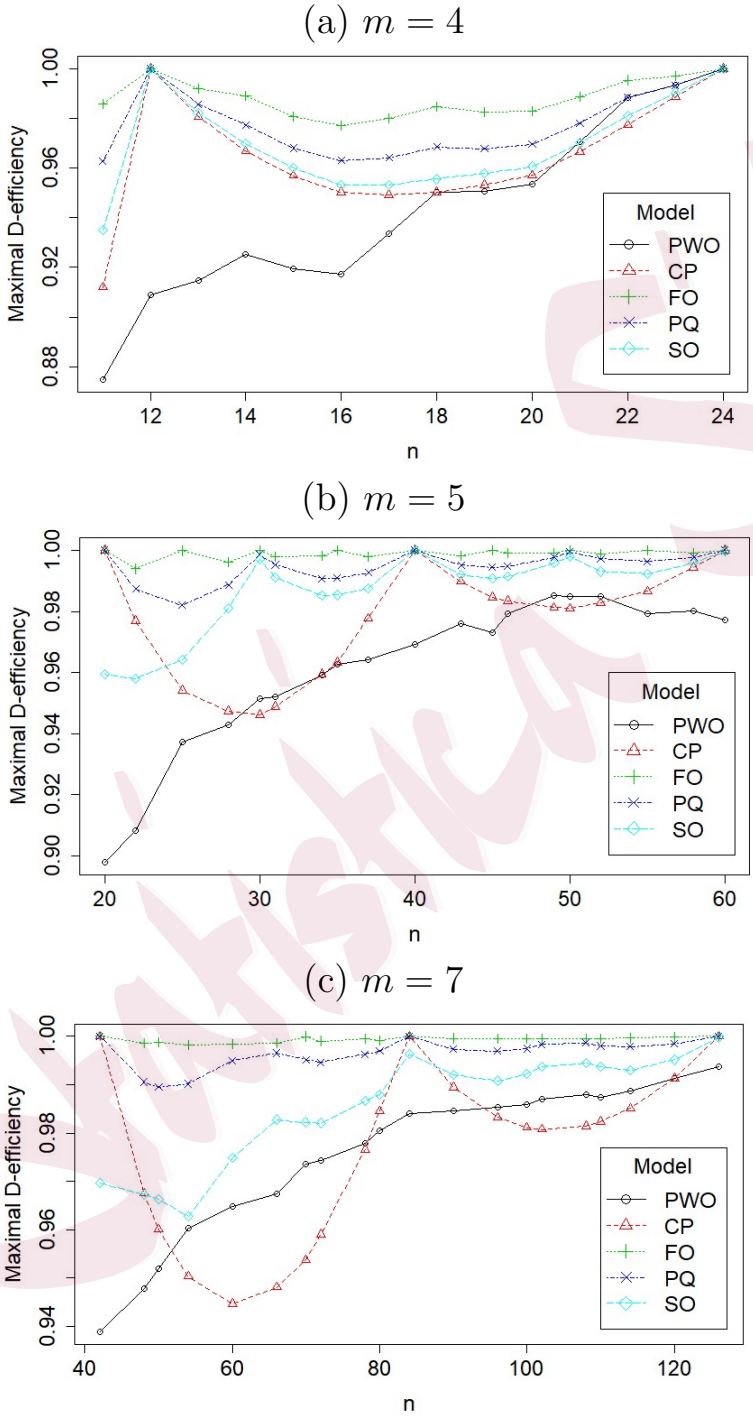

Figure 2: Maximal D-efficiency of $\mathbf{F}^*_{n,m}$ under column permutations for variable run sizes for (a) $m = 4$, (b) $m = 5$, and (c) $m = 7$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

have lower efficiency under the CP model, while our designs $\mathbf{F}_{n,m}$ have lower efficiency for the PWO model. If we design our experiment under the assumption that one of these models fits the experimental data, when in reality a different model captures the trend, then we run the risk of choosing an inefficient design. To test the ability of our designs to withstand such an error, we consider the trade-off in efficiency for different levels of confidence in our selection of the true model. Here, we find that our designs are robust to misspecifcation, and specifically to the assumption that relative/absolute position effects are most relevant. The details of this study can be found in Section S4.

6. Conclusion

Researchers are often interested in understanding the relationship between the process order of their experiment and the measured response. However, statistical techniques for efficiently studying this effect are largely absent from the literature. In this work, we have proposed succinct models and cost-effective designs for accurately capturing important trends. Through careful research, we have seen that our models yield a superior fit and interpretable estimates, while our designs are optimal in many cases and robust to model misspecification. Note, however, that all of the models we

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

consider, with the exception of the PWO model, are based on the absolute position effects assumption. Were we to consider further extensions of the PWO model, such as those proposed in Voelkel and Gallagher (2019) or Mee (2020), we may see different results. Furthermore, Schoen and Mee (2020) have recently found designs for $m = 5, 6, 7$ that are optimal under the PWO model and exhibit stronger balance than Voelkel's. Such designs may be more appropriate if there is strong confidence in the relative position assumption, but we do not consider them here. Further investigation is also needed into the robustness of the designs to omitted higher-order terms and interactions in both absolute and relative position-based models.

Applications to sequential drug administration have further demonstrated the scientific value of these methods to the broader research community. However, there is still much work left to be done in this field. The construction of efficient designs with large, nonprime m remains a challenging open problem. Existing algorithms may be sufficient for constructing nearly optimal designs for small, nonprime m. Mee (2020) briefly discussed the idea of ordering restrictions, yet there are many constrained situations for which no appropriate designs exist. Additionally, standard approaches for combining designs for the ordering effect with those for additional covariates result in experiments too large to be of much practical use. Through

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

further development of our techniques and subsequent research into these and other related problems, we hope that meaningful guidelines will be produced for scientists conducting order-of-addition experiments.

Supplementary Material

The online Supplementary Material includes proofs of the theorems, a discussion of the A-optimality results, and the model robustness of the proposed order-of-addition designs.

Acknowledgments

The authors thank the two reviewers for their helpful comments.

References

Anderson-Cook, C. M. and Lu, L. (2019). Discussion on "Order-of-Addition Experiments: A review and some new thoughts" *Quality Engineering* **31**, 64–68.

Barker, H. A. (1986). Sum and product tables for Galois fields. *International Journal of Mathematical Education in Science and Technology* **17**, 473–485.

Cheng, C.-S., Deng, L.-Y., and Tang, B. (2002). Generalized minimum aberration and design efficiency for nonregular fractional factorial designs. *Statist. Sinica* **12**, 991–1000.

Ding, X., Matsuo, K., Xu, L., Yang, J. and Zheng, L. (2015). Optimized combinations of

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

bortezomib, camptothecin, and doxorubicin show increased efficacy and reduced toxicity in treating oral cancer. *Anti-Cancer Drugs* **26**, 547–554.

Ding, X., Xu, H., Hopper, C., Yang, J., and Ho, C.-M. (2013). Use of fractional factorial designs in antiviral drug studies. *Quality and Reliability Engineering International* **29**, 299–304.

Fries, A. and Hunter, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics* **22**, 601–608.

Groemping, U., Amarov, B., and Xu, H. (2014). DoE.base: Full factorials, orthogonal arrays and base utilities for DoE packages. Package version 0.25-3.

Hedayat, A. S., Sloane, N. J. A. and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer.

Jaynes, J., Ding, X., Xu, H., Wong, W. K., and Ho, C.-M. (2013). Application of fractional factorial designs to study drug combinations. *Statistics in Medicine* **32**, 307–318.

Lin, D. K. J. and Peng, J. (2019). Order-of-addition experiments: A review and some new thoughts. *Quality Engineering* **31**, 49–59.

MacBeath, G., and Yaffe, M. B. (2012). Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780–794.

Mandal, A. and Mukerjee, R. (2005). Design efficiency under model uncertainty for nonregular fractions of general factorials. *Statist. Sinica* **15**, 697–707.

Mee, R. W. (2020). Order of addition modeling. *Statist. Sinica* **30**, 1543–1559.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

Peng, J. Y., Mukerjee, R. and Lin, D. K. J. (2019). Design of order-of-addition experiments. *Biometrika* **106**, 683–694.

Rao, C. R. (1961). Combinatorial arrangements analogous to orthogonal arrays. *Sankhya: The Indian Journal of Statistics, Series A* **23**, 283–286.

Schoen, E. D. and Mee, R. W. (2020). Order-of-addition orthogonal arrays to elucidate the sequence effects of treatments. Preprint.

Silva, A., Lee, B. Y., Clemens, D. L., Kee, T., Ding, X., Ho, C. M. and Horwitz, M. A. (2016). Output-driven feedback system control platform optimizes combinatorial therapy of tuberculosis using a macrophage cell culture model. *Proceedings of the National Academy of Sciences* **113**, E2172–E2179.

Silvey, S. D. (1980). *Optimal Design.* Chapman & Hall, London.

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

Van Nostrand, R. C. (1995). Design of experiments where the order of addition is important. In *ASA Proceeding of the Section on Physical and Engineering Sciences*, 155–160. Alexandria, VA: American Statistical Association.

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

Voelkel, J. G. and Gallagher, K. P. (2019). The design and analysis of order-of-addition exper-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

iments: An introduction and case study. *Quality Engineering* **31**, 627–638.

Wang, A., Xu, H. and Ding, X. (2020). Simultaneous optimization of drug combination dose-ratio-sequence with innovative design and active learning. *Advanced Therapeutics* **3**, 1900135.

Wu, C. F. J. and Hamada, M. S. (2009). *Experiments: Planning, Analysis and Optimization*. 2nd edition, New York: Wiley.

Xiao, Q., Wang, L., and Xu, H. (2019). Application of Kriging models for a drug combination experiment on lung cancer. *Statistics in Medicine* **38**, 236–246.

Xu, H. (2003). Minimum moment aberration for nonregular designs and supersaturated designs. *Statist. Sinica* **13**, 691–708.

Xu, H., Phoa, F. K. H., and Wong, W. K. (2009). Recent developments in nonregular fractional factorial designs. *Statistics Surveys* **3**, 18–46.

Xu, H. and Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Ann. Statist.* **29**, 1066–1077.

Yang, J., Sun, F., and Xu, H. (2020). A component-position model, analysis and design for order-of-addition experiments, *Technometrics*, doi: 10.1080/00401706.2020.1764394.

Zhao, Y. N., Lin, D. K. J. and Liu, M. Q. (2020). Designs for order-of-addition experiments. *Journal of Applied Statistics*, DOI: 10.1080/02664763.2020.1801607.

Department of Statistics, University of California, Los Angeles, CA 90095, USA

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0178

zstokes@ucla.edu

hqxu@stat.ucla.edu