HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2018 June 27.

Published in final edited form as:

Biometrics. 2018 June ; 74(2): 538–547. doi:10.1111/biom.12777.

Experimental Design for Multi-drug Combination Studies Using Signaling Networks

Hengzhen Huang[1], **Hong-Bin Fang**[2,*], and **Ming T. Tan**[2,*]

[1]College of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, China

[2]Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington, DC 20057, USA

Summary

Combinations of multiple drugs are an important approach to maximize the chance for therapeutic success by inhibiting multiple pathways/targets. Analytic methods for studying drug combinations have received increasing attention because major advances in biomedical research have made available large number of potential agents for testing. The preclinical experiment on multi-drug combinations plays a key role in (especially cancer) drug development because of the complex nature of the disease, the need to reduce development time and costs. Despite recent progresses in statistical methods for assessing drug interaction, there is an acute lack of methods for designing experiments on multi-drug combinations. The number of combinations grows exponentially with the number of drugs and dose-levels and it quickly precludes laboratory testing. Utilizing experimental dose-response data of single drugs and a few combinations along with pathway/network information to obtain an estimate of the functional structure of the dose-response relationship *in silico*, we propose an optimal design that allows exploration of the dose-effect surface with the smallest possible sample size in this paper. The simulation studies show our proposed methods perform well.

Keywords

Experimental design; Dose-response; Drug combinations; Functional ANOVA; Signaling network

1. Introduction

Multi-drug combination is an important therapeutic approach for diseases such as cancer, viral or microbial infections, hypertension and other diseases involving complex biological pathways. Synergistic drug combinations, which are more effective than expected from summing effects of individual drugs, offer the potential for improved therapeutic index. The past decade has seen significant progresses in developing proper design and analysis methods for two/three-drug combination studies, which have increased the chances of identifying optimized combinations for further therapeutic opportunities (Tan et al., 2003, 2009; Kong and Lee 2006; Fang et al. 2008, 2015) as well as adaptive phase I clinical trial

* mtt34@georgetown.edu & hf183@georgetown.edu for correspondence.

designs that attempt to identify the best possible maximum tolerated doses through modeling of the joint dose-toxicity relationship (see, e.g., Yuan and Yin, 2008; Yin and Yuan, 2009a, 2009b; Tan et al., 2016; Yang et al., 2016).

However, combination drug therapy targeting just a few gene products may be ineffective (Chen and Dancey, 2008; Jones, et al., 2008; Parson et al., 2008). Increasing the number of agents in a combination has been another strategy for increasing the level or type of interaction produced. In the past decade, the approach to cancer therapy has been revolutionized by the identification of a variety of novel signal transduction targets amenable to therapeutic intervention. These targets were identified based on improved understanding of the molecular mechanisms of action of second messengers, other components of signal transduction pathways, and systems biology. These advances have also made available large number of potential agents and call for new quantitative approaches for combination therapy (Fitzgerald et al., 2006; Hopkins, 2008; Xavier and Sander, 2010). Despite the changing paradigm to target multiple pathways, methodological advances in accurately identifying drug interactions have fallen behind, as shown by a paucity of literature on the design and analysis of multi-drug combinations.

The design of multi-drug combination experiment presents exceptional challenges and a high dimensional statistical problem. The number of combinations reaches 59049 for a study of 10 drug combinations with only 3 dose-levels for each drug. Since the number of combinations grows exponentially with dose-levels, it quickly precludes laboratory testing. In spite of the biological advances mentioned above and the significance of multi-agent combinations, current methods remain to be descriptive, thus, fail to address dimensionality and often violate statistical assumptions. As a result, many multi-drug combination studies are designed suboptimally by studying only pairwise combinations while fixing the dose of one or more drugs.

To determine the interaction among multi-drugs, the dose-response surface provides a comprehensive description on dose effects. For estimating the high dimensional dose-response surface, experimental designs are required that provide selected concentrations or dose-levels of combinations, which allow exploration of the dose-effect surface with high accuracy at reasonable sample sizes. Recently, we developed a novel method to screen the large number of combinations and identify the functional structure of the dose-response relationship by using the dose-response data of single drugs and pathway/network knowledge (Fang et al., 2016). That is, data from experiments of single drugs and a few combinations as well as existing signaling network knowledge from sources such as KEGG are utilized to develop a statistical re-scaling model to describe the effects of drugs on network topology. The system comprises a series of statistical models with biological considerations, such as Hill equations, generic enzymatic rate equations and a regression model, to represent the cumulative effect of genes implicated in activation of the cell death machinery. In other words, a quantitative model can be established upon existing network topology along with meaningful signal propagation rules, and the significant drug interactions as well as the single drug effects are expected to “hide” in such a model. How to extract the significant drug interactions and single drug effects from the model will be described in Section 2. The method of Fang et al. (2016) is highly beneficial in bringing

forth a statistical framework for selecting drug interactions, and developing experimental designs and statistical analysis to estimate the high dimensional dose-response surface.

The purpose of this article is to derive designs of combinations of multi-drugs based on the functional structure of the dose-response relationship. The remainder of this paper is organized as follows. Section 2 briefly describes model formulation for the dose-response of multi-drug combinations. The experimental design of maximizing the posterior information on the dose-response is proposed in Section 3. Section 4 gives the design construction algorithm and sample size calculation. Simulation studies are conducted in Section 5. Concluding remarks and further discussions in Section 6 close this paper.

2. Model Formulation

Consider a combination study of s drugs $A_1, A_2, \ldots, A_s$ inhibiting some cell line or against some cancer tumor. Assume the dose-response surface to be

$$y(\mathbf{x}) = g(x_1, \ldots, x_s), \quad \text{for } \mathbf{x} = (x_1, \ldots, x_s)^\top \in \mathfrak{D}, \quad (1)$$

where x_i is the dose-level of drug A_i, y is the dose-effect scaled to be a viability (proportion of cells surviving) or a tumor volume (with some transformation), and $\mathfrak{D}$ is the dose region. Without loss of generality, we assume that $\mathfrak{D} = C^s = [0, 1]^s$ in this paper. Using the functional ANOVA decomposition (see, Sobol', 2001), the dose-response $y(\mathbf{x})$ has the following unique decomposition,

$$y(x_1, \ldots, x_s) = g_0 + \sum_{i=1}^{s} g_i(x_i) + \sum_{1 \le i < j \le s} g_{ij}(x_i, x_j) + \cdots + g_{1,2,\ldots,s}(x_1, \ldots, x_s), \quad (2)$$

where $g_0 = \int_{C^s} y(\mathbf{x}) d\mathbf{x}$ is the overall mean of $y(\mathbf{x})$, and

$$\int_0^1 g_{i_1,\ldots,i_u}(x_{i_1}, \ldots, x_{i_u}) dx_{i_k} = 0, \quad \text{for any } 1 \le u \le \text{ and } 1 \le k \le u, \quad (3)$$

$$\int_{C^s} g_{i_1,\ldots,i_u}(x_{i_1}, \ldots, x_{i_u}) g_{j_1,\ldots,j_v}(x_{j_1}, \ldots, x_{j_v}) dx_1 \cdots dx_s = 0, \quad \text{if } (i_1, \ldots, i_u) \ne (j_1, \ldots, j_v). \quad (4)$$

Eq.(3) implies that the terms on the right-hand side of Eq.(2) are centered, whereas Eq.(4) implies that the terms on the right-hand side of Eq.(2) are mutually orthogonal. Furthermore, if we denote $D = \int_{C^s} [y(\mathbf{x})]^2 d\mathbf{x} - g_0^2$ and $D_I = \int_{C^s} [g_I(\mathbf{x}_I)]^2 d\mathbf{x}$ for $I \subset \{1, \ldots, s\}$, then due to Eqs.(3) and (4) it follows

$$D = \sum_{I \subset \{1, \ldots, s\}} D_I . \quad (5)$$

Therefore, the influence of g_I on $y(\mathbf{x})$ can be measured by the ratio $S_I = D_I / D$, which is called the global sensitivity index and satisfies

$$\sum_{i=1}^{s} S_i + \sum_{1 \le i < j \le s} S_{ij} + \cdots + S_{1,2,\ldots,s} = 1 . \quad (6)$$

The variances D and D_I's and, hence, the global sensitivity indices can be approximated by the quasi Monte Carlo method (Fang, Li and Sudijianto 2006).

The global sensitivity indices are often used to rank the importance of the $g_I(x_I)$'s appearing on the right-hand side of Eq.(2). The larger the index S_I is, the more significant the effect of $g_I(x_I)$ in the dose-response is. Thus, the functional structure of $y(\mathbf{x})$ can be studied by calculating the indices. Since there are 2^s terms on the right-hand side of Eq.(2), the estimation of the corresponding coefficients of those terms will not be possible, with limited sample size and/or wrong design settings. Recently, Fang et al. (2016) proposed a novel procedure to identify the most significant $g_I(x_I)$'s by utilizing data from experiments of single drugs (and a few combinations) and existing signaling network knowledge. The simulation studies showed that most contributions of single drugs and drug-interactions in the dose-response yielding a total of global sensitivity indices over 85%, are consistent with those from the true dose-response. Denote the vector of the dominating terms (e.g., those terms with their total global sensitivity indices more than 80%) by $\mathbf{z}(\mathbf{x}) = [1, z_1(\mathbf{x}), \ldots, z_p(\mathbf{x})]^\top$, where 1 corresponds to the overall mean and each of the functions z_i corresponds to some g_I, and the corresponding vector of regression coefficients is denoted by $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)^\top$. Then the dose-response Eq.(1) becomes as

$$y(\mathbf{x}) = \mathbf{z}(\mathbf{x})^\top \boldsymbol{\theta} + f(\mathbf{x}), \quad \text{for } \mathbf{x} = (x_1, \ldots, x_s)^\top \in \mathfrak{D}, \quad (7)$$

where $f(\mathbf{x})$ is an unknown function and its global sensitivity index should be less than 20%. Since $g_I(x_I)$ is determined by integrating $y(\mathbf{x})$ with respect to some specific coordinates (see, Santner, Williams and Notz 2003, pages 193–194; Fang, Li and Sudjianto 2006, pages 193–194), it typically has no closed form representation. As a result, a reduced form needs to be specified for the corresponding $z_I(\mathbf{x})$. Throughout this paper $z_I(\mathbf{x})$ is taken to be the product of the centered variables involved to represent the corresponding drug interaction. The purpose of variable centralization is to make Eqs.(3) and (4) being satisfied among $z_I(\mathbf{x})$'s and it will not change the meanings of $z_I(\mathbf{x})$'s. In this case, $f(\mathbf{x})$ is interpreted as the effects that cannot be captured by the $z_I(\mathbf{x})$'s. Below is a simple illustration.

Example 1

Suppose that $s = 3$ and $y(\mathbf{x})$ with $\mathbf{x} = (x_1, x_2, x_3)$ is the *in silico* model describing the does-effects on the response. Then by the functional ANOVA one can obtain that

$$y(x_1, x_2, x_3) = g_0 + g_1(x_1) + g_2(x_2) + g_3(x_3) + g_{12}(x_1, x_2) + g_{13}(x_1, x_3) + g_{23}(x_2, x_3) + g_{123}(x_1, x_2, x_3),$$

where each $g_I(x_I)$ is uniquely determined by integrating $y(\mathbf{x})$ with respect to some specific coordinates (please see, Santner, Williams and Notz 2003, pages 193–194; Fang, Li and Sudjianto 2006, pages 193–194, for details). Furthermore, if the sensitivity indices of $g_1(x_1)$, $g_{12}(x_1, x_2)$ and $g_{123}(x_1, x_2, x_3)$ are the largest three and their total sensitivity indices is more that 80%, $y(\mathbf{x})$ is then re-written as

$$y(\mathbf{x}) = \theta_0 + \theta_1(x_1 - 0.5) + \theta_{12}(x_1 - 0.5)(x_2 - 0.5) + \theta_{123}(x_1 - 0.5)(x_2 - 0.5)(x_3 - 0.5) + f(\mathbf{x}),$$

where $f(\mathbf{x})$ is interpreted as the effects that cannot be captured by $z_1(\mathbf{x}) = (x_1 - 0.5)$, $z_2(\mathbf{x}) = (x_1 - 0.5)(x_2 - 0.5)$ and $z_3(\mathbf{x}) = (x_1 - 0.5)(x_2 - 0.5)(x_3 - 0.5)$. It is easy to check that Eqs.(3) and (4) are satisfied among $z_1(\mathbf{x})$, $z_2(\mathbf{x})$ and $z_3(\mathbf{x})$.

As will be seen in the next section, Eq.(7) is served as the model basis for the combination experiments. Notice that according to the functional structure obtained from functional ANOVA, the function $f(\mathbf{x})$ should satisfy

$$\int_{C^s} f(\mathbf{x}) d\mathbf{x} = 0 \text{ and } \int_{C^s} z(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0}. \quad (8)$$

The orthogonality requirement in (8) makes many conventional nonparametric methods such as spline technique and multivariate kernel smoothing not convenient to use. In fact, another way of thinking about the orthogonality is to use it to guarantee the identifiability of the regression parameter $\boldsymbol{\theta}$ (Wiens 1991; Tan, Fang and Tian 2009). Finding a function that is orthogonal to each of the regression functions, however, may be too restricted. Our take is that a function that can address the identifiability problem is more general than the one that is exactly orthogonal to the regression functions.

3. Experimental Design

If the interest is the estimation of regression coefficients, the D-optimal design (Wu and Hamada, 2009) and Bayesian D-optimal design (Chaloner and Verdinelli, 1995) may be useful. Since the purpose of a drug combination study is to discover the promising dose-level combinations among the agents (e.g., identify the synergistic dose region), a prediction-based design appears to be more desirable. In this section, a maximum entropy design is proposed for the combination experiments.

Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ be the candidate set of design points in the experimental domain, e.g., $\mathcal{X}$ is typically chosen to be a set of lattice points over the experimental domain. The aim is

to choose n points ($n \ll k$) from $\mathscr{X}$ as the experimental points such that the prediction variability at un-experimental points, conditionally on the experimental points, is minimized. Based on Eq.(7), the dose-response can be formulated as

$$Y_j(\mathbf{x}_i) = z(\mathbf{x}_i)^\top \boldsymbol{\theta} + f(\mathbf{x}_i) + \varepsilon_{ij}, i = 1, \ldots, k, j = 1, \ldots, n_i, \quad (9)$$

where $Y_j(\mathbf{x}_i)$ is the response value of the jth replication at the point $\mathbf{x}_i$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is the measurement error and n_i is the number of replications at $\mathbf{x}_i$. The unknown function $f(\mathbf{x})$ is modeled as a *Gaussian random function* with zero-mean and global covariance matrix $\mathrm{Cov}[(f(\mathbf{x}_1), \ldots, f(\mathbf{x}_k))^\top] = \sigma_f^2 \mathbf{V}_f^k$. In other words, $\mathbf{F}_{\mathscr{X}} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_k)]^\top$ is regarded as a realization of $f(\mathbf{x})$. Similarly, a Gaussian prior is placed on the regression parameters, i.e., $\boldsymbol{\theta} \sim N_{p+1}(\boldsymbol{\beta}_{\boldsymbol{\theta}}, \sigma_{\boldsymbol{\theta}}^2 \mathbf{I}_{p+1})$. Generally, we need the following assumptions.

Assumption 1: $\boldsymbol{\theta} \sim N_{p+1}(\boldsymbol{\beta}_{\boldsymbol{\theta}}, \sigma_{\boldsymbol{\theta}}^2 \mathbf{I}_{p+1})$, $\mathbf{F}_{\mathscr{X}} \sim N_k(\mathbf{0}, \sigma_f^2 \mathbf{V}_f^k)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ for $i = 1, \ldots, k, j = 1, \ldots, n_i$.

Assumption 2: $z_l(\mathbf{x})$ is the product of the centered variables involved, $l = 1, \ldots, p$.

Assumption 3: The distributions of $\boldsymbol{\theta}$, $\mathbf{F}_{\mathscr{X}}$ and $\boldsymbol{\varepsilon}_{ij}$'s are independent.

The above assumptions allow the experimenter to write down the model as follows

$$\mathbf{Y}_{\mathscr{X}} = \mathbf{Z}_{\mathscr{X}} \boldsymbol{\theta} + \mathbf{F}_{\mathscr{X}} + \varepsilon, \text{ with}$$

$$\mathbf{Y}_{\mathscr{X}} = [Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_k)]^\top \text{ and}$$

$$\mathbf{Z}_{\mathscr{X}} = [z(\mathbf{x}_1), \ldots, z(\mathbf{x}_k)]^\top \text{ and}$$

$$\boldsymbol{\theta} \sim N_{p+1}(\boldsymbol{\beta}_{\boldsymbol{\theta}}, \sigma_{\boldsymbol{\theta}}^2 \mathbf{I}_{p+1}) \text{ and}$$

$$\mathbf{F}_{\mathscr{X}} \sim N_k(\mathbf{0}, \sigma_f^2 \mathbf{V}_f^k) \text{ and}$$

$$\varepsilon \sim N_k(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{W}^{-1}) \quad (10)$$

where $\mathbf{W} = \mathrm{diag}\{n_1, \ldots, n_k\}$ is a diagonal matrix defining the replications on *each* of the candidate design points. In applications, the prior knowledge regarding which candidate

design point is more important is typically unavailable, setting the same number of replications for each candidate design point, say $n_1 = \cdots n_k = m$, is a fair practice. The number m is called hereafter as the number of replications for simplicity and its value may be determined by knowledge about the run-to-run variation. For example, if the variation among the animals is known to be substantial, the number of replications may need to be a relatively large one. It is worthwhile to note that the model in (10) is defined at *all* candidate design points, i.e., no matter whether the point $\mathbf{x}_i$ is actually selected as the design point or not, a number of replications is assigned to it before the experiment starts.

3.1 Design Criterion

In this section a design criterion is proposed under the special case that the model parameters $\sigma^2_{\boldsymbol{\theta}}, \sigma^2_f, \sigma^2_\varepsilon$ and $\mathbf{V}^k_f$ are known. This setting allows us to demonstrate why the proposed design criterion is desirable without cluttering the discussion with estimation issues, which are resolved in Section 3.2. Without loss of generality, let e be a n-run experiment selected from $\mathscr{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$, i.e., e has n distinct support points selected from $\mathscr{X}$. Denote $\mathbf{Y}_e$ as the vector of response values at e, $\mathbf{Y}_{\bar{e}}$ as the vector of response values at $\bar{e} = \mathscr{X} - e$. In other words, e is the set of experimental points whereas $\bar{e}$ is the set of un-experimental points such that $e \cup \bar{e} = \mathscr{X}$ and $e \cap \bar{e} = \varnothing$. Let $p_{\mathbf{Z}}(\cdot)$ be the probability density function of the random vector $\mathbf{Z}$, the *entropy* of $\mathbf{Z}$ is then defined by

$$\mathrm{Ent}(\mathbf{Z}) = -\int p_{\mathbf{Z}}(\mathbf{z}) \log p_{\mathbf{Z}}(\mathbf{z}) d\mathbf{z}.$$

Entropy is a measure of unpredictability of a random vector, i.e., the larger the value of Ent($\mathbf{Z}$), the more uniform the distribution of the random vector $\mathbf{Z}$—which in turns implies that the more unpredictable $\mathbf{Z}$ is likely to be (Shewry and Wynn 1987). The standard formula from information theory suggests that (cf., Lindley, 1956)

$$\mathrm{Ent}(\mathbf{Y}_{\mathscr{X}}) = \mathrm{Ent}(\mathbf{Y}_e) + E_{\mathbf{Y}_e}\{\mathrm{Ent}(\mathbf{Y}_{\bar{e}} | \mathbf{Y}_e)\}, \quad (11)$$

where $\mathbf{Y}_{\mathscr{X}} = (\mathbf{Y}_e, \mathbf{Y}_{\bar{e}})$ and the expectation is with respect to the marginal distribution of $\mathbf{Y}_e$. Obviously, it is desirable for a combination experiment to minimize the second term on the right-hand side of Eq.(11) because this term represents the average prediction variability of the unsampled vector given the experimental design. By the model in (10), $\mathbf{Y}_{\mathscr{X}}$ is a k-dimensional Gaussian vector with mean and covariance matrix being

$$\mu_{\mathbf{Y}_{\mathscr{X}}} = \mathbf{Z}_{\mathscr{X}}\boldsymbol{\beta}_{\boldsymbol{\theta}} \text{ and } \sum_{\mathbf{Y}_{\mathscr{X}}} = \sigma^2_{\boldsymbol{\theta}}\mathbf{Z}_{\mathscr{X}}\mathbf{Z}^{\top}_{\mathscr{X}} + \sigma^2_f\mathbf{V}^k_f + \sigma^2_\varepsilon\mathbf{W},$$

respectively. Also, for a Gaussian random vector $\Gamma \sim N_n(\mu_\Gamma, \Sigma_\Gamma)$ one can verify

$$\mathrm{Ent}(\Gamma) = \frac{1}{2}[\log\{\det(\sum\nolimits_\Gamma)\} + n\log(2\pi) + n], \quad (12)$$

which implies that Ent($\mathbf{Y}_{\mathscr{X}}$) is a constant. Therefore, minimizing the value of $E_{\mathbf{Y}_e}\{\text{Ent}(\mathbf{Y}_{\bar{e}}|\mathbf{Y}_e)\}$ is equivalent to maximizing the value of Ent($\mathbf{Y}_e$). The optimal design, denoted by e^*, obtained by solving the following optimization problem

$$e^* = \arg\max_{e \subset \mathscr{X}} \text{Ent}(\mathbf{Y}_e) \quad (13)$$

is referred to as the *maximum entropy design* in the literature (Santner, Williams and Notz 2003; Fang, Li and Sudijianto 2006). That is, by using the experimental design obtained from Eq.(13), the average prediction variability at un-experimental design points is minimized.

Further denoting $e = \{\mathbf{x}_1^e, \ldots, \mathbf{x}_n^e\}$ and $\mathbf{Y}_e = [Y(\mathbf{x}_1^e), \ldots, Y(\mathbf{x}_n^e)]^\top$. Then, by the model in (10) $\mathbf{Y}_e$ is a n-dimensional Gaussian vector with mean and covariance matrix being

$$\mu_{\mathbf{Y}_e} = \mathbf{Z}_e\boldsymbol{\beta}_{\boldsymbol{\theta}} \text{ and } \sum_{\mathbf{Y}_e} = \sigma_{\boldsymbol{\theta}}^2\mathbf{Z}_e\mathbf{Z}_e^\top + \sigma_f^2\mathbf{V}_f^e + \sigma_\varepsilon^2\mathbf{W}_e,$$

respectively, where $\mathbf{Z}_e = [\mathbf{z}(\mathbf{x}_1^e), \ldots, \mathbf{z}(\mathbf{x}_n^e)]^\top$, $\mathbf{V}_f^e$ is a submatrix of $\mathbf{V}_f^k$, determined by the experiment e, and $\mathbf{W}_e$ is diagonal matrix also determined by e. According to Eqs. (12) and (13), the proposed design criterion is to find a design e^* such that

$$e^* = \arg\max_{e \subset \mathscr{X}} \det(\sigma_{\boldsymbol{\theta}}^2\mathbf{Z}_e\mathbf{Z}_e^\top + \sigma_f^2\mathbf{V}_f^e + \sigma_\varepsilon^2\mathbf{W}_e). = \arg\max_{e \subset \mathscr{X}} \det\{(\sigma_{\boldsymbol{\theta}}^2/\sigma_f^2)\mathbf{Z}_e\mathbf{Z}_e^\top + \mathbf{V}_f^e + (\sigma_\varepsilon^2/\sigma_f^2)\mathbf{W}_e\}. \quad (14)$$

It is easy to see that the maximum entropy design criterion only depends on the experimental design points. This feature makes it convenient to use. In contrast, conventional prediction-based criteria, such as c-optimality, E-optimality, G-optimality, Q-optimality, I-optimality and their Bayesian versions, not only depend on the experimental design points but also on the points to be predicted. Consequently, their closed form representations are usually not easy to obtain for high dimensional cases or only target some very special points to be predicted. It shall be pointed out that only *exact* designs are considered here because n_i's (they are set to be the same in this paper) are required to be integer-valued for drug combination experiments. *Continuous* designs, which relax the requirement for n_i's to be integer-valued and may regard any probability measure as a design, are beyond the scope of this paper and readers are referred to Kiefer (1959) and Fang, Liu and Zhou (2011) for detailed treatments of this topic.

3.2 Parameter Estimation

The maximum entropy design criterion (14) is relative to the variance ratios $\sigma^2_{\boldsymbol{\theta}}/\sigma^2_f, \sigma^2_\varepsilon/\sigma^2_f$ and the correlation matrix of the random function $\mathbf{V}^e_f$. As mentioned in Section 2, the total global sensitivity indices of the dominating terms is usually more than 80%, whereas the global sensitivity index of $f(\mathbf{x})$ is less than 20%. This suggests that the variance ratio $\sigma^2_{\boldsymbol{\theta}}/\sigma^2_f \approx (\geq)4$. As is shown at the end of this section, the value of the ratio $\sigma^2_\varepsilon/\sigma^2_f$ is also convenient to specify. Then, the primary matter is to estimate the correlation matrix $\mathbf{V}^e_f$.

The idea to estimate $\mathbf{V}^e_f$ is to use the single drug dose-effect curves which are estimated from the experimental data of single drugs. Let the covariance function between $f(\mathbf{x}_i)$ and $f(\mathbf{x}_j)$ be defined by $\mathrm{cov}[f(\mathbf{x}_i), f(\mathbf{x}_j)] = \sigma^2_f R(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{x}_i$ and $\mathbf{x}_j$ are two design points and $R(\mathbf{x}_i, \mathbf{x}_j)$ is the correlation function. The most commonly-used correlation function is the power exponential correlation (Cressie, 1993; Santner, Williams and Notes, 2003),

$$R(\mathbf{x}_i, \mathbf{x}_j) = \prod_{u=1}^{s} \exp\left\{-\left(\frac{|x_{iu} - \mathbf{x}_{ju}|}{\phi_u}\right)^{p_u}\right\}, \quad (15)$$

where x_{iu} is the uth element of $\mathbf{x}_i$, $\phi_u > 0$ and $0 < p_u \leq 2$, $u = 1, \ldots, s$. In order to alleviate the computational complexity and make the design easier to interpret, the value of p_u is here considered to be fixed and given as 2, and $\phi_1 = \cdots = \phi_s = \phi$ is specified. The correlation in Eq.(15) is hence defined by the parameter ϕ.

Denote the set of design points by $se = \{\mathbf{x}^{se}_1, \ldots, \mathbf{x}^{se}_h\}$, where each $\mathbf{x}^{se}_i$ has at least one component not equivalent to 0, and the vector of observed response values by $\mathbf{Y}_{se} = [Y(\mathbf{x}^{se}_1), \ldots, Y(\mathbf{x}^{se}_h)]^\top$. Using the Gaussian assumption on $f(\mathbf{x})$, the log-likelihood is proportional to

$$l(\boldsymbol{\theta}, \sigma^2_f, \phi) = -\frac{1}{2}\{h \log \sigma^2_f + \log[\det(\mathbf{V}^{se}_f(\phi))] + (\mathbf{Y}_{se} - \mathbf{Z}_{se}\boldsymbol{\theta})^\top [\mathbf{V}^{se}_f(\phi)]^{-1}(\mathbf{Y}_{se} - \mathbf{Z}_{se}\boldsymbol{\theta})/\sigma^2_f\},$$

(16)

where $\mathbf{V}^{se}_f(\phi) = [R(\mathbf{x}^{se}_i, \mathbf{x}^{se}_j)]_{h\times h}$ and $\mathbf{Z}_{se} = [\mathbf{z}(\mathbf{x}^{se}_1), \ldots, \mathbf{z}(\mathbf{x}^{se}_h)]^\top$. Given ϕ, the maximum likelihood estimator (MLE) for $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\phi) = \{\mathbf{Z}^\top_{se}[\mathbf{V}^{se}_f(\phi)]^{-1}\mathbf{Z}_{se}\}^{-1}\mathbf{Z}^\top_{se}[\mathbf{V}^{se}_f(\phi)]^{-1}\mathbf{Y}_{se} \quad (17)$$

and the MLE for σ_f^2 is given by

$$\widehat{\sigma_f^2} = \widehat{\sigma_f^2}(\phi) = \frac{1}{h}(\mathbf{Y}_{se} - \mathbf{Z}_{se}\hat{\boldsymbol{\gamma}})^\top \mathbf{V}_{fse}^{-1}(\phi)(\mathbf{Y}_{se} - \mathbf{Z}_{se}\hat{\boldsymbol{\gamma}}). \quad (18)$$

Substituting Eqs.(17) and (18) into Eq.(16), one can obtain that the maximum of the likelihood over $\boldsymbol{\theta}$ and σ_f^2 is

$$l(\hat{\boldsymbol{\theta}}, \widehat{\sigma_f^2}, \phi) = -\frac{1}{2}\{h \log \widehat{\sigma_f^2}(\phi) + \log[\det(\mathbf{V}_f^{se}(\phi))] + h\},$$

which depends on $\boldsymbol{\phi}$ alone. Thus the MLE of $\boldsymbol{\phi}$ is given by

$$\hat{\phi} = \arg\min_{\mathscr{R}_+^s} \{h \log \widehat{\sigma_f^2}(\phi) + \log[\det(\mathbf{V}_f^{se}(\phi))]\}, \quad (19)$$

where $\widehat{\sigma_f^2}(\boldsymbol{\phi})$ is given by (18). Hence, the correlation $\mathbf{V}_f^e$ in the design criterion (14) can be estimated by the parameter $\hat{\boldsymbol{\phi}}$ obtained from Eq.(19). At this point, the value of $\sigma_\varepsilon^2/\sigma_f^2$ can be readily specified: on one hand, the value of σ_f^2 is already estimated by Eq.(18); on the other hand, the measurement error variance σ_ε^2 can be estimated by the pooled variance from the single drug experimental data (Tan et al., 2003, 2009; Fang et al., 2008, 2015).

4. Computational Algorithms

According to the maximum entropy design criterion in (14), we propose a computational algorithm for design construction in this section. An algorithm for sample size determination is also given. As noted in Section 3, the number of replications m is often determined by knowledge about the run-to-run variation. Therefore, the determination of the sample size essentially relies upon the determination of the number of design points.

4.1 Computational Algorithm for Design Construction

Recall that $\mathscr{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is the candidate set of the design points. In order to cover the dose region $\mathscr{C}^s$ thoroughly, the lattice method is often used to construct $\mathscr{X}$ (Fang, Li and Sudijianto 2006). That is, q equidistant dose-levels are first selected for each single drug, then all the level-combinations among the single drugs constitute $\mathscr{X}$. The n-run maximum entropy design e^* is selected from $\mathscr{X}$ by solving the optimization problem (14). Typically, q is not a small number so $n \ll k = q^s$, which means that candidate-set dependent algorithms such as row exchange algorithm may be time-consuming. We use the candidate-set free coordinate exchange algorithm, as described by Meyer and Nachtsheim (1995), to address the optimization problem (14).

At the beginning, a $n \times s$ starting design is randomly generated in the dose region C^s. That is, each entry of the starting design is randomly generated from the interval (0, 1). The starting design is improved by sequentially updating its entries. For each entry, the algorithm evaluates the effect of changing that entry to the q levels. If the objective function in (14) improves for at least one of these operations, then the current entry is updated with the level that results in the maximum entropy. After the first pass through each entry in the design matrix, the algorithm takes a second pass. If any entry of the design changes in the second pass, then the algorithm performs another pass. This process continues until there are no changes in any pass through the design or when a maximum iteration limit is reached. For practical purpose, a step-by-step algorithm for design construction is provided in Table 1.

The resulting design **D** from Table 1 may be local optimal with respect to the maximum entropy criterion, so multiple random starting designs may be used.

It is noteworthy that the number of replications, irrespective of whether it is the same or not for each candidate design point, needs to be preassigned for Algorithm 1. This is because if the number of replications changes, so dose the value of Ent($\mathbf{Y}_{\mathscr{X}}$) that appears on the left-hand side of Eq.(11). As discussed in Section 3.1, Ent($\mathbf{Y}_{\mathscr{X}}$) needs to be kept as a constant to justify the proposed maximum entropy design. Preassigning a number of replications at each candidate design point is the limitation of the proposed design. Also, an approach to verify whether the resulting design is optimal with respect to the maximum entropy criterion may need further study. For instance, a tight upper bound may exist for the entropy value given the number of replications. If one can find a tight upper bound of the entropy over the design region, this upper bound can be served as a benchmark for construction of the maximum entropy design and may speed up the searching process.

4.2 Computational Algorithm for Sample Size Determination

For any n, the maximum entropy design can be constructed via Algorithm 1. However, an exact value of n shall be determined before the combination experiments. In this section, we first propose a criterion for sample size determination. Similar to the practice used in Section 3.1, it is tentatively assumed that parameters $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, σ_ε^2 and σ_f^2 are known. Such a practice would facilitate the derivation of the sample size criterion without cluttering the discussion with estimation issues. It turns out that only the parameters $\boldsymbol{\phi}$, σ_ε^2 and σ_f^2 need to be estimated, which is the same requirement for the design criterion in (14).

Consider the estimation of $Y(\mathbf{x}_0) = \mathbf{z}(\mathbf{x}_0)^\top \boldsymbol{\theta} + f(\mathbf{x}_0)$ at any $\mathbf{x}_0 \in \bar{e}$. By **Assumptions** 1–3 and the model in (10), $[Y(\mathbf{x}_0), \mathbf{Y}_e^\top]^\top$ follows a Gaussian distribution with mean and covariance matrix being

$$\mu^* = \begin{pmatrix} z(\mathbf{x}_0)^\top \boldsymbol{\theta} \\ \mathbf{Z}_e \boldsymbol{\theta} \end{pmatrix} \text{ and } \sum{}^* = \begin{pmatrix} \sigma_f^2 & \sigma_f^2 \mathbf{r}(\mathbf{x}_0)^\top \\ \sigma_f^2 \mathbf{r}(\mathbf{x}_0) & \sigma_f^2 \mathbf{V}_f^e + \sigma_\varepsilon^2 \mathbf{W}_e \end{pmatrix},$$

respectively, where $e = \{\mathbf{x}_1^e, \ldots, \mathbf{x}_n^e\}$ is the experiment and $\mathbf{r}(\mathbf{x}_0) = [R(\mathbf{x}_0, \mathbf{x}_1^e), \ldots, R(\mathbf{x}_0, \mathbf{x}_n^e)]^\top$. It is known that the MSE-optimal predictor of $Y(\mathbf{x}_0)$, conditional on the experiment e, is given by $E[Y(\mathbf{x}_0)|\mathbf{Y}_e]$ (see, for example, page 40 of Shao (2003)). Let $\hat{Y}_e(\mathbf{x}_0) = E[Y(\mathbf{x}_0)|\mathbf{Y}_e]$, then according to the property of multivariate normal distribution (see, for example, page 248 of Fang, Li and Sudijianto (2006)) the closed form expression of $\hat{Y}_e(\mathbf{x}_0)$ is given by

$$\hat{Y}_e(\mathbf{x}_0) = \mathbf{z}(\mathbf{x}_0)^\top \boldsymbol{\theta} + \sigma_f^2 \mathbf{r}(\mathbf{x}_0)^\top [\sigma_f^2 \mathbf{V}_f^e(\boldsymbol{\phi}) + \sigma_\varepsilon^2 \mathbf{W}_e]^{-1} (\mathbf{Y}_e - \mathbf{Z}_e \boldsymbol{\theta}), \quad (20)$$

and the MSE of $\hat{Y}_e(\mathbf{x}_0)$ is

$$\mathrm{MSE}[\hat{Y}_e(\mathbf{x}_0); n] = E[Y_e(\mathbf{x}_0) - \hat{Y}_e(\mathbf{x}_0)]^2 = \sigma_f^2 - \sigma_f^2 \mathbf{r}(\mathbf{x}_0)^\top [\mathbf{V}_f^e(\boldsymbol{\phi}) + (\sigma_\varepsilon^2/\sigma_f^2)\mathbf{W}_e]^{-1} \mathbf{r}(\mathbf{x}_0).$$

The detailed derivation of the MSE is provided in Web Appendix A for interested readers. Define the relative MSE (RMSE) as to be

$$\mathrm{RMSE}[\hat{Y}_e(\mathbf{x}_0); n] = \mathrm{MSE}[\hat{Y}_e(\mathbf{x}_0); n]/\sigma_f^2 = 1 - \mathbf{r}(\mathbf{x}_0)^\top [\mathbf{V}_f^e(\boldsymbol{\phi}) + (\sigma_\varepsilon^2/\sigma_f^2)\mathbf{W}_e]^{-1} \mathbf{r}(\mathbf{x}_0)$$

and the average RMSE (ARMSE) as

$$\mathrm{ARMSE}(e; n) = \frac{\sum_{\mathbf{x}_0 \in \bar{e}} \mathrm{RMSE}[\hat{Y}_e(\mathbf{x}_0); n]}{k - n}.$$

It is easy to see that $0 < ARMSE(e, n) < 1$, thus our idea for sample size determination is to find the smallest number of design points, say n^*, such that

$$n^* = \arg \min_{n \in \mathcal{Z}^+} ARMSE(e_n; n) \le \delta,$$

where $0 < \delta < 1$ is a user-specified targeted precision, e_n is the n-run maximum entropy design constructed by Algorithm 1. The ARMSE depends on the parameters $\boldsymbol{\phi}$, σ_ε^2 and σ_f^2. As discussed in Section 3.2, $\boldsymbol{\phi}$ and σ_f^2 can be respectively estimated through Eqs.(19) and (18), whereas the measurement error variance σ_ε^2 can be estimated by the pooled variance from the single drug experimental data. Once $\boldsymbol{\phi}$, σ_ε^2 and σ_f^2 are estimated from the previous studies, the empirical MSE-optimal predictor can be obtained by plugging $\hat{\boldsymbol{\theta}}(\boldsymbol{\phi})$ in Eq.(17) or, alternatively, the posterior mean $E[\boldsymbol{\theta}|\mathbf{Y}_e]$ into Eq.(20). Such a predictor is more general than the regression predictor $\mathbf{z}(\mathbf{x}_0)^\top \hat{\boldsymbol{\theta}}(\boldsymbol{\phi})$ (or $\mathbf{z}(\mathbf{x}_0)^\top E[\boldsymbol{\theta}|\mathbf{Y}_e]$) when the random function $f(\mathbf{x})$ is presented in the dose-response model, as it not only accounts for the regression predictor but also the "correction" caused by the random function $f(\mathbf{x})$.

The computational algorithm we use for sample size determination is a root-finding algorithm, which proceeds as follows. For given m, let n_{min} and n_{max} be respectively the minimum and maximum numbers of design points that can be afforded. First of all, calculate ARMSE(e_{min}; n_{min}), where e_{min} is an n_{min}-run maximum entropy design. If ARMSE(e_{min}; n_{min}) ≤ δ, output $n = n_{min}$; otherwise increase n_{min} by Δ (e.g., Δ = 10) and calculate ARMSE(e_1; n_{min}+Δ), where e_1 is an (n_{min} + Δ)-run maximum entropy design. If ARMSE(e_1; n_{min} + Δ) ≤ δ, output $n = n_{min}$ + Δ; otherwise continue increasing the number of design points until the ARMSE no larger than δ or n_{max} is reached.

5. Numerical Illustration

In this section, we revisit one numerical experiment of Fang et al. (2016) to demonstrate how to construct the proposed maximum entropy design for a given multi-drug combination study and compare its efficiency with some other conventional designs. We use the example of Fang et al. (2016) as illustration because functional ANOVA, which is introduced in Section 2, had been applied to their *in silico* model. Based on our experience, the establishment of an *in silico* model upon a signaling network requires some carefully selected combination data and is computationally demanding. To save the efforts for data collection and computation time, the sensitivity indices calculated by Fang et al. (2016) are directly presented here.

In the numerical experiment of Fang et al. (2016), 10 single drugs, denoted by A_1, …, A_{10}, were considered and their single drug curves are provided in Web Appendix B. Using the single drug information plus some combination data, Fang et al. (2016) established an *in silico* model based on the apoptosis signaling network (hsa04210). Then, functional ANOVA was applied to this *in silico* model. Their results showed that A_1, $A_1A_2A_3$, $A_1A_2A_3A_4A_5$ and $A_1A_2A_3A_4A_5A_6A_7$ are significant single drug effect and interaction effects whose total sensitivity indices is about 80%. As a result, $\mathbf{z}(\mathbf{x}) = (x_1 - 0.5, (x_1 - 0.5)(x_2 - 0.5)(x_3 - 0.5), (x_1 - 0.5)(x_2 - 0.5)(x_3 - 0.5)(x_4 - 0.5)(x_5 - 0.5), (x_1 - 0.5)(x_2 - 0.5)(x_3 - 0.5)(x_4 - 0.5)(x_5 - 0.5)(x_6 - 0.5)(x_7 - 0.5))$ is taken for this example.

To estimate the correlation matrix $\mathbf{V}_f^e$, 8 equidistant dose-levels are selected from the interval [0.01, 0.99] for each single drug curve. The data generated from the single drug curves are provided in Web Table 1. Using the data in Web Table 1, the maximum likelihood estimates of the correlation parameter and the random function variance are $\hat{\boldsymbol{\phi}} = 1.5515$ and $\hat{\sigma}_f^2 = 466.5115$, respectively. In addition, the measurement error variance (pooled variance) is estimated to be $\hat{\sigma}_\varepsilon^2 = 254.8211$. Based on these estimates, one can use the design criterion (14) (with $\sigma_{\boldsymbol{\theta}}^2/\sigma_f^2 = 4$) and the computational algorithms described in Section 4 to construct the maximum entropy designs and determine the corresponding sample sizes. In particular, three values of $\delta = 0.30, 0.20, 0.10$ and four values of $m = 2, 4, 6, 8$ are examined. Under each tuple of (δ, m), the sample sizes are plotted in Figure 1 with mark "○". As expected, for given number of replications, the sample size increases as δ becomes small. In addition, for given δ the more replications are, the more is the sample size. We stress that for different drug combination studies, the sample sizes and the corresponding experimental designs may

vary significantly. However, one can always follow our approach to determine the sample sizes and the corresponding experimental designs and, then, an appropriate tuple of (δ, m) may be selected for the combination experiments. For instance, if the sample size could not be more than 300 in a study, the tuple (δ, m) = (0.20, 4) might be selected because this choice yields the highest precision with the largest sample size ≤ 300.

To assess the effectiveness of using the design criterion (14), consider another design criterion below

$$e^* = \arg \max_{e \subset \mathcal{X}} \det\{\mathbf{V}_f^e + (\sigma_\varepsilon^2/\sigma_f^2)\mathbf{W}_e\} \quad (21)$$

The above criterion represents the case where no regression terms are identified for the dose-response. Based on criterion (21) and the estimates previously obtained, the sample sizes under various tuples of (δ, m) are plotted in Figure 1 with mark "×". By comparing the sample sizes marked with "○" and those marked with "×", it is easy to see that for any given (δ, m) the sample size determined from criterion (21) is considerably larger than that determined from criterion (14). This indicates that by incorporating significant regression terms into the design criterion, the sample size for achieving a given model accuracy can be significantly reduced. Furthermore, the sample sizes of the D-optimal design (marked with "△"), which maximizes $\det[\mathbf{Z}_e^\top(\sigma_f^2\mathbf{V}_f^e + \sigma_\varepsilon^2\mathbf{W}_e)^{-1}\mathbf{Z}_e]$, and the Bayesian D-optimal design (marked with "+"), which maximizes $\det\left\{(\sigma_{\boldsymbol{\theta}}^2/\sigma_f^2)\mathbf{Z}_e^\top[\mathbf{V}_f^e + (\sigma_\varepsilon^2/\sigma_f^2)\mathbf{W}_e]^{-1}\mathbf{Z}_e + \mathbf{I}_p\right\}$, are also plotted in Figure 1 for comparison purpose. In summary, the Bayesian D-optimal design is a bit more efficient than the D-optimal design and both of them are clearly more efficient than the design under criterion (21). However, the proposed maximum entropy design performs the best across all tuples of (δ, m).

From the simulation study, we have

i. To give the experimenters an impression that how the design points distribute over the design region, bivariate projections of the compared designs with (n, m) = (50, 2) are presented in Figures 2 and 3. The bivariate projections of the design under criterion (21) (see Figure 2 (a)) look similar for any pair of variables. This is due to the isotropy (i,e., $\boldsymbol{\phi}_1 = \cdots = \boldsymbol{\phi}_s = \boldsymbol{\phi}$) specified for the correlation function (15). The bivariate projections of the Bayesian D-optimal design (see Figure 2 (b)) show that more and more no-extreme levels tends to emerge as the variable subscript increases. This indicates that the less frequent a variable appears in the regression functions the more no-extreme levels it tends to take. The bivariate projections of the proposed maximum entropy design (see Figure 2 (c)) can be viewed as a compromise between Figures 2 (a) and (b) in the sense that the less frequent a variable appears in the regression functions the more no-extreme levels it tends to take, but the design still tries to keep the projections similar especially for the first a few variables. In fact, by using the standard formula $\det(A + BC) = \det(A)\det(I + CA^{-1}B)$ one can obtain that

$$\det\left\{\frac{\sigma_{\boldsymbol{\theta}}^2}{\sigma_f^2}\mathbf{Z}_e\mathbf{Z}_e^\top+\mathbf{V}_f^e+\frac{\sigma_\varepsilon^2}{\sigma_f^2}\mathbf{W}_e\right\}=\det\left\{\frac{\sigma_{\boldsymbol{\theta}}^2}{\sigma_f^2}\mathbf{Z}_e^\top\left[\mathbf{V}_f^e+\frac{\sigma_\varepsilon^2}{\sigma_f^2}\mathbf{W}_e\right]^{-1}\mathbf{Z}_e+\mathbf{I}_p\right\} \quad (22)$$
$$\times\det\left\{\mathbf{V}_f^e+\frac{\sigma_\varepsilon^2}{\sigma_f^2}\mathbf{W}_e\right\}.$$

This means that the criterion value of the proposed maximum entropy design is the product of the criterion value in (21) and that of the Bayesian D-optimal design. As a result, Figure 2 (c) looks like a compromise between Figures 2 (a) and (b). Finally, the D-optimal design (see Figure 3 (a)) pushes the variables appearing in the regression functions towards their extreme levels while allows the remaining variables to take many non-extreme levels. The driver for the higher number of levels on drugs 8, 9 and 10 is that the regression functions currently do not include drugs 8, 9 and 10, such that all information on them may only be acquired via the correlation function of $f(\mathbf{x})$. The bivariate projections under other values of (n, m) are similar so the details are omitted.

ii. A sensitivity analysis for the variance ratio $\sigma_{\boldsymbol{\theta}}^2/\sigma_f^2$ may be needed if its value is suspicious not to be around 4. For large $\sigma_{\boldsymbol{\theta}}^2/\sigma_f^2$ Eq.(22) indicates that the design criterion value behaves like $\det\left\{\mathbf{Z}_e^\top[\mathbf{V}_f^e+(\sigma_\varepsilon^2/\sigma_f^2)\mathbf{W}_e]^{-1}\mathbf{Z}_e\right\}$ which is D-optimality, whereas a small $\sigma_{\boldsymbol{\theta}}^2/\sigma_f^2$ means that the design criterion value behaves like that in Eq.(21). Figures 3 (b) and (c) present the bivariate projections of the maximum entropy design with $(n, m, \sigma_{\boldsymbol{\theta}}^2/\sigma_f^2) = (50, 2, 0.01)$ and (50, 2, 100), respectively. It is not difficult to see that Figure 3 (b) is similar to Figure 2 (a) while Figure 3 (c) looks similar to Figure 3 (a). Although only two values are examined here, such a sensitivity analysis demonstrates that how the distribution of design points varies as the value of $\sigma_{\boldsymbol{\theta}}^2/\sigma_f^2$ changes.

iii. Increasing the number of replications may not significantly reduce the number of design points. For example, the sample size of $m = 4$ almost doubles that of $m = 2$. As pointed out by one referee, the number of required drugs and doses to estimate the regression parameters is relatively small and they will provide no information from the model on drugs 8, 9 and 10. Doubling the number of replications adds information to the model. However, as the regression predictor will not forward much information to drugs 8, 9 and 10, the number of design points would not significantly decrease, as design points are still needed to cover much of the design region for the modeling via the correlation function of $f(\mathbf{x})$.

6. Concluding Remarks and Discussions

The proposed novel combination of functional ANOVA with maximum entropy designs utilizes both the biological pathway and single drug data for the selection of optimal drug

combinations in multi-drug combination studies. The proposed study designs for drug-combinations provide a basis for future developments on statistical methods and experimental designs for complex multi-drug dose-finding problems. The simulation studies showed that the proposed experimental design (dose-level selection and sample size determination) is efficient for combination studies and statistical procedures to fit the high dimensional dose-response surface.

7. Supplementary Materials

Web Appendices A, B and Web Table 1 referenced in Sections 4–5 are available with this paper at *Biometrics* website on Wiley Online Library.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the editor, the associate editor and two reviewers for their constructive comments which improved the manuscript significantly. Huang's research was completed at Georgetown University where he was a post-doctoral research fellow. This work was partially supported by the National Cancer Institute (NCI) grant R01CA164717 and the National Natural Science Foundation of China (Grant No. 11701109).

References

Chen, HX., Dancey, JE. Combinations of Molecular-Targeted Therapies: Opportunities and Challenges. In: Kaufman, HL.Wadler, S., Antman, K., editors. Molecular Targeting in Oncology. Totowa, New Jersey: Humana Press; 2008. p. 693-705.

Cressie, NAC. Statistics for Spatial DATA. New York: Wiley; 1993.

Chaloner K, Verdinelli I. Bayesian experimental design: a review. Statistical Science. 1995; 10:273–304.

Fang HB, Huang H, Clarke R, Tan M. Predicting multi-drug inhibition interactions based on signaling networks and single drug dose-response information. Journal of Computational Systems Biology. 2016; 2:101.

Fang HB, Ross DD, Sausville E, Tan M. Experimental design and interaction analysis of combination studies of drugs with log-linear dose responses. Statistics in Medicine. 2008; 27:3071–3083. [PubMed: 18186545]

Fang HB, Chen XR, Pei XY, Grant S, Tan M. Experimental design and statistical analysis for three-drug combination studies. Statistical Methods in Medical Research. 2017; 26:1261–1280. [PubMed: 25744107]

Fang, KT., Li, R., Sudjianto, A. Design and Modeling for Computer Experiments. New York: Chapman & Hall/CRC; 2006.

Fang, KT., Liu, MQ., Zhou, YD. Design and Modeling of Experiments. Beijing: Higher Education Press; 2011.

Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK. Systems biology and combination therapy in the quest for clinical efficacy. Nature Chemical Biology. 2006; 2:458–466. [PubMed: 16921358]

Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nature Chemical Biology. 2008; 4:682–690. [PubMed: 18936753]

Lindley DV. On a measure of information provided by an experiment. Annals of Mathematical Statistics. 1956; 27:986–1005.

Jones S, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analysis. Science. 2008; 321:1801–1806. [PubMed: 18772397]

Kiefer J. Optimum experimental designs (with discussion). Journal of the Royal Statistical Society B. 1959; 21:272–319.

Kong M, Lee JJ. A generalized response surface model with varying relative potency for assessing drug interaction. Biometrics. 2006; 62:986–995. [PubMed: 17156272]

Meyer RK, Nachisheim CJ. The coordinate-exchange algorithm for constructing exact optimal experimental designs. Technometrics. 1995; 37:60–69.

Parsons DW, et al. An integrated genomic analysis of human glioblastoma multiforme. Science. 2008; 321:1807–1812. [PubMed: 18772396]

Santner, TJ., Williams, BJ., Notz, WI. The Design and Analysis of Computer Experiments. New York: Springer; 2003.

Shao, J. Mathematical Statistics. 2. New York: Springer; 2003.

Shewry MC, Wynn HP. Maximum entropy sampling. Journal of Applied Statistics. 1987; 14:165–170.

Sobol' IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation. 2001; 55:271–280.

Tan, M., Fang, HB., Huang, H., Yang, Y. Design and statistical analysis of multidrug combinations in preclinical studies and clinical trials. In: Lin, J.Wang, B.Hu, X.Chen, K., Liu, R., editors. Statistical Applications from Clinical Trials and Personalized Medicine to Finance and Business Analytics. Springer; New York & Switzerland: 2016. p. 215-234.

Tan M, Fang HB, Tian GL, Houghton PJ. Experimental design and sample size determination for testing synergy in drug combination studies based on uniform measures. Statistics in Medicine. 2003; 22:2091–2100. [PubMed: 12820275]

Tan M, Fang HB, Tian GL. Dose and sample size determination for multidrug combination studies. Statistics in Biopharmaceutical Research. 2009; 1:301–316.

Wiens DP. Designs for approximately linear regression: two optimality properties of uniform designs. Statistics & Probability letters. 1991; 12:217–221.

Wu, CFJ., Hamada, M. Experiments: Planning, Analysis, and Optimization. 2. New York: Wiley; 2009.

Xavier JB, Sander C. Principle of System Balance for Drug Interactions. The New England Journal of Medicine. 2010; 362:1339–1340. [PubMed: 20375413]

Yang Y, Fang HB, Roy A, Tan M. An adaptive Bayesian dose finding approach for drug combinations with drug-drug interaction. Statistics and Its interface. 2017 (**In press**).

Yin G, Yuan Y. A latent contingency table approach to dose finding for combinations of two agents. Biometrics. 2009a; 65:866–875. [PubMed: 18759848]

Yin G, Yuan Y. Bayesian dose finding in oncology for drug combinations by copula regression. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2009b; 58:211–224.

Yuan Y, Yin G. Sequential continual reassessment method for two-dimensional dose finding. Statistics in medicine. 2008; 27:5664–5678. [PubMed: 18618901]

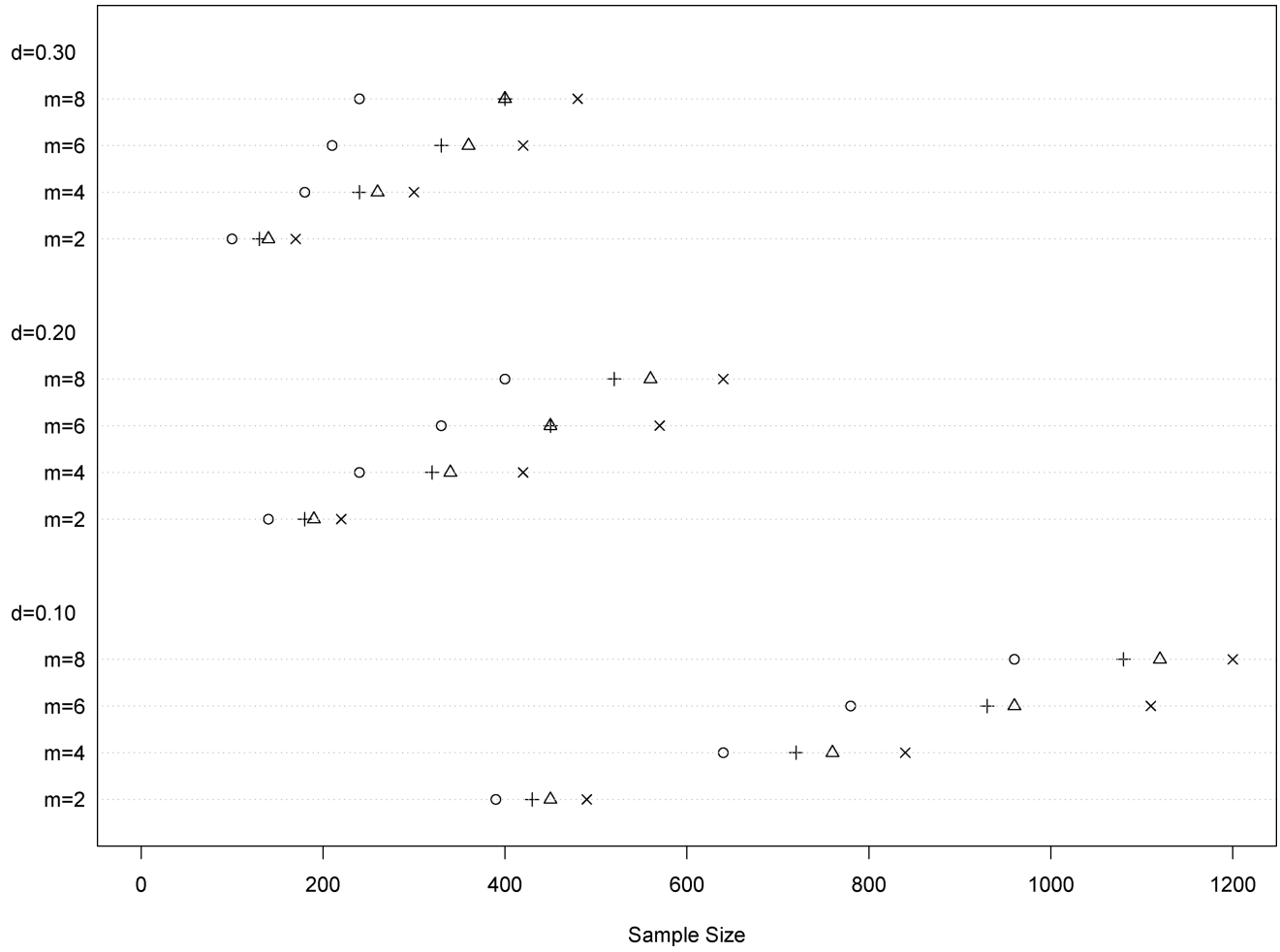

Figure 1.

Sample sizes of the compared designs: "○"–the proposed maximum entropy design; "+"–the Bayesian D-optimal design; "△"–the D-optimal design; "×"–the design under criterion (21). To facilitate the comparison, $\Delta = 5$, $n_{min} = 50$ and $n_{max} = 1500$ are used in the computational algorithm for sample size determination.

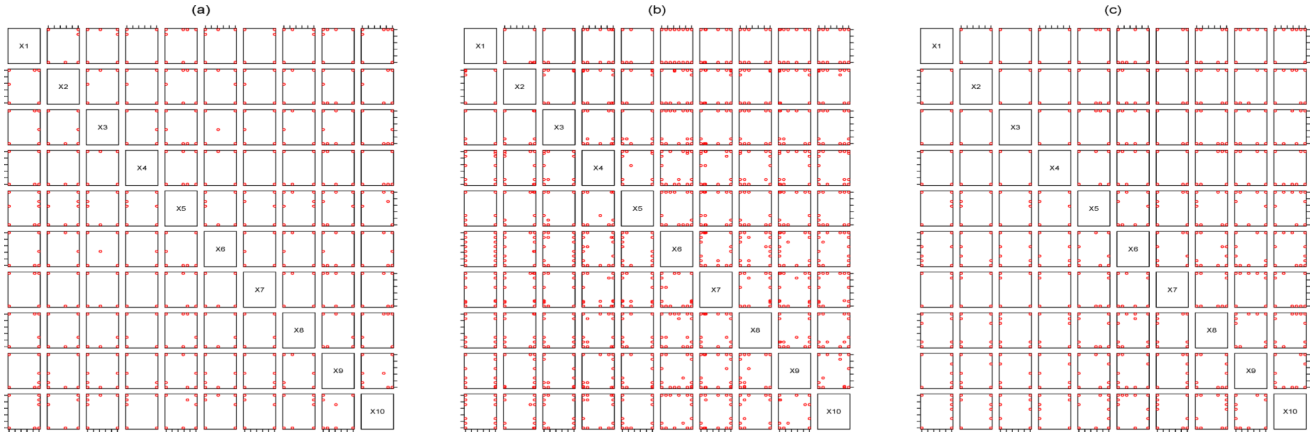

Figure 2.

(a): Bivariate projections of the design under criterion (21) with (*n, m*) = (50, 2). (b): Bivariate projections of the Bayesian D-optimal design with (*n, m*) = (50, 2). (c): Bivariate projections of the maximum entropy design with (*n, m*) = (50, 2).

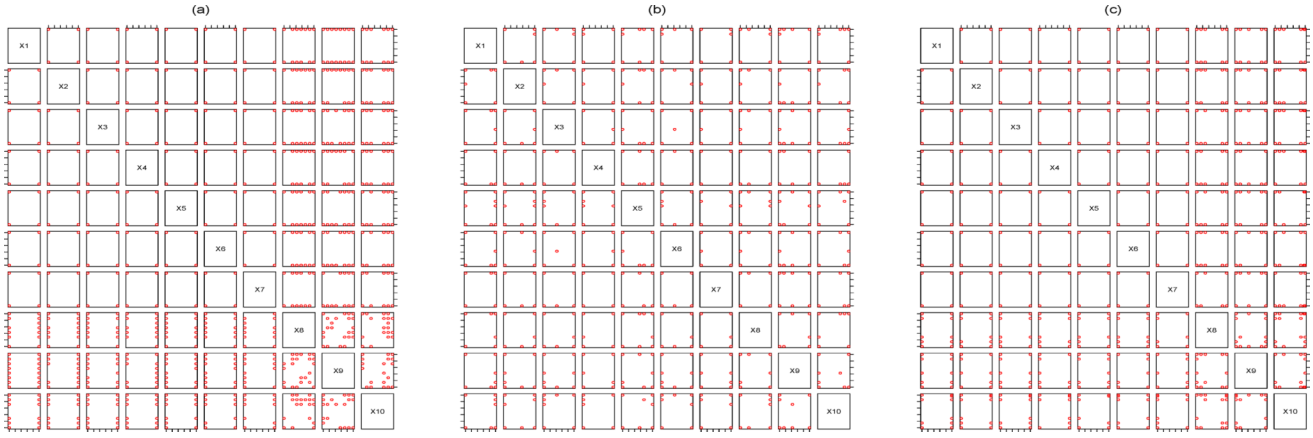

Figure 3.

(a): Bivariate projections of the D-optimal design with (*n, m*) = (50, 2). (b): Bivariate projections of the maximum entropy design with $(n, m, \sigma_{\boldsymbol{\theta}}^2/\sigma_f^2) = (50, 2, 0.01)$. (c): Bivariate projections of the maximum entropy design with $(n, m, \sigma_{\boldsymbol{\theta}}^2/\sigma_f^2) = (50, 2, 100)$.

Table 1

Algorithm 1 [Design Construction]

Initialize an $n \times s$ matrix **D** with each entry generated from (0, 1), a large positive integer T (e.g., $T = 1000$), the number of dose-levels q and $t = 1$.

Denote the entries of **D** by $d(1), \ldots, d(n \times s)$ using a row-by-row order, and for simplicity, denote $D = \{d(1), \ldots, d(n \times s)\}$.

```
while t ≤ T do
  for (i in 1 : n × s)
    for (j in 0 : 1/(q − 1) : 1)
      D* = {d(1), …, d(i − 1), j, d(i + 1), …, d(n × s)};
      if Ent(D*) > Ent(D)
        D = D*;
      end if
    end for
  end for
  t = t + 1;
end while
Output D as the optimal design.
```