



Analyzing dispersion effects from replicated order-of-addition experiments

Shin-Fu Tsai

To cite this article: Shin-Fu Tsai (2023) Analyzing dispersion effects from replicated order-of-addition experiments, Journal of Quality Technology, 55:3, 271-288, DOI: [10.1080/00224065.2022.2110024](https://doi.org/10.1080/00224065.2022.2110024)

To link to this article: <https://doi.org/10.1080/00224065.2022.2110024>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC



[View supplementary material](#)



Published online: 22 Aug 2022.



[Submit your article to this journal](#)



Article views: 1728



[View related articles](#)




[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)

Analyzing dispersion effects from replicated order-of-addition experiments

Shin-Fu Tsai 

Department of Agronomy, National Taiwan University, Taipei, Taiwan

ABSTRACT

Dispersion effects may play a vital role, in addition to location effects, in exploring optimal addition orders of several materials in some chemical, industrial and pharmaceutical studies. Two replication-based statistical methods developed using frequentist and fiducial probability arguments are introduced in this paper to identify active dispersion effects from replicated order-of-addition experiments. Simulation results show that both approaches can maintain empirical sizes sufficiently close to the nominal level while their finite-sample performances are very similar. From a statistical perspective, the fiducial method can provide a unified probability framework to analyze dispersion effects as well as location effects. However, it is computationally more expensive than the frequentist method. Consequently, the frequentist method is recommended for real-world applications due to its low computational cost. A drug combination study is used to illustrate these two approaches. In addition, some eligible order-of-addition designs are collected in a catalogue for future work.



KEYWORDS


experimental design; orthogonal array; pairwise order model; quality improvement; screening experiment; variation reduction

1. Introduction

Physical and/or chemical properties of process outputs can be very different when materials are added in different orders in some chemical, industrial and pharmaceutical studies. To explore the optimal addition order, researchers may conduct several trials in which the materials are deliberately added in different orders. Often, these trials are called order-of-addition experiments, and the materials of interest are called the components. After analyzing the observed data, active order effects can be first identified to study how the process output depends on the addition order of these components. Next, based on the identified effects, an empirical model can be built to predict the process outputs of those untested addition orders for the purpose of determining the optimal addition order. Efficient experimental designs and analysis methods for order-of-addition experiments have received considerable attention in recent years. Van Nostrand (1995) first established the relationship between order-of-addition designs and two-level factorial designs using the concept of pairwise order factors. Based on this key concept, Voelkel (2019) defined order-of-addition orthogonal arrays for

designing such experiments. By definition, a pairwise order matrix, whose entries are either $+1$ or -1 determined by all component addition orders, is called an order-of-addition orthogonal array of strength two if, for any two columns, the frequencies of the ordered pairs $(-1, -1)$, $(+1, -1)$, $(-1, +1)$ and $(+1, +1)$ are proportional to those of the full design. Furthermore, Peng, Mukerjee, and Lin (2019) developed several theoretical results to characterize and construct optimal order-of-addition designs. Most remarkably, they showed that an order-of-addition design is optimal for estimating the overall mean and all main effects of pairwise order factors with respect to a wide range of optimality criteria if its pairwise order matrix is an order-of-addition orthogonal array of strength two. Later, many combinatorial and computational methods were used to construct optimal or near-optimal order-of-addition designs. Recent proposals include those by Chen, Mukerjee, and Lin (2020), Winker, Chen, and Lin (2020), Zhao, Lin, and Liu (2021), Tsai (2022), Wang and Mee (2022) and Zhao, Lin, and Liu (2022). The reader is referred to Lin and Peng (2019) for a comprehensive review of order-of-addition experiments.

CONTACT Shin-Fu Tsai  shinfu@ntu.edu.tw  Department of Agronomy, National Taiwan University, Taipei 10617, Taiwan.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/00224065.2022.2110024>

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Conventionally, active effects can be classified into two types. If an effect has a significant impact on the response mean, then it is called an active location effect. After characterizing the relationship between response mean and all active location effects, the expected process output can be adjusted by a practitioner toward a target value to address a nominal-the-best problem. Similarly, a practitioner can determine a setting such that the expected process output will be as large/small as possible to address a larger-/smaller-the-better problem. On the other hand, an effect is said to be an active dispersion effect if it has a significant impact on the response variance. To address a nominal-the-best problem, after relating the response variance to all active dispersion effects, a practitioner can reduce process variability around a target value. Those null effects can be set to their most economical levels for the purpose of reducing the cost. With these two kinds of important effects in hand, a practitioner can optimize the process to address a nominal-the-best problem, that is, the expected process output is arbitrarily close to a target value with minimum run-to-run variability. By modeling the response mean and response variance as linear functions of pairwise order factors, Chen, Zhang, and Lin (2021) developed a systematic procedure to determine the optimal addition order in the presence of active dispersion effects. However, it is not clear how to identify these active dispersion effects. As far as I know, statistical methods for identifying active dispersion effects from order-of-addition experiments have not yet been studied systematically. The following is a brief review of some existing methods for factorial experiments. Box and Meyer (1986) introduced a residual-based test to identify active dispersion effects using unreplicated two-level factorial designs. Bergman and Hynén (1997) and McGrath and Lin (2001a) proposed further refinements. Pan (1999) and McGrath and Lin (2001b) noted that location effects could be confounded with dispersion effects in unreplicated experiments. This problem can be readily solved by conducting a replicated experiment. Nair and Pregibon (1988) compared three different statistical methods for estimating dispersion effects using replicated two-level factorial designs. Nelder and Lee (1991) and Lee and Nelder (1998) used different mean and variance models to achieve the same goal. Additionally, related inference procedures for these generalized linear models have been developed using asymptotic theory. Consequently, their performance can be unsatisfactory when the sample size is relatively small. Based on the concept of fiducial generalized pivotal quantities, Tsai, Liao, and Chai (2015) proposed a replication-based method to identify active dispersion effects from fully

replicated and partially replicated two-level factorial experiments. A necessary prerequisite for applying their approach is that the experimental design must be a two-level orthogonal array of strength two. Specifically, a matrix, whose entries are either $+1$ or -1 , is called a two-level orthogonal array of strength two if, for any two columns, the ordered pairs $(-1, -1)$, $(+1, -1)$, $(-1, +1)$ and $(+1, +1)$ occur equally often. Orthogonal arrays are important combinatorial tools that have been widely used to design multifactorial experiments. The reader is referred to Hedayat, Sloane, and Stufken (1999) for a comprehensive introduction to orthogonal arrays and their applications. However, when pairwise order factors are used to decode component addition orders into real-valued levels, the pairwise order matrix cannot be column-orthogonal, that is, it cannot be a two-level orthogonal array of strength two. As a result, the fiducial method developed by Tsai, Liao, and Chai (2015) cannot be applied directly. This paper aims to provide feasible solutions to identify active dispersion effects from replicated order-of-addition experiments.

The remainder of this paper is organized as follows. Section 2 presents the pairwise order mean and variance models. Section 3 introduces two replication-based statistical methods to test whether or not a dispersion effect has a significant impact on the response variance. These two methods are evaluated in Section 4 via a series of simulation studies. Section 5 uses a drug combination study to show that active dispersion effects may exist in real-world order-of-addition experiments. Some eligible order-of-addition designs are provided in Section 6 for future work. Concluding remarks are given in the final section.

2. Pairwise order mean and variance models

Suppose that an n -treatment order-of-addition experiment is conducted to study m components of interest. Every treatment has r replicates, so the total run size is equal to $N = n \times r$. Specifically, a treatment can be represented by a permutation of the component labels $1, 2, \dots, m$. To implement the h th treatment t_h , a practitioner must add these m components sequentially according to the order specified by t_h . Assume that the responses $Y_{h1}, Y_{h2}, \dots, Y_{hr}$ are independent and identically distributed normal random variables, denoted by

$$Y_{h1}, Y_{h2}, \dots, Y_{hr} \stackrel{iid}{\sim} N(\mu_h, \sigma_h^2).$$

The observations of different treatments are also assumed to be independent. To characterize the potential impact of location effects on the response mean, assume that

$$\mu_h = \gamma_0 + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \gamma_{ij} z_{h,ij}, \quad [1]$$

where γ_0 represents the overall mean, γ_{ij} denotes the pairwise order location effect of components i and j , and $z_{h,ij}$ represents the pairwise order factor given by

$$z_{h,ij} = \begin{cases} +1 & \text{if } i \text{ precedes } j \text{ in } t_h; \\ -1 & \text{if } j \text{ precedes } i \text{ in } t_h. \end{cases}$$

The concept of pairwise order factors was proposed by Van Nostrand (1995). Based on this key concept, the pairwise order mean model in [1] was first used by Voelkel (2019) to construct order-of-addition orthogonal arrays. Later, Peng, Mukerjee, and Lin (2019) proposed further generalizations of this model to take component distances into account. In fact, pairwise order factors are pseudo factors which are designed for the purpose of decoding component addition orders into real-valued levels. This intuitive and simple decoding scheme can be particularly useful for identifying active order effects. Given the number of components m , there are $q = m(m-1)/2$ pairwise order factors in [1], with the result that the pairwise order mean model has $p = q + 1$ parameters in total. Suppose that the pairwise order location effect γ_{ij} is positive. The expected response μ_h may increase by γ_{ij} if component i is added before component j in t_h . Similarly, the expected response μ_h may decrease by γ_{ij} if component i is added after component j in t_h . Because all parameters have straightforward interpretations, this model is often used in practice. Some real-world applications can be found in the papers by Voelkel and Gallagher (2019) and Mee (2020).

To establish the relationship between response variance and pairwise order factors, assume that

$$\sigma_h^2 = \delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{z_{h,ij}}, \quad [2]$$

where δ_0 represents the baseline variance, and δ_{ij} denotes the pairwise order dispersion effect of components i and j . Accordingly, if some pairwise order dispersion effects are not equal to one, then the response variance would be treatment-dependent. Suppose that $\delta_{ij} = 4$ and the remaining pairwise order dispersion effects are all equal to one. The response variance σ_h^2 would be twice as large as the baseline variance δ_0 if component i is added before component j in t_h . Similarly, if component i is added after component j in t_h , then the response variance σ_h^2 would be half as large as the baseline variance δ_0 . The pairwise order variance model in [2] also provides straightforward interpretations of all parameters. A similar multiplicative model

was used by McGrath and Lin (2001a) and Tsai, Liao, and Chai (2015) to analyze two-level factorial experiments. In particular, the pairwise order mean and variance models in [1] and [2] are both main effects models. They take only the main effects of pairwise order factors into account. In practice, incorporating interaction effects into these main effects models may yield better fitting results. This analysis strategy will be briefly discussed in the final section. In addition to statistical models, another key to the success of identifying active effects is the experimental design. A well-designed experiment enables us to extract as much information as possible from the observed data. Related design issues will be addressed in Section 6.

3. Two replication-based statistical methods

Two replication-based statistical methods developed using different probability arguments are introduced in this section to identify active dispersion effects from replicated order-of-addition experiments.

3.1. Frequentist method

Let W_h denote the h th logarithmic sample variance given by

$$W_h = \log S_h^2,$$

where S_h^2 represents the sample variance of all observations of the h th treatment t_h . Based on the first-order approximation, the logarithmic sample variance $W_h = \log S_h^2$ has approximately mean $\log \sigma_h^2$. Furthermore, based on Bartlett and Kendall (1946) approximation, the random variable W_h approximately follows a normal distribution, denoted by

$$W_h \approx N\left(\log \sigma_h^2, \psi'\left(\frac{r-1}{2}\right)\right), \quad [3]$$

where $\psi'(\cdot)$ represents the trigamma function. Note that the R function `trigamma` can be used to compute the trigamma function. A similar approximation to [3] can be found in (4.29) of Wu and Hamada (2021) in which the variance of W_h is approximated by $2/(r-1)$. Taking the logarithm of the multiplicative model in [2] yields

$$\log \sigma_h^2 = \alpha_0 + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \alpha_{ij} z_{h,ij}, \quad [4]$$

where $\alpha_0 = \log \delta_0$ and $\alpha_{ij} = (\log \delta_{ij})/2$. The logarithmic transformation conducted in [4] is a frequently used approach to induce additivity and variance stabilization. Let \mathbf{W} represent the $n \times 1$ vector consisting

of all logarithmic sample variances W_1, W_2, \dots, W_n . Based on the normal approximation in [3] and additive model in [4], the distribution of \mathbf{W} can be approximated using a multivariate normal distribution, denoted by

$$\mathbf{W} \approx N\left(\mathbf{X}\boldsymbol{\alpha}, \psi'\left(\frac{r-1}{2}\right)\mathbf{I}\right),$$

where \mathbf{X} represents the $n \times p$ model matrix, $\boldsymbol{\alpha}$ denotes the $p \times 1$ vector consisting of all parameters on the right-hand-side of [4], and \mathbf{I} represents the identity matrix of order n . Specifically, the first column of \mathbf{X} is an $n \times 1$ vector of ones, and the last q columns of \mathbf{X} constitute the $n \times q$ pairwise order matrix. Accordingly, the least squares estimator of $\boldsymbol{\alpha}$ can be expressed in the following matrix form:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W},$$

where its variance-covariance matrix has the form

$$\text{Var}(\hat{\boldsymbol{\alpha}}) = (\mathbf{X}^\top \mathbf{X})^{-1} \psi'\left(\frac{r-1}{2}\right).$$

Given a pairwise order dispersion effect δ_{st} , where $1 \leq s < t \leq m$, because the null and alternative hypotheses given by

$$H_0 : \delta_{st} = 1 \quad \text{and} \quad H_1 : \delta_{st} \neq 1$$

can be reformulated as

$$H_0 : \alpha_{st} = 0 \quad \text{and} \quad H_1 : \alpha_{st} \neq 0,$$

the frequentist test variable given by

$$Z_{\alpha_{st}} = \frac{\hat{\alpha}_{st}}{\sqrt{c_{st,st} \psi'\left(\frac{r-1}{2}\right)}}$$

can be used to evaluate whether or not α_{st} is significantly different from zero, where $\hat{\alpha}_{st}$ denotes the least squares estimator of α_{st} , and $c_{st,st}$ represents the diagonal entry of $(\mathbf{X}^\top \mathbf{X})^{-1}$ corresponding to $\hat{\alpha}_{st}$. The distribution of $Z_{\alpha_{st}}$ can be approximated by the standard normal distribution. According to Theorem 2 of Peng, Mukerjee, and Lin (2019), when the pairwise order matrix is an order-of-addition orthogonal array of strength two, the information matrix has the form

$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} n & \mathbf{0}^\top \\ \mathbf{0} & n\mathbf{I} + \frac{n}{3}\mathbf{V} \end{bmatrix}.$$

Specifically, $\mathbf{0}$ is the $q \times 1$ zero vector, \mathbf{I} is the identity matrix of order q , and \mathbf{V} is a $q \times q$ symmetric matrix with rows and columns indexed by ij and kl , where $1 \leq i < j \leq m$, $1 \leq k < l \leq m$, and an entry of \mathbf{V} is

$$v_{ij,kl} = \begin{cases} +1 & \text{if } i = k, j \neq l \text{ or } i \neq k, j = l; \\ -1 & \text{if } i = l \text{ or } j = k; \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, it is straightforward to verify that the inverse of information matrix has the form

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n} & \mathbf{0}^\top \\ \mathbf{0} & \frac{3(m-1)}{n(m+1)}\mathbf{I} - \frac{3}{n(m+1)}\mathbf{V} \end{bmatrix}.$$

Therefore, the constant term $c_{st,st}$ can be explicitly characterized, so the frequentist test variable $Z_{\alpha_{st}}$ can be further simplified as

$$Z_{\alpha_{st}} = \frac{\hat{\alpha}_{st}}{\sqrt{\frac{3(m-1)}{n(m+1)} \psi'\left(\frac{r-1}{2}\right)}}. \quad [5]$$

Sometimes, a research question can be addressed by testing whether or not the baseline variance δ_0 is equal to a particular constant. To achieve this goal, a frequentist test variable for $\alpha_0 = \log \delta_0$ is provided in [Appendix A](#). Bartlett and Kendall (1946) noted that the normal approximation in [3] would be inaccurate if the number of replicates r is smaller than five. However, it is straightforward to see that $\hat{\alpha}_{st}$ is a linear combination of all logarithmic sample variances W_1, W_2, \dots, W_n . The normal approximation for the distribution of $Z_{\alpha_{st}}$ in [5] can be further strengthened by a central limit theorem-like argument. Further details are shown in [Appendix B](#). Based on the simulation results in Section 4, this approximation often yields satisfactory results. Most importantly, the frequentist test variable $Z_{\alpha_{st}}$ in [5] is easy to compute.

3.2. Fiducial method

Let \mathbf{Y} denote a random sample obtained from a distribution indexed by a parameter vector $\boldsymbol{\xi}$, where a realization of \mathbf{Y} is denoted by \mathbf{y} . Suppose that a research question can be addressed by estimating and/or testing a complicated function of $\boldsymbol{\xi}$, denoted by $\theta = g(\boldsymbol{\xi})$.

Definition 1. A random quantity $R_\theta = R(\mathbf{Y}, \mathbf{y}, \boldsymbol{\xi})$ is called a fiducial generalized pivotal quantity, abbreviated as FGPQ hereinafter, for θ if it satisfies the following two conditions.

- The distribution of R_θ is free of all unknown parameters.
- $R(\mathbf{y}, \mathbf{y}, \boldsymbol{\xi}) = \theta$ for every allowable \mathbf{y} .

Given an FGPQ R_θ , a fiducial test variable given by $T_\theta = R_\theta - \theta$ or $T_\theta = R_\theta/\theta$ can be used to evaluate hypotheses about θ . The concept of FGPQs has been

widely used to develop inference procedures for complicated functions of parameters. Recent applications include those by Wang and Wu (2018), Al-Sarraj, von Brömssen, and Forkman (2019), Malekzadeh and Kharrati-Kopaei (2020), and Wang et al. (2021). The reader is referred to Hannig, Iyer, and Patterson (2006) for a comprehensive introduction to FGPs. Before presenting the fiducial method for testing whether or not δ_{st} is significantly different from one, the replication-based statistical method proposed by Tsai, Liao, and Chai (2015) is introduced first. Let

$$\mathcal{P}_{st}^* = \{h : z_{h,st} = +1\},$$

and

$$\mathcal{N}_{st}^* = \{h : z_{h,st} = -1\},$$

where \mathcal{P} represents “positive” and \mathcal{N} represents “negative.” The numbers of elements in \mathcal{P}_{st}^* and \mathcal{N}_{st}^* are denoted by p_{st}^* and n_{st}^* , respectively. These two index sets \mathcal{P}_{st}^* and \mathcal{N}_{st}^* can be used to classify the n sample variances into two groups according to the levels of pairwise order factor z_{st} . Under the assumption of normality, when the pairwise order variance model in [2] is used, one has

$$\begin{aligned} V_h &= \frac{(r-1)S_h^2}{\sigma_h^2} \\ &= \frac{(r-1)S_h^2}{\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{z_{h,ij}}} \sim \chi_{r-1}^2. \end{aligned} \quad [6]$$

Tsai, Liao, and Chai (2015) verified that the random quantity

$$R_{\delta_{st}}^* = \frac{\left[\prod_{h \in \mathcal{P}_{st}^*} \frac{(r-1)S_h^2}{V_h} \right]^{\frac{1}{p_{st}^*}}}{\left[\prod_{h \in \mathcal{N}_{st}^*} \frac{(r-1)S_h^2}{V_h} \right]^{\frac{1}{n_{st}^*}}} \quad [7]$$

would be an FGPQ for δ_{st} if

$$\sum_{h \in \mathcal{P}_{st}^*} z_{h,ij} = \sum_{h \in \mathcal{N}_{st}^*} z_{h,ij} = 0 \quad [8]$$

for every $ij \neq st$. Based on the FGPQ $R_{\delta_{st}}^*$, the fiducial test variable given by

$$T_{\delta_{st}}^* = \frac{R_{\delta_{st}}^*}{\delta_{st}} \quad [9]$$

can be used to identify active dispersion effects from fully replicated and partially replicated two-level factorial experiments. Although the pairwise order factors are two-level factors, the conditions in [8] can never be satisfied for an order-of-addition design due to the transitive property of pairwise order factors. The reader can consult Lin and Peng (2019) for a

brief discussion regarding the transitive property. Consequently, when the pairwise order factors are used to decode an order-of-addition design into a two-level design, the random quantity $R_{\delta_{st}}^*$ in [7] does not satisfy condition (B) of Definition 1, so some nuisance parameters may have a substantial impact on the fiducial test variable $T_{\delta_{st}}^*$ in [9]. This negative impact will be observed later from the simulation results in Section 4. Actually, the fiducial test variable $T_{\delta_{st}}^*$ in [9] is not recommended for analyzing real-world order-of-addition experiments. It is introduced only for comparative purposes. Below, another two sets are proposed to replace \mathcal{P}_{st}^* and \mathcal{N}_{st}^* , respectively, to eliminate the potential impact of nuisance parameters.

Definition 2. Given two treatments t_a and t_b , their pairwise order vectors z_a and z_b are called a pair of positive quasi-foldover vectors with respect to a pairwise order factor z_{st} if $z_{a,st} = z_{b,st} = +1$ and $z_{a,ij} + z_{b,ij} = 0$ for all $ij \neq st$. Similarly, z_a and z_b are called a pair of negative quasi-foldover vectors with respect to z_{st} if $z_{a,st} = z_{b,st} = -1$ and $z_{a,ij} + z_{b,ij} = 0$ for all $ij \neq st$.

Given a pair of positive quasi-foldover vectors z_a and z_b with respect to z_{st} , component s is added before component t both in t_a and t_b , and the remaining pairs of components are added in reverse order. A pair of negative quasi-foldover vectors with respect to z_{st} has similar symmetric properties. Therefore, when an inference procedure for δ_{st} is developed based on positive and negative quasi-foldover vectors, the potential impact of nuisance parameters can be eliminated systematically. Let \mathcal{P}_{st} and \mathcal{N}_{st} denote the sets consisting of all index pairs of positive and negative quasi-foldover vectors with respect to z_{st} . Formally, \mathcal{P}_{st} and \mathcal{N}_{st} can be written as

$$\begin{aligned} \mathcal{P}_{st} &= \{(a, b) : z_{a,st} = z_{b,st} = +1 \text{ and } z_{a,ij} \\ &\quad + z_{b,ij} = 0 \text{ for every } ij \neq st\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{N}_{st} &= \{(a, b) : z_{a,st} = z_{b,st} = -1 \text{ and } z_{a,ij} \\ &\quad + z_{b,ij} = 0 \text{ for every } ij \neq st\}. \end{aligned}$$

Given an order-of-addition design, the two sets \mathcal{P}_{st} and \mathcal{N}_{st} can be uniquely determined for every pairwise order factor z_{st} , where all index pairs (a, b) in \mathcal{P}_{st} and \mathcal{N}_{st} are unordered pairs. The following example is given to illustrate positive and negative quasi-foldover vectors.

Example 1. Suppose that the full design is conducted to explore the optimal addition order of three

components of interest. All treatments and their pairwise order vectors are listed in Table 1. It is straightforward to see that the two vectors z_1 and z_5 constitute a pair of positive quasi-foldover vectors with respect to z_{12} . Additionally, another two vectors z_3 and z_6 constitute a pair of negative quasi-foldover vectors with respect to z_{12} . Consequently, one has $\mathcal{P}_{12} = \{(1, 5)\}$ and $\mathcal{N}_{12} = \{(3, 6)\}$. Similarly, the sets $\mathcal{P}_{13} = \{(2, 3)\}$, $\mathcal{N}_{13} = \{(4, 5)\}$, $\mathcal{P}_{23} = \{(1, 4)\}$ and $\mathcal{N}_{23} = \{(2, 6)\}$ can be obtained by searching for positive or negative quasi-foldover vectors over all pairwise order vectors.

Assume that an order-of-addition design with non-empty sets \mathcal{P}_{st} and \mathcal{N}_{st} is conducted, where every treatment has r replicates. The two sets \mathcal{P}_{st} and \mathcal{N}_{st} would be nonempty if the experiment is designed carefully. This design issue will be addressed in Section 6. By replacing \mathcal{P}_{st}^* and \mathcal{N}_{st}^* in [7] with \mathcal{P}_{st} and \mathcal{N}_{st} , respectively, and replacing h with either a or b , a random quantity can be obtained as

$$R_{\delta_{st}} = \frac{\prod_{(a,b) \in \mathcal{P}_{st}} \left[\frac{(r-1)s_a^2}{V_a} \frac{(r-1)s_b^2}{V_b} \right]^{\frac{1}{2p_{st}}}}{\prod_{(a,b) \in \mathcal{N}_{st}} \left[\frac{(r-1)s_a^2}{V_a} \frac{(r-1)s_b^2}{V_b} \right]^{\frac{1}{2n_{st}}}}, \quad [10]$$

where p_{st} and n_{st} denote the numbers of index pairs in \mathcal{P}_{st} and \mathcal{N}_{st} , respectively. First, it is straightforward to see that the distribution of $R_{\delta_{st}}$ is free of all parameters. Therefore, condition (A) of Definition 1 is satisfied. Next, replacing S_h^2 in [6] with its realization s_h^2 and replacing h with either a or b . Plugging the resulting term into [10] yields

$$\begin{aligned} & \frac{\prod_{(a,b) \in \mathcal{P}_{st}} \left[\frac{(r-1)s_a^2}{V_a} \frac{(r-1)s_b^2}{V_b} \right]^{\frac{1}{2p_{st}}}}{\prod_{(a,b) \in \mathcal{N}_{st}} \left[\frac{(r-1)s_a^2}{V_a} \frac{(r-1)s_b^2}{V_b} \right]^{\frac{1}{2n_{st}}}} \\ &= \frac{\prod_{(a,b) \in \mathcal{P}_{st}} \left[\left(\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{\frac{z_{a,ij}}{2}} \right) \left(\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{\frac{z_{b,ij}}{2}} \right) \right]^{\frac{1}{2p_{st}}}}{\prod_{(a,b) \in \mathcal{N}_{st}} \left[\left(\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{\frac{z_{a,ij}}{2}} \right) \left(\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{\frac{z_{b,ij}}{2}} \right) \right]^{\frac{1}{2n_{st}}}} \\ &= \delta_{st} \frac{\prod_{ij \neq st} \left[\delta_{ij}^{\frac{1}{2} \sum_{(a,b) \in \mathcal{P}_{st}} (z_{a,ij} + z_{b,ij})} \right]^{\frac{1}{2p_{st}}}}{\prod_{ij \neq st} \left[\delta_{ij}^{\frac{1}{2} \sum_{(a,b) \in \mathcal{N}_{st}} (z_{a,ij} + z_{b,ij})} \right]^{\frac{1}{2n_{st}}}} \\ &= \delta_{st}, \end{aligned}$$

because

$$\sum_{(a,b) \in \mathcal{P}_{st}} (z_{a,ij} + z_{b,ij}) = \sum_{(a,b) \in \mathcal{N}_{st}} (z_{a,ij} + z_{b,ij}) = 0$$

Table 1. Treatments and pairwise order vectors of the three-component full design.

	Treatment				Pairwise order factor		
					z_{12}	z_{13}	z_{23}
t_1	1	2	3	z_1	+1	+1	+1
t_2	1	3	2	z_2	+1	+1	-1
t_3	2	1	3	z_3	-1	+1	+1
t_4	2	3	1	z_4	-1	-1	+1
t_5	3	1	2	z_5	+1	-1	-1
t_6	3	2	1	z_6	-1	-1	-1

for every $ij \neq st$. That is, condition (B) of Definition 1 is also satisfied. Consequently, $R_{\delta_{st}}$ in [10] is an FG PQ for δ_{st} . Based on the FG PQ $R_{\delta_{st}}$, the fiducial test variable given by

$$T_{\delta_{st}} = \frac{R_{\delta_{st}}}{\delta_{st}} \quad [11]$$

can be used to evaluate $H_0 : \delta_{st} = 1$ versus $H_1 : \delta_{st} \neq 1$. Because the distribution of $T_{\delta_{st}}$ in [11] has no explicit form, the following Monte-Carlo algorithm is designed to estimate the required p -value.

- Step 1: Choose a Monte-Carlo sample size K , e.g., set $K = 5,000$. Generate two independent chi-square random variables $V_{a,k}$ and $V_{b,k}$ with $r-1$ degrees of freedom for every index pair (a, b) in \mathcal{P}_{st} and \mathcal{N}_{st} for $1 \leq k \leq K$, respectively.
- Step 2: Calculate the fiducial test variable $T_{\delta_{st},k}$ in [11], where $\delta_{st} = 1$.
- Step 3: The required p -value can be estimated using $P_{\delta_{st}} = 2Q_{\delta_{st}}/K$, where

$$Q_{\delta_{st}} = \min \left[\sum_{k=1}^K I(T_{\delta_{st},k} > 1), \sum_{k=1}^K I(T_{\delta_{st},k} < 1) \right],$$

and $I(\cdot)$ represents the indicator function. Given a nominal level α , the null hypothesis $H_0 : \delta_{st} = 1$ should be rejected if the corresponding p -value $P_{\delta_{st}}$ is smaller than α . Specifically, a fiducial test variable for δ_0 is presented in Appendix A to infer the baseline variance.

4. Simulation studies

To evaluate the frequentist and fiducial methods, the eight scenarios in Table 2 with different combinations of active effects are chosen for simulation studies. These scenarios include all possible combinations of two mean models and four variance models. The four-component full design with $n = 24$ treatments is used to simulate experimental data, where the number of replicates r is set to either 2 or 3. Consequently,

Table 2. Active effects in pairwise order mean and variance models for simulation studies.

Scenario	Mean model	Variance model
1	$\{\gamma_0, \gamma_{12}\}$	$\{\delta_0, \delta_{12}\}$
2	$\{\gamma_0, \gamma_{12}\}$	$\{\delta_0, \delta_{12}, \delta_{13}\}$
3	$\{\gamma_0, \gamma_{12}\}$	$\{\delta_0, \delta_{12}, \delta_{34}\}$
4	$\{\gamma_0, \gamma_{12}\}$	$\{\delta_0, \delta_{12}, \delta_{14}, \delta_{24}\}$
5	$\{\gamma_0, \gamma_{12}, \gamma_{14}, \gamma_{24}\}$	$\{\delta_0, \delta_{12}\}$
6	$\{\gamma_0, \gamma_{12}, \gamma_{14}, \gamma_{24}\}$	$\{\delta_0, \delta_{12}, \delta_{13}\}$
7	$\{\gamma_0, \gamma_{12}, \gamma_{14}, \gamma_{24}\}$	$\{\delta_0, \delta_{12}, \delta_{34}\}$
8	$\{\gamma_0, \gamma_{12}, \gamma_{14}, \gamma_{24}\}$	$\{\delta_0, \delta_{12}, \delta_{14}, \delta_{24}\}$

the run sizes N are equal to 48 and 72, respectively. The overall mean γ_0 is fixed at 100, and the effect sizes of all active location effects γ are set to 25 over all scenarios. Additionally, the baseline variance δ_0 is fixed at 10, and the effect sizes of all active dispersion effects δ are set to 4, 6 and 8, respectively. After specifying all parameter values, an $N \times 1$ response vector \mathbf{Y} can be simulated from a multivariate normal distribution, denoted by

$$\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

where the entries of mean vector $\boldsymbol{\mu}$ are determined by the pairwise order mean model in [1], the diagonal entries of variance-covariance matrix $\boldsymbol{\Sigma}$ are determined by the pairwise order variance model in [2], and all off-diagonal entries of $\boldsymbol{\Sigma}$ are equal to zero. Specifically, each entry of \mathbf{Y} is generated via

$$Y_h = \mu_h + \sigma_h Z_h, \quad [12]$$

where Z_h represents a standard normal random variable. Based on this data-generating scheme, 5,000 simulated response vectors are generated for each set of simulation parameters consisting of scenario, r and δ . All simulated data are then analyzed using the test variables $Z_{\alpha_{st}}$ in [5], $T_{\delta_{st}}^*$ in [9] and $T_{\delta_{st}}$ in [11], respectively. In fact, as noted in Section 3.2, the fiducial test variable $T_{\delta_{st}}^*$ in [9] is not valid for analyzing order-of-addition experiments. It is included only for comparative purposes. The empirical powers are calculated for all pairwise order dispersion effects, where the nominal level is set to 0.05. The empirical power for each effect is defined as the proportion of the 5,000 p -values that are smaller than the nominal level. In addition, the empirical power of a null effect is called the empirical size. All simulation results are presented in Appendix C.

By comparing the empirical powers of the first four and last four scenarios in Tables C1–C3, the finite-sample performances of these test variables are observed to be independent of the pairwise order mean models because they are all replication-based methods. Although the fiducial test variable $T_{\delta_{st}}^*$ in [9] provides larger empirical powers for active effects, as

shown in Table C2, it cannot maintain the empirical sizes for null effects. Because some nuisance parameters may have a substantial impact on $T_{\delta_{st}}^*$ in [9], many null effects are mistakenly identified as active effects, such that their empirical sizes are considerably larger than the nominal level. Most often, such high false positive rates are not acceptable. To obtain more reliable results, the fiducial test variable $T_{\delta_{st}}^*$ in [9] should be superseded by another two test variables. On the other hand, as shown in Tables C1 and C3, the frequentist test variable $Z_{\alpha_{st}}$ in [5] and the fiducial test variable $T_{\delta_{st}}$ in [11] can maintain empirical sizes sufficiently close to the nominal level. Moreover, as expected, their empirical powers increase as the dispersion effect size δ increases and as the number of replicates r increases. These two test variables have very similar finite-sample behaviors in all scenarios in Table 2. In conclusion, both test variables are feasible for identifying active dispersion effects.

To evaluate whether or not the frequentist and fiducial methods are robust to the assumption of normality, a similar data-generating scheme to [12] given by

$$Y_h = \mu_h + \sigma_h \left[\frac{U_h - E(U_h)}{\sqrt{\text{Var}(U_h)}} \right] \quad [13]$$

is used to simulate non-normal data. In the brackets of [13], the non-normal random variable U_h is standardized to ensure that the simulated data generated by [12] and [13] have the same mean and variance, such that the additional simulation results can be compared with those presented in Tables C1 and C3. Five different distributions are assumed for the non-normal random variable U_h in [13], including the t -distribution with three degrees of freedom, denoted by T_3 , and four gamma distributions, denoted by $\Gamma_{1,2}$, $\Gamma_{2,2}$, $\Gamma_{3,2}$ and $\Gamma_{6,2}$, respectively, where the subscripts indicate the shape and scale parameters. These gamma distributions, with excess kurtosis parameters equal to 6, 3, 2 and 1, respectively, are chosen to explore the potential impact of the excess kurtosis parameter on these two methods. Given a non-normal distribution for U_h , 5,000 simulated response vectors are generated for each set of simulation parameters. All simulated data are then analyzed using the test variables $Z_{\alpha_{st}}$ in [5] and $T_{\delta_{st}}$ in [11], respectively, to calculate their average sizes and average powers. As shown in Table C4, when the assumption of normality is violated, the empirical sizes are all larger than the nominal level 0.05, and the frequentist and fiducial methods have very similar behaviors. The differences between empirical sizes and nominal level are relatively large when

the excess kurtosis parameter is equal to 6. However, the empirical size approaches the nominal level as the excess kurtosis parameter is decreased in every scenario in Table 2. In particular, when the simulated data are generated from the gamma distribution $\Gamma_{6,2}$, where the excess kurtosis parameter is equal to 1, the differences between empirical sizes and nominal level can be negligible. Additionally, as shown in Table C5, these two methods have similar empirical powers for analyzing non-normal data. Overall, both frequentist and fiducial methods appear to be robust to the assumption of normality.

5. Drug combination data

Wang, Xu, and Ding (2020) conducted a series of in vitro experiments to study whether or not treatment order of chemotherapeutics and their dosages have a significant impact on lymphoma cell growth. Three FDA-approved chemotherapeutics for clinical trials, namely, paclitaxel, doxorubicin and mitoxantrone, denoted by 1, 2 and 3, respectively, were sequentially added into lymphoma cell cultures in the first experiment. In particular, chemotherapeutics 1 and 2 were set at two different concentration levels and the level of chemotherapeutic 3 was held fixed throughout the experiment. Accordingly, six treatment orders and four level combinations were taken into account. A 24-run cross-product array was used to design this experiment, that is, all order-level combinations were conducted, where each treatment had three replicates. All responses were then measured by a tumor inhibition index. Based on Table S3 of the original paper, the dosage effects are negligible. Therefore, observations with the same treatment order are further merged, such that each treatment has 12 replicates. In addition, this experiment had some observations generated by adding these three chemotherapeutics into lymphoma cell cultures simultaneously. These observations are not included in the later analysis. The adapted design and related summary statistics are shown in Table 3.

As shown in Table 3, when the three chemotherapeutics are added in different orders, sample averages and sample variances are quite different. This

interesting finding indicates that treatment order may have a substantial impact on both the response mean and response variance. To understand how the response mean and response variance depend on the treatment order, identifying active effects can be practical. In particular, it is straightforward to see that the three sample variances s_1^2 , s_3^2 and s_4^2 are larger than three other sample variances s_2^2 , s_5^2 and s_6^2 . One might suspect that some active dispersion effects may affect the response variance. The frequentist test variable $Z_{\alpha_{st}}$ in [5] and fiducial test variable $T_{\delta_{st}}$ in [11] are used to analyze the drug combination data. The positive and negative quasi-foldover vectors required by the fiducial method are given in Example 1, and the corresponding p -values are estimated using the Monte-Carlo algorithm introduced in Section 3.2. When the nominal level is set to 0.05, as shown in Table 4, both methods indicate that the pairwise order dispersion effect δ_{23} differs significantly from one. To reduce patient-to-patient variability in tumor cell inhibition, chemotherapeutic 3 should be administered before chemotherapeutic 2.

After exploring potentially active dispersion effects, two other methods are used to identify active location effects under the heterogeneous variance assumption, where the response variances are related to the levels of pairwise order factor z_{23} . The first method is the Student's t -test, where the degrees of freedom is determined according to the Satterthwaite's (1946) approximation. The second method is the fiducial test developed by Tsai and Liao (2020), where the reduced pairwise order variance model is set to

$$\sigma_h^2 = \delta_0 \delta_{23}^{\frac{z_{h,23}}{2}}.$$

Under the pairwise order mean and variance models in [1] and [2], this fiducial test remains valid for identifying active location effects from fully replicated order-of-addition experiments. The reader is referred to the original paper for more details regarding this approach. According to the p -values in Table 4, the pairwise order location effects γ_{12} and γ_{23} are identified as active effects by both methods when the nominal level is set to 0.05. To maximize the expected response, chemotherapeutic 1 should be administered before chemotherapeutic 2 and chemotherapeutic 3

Table 3. Adapted drug combination data.

Treatment	Time 1	Time 2	Time 3	r	\bar{y}_h	s_h^2
1	1	2	3	12	39.78	41.85
2	1	3	2	12	43.42	16.12
3	2	1	3	12	22.62	22.96
4	2	3	1	12	27.11	39.33
5	3	1	2	12	34.94	16.26
6	3	2	1	12	37.97	14.73

Table 4. p -values for adapted drug combination data

Method	Dispersion effect			Location effect		
	δ_{12}	δ_{13}	δ_{23}	γ_{12}	γ_{13}	γ_{23}
Frequentist method	0.4336	0.5402	0.0301	<0.0001	0.1071	<0.0001
Fiducial method	0.4295	0.5361	0.0314	<0.0001	0.1995	<0.0001

should be administered before chemotherapeutic 2. By integrating all screening results in Table 4, conducting the second treatment is expected to yield optimal efficacy of tumor cell inhibition with minimum patient-to-patient variability. This statistical analysis provides a possible explanation of the adapted drug combination data in Table 3. The analysis results may be useful for researchers to develop a novel combination therapy for treating lymphoma. Most interestingly, this example shows that active dispersion effects may exist in order-of-addition experiments. Therefore, related statistical methods can be practical for real-world applications.

6. Eligible designs

Given an order-of-addition design, if either \mathcal{P}_{st} or \mathcal{N}_{st} is empty, then the dispersion effect δ_{st} would not be testable by the fiducial test variable $T_{\delta_{st}}$ in [11]. Therefore, an order-of-addition design is said to be eligible for testing all dispersion effects if these two sets \mathcal{P}_{st} and \mathcal{N}_{st} are nonempty for every pairwise order factor z_{st} . From a practical perspective, conducting an eligible design allows us to analyze the observed data using both frequentist and fiducial methods, such that more insights can be gained by comparing their results. To construct eligible designs, an intuitive but important result is introduced first.

Theorem 1. *Given the number of components m , the full design is eligible for testing all pairwise order dispersion effects using the fiducial test variable $T_{\delta_{st}}$ in [11], where $p_{st} = n_{st} = (m-1)!/2$ for every pairwise order factor z_{st} .*

Proof. Given a treatment t_a , its pairwise order vector z_a belongs to a pair of positive quasi-foldover vectors with respect to a pairwise order factor z_{st} only if component s is added right before component t in t_a . Given such a treatment t_a , let t_b denote the treatment

obtained by reversing the treatment t_a except for the two components s and t . It is straightforward to verify that their pairwise order vectors z_a and z_b constitute a pair of positive quasi-foldover vectors with respect to z_{st} . There are $(m-1)!$ such treatments in the full design for every pairwise order factor z_{st} , with the result that $p_{st} = (m-1)!/2$. By similar arguments, one gets $n_{st} = (m-1)!/2$. This completes the proof. \square

According to Theorem 1, when analyzing dispersion effects using the full design, the number of treatments involving in the fiducial test variable $T_{\delta_{st}}$ in [11] is equal to $2(m-1)!$, with the result that $2/m$ of the treatments are used to test a particular dispersion effect. The proportion $2/m$ decreases as the number of components m increases. However, as noted by Voelkel (2019) and Peng, Mukerjee, and Lin (2019), when m is moderately large, the full design can be expensive and/or time-consuming due to its excessively large run size. For example, when there are seven components to be investigated, i.e., $m=7$, a practitioner must add these seven components sequentially in $7! = 5,040$ different orders to conduct the full design. Because all treatments must be replicated two or more times to apply a replication-based method, at least 10,080 runs must be conducted. Such a large-scale experiment is rarely feasible in practice. Given resource constraints, an eligible fractional design consisting of a subset of selected treatments is a cost-effective alternative. The following example is given to illustrate ineligible fractional designs.

Example 2. Suppose that an eligible fractional design is required to study four components of interest. Each treatment will be replicated several times for the purpose of identifying active dispersion effects. Designs 1 and 2 in Table 2 of Voelkel (2019) are considered for this experiment, where their treatments and pairwise order vectors are listed in Table 5. In particular, the

Table 5. Designs 1 and 2 in Table 2 of Voelkel (2019).

Design 1				Pairwise order factor						Design 2				Pairwise order factor					
				z_{12}	z_{13}	z_{14}	z_{23}	z_{24}	z_{34}					z_{12}	z_{13}	z_{14}	z_{23}	z_{24}	z_{34}
1	2	4	3	+1	+1	+1	+1	+1	-1	1	3	2	4	+1	+1	+1	-1	+1	+1
1	3	2	4	+1	+1	+1	-1	+1	+1	1	3	4	2	+1	+1	+1	-1	-1	+1
1	4	2	3	+1	+1	+1	+1	-1	-1	1	4	2	3	+1	+1	+1	+1	-1	-1
2	1	3	4	-1	+1	+1	+1	+1	+1	2	1	4	3	-1	+1	+1	+1	+1	-1
2	3	4	1	-1	-1	-1	+1	+1	+1	2	3	1	4	-1	-1	+1	+1	+1	+1
2	4	3	1	-1	-1	-1	+1	+1	-1	2	3	4	1	-1	-1	-1	+1	+1	+1
3	1	4	2	+1	-1	+1	-1	-1	+1	2	4	1	3	-1	+1	-1	+1	+1	-1
3	2	1	4	-1	-1	+1	-1	+1	+1	3	1	2	4	+1	-1	+1	-1	+1	+1
3	4	1	2	+1	-1	-1	-1	-1	+1	3	4	2	1	-1	-1	-1	-1	-1	+1
4	1	3	2	+1	+1	-1	-1	-1	-1	4	1	2	3	+1	+1	-1	+1	-1	-1
4	2	1	3	-1	+1	-1	+1	-1	-1	4	3	1	2	+1	-1	-1	-1	-1	-1
4	3	2	1	-1	-1	-1	-1	-1	-1	4	3	2	1	-1	-1	-1	-1	-1	-1

pairwise order matrices of these two designs are order-of-addition orthogonal arrays of strength two. After searching for positive and negative quasi-foldover vectors with respect to every pairwise order factor, Design 2 is found to have empty \mathcal{P}_{14} and \mathcal{N}_{14} . Accordingly, the pairwise order dispersion effect δ_{14} would not be testable by the fiducial test variable $T_{\delta_{st}}$ in [11] if Design 2 is chosen for the experiment. On the other hand, Design 1 is found to be eligible for testing all dispersion effects. Most interestingly, as noted by Voelkel (2019), Design 1 is superior to Design 2 according to some other optimality criteria. Therefore, Design 1 is recommended for this experiment.

Given an arbitrary fractional design, as shown by Example 2, it can be ineligible for testing certain dispersion effects. Therefore, generating a series of eligible fractional designs can be practical. Because order-of-addition orthogonal arrays are well-studied in the current literature, they are a good starting point when searching for eligible fractional designs. Below, two software packages are used to generate eligible fractional designs. Given the number of components m and the number of treatments n , the R function `optFederov` in the package `AlgDesign` developed by Wheeler (2022) is used to generate 10,000 candidate designs whose pairwise order matrices are order-of-addition orthogonal arrays of strength two. Typically, `optFederov` is a time-consuming procedure to generate candidate designs for those cases with more than five components. Another program `Latinoa` developed by Tsai (2022) is implemented for such cases. Next, an eligible design is identified from these candidate designs by searching for positive and negative quasi-foldover vectors with respect to every pairwise order factor. All obtained designs are collected in the [supplementary data](#) of this paper and their parameters are shown in Table 6. Based on this design catalogue, researchers can choose eligible fractional designs with the most suitable run sizes for their experiments.

The numbers of quasi-foldover vector pairs in each design in Table 6 are also reported in the [supplementary data](#). Because p_{st} and n_{st} are often not equal over all pairwise order factors, the minimum number of quasi-foldover vector pairs, denoted by s , is reported in Table 6. Specifically, s indicates that at least s pairs of positive and negative quasi-foldover vectors can be used to test a particular dispersion effect using the fiducial test variable $T_{\delta_{st}}$ in [11]. Given an order-of-addition orthogonal array of strength two, quasi-foldover vector pairs can be very sparse, especially when m is moderately large and n is relatively small. A theoretical upper bound for the number of quasi-foldover

Table 6. Parameters of eligible order-of-addition orthogonal arrays of strength two.

m	n	s	Generator
4	12	1	optFederov
5	36	1	optFederov
5	48	2	optFederov
5	60	3	optFederov
5	72	4	optFederov
5	84	6	optFederov
5	96	7	optFederov
5	108	8	optFederov
6	72	1	Latinoa
6	96	1	Latinoa
6	120	2	Latinoa
6	144	3	Latinoa
6	168	4	Latinoa
6	192	5	Latinoa
7	168	1	Latinoa
7	252	1	Latinoa

Table 7. Full and reduced positive quasi-foldover vectors.

						Pairwise order factor					
						Z ₁₂	Z ₁₃	Z ₁₄	Z ₂₃	Z ₂₄	Z ₃₄
Treatment											
t ₁	1	2	3	4	z ₁	+1	+1	+1	+1	+1	+1
					z ₁ [*]	+1	+1	.	+1	.	.
t ₂	4	3	1	2	z ₂	+1	-1	-1	-1	-1	-1
					z ₂ [*]	+1	-1	.	-1	.	.

vector pairs can be practical in searching for eligible fractional designs. This topic is beyond the scope of this paper. However, it deserves further investigation.

7. Concluding remarks

Because both frequentist and fiducial methods require an n -treatment order-of-addition design to be replicated r times, the run size is equal to $N = n \times r$. The run size N can be quite large when n is moderately large, with the result that it can be infeasible to conduct such a large-scale experiment. As pointed out by an anonymous paper reviewer, a practitioner can project an unreplicated design onto a subset of components, such that all reduced treatments have two or more replicates. According to Theorem 3 of Zhao, Lin, and Liu (2022), after projecting onto any two or three components of an order-of-addition orthogonal array of strength two, every reduced treatment has two or more replicates. By fitting a reduced pairwise order variance model to the observed data, the frequentist method is then applicable, where the observations of each reduced treatment must be assumed to have the same mean. On the other hand, the fiducial method is also applicable under such an assumption. To see this, consider the two four-component treatments t_1 and t_2 in Table 7. First, it is straightforward to see that the two vectors z_1 and z_2 constitute a pair of positive quasi-foldover vectors with respect to z_{12} .

Next, assume that $\delta_{14} = \delta_{24} = \delta_{34} = 1$, that is, all dispersion effects involving component 4 are inactive. After deleting the three pairwise order factors z_{14} , z_{24} and z_{34} , the two reduced vectors \mathbf{z}_1^* and \mathbf{z}_2^* constitute a pair of positive quasi-foldover vectors with respect to z_{12} . Note that the symbol \cdot in Table 7 is used to indicate that the pairwise order factor is deleted. Based on this observation, it is straightforward to verify that an eligible design for m components is still eligible for m' components after leaving out $m - m'$ components. Consequently, the fiducial method can be used when some components are left out of the original design.

Mee (2020) noted that adding interaction effects to the pairwise order mean model in [1] may yield better fitting results. Some real-world examples can be found in his paper. Because the frequentist and fiducial methods are based on sample variances of replicated treatments which are free of the location parameters $\mu_1, \mu_2, \dots, \mu_n$, these two methods can still be used to identify active dispersion effects when some interaction effects are incorporated into the pairwise order mean model in [1]. Sometimes, when conducting order-of-addition experiments, different treatments may have different numbers of replicates due to resource constraints. The fiducial method can be easily extended to analyze such unequi-replicated data. Suppose that the treatment t_h now has r_h replicates. Under the assumption of normality, based on the pairwise order variance model in [2], one has

$$\begin{aligned} V_h &= \frac{(r_h - 1)S_h^2}{\sigma_h^2} \\ &= \frac{(r_h - 1)S_h^2}{\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{z_{h,ij}}} \sim \chi_{r_h-1}^2. \end{aligned} \quad [14]$$

The key difference between [14] and [6] is that the constant term r is replaced with r_h , with the result that the degrees of freedom of V_h can now be varied according to the number of replicates. Based on the chi-square random variables V_1, V_2, \dots, V_n having the form in [14], when r_1, r_2, \dots, r_n are different, the random quantity $R_{\delta_{st}}$ in [10] can be verified as an FG PQ for δ_{st} using similar arguments. That is, the fiducial method remains valid for analyzing unequi-replicated data. On the other hand, when r_1, r_2, \dots, r_n are not equal, the distributions of all logarithmic sample variances W_1, W_2, \dots, W_n can be first approximated by normal distributions with heterogeneous variances. Next, the weighted least squares method can be used to estimate all parameters on the right-hand-side of [4]. However, when the pairwise order matrix is an order-of-addition orthogonal array of strength two, related

frequentist test variables have no compact form such as that in [5]. In my view, the fiducial method appears to be more conceptually systematic for analyzing both equi-replicated and unequi-replicated data.

In order-of-addition experiments, as pointed out by another anonymous paper reviewer, active dispersion effects could be due to noise factors whose values are hard to control. For example, when exploring the optimal addition order of several food ingredients, certain chemical properties of these food ingredients can be very different in various difficult-to-control environments, with the result that the response variances can be heterogeneous. To address such order-of-addition problems, robust parameter design methodology with manipulated noise factors can be applied. Two replication-based statistical methods developed using frequentist and fiducial probability arguments are introduced in this paper to identify active dispersion effects from replicated order-of-addition experiments. Simulation results show that both approaches can maintain empirical sizes sufficiently close to the nominal level while their finite-sample performances are very similar. Because the fiducial test proposed by Tsai and Liao (2020) remains valid to identify active location effects under the pairwise order mean and variance models in [1] and [2], it can be incorporated into the experimental data analysis as shown in Section 5. From a statistical perspective, the fiducial method can provide a unified probability framework to analyze dispersion effects as well as location effects. However, its computational cost is larger than that of the frequentist method. Therefore, the frequentist method is further recommended for real-world applications due to its low computational cost.

About the author

Shin-Fu Tsai is an associate professor in the Department of Agronomy at National Taiwan University. His email address is shinfu@ntu.edu.tw.

Acknowledgements

I would like to express my thanks to the editor and two anonymous referees for their valuable comments and constructive suggestions. I am grateful to Computer and Information Networking Center of National Taiwan University for the support of high-performance computing facilities. Thanks also go to Xianting Ding (Professor, Shanghai Jiao Tong University) for providing the original drug combination data.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

Shin-Fu Tsai's research was supported in part by National Science and Technology Council of Taiwan (Grant Number NSTC 111-2118-M-002-005-MY2).

ORCID

Shin-Fu Tsai  <http://orcid.org/0000-0003-1553-282X>

References

- Al-Sarraj, R., C. von Brömssen, and J. Forkman. 2019. Generalized prediction intervals for treatment effects in random-effects models. *Biometrical Journal. Biometrische Zeitschrift* 61 (5):1242–57. doi: [10.1002/bimj.201700255](https://doi.org/10.1002/bimj.201700255).
- Bartlett, M. S., and D. G. Kendall. 1946. The statistical analysis of variance-heterogeneity and the logarithmic transformation. *Supplement to the Journal of the Royal Statistical Society* 8 (1):128–38. doi: [10.2307/2983618](https://doi.org/10.2307/2983618).
- Bergman, B., and A. Hynén. 1997. Dispersion effects from unreplicated designs in the 2^{k-p} series. *Technometrics* 39: 191–8.
- Box, G. E. P., and D. Meyer. 1986. Dispersion effects from fractional designs. *Technometrics* 28 (1):19–27. doi: [10.1080/00401706.1986.10488094](https://doi.org/10.1080/00401706.1986.10488094).
- Chen, J., R. Mukerjee, and D. K.-J. Lin. 2020. Construction of optimal fractional order-of-addition designs via block designs. *Statistics and Probability Letters* 161:108728. doi: [10.1016/j.spl.2020.108728](https://doi.org/10.1016/j.spl.2020.108728).
- Chen, J., X.-R. Zhang, and D. K.-J. Lin. 2021. Analysis of replicated order-of-addition experiments. *Statistics and Applications* 19:453–66.
- Hannig, J., H. Iyer, and P. Patterson. 2006. Fiducial generalized confidence intervals. *Journal of the American Statistical Association* 101 (473):254–69. doi: [10.1198/016214505000000736](https://doi.org/10.1198/016214505000000736).
- Hedayat, A. S., N. J. A. Sloane, and J. Stufken. 1999. *Orthogonal arrays: Theory and applications*. New York, NY: Springer-Verlag.
- Lee, Y., and J. A. Nelder. 1998. Generalized linear models for the analysis of quality-improvement experiments. *Canadian Journal of Statistics* 26 (1):95–105. doi: [10.2307/3315676](https://doi.org/10.2307/3315676).
- Lin, D. K.-J., and J. Peng. 2019. Order-of-addition experiments: A review and some new thoughts. *Quality Engineering* 31 (1):49–59. doi: [10.1080/08982112.2018.1548021](https://doi.org/10.1080/08982112.2018.1548021).
- Malekzadeh, A., and M. Kharrati-Kopaei. 2020. Simultaneous confidence intervals for the quantile differences of several two-parameter exponential distributions under the progressive type II censoring scheme. *Journal of Statistical Computation and Simulation* 90 (11):2037–56. doi: [10.1080/00949655.2020.1762084](https://doi.org/10.1080/00949655.2020.1762084).
- McGrath, R. N., and D. K.-J. Lin. 2001a. Testing multiple dispersion effects in unreplicated fractional factorial designs. *Technometrics* 43 (4):406–14. doi: [10.1198/00401700152672492](https://doi.org/10.1198/00401700152672492).
- McGrath, R. N., and D. K.-J. Lin. 2001b. Confounding of location and dispersion effects in unreplicated fractional factorials. *Journal of Quality Technology* 33 (2):129–39. doi: [10.1080/00224065.2001.11980062](https://doi.org/10.1080/00224065.2001.11980062).
- Mee, R. W. 2020. Order-of-addition modeling. *Statistica Sinica* 30:1543–59. doi: [10.5707/ss.202018.0210](https://doi.org/10.5707/ss.202018.0210).
- Nelder, J. A., and Y. Lee. 1991. Generalized linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis* 7 (1):107–20. doi: [10.1002/asm.3150070110](https://doi.org/10.1002/asm.3150070110).
- Pan, G. 1999. The impact of unidentified location effects on dispersion-effects identification from unreplicated factorial designs. *Technometrics* 41 (4):313–26. doi: [10.1080/00401706.1999.10485931](https://doi.org/10.1080/00401706.1999.10485931).
- Peng, J., R. Mukerjee, and D. K.-J. Lin. 2019. Design of order-of-addition experiments. *Biometrika* 106 (3): 683–94. doi: [10.1093/biomet/asz025](https://doi.org/10.1093/biomet/asz025).
- Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2 (6):110–4. doi: [10.2307/3002019](https://doi.org/10.2307/3002019).
- Tsai, S.-F. 2022. Generating optimal order-of-addition designs with flexible run sizes. *Journal of Statistical Planning and Inference* 218:147–63. doi: [10.1016/j.jspi.2021.11.001](https://doi.org/10.1016/j.jspi.2021.11.001).
- Tsai, S.-F., and C.-T. Liao. 2020. Detection of location and dispersion effects from partially replicated two-level factorial designs. *Journal of Quality Technology* 52 (3): 281–92. doi: [10.1080/00224065.2019.1571349](https://doi.org/10.1080/00224065.2019.1571349).
- Tsai, S.-F., C.-T. Liao, and F.-S. Chai. 2015. Identification of dispersion effects from partially replicated two-level factorial designs. *Journal of Quality Technology* 47 (1):43–53. doi: [10.1080/00224065.2015.11918105](https://doi.org/10.1080/00224065.2015.11918105).
- Van Nostrand, R. C. 1995. Design of experiments where the order of addition is important. In *ASA Proceedings of the Section on Physical and Engineering Sciences*, 155–60. Alexandria, VA: American Statistical Association.
- Voelkel, J. G. 2019. The design of order-of-addition experiments. *Journal of Quality Technology* 51 (3):230–41. doi: [10.1080/00224065.2019.1569958](https://doi.org/10.1080/00224065.2019.1569958).
- Voelkel, J. G., and K. P. Gallagher. 2019. The design and analysis of order-of-addition experiments: An introduction and case study. *Quality Engineering* 31 (4):627–38. doi: [10.1080/08982112.2019.1578374](https://doi.org/10.1080/08982112.2019.1578374).
- Wang, A., H. Xu, and X. Ding. 2020. Simultaneous optimization of drug combination dose-ratio sequence with innovative design and active learning. *Advanced Therapeutics* 3 (4):1900135. doi: [10.1002/adtp.201900135](https://doi.org/10.1002/adtp.201900135).
- Wang, B. X., and F. Wu. 2018. Inference on the gamma distribution. *Technometrics* 60 (2):235–44. doi: [10.1080/00401706.2017.1328377](https://doi.org/10.1080/00401706.2017.1328377).
- Wang, C., and R. W. Mee. 2022. Saturated and supersaturated order-of-addition designs. *Journal of Statistical Planning and Inference* 219:204–15. doi: [10.1016/j.jspi.2021.12.006](https://doi.org/10.1016/j.jspi.2021.12.006).
- Wang, X., B. X. Wang, Y. Hong, and P. H. Jiang. 2021. Degradation data analysis based on gamma process with random effects. *European Journal of Operational Research* 292 (3):1200–8. doi: [10.1016/j.ejor.2020.11.036](https://doi.org/10.1016/j.ejor.2020.11.036).
- Wheeler, B. 2022. *AlgDesign: Algorithmic experimental design*. R package version 1.2.1. <https://CRAN.R-project.org/package=AlgDesign>.
- Winker, P., J. Chen, and D. K.-J. Lin. 2020. The construction of optimal design for order-of-addition experiment via threshold accepting. In *Contemporary Experimental Design, Multivariate Analysis and Data Mining*, ed. J. Fan and J. Pan, 93–109. Cham, Switzerland: Springer-Verlag.

Wu, C. F. J., and M. S. Hamada. 2021. *Experiments: Planning, analysis, and optimization*. 3rd ed. Hoboken, NJ: John Wiley and Sons.

Zhao, Y., D. K.-J. Lin, and M.-Q. Liu. 2021. Designs for order-of-addition experiments. *Journal of Applied Statistics* 48 (8):1475–95. doi: 10.1080/02664763.2020.1801607.

Zhao, Y., D. K.-J. Lin, and M.-Q. Liu. 2022. Optimal designs for order-of-addition experiments. *Computational Statistics and Data Analysis* 165:107320. doi: 10.1016/j.csda.2021.107320.

Appendix A: Test variables for the baseline variance

Based on the inverse of information matrix in Section 3.1, when the pairwise order matrix is an order-of-addition orthogonal array of strength two, one has $c_{0,0} = 1/n$. The frequentist test variable given by

$$Z_{\alpha_0} = \frac{\hat{\alpha}_0 - e}{\sqrt{\frac{1}{n} \psi' \left(\frac{r-1}{2} \right)}}$$

can be used to evaluate $H_0 : \alpha_0 = e$ versus $H_1 : \alpha_0 \neq e$, where e represents an arbitrary constant. Similarly, the distribution of Z_{α_0} can be approximated by the standard normal distribution. On the other hand, a fiducial test variable for δ_0 is introduced to address this testing problem. Let \mathcal{F} denote the set consisting of all index pairs of full-foldover vectors over all pairwise order vectors. Formally, \mathcal{F} can be written as

$$\mathcal{F} = \{(a, b) : z_{a,ij} + z_{b,ij} = 0 \text{ for every } ij\}.$$

Define

$$R_{\delta_0} = \prod_{(a,b) \in \mathcal{F}} \left[\frac{(r-1)s_a^2}{V_a} \frac{(r-1)s_b^2}{V_b} \right]^{\frac{1}{2f}}, \quad [\text{A1}]$$

where f denotes the number of index pairs in \mathcal{F} . First, it is straightforward to see that the distribution of R_{δ_0} is free of all parameters. Therefore, condition (A) of Definition 1 is satisfied. Next, replacing S_h^2 in [6] with its realization s_h^2 and replacing h with either a or b . Plugging the resulting term into [A1] yields

$$\begin{aligned} & \prod_{(a,b) \in \mathcal{F}} \left[\frac{(r-1)s_a^2}{V_a} \frac{(r-1)s_b^2}{V_b} \right]^{\frac{1}{2f}} \\ &= \prod_{(a,b) \in \mathcal{F}} \left[\left(\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{z_{a,ij}} \right) \left(\delta_0 \prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{z_{b,ij}} \right) \right]^{\frac{1}{2f}} \\ &= \delta_0 \left[\prod_{i=1}^{m-1} \prod_{j=i+1}^m \delta_{ij}^{\frac{1}{2} \sum_{(a,b) \in \mathcal{F}} (z_{a,ij} + z_{b,ij})} \right]^{\frac{1}{2f}} \\ &= \delta_0, \end{aligned}$$

because

$$\sum_{(a,b) \in \mathcal{F}} (z_{a,ij} + z_{b,ij}) = 0$$

for every ij . That is, condition (B) of Definition 1 is also satisfied. Consequently, R_{δ_0} is an FGPQ for δ_0 . Based on the FGPQ R_{δ_0} , the fiducial test variable given by

$$T_{\delta_0} = \frac{R_{\delta_0}}{\delta_0}$$

can be used to evaluate $H_0 : \delta_0 = d$ versus $H_1 : \delta_0 \neq d$, where d denotes an arbitrary constant. Similarly, the corresponding p -value can be estimated using a Monte-Carlo algorithm.

Appendix B: Non-matrix forms of least squares estimators

Based on the inverse of information matrix in Section 3.1, when the pairwise order matrix is an order-of-addition orthogonal array of strength two, one has

$$\hat{\alpha} = (X^T X)^{-1} X^T W = \begin{bmatrix} \frac{1}{n} & \mathbf{0}^T \\ \mathbf{0} & \frac{3(m-1)}{n(m+1)} \mathbf{I} - \frac{3}{n(m+1)} \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{1}^T W \\ \mathbf{Z}^T W \end{bmatrix},$$

where \mathbf{Z} denotes the $n \times q$ pairwise order matrix. First, it is straightforward to see that

$$\hat{\alpha}_0 = \frac{1}{n} \sum_{h=1}^n W_h.$$

The intercept estimator can be expressed in non-matrix form. Next, let

$$\mathbf{U} = \mathbf{Z}^T \mathbf{W}.$$

Given a pairwise order factor z_{st} , its corresponding entry in \mathbf{U} can be expressed as

$$U_{st} = \sum_{h \in \{h: z_{h,st}=+1\}} W_h - \sum_{h \in \{h: z_{h,st}=-1\}} W_h.$$

Define

$$\mathcal{A}_{st} = \{kl : k = s, l \neq t \text{ or } k \neq s, l = t\}$$

and

$$\mathcal{B}_{st} = \{kl : k = t \text{ or } l = s\}.$$

Based on these two index sets, the least squares estimator of α_{st} can be written as

$$\hat{\alpha}_{st} = \frac{3(m-1)}{n(m+1)} U_{st} - \frac{3}{n(m+1)} \left(\sum_{kl \in \mathcal{A}_{st}} U_{kl} - \sum_{kl \in \mathcal{B}_{st}} U_{kl} \right).$$

The least squares estimator of α_{st} can also be expressed in non-matrix form. Furthermore, if the number of components m is very large, the term $\frac{3}{n(m+1)}$ would be very small, with the result that all logarithmic sample variances W_1, W_2, \dots, W_n would be nearly equally weighted up to sign in the least squares fit. Therefore, as n goes to infinity, the normal approximation for the distribution of $Z_{\alpha_{st}}$ in [5] can be further strengthened by a central limit theorem-like argument.

Appendix C: Simulation results

Table C1. Empirical powers of the frequentist test variable $Z_{\alpha t}$.

Scenario	r	δ	Dispersion effect						Average	
			δ_{12}	δ_{13}	δ_{14}	δ_{23}	δ_{24}	δ_{34}	Size	Power
1	2	4	0.219	0.050	0.055	0.054	0.052	0.058	0.054	0.219
		6	0.308	0.045	0.053	0.056	0.055	0.050	0.052	0.308
		8	0.405	0.050	0.052	0.057	0.055	0.047	0.052	0.405
	3	4	0.510	0.063	0.054	0.054	0.047	0.051	0.054	0.510
		6	0.725	0.047	0.051	0.050	0.048	0.051	0.049	0.725
		8	0.842	0.053	0.049	0.052	0.053	0.050	0.051	0.842
2	2	4	0.204	0.200	0.054	0.051	0.051	0.053	0.052	0.202
		6	0.302	0.303	0.047	0.054	0.047	0.055	0.051	0.303
		8	0.401	0.393	0.053	0.046	0.043	0.050	0.048	0.397
	3	4	0.496	0.500	0.051	0.054	0.051	0.060	0.054	0.498
		6	0.721	0.719	0.056	0.051	0.052	0.057	0.054	0.720
		8	0.841	0.850	0.051	0.051	0.054	0.054	0.053	0.846
3	2	4	0.206	0.054	0.049	0.054	0.055	0.202	0.053	0.204
		6	0.302	0.051	0.051	0.054	0.056	0.307	0.053	0.305
		8	0.393	0.058	0.054	0.047	0.054	0.402	0.053	0.398
	3	4	0.503	0.054	0.057	0.055	0.054	0.507	0.055	0.505
		6	0.723	0.054	0.053	0.051	0.053	0.724	0.053	0.724
		8	0.847	0.055	0.059	0.058	0.053	0.845	0.056	0.846
4	2	4	0.193	0.047	0.212	0.049	0.206	0.050	0.049	0.204
		6	0.313	0.050	0.313	0.050	0.303	0.049	0.050	0.310
		8	0.395	0.049	0.399	0.050	0.408	0.055	0.051	0.401
	3	4	0.499	0.056	0.505	0.050	0.503	0.048	0.051	0.502
		6	0.735	0.051	0.719	0.051	0.717	0.050	0.051	0.724
		8	0.846	0.054	0.848	0.051	0.853	0.054	0.053	0.849
5	2	4	0.203	0.048	0.056	0.047	0.056	0.056	0.053	0.203
		6	0.316	0.051	0.051	0.058	0.056	0.048	0.053	0.316
		8	0.398	0.054	0.061	0.051	0.053	0.052	0.054	0.398
	3	4	0.510	0.047	0.050	0.051	0.053	0.057	0.052	0.510
		6	0.721	0.047	0.052	0.057	0.055	0.054	0.053	0.721
		8	0.847	0.051	0.055	0.056	0.052	0.053	0.053	0.847
6	2	4	0.212	0.204	0.053	0.055	0.051	0.050	0.052	0.208
		6	0.296	0.308	0.054	0.054	0.054	0.054	0.054	0.302
		8	0.401	0.398	0.054	0.055	0.052	0.055	0.054	0.400
	3	4	0.506	0.508	0.047	0.053	0.054	0.051	0.051	0.507
		6	0.724	0.733	0.051	0.055	0.051	0.048	0.051	0.729
		8	0.853	0.855	0.053	0.050	0.054	0.048	0.051	0.854
7	2	4	0.206	0.056	0.051	0.051	0.045	0.212	0.051	0.209
		6	0.305	0.058	0.047	0.053	0.054	0.310	0.053	0.308
		8	0.401	0.050	0.048	0.056	0.054	0.392	0.052	0.397
	3	4	0.512	0.052	0.051	0.045	0.050	0.509	0.050	0.511
		6	0.735	0.046	0.046	0.047	0.053	0.710	0.048	0.723
		8	0.844	0.053	0.047	0.045	0.047	0.846	0.048	0.845
8	2	4	0.203	0.056	0.202	0.053	0.203	0.049	0.053	0.203
		6	0.307	0.050	0.316	0.053	0.301	0.051	0.051	0.308
		8	0.402	0.051	0.399	0.047	0.405	0.052	0.050	0.402
	3	4	0.507	0.049	0.519	0.051	0.501	0.054	0.051	0.509
		6	0.729	0.042	0.730	0.051	0.735	0.052	0.048	0.731
		8	0.833	0.055	0.851	0.054	0.844	0.051	0.053	0.843

Empirical powers of active effects are shown in boldface.

Table C2. Empirical powers of the fiducial test variable $T_{\delta_{\text{st}}}^*$.

Scenario	r	δ	Dispersion effect						Average	
			δ_{12}	δ_{13}	δ_{14}	δ_{23}	δ_{24}	δ_{34}	Size	Power
1	2	4	0.325	0.076	0.081	0.083	0.079	0.052	0.074	0.325
		6	0.503	0.094	0.102	0.097	0.093	0.053	0.088	0.503
		8	0.622	0.119	0.112	0.108	0.114	0.052	0.101	0.622
	3	4	0.750	0.136	0.140	0.138	0.147	0.052	0.123	0.750
		6	0.931	0.207	0.196	0.206	0.205	0.050	0.173	0.931
		8	0.977	0.261	0.263	0.262	0.264	0.047	0.219	0.977
2	2	4	0.527	0.526	0.169	0.051	0.077	0.074	0.093	0.527
		6	0.740	0.735	0.249	0.055	0.100	0.099	0.126	0.738
		8	0.862	0.859	0.326	0.051	0.116	0.119	0.153	0.861
	3	4	0.947	0.936	0.423	0.055	0.141	0.136	0.189	0.942
		6	0.995	0.995	0.619	0.048	0.205	0.210	0.271	0.995
		8	0.999	0.999	0.752	0.053	0.265	0.265	0.334	0.999
3	2	4	0.330	0.051	0.176	0.173	0.050	0.324	0.113	0.327
		6	0.501	0.051	0.236	0.253	0.049	0.496	0.147	0.499
		8	0.622	0.044	0.314	0.322	0.046	0.619	0.182	0.621
	3	4	0.755	0.053	0.428	0.428	0.051	0.754	0.240	0.755
		6	0.929	0.049	0.638	0.624	0.048	0.931	0.340	0.930
		8	0.976	0.047	0.750	0.752	0.053	0.975	0.401	0.976
4	2	4	0.325	0.170	0.714	0.052	0.332	0.167	0.130	0.457
		6	0.502	0.261	0.909	0.052	0.499	0.255	0.189	0.637
		8	0.629	0.332	0.969	0.051	0.631	0.319	0.234	0.743
	3	4	0.745	0.422	0.993	0.054	0.754	0.414	0.297	0.831
		6	0.929	0.629	0.999	0.048	0.929	0.614	0.430	0.952
		8	0.976	0.763	0.999	0.052	0.978	0.758	0.524	0.984
5	2	4	0.315	0.076	0.075	0.077	0.080	0.052	0.072	0.315
		6	0.506	0.103	0.094	0.100	0.097	0.049	0.089	0.506
		8	0.627	0.118	0.125	0.119	0.119	0.052	0.107	0.627
	3	4	0.759	0.134	0.134	0.142	0.143	0.051	0.121	0.759
		6	0.925	0.199	0.204	0.206	0.201	0.054	0.173	0.925
		8	0.976	0.250	0.249	0.252	0.265	0.049	0.213	0.976
6	2	4	0.533	0.516	0.166	0.055	0.079	0.087	0.097	0.525
		6	0.745	0.736	0.258	0.049	0.097	0.096	0.125	0.741
		8	0.861	0.862	0.317	0.052	0.116	0.109	0.149	0.862
	3	4	0.941	0.935	0.412	0.049	0.138	0.146	0.186	0.938
		6	0.994	0.995	0.620	0.053	0.204	0.197	0.269	0.995
		8	0.999	0.999	0.758	0.046	0.257	0.258	0.330	0.999
7	2	4	0.332	0.043	0.179	0.172	0.044	0.332	0.110	0.332
		6	0.491	0.048	0.248	0.263	0.051	0.506	0.153	0.499
		8	0.641	0.047	0.333	0.333	0.041	0.623	0.189	0.632
	3	4	0.761	0.050	0.424	0.423	0.053	0.765	0.238	0.763
		6	0.919	0.048	0.625	0.620	0.045	0.927	0.335	0.923
		8	0.977	0.049	0.760	0.755	0.056	0.977	0.405	0.977
8	2	4	0.327	0.165	0.724	0.048	0.321	0.169	0.127	0.457
		6	0.504	0.238	0.907	0.047	0.497	0.269	0.185	0.636
		8	0.615	0.316	0.964	0.045	0.631	0.325	0.229	0.737
	3	4	0.763	0.415	0.992	0.055	0.753	0.421	0.297	0.836
		6	0.928	0.642	0.999	0.051	0.928	0.626	0.440	0.952
		8	0.982	0.753	0.999	0.051	0.979	0.753	0.519	0.987

Empirical powers of active effects are shown in boldface.

Table C3. Empirical powers of the fiducial test variable $T_{\delta_{3t}}$.

Scenario	r	δ	Dispersion effect						Average	
			δ_{12}	δ_{13}	δ_{14}	δ_{23}	δ_{24}	δ_{34}	Size	Power
1	2	4	0.173	0.048	0.052	0.053	0.050	0.049	0.050	0.173
		6	0.274	0.048	0.050	0.054	0.048	0.048	0.050	0.274
		8	0.349	0.053	0.051	0.055	0.054	0.052	0.053	0.349
	3	4	0.464	0.052	0.048	0.045	0.051	0.046	0.048	0.464
		6	0.682	0.049	0.051	0.052	0.045	0.050	0.049	0.682
		8	0.804	0.051	0.051	0.055	0.050	0.053	0.052	0.804
2	2	4	0.174	0.186	0.056	0.048	0.052	0.048	0.051	0.180
		6	0.280	0.282	0.052	0.053	0.051	0.053	0.052	0.281
		8	0.357	0.355	0.049	0.049	0.047	0.053	0.050	0.356
	3	4	0.463	0.454	0.056	0.050	0.049	0.048	0.051	0.459
		6	0.675	0.682	0.053	0.052	0.048	0.049	0.051	0.679
		8	0.787	0.811	0.047	0.055	0.050	0.046	0.050	0.799
3	2	4	0.172	0.050	0.047	0.049	0.050	0.181	0.049	0.177
		6	0.271	0.046	0.046	0.049	0.052	0.280	0.048	0.276
		8	0.347	0.050	0.055	0.053	0.050	0.353	0.052	0.350
	3	4	0.454	0.054	0.048	0.045	0.047	0.463	0.049	0.459
		6	0.676	0.049	0.049	0.046	0.050	0.677	0.049	0.677
		8	0.802	0.049	0.050	0.048	0.053	0.806	0.050	0.804
4	2	4	0.180	0.047	0.178	0.051	0.179	0.049	0.049	0.179
		6	0.263	0.052	0.272	0.044	0.280	0.050	0.049	0.272
		8	0.351	0.049	0.353	0.051	0.355	0.057	0.052	0.353
	3	4	0.455	0.047	0.458	0.048	0.465	0.050	0.048	0.459
		6	0.683	0.041	0.683	0.052	0.674	0.051	0.048	0.680
		8	0.810	0.050	0.801	0.051	0.797	0.047	0.049	0.803
5	2	4	0.180	0.053	0.045	0.052	0.049	0.051	0.050	0.180
		6	0.264	0.052	0.045	0.052	0.051	0.053	0.051	0.264
		8	0.353	0.047	0.046	0.052	0.049	0.047	0.048	0.353
	3	4	0.466	0.048	0.056	0.051	0.050	0.056	0.052	0.466
		6	0.674	0.048	0.052	0.051	0.047	0.057	0.051	0.674
		8	0.807	0.052	0.047	0.054	0.048	0.053	0.051	0.807
6	2	4	0.183	0.175	0.049	0.047	0.056	0.049	0.050	0.179
		6	0.284	0.275	0.052	0.052	0.045	0.047	0.049	0.280
		8	0.355	0.355	0.053	0.054	0.057	0.049	0.053	0.355
	3	4	0.453	0.455	0.051	0.053	0.049	0.052	0.051	0.454
		6	0.681	0.674	0.050	0.050	0.052	0.053	0.051	0.678
		8	0.809	0.810	0.054	0.048	0.049	0.052	0.051	0.810
7	2	4	0.182	0.050	0.051	0.050	0.056	0.180	0.052	0.181
		6	0.278	0.050	0.048	0.053	0.057	0.273	0.052	0.276
		8	0.347	0.053	0.051	0.055	0.052	0.362	0.053	0.355
	3	4	0.459	0.055	0.048	0.049	0.047	0.463	0.050	0.461
		6	0.678	0.056	0.056	0.045	0.053	0.667	0.053	0.673
		8	0.805	0.052	0.048	0.049	0.051	0.816	0.050	0.811
8	2	4	0.178	0.043	0.176	0.053	0.184	0.048	0.048	0.179
		6	0.273	0.054	0.276	0.051	0.267	0.051	0.052	0.272
		8	0.361	0.048	0.344	0.048	0.349	0.048	0.048	0.351
	3	4	0.455	0.049	0.459	0.051	0.457	0.048	0.049	0.457
		6	0.676	0.049	0.682	0.049	0.675	0.051	0.050	0.678
		8	0.803	0.047	0.804	0.049	0.816	0.048	0.048	0.808

Empirical powers of active effects are shown in boldface.

Table C4. Average sizes of frequentist and fiducial methods under non-normal distributions.

Scenario	r	δ	Frequentist test variable $Z_{\delta_{st}}$					Fiducial test variable $T_{\delta_{st}}$				
			T_3	$\Gamma_{1,2}$	$\Gamma_{2,2}$	$\Gamma_{3,2}$	$\Gamma_{6,2}$	T_3	$\Gamma_{1,2}$	$\Gamma_{2,2}$	$\Gamma_{3,2}$	$\Gamma_{6,2}$
1	2	4	0.067	0.088	0.066	0.058	0.055	0.066	0.088	0.066	0.059	0.054
		6	0.069	0.089	0.068	0.063	0.054	0.062	0.086	0.062	0.058	0.055
		8	0.069	0.092	0.066	0.064	0.055	0.063	0.083	0.065	0.057	0.053
	3	4	0.098	0.129	0.084	0.070	0.058	0.092	0.127	0.081	0.070	0.058
		6	0.095	0.132	0.083	0.068	0.058	0.092	0.130	0.083	0.070	0.059
		8	0.099	0.128	0.084	0.067	0.060	0.094	0.123	0.082	0.071	0.058
2	2	4	0.071	0.089	0.067	0.058	0.057	0.062	0.090	0.061	0.056	0.052
		6	0.069	0.087	0.067	0.061	0.054	0.065	0.079	0.063	0.056	0.056
		8	0.062	0.088	0.067	0.059	0.060	0.067	0.088	0.066	0.057	0.055
	3	4	0.095	0.131	0.087	0.071	0.061	0.094	0.129	0.084	0.072	0.058
		6	0.095	0.133	0.079	0.070	0.060	0.096	0.133	0.079	0.067	0.062
		8	0.095	0.128	0.082	0.073	0.059	0.094	0.126	0.080	0.072	0.060
3	2	4	0.070	0.093	0.069	0.059	0.060	0.067	0.086	0.063	0.055	0.056
		6	0.067	0.090	0.065	0.060	0.055	0.063	0.088	0.064	0.056	0.053
		8	0.068	0.089	0.065	0.060	0.056	0.063	0.086	0.066	0.059	0.052
	3	4	0.097	0.131	0.082	0.072	0.061	0.095	0.126	0.079	0.069	0.058
		6	0.095	0.130	0.087	0.072	0.061	0.098	0.129	0.082	0.068	0.060
		8	0.095	0.131	0.086	0.071	0.060	0.097	0.123	0.080	0.073	0.061
4	2	4	0.066	0.092	0.067	0.064	0.054	0.062	0.088	0.066	0.057	0.051
		6	0.063	0.086	0.073	0.060	0.057	0.066	0.091	0.062	0.059	0.051
		8	0.068	0.091	0.066	0.059	0.056	0.066	0.082	0.067	0.059	0.053
	3	4	0.100	0.134	0.086	0.071	0.063	0.096	0.127	0.081	0.067	0.061
		6	0.093	0.132	0.083	0.070	0.063	0.096	0.128	0.083	0.065	0.062
		8	0.098	0.131	0.088	0.072	0.061	0.100	0.126	0.085	0.071	0.060
5	2	4	0.066	0.088	0.063	0.063	0.053	0.064	0.087	0.065	0.058	0.056
		6	0.066	0.091	0.065	0.059	0.057	0.066	0.088	0.062	0.058	0.055
		8	0.072	0.089	0.069	0.060	0.053	0.065	0.086	0.065	0.059	0.057
	3	4	0.098	0.131	0.086	0.070	0.061	0.092	0.125	0.081	0.072	0.060
		6	0.097	0.132	0.084	0.074	0.062	0.095	0.128	0.081	0.069	0.057
		8	0.095	0.136	0.087	0.073	0.060	0.094	0.126	0.082	0.073	0.057
6	2	4	0.065	0.092	0.065	0.065	0.054	0.065	0.088	0.067	0.057	0.054
		6	0.068	0.091	0.063	0.061	0.056	0.063	0.083	0.064	0.058	0.053
		8	0.065	0.089	0.066	0.059	0.056	0.065	0.083	0.063	0.061	0.055
	3	4	0.095	0.135	0.084	0.073	0.063	0.093	0.132	0.085	0.070	0.060
		6	0.094	0.133	0.085	0.073	0.056	0.096	0.124	0.082	0.071	0.061
		8	0.094	0.130	0.083	0.070	0.055	0.097	0.130	0.083	0.072	0.059
7	2	4	0.069	0.090	0.066	0.060	0.056	0.065	0.085	0.068	0.057	0.052
		6	0.069	0.088	0.068	0.061	0.059	0.065	0.083	0.064	0.059	0.056
		8	0.066	0.089	0.065	0.062	0.058	0.064	0.083	0.062	0.053	0.055
	3	4	0.100	0.134	0.087	0.070	0.059	0.095	0.125	0.083	0.069	0.061
		6	0.092	0.126	0.084	0.072	0.061	0.092	0.129	0.079	0.067	0.058
		8	0.099	0.129	0.083	0.073	0.061	0.092	0.129	0.083	0.068	0.063
8	2	4	0.067	0.088	0.065	0.061	0.056	0.066	0.085	0.061	0.061	0.053
		6	0.070	0.093	0.064	0.063	0.058	0.071	0.085	0.062	0.057	0.053
		8	0.068	0.088	0.064	0.060	0.059	0.061	0.086	0.065	0.055	0.052
	3	4	0.093	0.129	0.088	0.074	0.059	0.094	0.123	0.085	0.070	0.057
		6	0.099	0.128	0.082	0.074	0.060	0.096	0.128	0.084	0.066	0.059
		8	0.101	0.131	0.082	0.072	0.062	0.094	0.129	0.081	0.069	0.057

The excess kurtosis parameters of the gamma distributions $\Gamma_{1,2}$, $\Gamma_{2,2}$, $\Gamma_{3,2}$ and $\Gamma_{6,2}$ are equal to 6, 3, 2 and 1, respectively.

Table C5. Average powers of frequentist and fiducial methods under non-normal distributions.

Scenario	r	δ	Frequentist test variable $Z_{\delta_{st}}$					Fiducial test variable $T_{\delta_{st}}$				
			T_3	$\Gamma_{1,2}$	$\Gamma_{2,2}$	$\Gamma_{3,2}$	$\Gamma_{6,2}$	T_3	$\Gamma_{1,2}$	$\Gamma_{2,2}$	$\Gamma_{3,2}$	$\Gamma_{6,2}$
1	2	4	0.215	0.251	0.217	0.199	0.214	0.191	0.225	0.199	0.188	0.188
		6	0.314	0.336	0.323	0.321	0.318	0.279	0.310	0.281	0.281	0.278
		8	0.409	0.421	0.396	0.408	0.402	0.356	0.376	0.359	0.365	0.360
	3	4	0.501	0.504	0.503	0.498	0.512	0.462	0.465	0.443	0.450	0.461
		6	0.701	0.673	0.710	0.705	0.705	0.653	0.639	0.652	0.671	0.665
		8	0.809	0.790	0.819	0.827	0.836	0.765	0.739	0.772	0.782	0.782
2	2	4	0.223	0.243	0.213	0.216	0.209	0.196	0.217	0.198	0.190	0.186
		6	0.320	0.336	0.325	0.316	0.316	0.289	0.298	0.288	0.284	0.279
		8	0.404	0.417	0.413	0.396	0.393	0.366	0.376	0.363	0.366	0.359
	3	4	0.507	0.507	0.503	0.508	0.507	0.475	0.467	0.455	0.472	0.457
		6	0.690	0.678	0.703	0.708	0.723	0.655	0.635	0.656	0.660	0.659
		8	0.797	0.790	0.811	0.821	0.831	0.770	0.745	0.780	0.786	0.792
3	2	4	0.221	0.242	0.219	0.220	0.206	0.196	0.218	0.194	0.181	0.186
		6	0.316	0.334	0.318	0.312	0.307	0.285	0.305	0.281	0.286	0.278
		8	0.406	0.412	0.411	0.410	0.404	0.368	0.369	0.361	0.356	0.356
	3	4	0.508	0.510	0.504	0.503	0.497	0.472	0.466	0.474	0.472	0.459
		6	0.692	0.681	0.707	0.710	0.724	0.647	0.633	0.658	0.665	0.667
		8	0.801	0.788	0.821	0.829	0.835	0.761	0.732	0.775	0.778	0.792
4	2	4	0.222	0.233	0.216	0.208	0.207	0.198	0.215	0.198	0.188	0.179
		6	0.322	0.339	0.319	0.323	0.316	0.283	0.302	0.280	0.283	0.272
		8	0.404	0.415	0.402	0.402	0.399	0.368	0.370	0.361	0.367	0.357
	3	4	0.508	0.511	0.501	0.504	0.499	0.471	0.464	0.457	0.456	0.460
		6	0.690	0.678	0.707	0.711	0.711	0.648	0.636	0.656	0.663	0.668
		8	0.804	0.780	0.816	0.818	0.837	0.763	0.741	0.765	0.782	0.789
5	2	4	0.214	0.232	0.227	0.216	0.210	0.193	0.209	0.190	0.184	0.189
		6	0.311	0.346	0.314	0.313	0.312	0.286	0.312	0.289	0.290	0.285
		8	0.402	0.414	0.409	0.409	0.402	0.365	0.373	0.362	0.368	0.356
	3	4	0.499	0.506	0.510	0.516	0.509	0.473	0.467	0.463	0.471	0.459
		6	0.700	0.675	0.706	0.708	0.724	0.652	0.644	0.661	0.654	0.667
		8	0.811	0.781	0.806	0.827	0.842	0.755	0.739	0.780	0.776	0.801
6	2	4	0.220	0.236	0.216	0.213	0.204	0.198	0.215	0.195	0.189	0.186
		6	0.327	0.337	0.318	0.308	0.309	0.286	0.305	0.292	0.285	0.280
		8	0.411	0.417	0.403	0.407	0.401	0.358	0.378	0.366	0.357	0.354
	3	4	0.505	0.503	0.506	0.501	0.507	0.474	0.468	0.463	0.462	0.468
		6	0.698	0.681	0.704	0.708	0.717	0.648	0.640	0.653	0.665	0.667
		8	0.804	0.778	0.809	0.821	0.831	0.756	0.744	0.769	0.783	0.792
7	2	4	0.224	0.245	0.212	0.215	0.204	0.194	0.221	0.199	0.188	0.178
		6	0.325	0.335	0.320	0.319	0.320	0.283	0.301	0.285	0.284	0.276
		8	0.410	0.409	0.394	0.410	0.403	0.370	0.369	0.366	0.365	0.356
	3	4	0.514	0.506	0.506	0.500	0.508	0.461	0.471	0.467	0.471	0.460
		6	0.694	0.672	0.695	0.710	0.724	0.644	0.649	0.663	0.659	0.673
		8	0.804	0.778	0.813	0.820	0.837	0.765	0.740	0.772	0.783	0.796
8	2	4	0.217	0.237	0.220	0.210	0.205	0.191	0.224	0.203	0.190	0.180
		6	0.321	0.336	0.322	0.318	0.310	0.292	0.308	0.286	0.285	0.275
		8	0.402	0.414	0.405	0.398	0.399	0.362	0.374	0.356	0.365	0.357
	3	4	0.510	0.506	0.506	0.514	0.507	0.471	0.472	0.467	0.463	0.458
		6	0.694	0.670	0.709	0.705	0.719	0.647	0.633	0.656	0.662	0.671
		8	0.799	0.779	0.816	0.820	0.835	0.762	0.740	0.769	0.783	0.788

The excess kurtosis parameters of the gamma distributions $\Gamma_{1,2}$, $\Gamma_{2,2}$, $\Gamma_{3,2}$ and $\Gamma_{6,2}$ are equal to 6, 3, 2 and 1, respectively.