

Nature-inspired Metaheuristics for finding Optimal Designs for the Continuation-Ratio Models

JIAHENG QIU AND WENG KEE WONG*

Abstract

The continuation-ratio (CR) model is frequently used in dose response studies to model a three-category outcome as the dose levels vary. Design issues for a CR model defined on an unrestricted dose interval have been discussed for estimating model parameters or a selected function of the model parameters. This paper uses metaheuristics to address design issues for a CR model defined on any compact dose interval when there are one or more objectives in the study and some are more important than others. Specifically, we use an exemplary nature-inspired metaheuristic algorithm called particle swarm optimization (PSO) to find locally optimal designs for estimating a few interesting functions of the model parameters, such as the most effective dose (*MED*), the maximum tolerated dose (*MTD*) and for estimating all parameters in a CR model. We demonstrate that PSO can efficiently find locally multiple-objective optimal designs for a CR model on various dose intervals and a small simulation study shows it tends to outperform the popular deterministic cocktail algorithm (CA) and another competitive metaheuristic algorithm called differential evolutionary (DE). We also discuss hybrid algorithms and their flexible applications to design early Phase 2 trials or tackle biomedical problems, such as different strategies for handling the recent pandemic.

KEYWORDS AND PHRASES: Dose response Models, Implicit function theorem, Maximin tolerated dose, Most effective dose, Multiple-objective optimal design.

1. INTRODUCTION: THE CONTINUATION-RATIO (CR) MODEL

In drug discovery, efficacy and toxicity of the drug are the two of the most important endpoints in early phase trials. In phase I clinical trials, a main goal is to describe the dose-limiting toxicities (*DLT*) and estimate the maximum tolerated dose (*MTD*). Phase II clinical trials focus on ascertaining the efficacy of the drug and further evaluate the safety of the drug. For example, the most effective dose (*MED*) that produces the maximum efficacy is a common quantity of interest. However, by itself it is not an adequate assessment of its usefulness without knowing its toxicity effects. For both economical and ethical and safety reasons, it is desirable to design a study that incorporates both efficacy and toxicity considerations at the onset.

To model trinomial responses above, [41] proposed the following constant slope proportional odds (PO) model for a patient’s response given dose x .

$$\begin{aligned} \log(\pi_3(x)/(1 - \pi_3(x))) &= a_1 + bx \\ \log((\pi_2(x) + \pi_3(x))/\pi_1(x)) &= a_2 + bx, \end{aligned}$$

where $\pi_1(x)$, $\pi_2(x)$ and $\pi_3(x)$ correspond to, respectively, probabilities of observing “no reaction”, “efficacy without toxicity” and “toxicity” at dose x . For any dose x ,

$\sum_{i=1}^3 \pi_i(x) = 1$. The model assumes a constant effect of dose exists across the cumulative logits. Such assumption may be violated for a trinomial model, and is unlikely to be valid when there are more than three types of responses.

The continuation-ratio (CR) model is a more flexible alternative because it does not have the restrictive assumption as the PO model does as described in Chapter 9 in [2]. In the CR model, slopes of the two equations b_1 and b_2 can be constant or not. Assuming $b_1, b_2 > 0$, the CR model is given by

$$\begin{aligned} \log(\pi_3(x)/(1 - \pi_3(x))) &= a_1 + b_1x \\ \log(\pi_2(x)/\pi_1(x)) &= a_2 + b_2x. \end{aligned} \tag{1.1}$$

It is straightforward to show for this model, probability of “no reaction” is

$$\pi_1(x) = \frac{1}{(1 + e^{a_1+b_1x})(1 + e^{a_2+b_2x})};$$

probability of “efficacy without toxicity” is

$$\pi_2(x) = \frac{e^{a_2+b_2x}}{(1 + e^{a_1+b_1x})(1 + e^{a_2+b_2x})};$$

and probability of “toxicity” is

$$\pi_3(x) = \frac{e^{a_1+b_1x}}{1 + e^{a_1+b_1x}}.$$

*Corresponding author.

[16] found optimal approximate designs for estimating all parameters in the CR model and also selected function of the model parameters. Approximate designs are large sample designs or simply probability measures on the given design space. They are characterized by the number of points the design has, where the design points are and the proportion of observations to take at each of the point. The formal is called a locally D -optimal design and is found by maximizing the logarithm of the determinant of the information matrix. In contrast, a c -optimal design minimizes the asymptotic variance of the estimated function of interest obtained from the Delta's method. In both cases, the optimization is over all designs on the given design space and the resulting design is termed locally optimal because the information matrix for a nonlinear model depends on the unknown model parameters. Hence to implement such designs, prior estimates or nominal values of the model parameters are required and they may come from a pilot study or from studies using a similar compound or drug. Clearly, poor estimates or incorrect nominal values for the model parameters can result in a different locally optimal design that is inefficient. To fix ideas, we focus on constructing locally optimal designs using metaheuristics but the technique can be easily amended to find other types of designs. The aim of this paper is to show that nature-inspired metaheuristic algorithms, which are already widely used in computer science and engineering, are both flexible and effective tools for finding multiple-objective optimal designs for clinical trials and tackling general medical problems. There are many appealing features of metaheuristics and the main ones are they virtually do not require technical assumptions for them to work well, they are generally fast, their codes are widely available and they are general-purpose optimization tools. Consequently, they are very flexible and can tackle very different types of optimization problems in increasingly many disciplines. Their meteoric rise in popularity in optimization, along with its reasons, are well documented; see for example, [49, 50].

2. PARTICLE SWARM OPTIMIZATION

Recently a class of algorithms called nature-inspired metaheuristic algorithms has proved very popular in the optimization literature. [49, 50] provided reasons for the rapid rise and interest in these algorithms. Early users of such algorithms to find exact optimal designs for linear models include [22, 4] who used an annealing algorithm to search for optimal designs for linear regression models, and [33] who used a genetic algorithm to construct exact D -optimal designs. Of particular note is the particle swarm optimization (PSO) introduced by [14] for tackling optimization problems. PSO is increasingly used across disciplines to solve hard optimization problems. PSO is essentially assumptions free and it searches in a simply and effective way. For example, unlike many algorithms, PSO does not require the

objective function to be differentiable or convex and can solve non-convex high-dimensional optimization problems.

PSO is a metaheuristic optimization algorithm inspired from the way animals, such as birds and fishes, search for food. The birds fly continuously in the sky to look for food on the ground. Each has its own perception where the food is (local optimum) but it communicates with the rest and collectively decide as a flock where the food is (global optimum). Accordingly, each bird flies toward the global optimum in the direction of the local optimum (not giving up completely where it thinks the food is). Birds are referred as particles and each bird represents a candidate solution to the optimization problem. Velocities and locations of each bird are adjusted at each iteration and if and when the flock converges, the perceived global optimum is found. In order to efficiently identify the optimal points, we initiate a flock of birds in the pre-defined search space. Let $X(k)$ be the locations of particles at the k -th iteration. Define $t_L(k-1)$ to be the locations with the best objective function values discovered by each particle before the k -th iteration, and $t_G(k-1)$ to be the locations of the best value found by the whole swarm before the k -th iteration. At the k -th iteration, the particles are updated by

$$X(k) = X(k-1) + V(k), \quad (2.1)$$

and

$$\begin{aligned} V(k) = & wV(k-1) + c_1 R_1 \otimes [t_L(k-1) - X(k-1)] \\ & + c_2 R_2 \otimes [t_G(k-1) - X(k-1)], \end{aligned} \quad (2.2)$$

where $V(k)$ is the velocity of the particle. There are several parameters in (2.2). The inertia weight represents how active the birds are and is denoted by w . This parameter may be chosen to be a positive constant but more typically its value changes over iteration and eventually decrease to 0. The parameters c_1 and c_2 are two positive constants which are recommended to be 2, and R_1 and R_2 are two random vectors whose components are independently drawn from the uniform variate on $[0, 1]$. In practice, the number of iterations and the swarm size are the most influential parameters in PSO. Large swarm size gives a better vision on the search area such that PSO could achieve the global optimum with a higher chance, and the more iterations allows the particles having search experience due to random perturbation. More details on PSO and the related metaheuristic optimization algorithms are available in [53]. Throughout, all computations were done on a PC with 3.50 GHz Intel(R) Core(TM) i7-4770K CPU.

3. DUAL-OBJECTIVE OPTIMAL DESIGNS

The current literature on optimal designs is replete with single objective optimal designs. In practice, there are frequently two or more objectives of interest in a study and they may be of unequal interest to the researcher. There is

some work on two and three-objective optimal design problems and we provide a brief review of them in this section.

When there are two objectives, the more important objective is the primary objective, and the other is the secondary objective. The sought dual-objective optimal design is one that optimizes the primary objective first, before optimizing the second objective and ensuring that the obtained design has a higher efficiency for the first objective. When there are multiple objectives, a similar idea applies; the objectives are first prioritized in terms of their importance and the sought design is the one that delivers higher efficiencies for the more important objectives. For example, in a model based dose response study, a researcher may wish to find a three-objective optimal design and in order of importance, the goals are to estimate the maximum tolerated dose, *MTD*, the most efficacious dose *MED*, and the third goal is to maximize the precision for all model parameters' estimates. The user-selected efficiencies for the sought design may be 90% and 80% for the first two objectives and subject to these constraints, the design should do as best possible under the tertiary objective. Does such a design exist and if it does, how do we find it?

For the simple case, when we have two-objective optimal designs and we have two linear and quadratic polynomials as competing regression models, [13] gave an analytical solution of such optimal designs. The dual-objective optimal design has to meet the user-specified efficiency requirement under the primary objective and is a constrained optimization problem. [13] showed that when the objectives are formulated in terms of concave functional of the information matrix, the constraint problem can be equivalently found under a simpler setup by optimizing a convex combination of the two (or more) concave function. This approach turns the dual-objective design problem into a single objective optimal design problem once the weights in the convex combination are fixed. Standard algorithms can then be applied to find the single-objective optimal design for each convex weight in the convex combination. The resulting designs are called compound optimal designs and Cook and Wong (1994) showed the class of compound optimal designs is the same as the class of constrained optimal designs for a dual-objective optimal design problem. To identify which compound optimal design coincides with the sought constrained optimal design with the user-specified efficiency under the primary criterion, they proposed plotting the efficiencies of each compound optimal design under the two criteria across all possible values of the weights from 0 to 1. The resulting plot is called an efficiency plot, which can then used to read off which convex combination weight will yield the sought constrained optimal design.

The simplicity of the above approach is valid for dual-objective optimal design problems. When there are multiple-objectives, the efficiency plot becomes high dimensional and it becomes difficult to visually identify the correspondence between which compound optimal design is the sought constrained optimal design. Unlike the two-objective optimal

design problems, there is no explicit easy analytical description of the multi-objective designs, especially when the models are nonlinear and complicated. Our experience is that standard algorithms to find them also becomes problematic.

4. MULTI-OBJECTIVE OPTIMAL DESIGNS FOR THE CR MODEL

We found a variety of optimal designs for the flexible continuation-ratio (CR) model, which has great potential for dose finding studies because it simultaneously models both probabilities of observing efficacy and adverse effect without having to assume the correlation between them.

There is work in the literature that simultaneously study efficacy and toxicity by postulating bivariate parametric models. For example, to model dose-response curves, [24] proposed a bivariate logistic model which simultaneously measures toxicity (yes/no), and efficacy (yes/no), assuming these two outcomes are correlated. A similar idea was presented by [18] who proposed a dichotomized outcome pair $(y_1, y_2)^T$ from a bivariate normal model, where y_1 is the efficacy response and y_2 is the toxicity response. It follows that, there are four possible outcomes and the utility function of interest is $\pi(x) = Pr(y_1 = 1, y_2 = 0|x)$. This model is complicated and difficult to use in practice because it requires user to specify the correlation structure between toxicity and efficacy. Alternatively, we may combine the two outcomes corresponding to the presence of toxicity together and reduce the four types of responses into three. Doing so greatly simplifies the model by eliminating the correlation between toxicity and efficacy. In the resulting trinomial models, the responses of patients are now exclusively and exhaustively classified into three categories: “no effect” (neither toxicity or efficacy found); “efficacy” with no toxicity; and “adverse reaction” for toxicity regardless of whether efficacy is presented or not.

The optimal designs that we are interested to construct is a three-objective compound optimal design that provides efficient estimates for the most effective dose (*MED*), the maximum tolerated dose (*MTD*) for a user specified toxicity rate, and for all parameters in the CR model. We show that PSO can successfully find multi-objective compound optimal designs for the CR model. Further, we investigate the proper choice of weights for three optimal criteria in multi-objective designs under different parameters settings. By using efficiency plots, researchers and practitioners can construct the desired compound optimal design through appropriate weights combination of three optimal criteria in a more flexible and informative way.

[16] focused on finding the dose x that maximizes probability of “efficacy without toxicity” $\pi_2(x)$, which is selected as the most effective dose (*MED*). If $b_1 = b_2$ then *MED* has a closed form given by

$$MED = -(a_1 + a_2)/2b_1,$$

but if $b_1 \neq b_2$, there is no closed form solution of *MED*. In their paper, we produce several locally c -optimal designs for estimating the *MED* using different nominal parameter settings on a user-specified dose interval. The latter is likely more practically relevant than having an unrestricted dose interval. This feature is helpful because some algorithms work only on certain types of intervals only, especially when there are many interacting factors involved and the dose interval for each factor varies considerably citepxu.

In dose-response studies, side effects and drug toxicity can be very serious and need to be well controlled. One of the main targets in clinical trials is to find out the “dose that is closest to an acceptable level of toxicity”. The target dose is defined as Maximum Tolerated Dose (*MTD*), which is the highest dose of a treatment agent that produces the dose limiting toxicity (*DLT*) in a proportion ρ of patients. In the CR model, it follows that

$$\pi_3(\text{MTD}, a_1, b_1) = \rho.$$

Here the probability ρ is user-specified and varies depending on what types of adverse effect is of interest. For life threatening side effects, ρ should be a small value. To find the *MTD*, practitioners usually start from a safe dose that produces a small ρ and gradually increase the dose to its highest acceptable level. The dose level at ρ is called lethal dose $LD100\rho$ ([57]). For example, a common choice is the median lethal dose denoted by $LD50$. For both economical and ethical reason, it is necessary to take both *MED* and *MTD* (or $LD100\rho$) into consideration at the design stage.

5. INFORMATION MATRIX OF THE CR MODEL WITH $b_1 \neq b_2$

The worth of a design is measured of the quality of the Fisher’s information matrix. This matrix is proportional to the negative of the expectation of the second derivatives of the total log likelihood function with respect to the model parameters. This section derives the matrix for the CR model and makes clear the explicit nature of the objective functions.

The response from a subject assigned to dose x has one of three possible outcomes: $y = (y_1, y_2, y_3)^T$ with $\sum_{i=1}^3 y_i = 1$. Its expected value is $E(y) = \pi(x)^T = (\pi_1(x), \pi_2(x), \pi_3(x))^T$ as previously defined. Denote the parameters in the CR model by $\theta = (a_1, b_1, a_2, b_2)^T$ and let δ_x be the design that puts all its weight at dose x . To find the information matrix $I(\delta_x, \theta)$ of δ_x for the CR model, we use the method developed by [58] for multi-nomial logistic models. Denote the left hand side of (1.1) by

$$\eta(x) = \begin{bmatrix} \log(\pi_3(x)/(\pi_1(x) + \pi_2(x))) \\ \log(\pi_2(x)/\pi_1(x)) \\ \log(\pi_1(x) + \pi_2(x) + \pi_3(x)) \end{bmatrix}$$

and its right hand side by

$$X\theta = \begin{bmatrix} 1 & x & 0 & 0 \\ 0 & 0 & 1 & x \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix}.$$

By the numerator-layout notation, we have

$$\begin{aligned} \frac{\partial \eta(x)}{\partial \pi(x)} &= \begin{bmatrix} \frac{\partial \eta(x)}{\partial \pi_1(x)} & \frac{\partial \eta(x)}{\partial \pi_2(x)} & \frac{\partial \eta(x)}{\partial \pi_3(x)} \end{bmatrix} \\ &= \begin{bmatrix} -(\pi_1(x) + \pi_2(x))^{-1} & -(\pi_1(x) + \pi_2(x))^{-1} & \pi_3(x)^{-1} \\ -\pi_1(x)^{-1} & \pi_2(x)^{-1} & 0 \\ 1 & 1 & 1 \end{bmatrix}. \end{aligned}$$

Therefore the derivative of $\pi(x)$ with respect to θ is derived by the chain rule

$$\begin{aligned} G(x) &= \frac{\partial \pi(x)}{\partial \theta} = \frac{\partial \pi(x)}{\partial \eta(x)} \frac{\partial \eta(x)}{\partial \theta} = \left(\frac{\partial \eta(x)}{\partial \pi(x)} \right)^{-1} X \\ &= \begin{bmatrix} -\pi_1(x)\pi_3(x) & -\frac{\pi_1(x)\pi_2(x)}{\pi_1(x)+\pi_2(x)} & \pi_1(x) \\ -\pi_2(x)\pi_3(x) & \frac{\pi_1(x)\pi_2(x)}{\pi_1(x)+\pi_2(x)} & \pi_2(x) \\ (\pi_1(x) + \pi_2(x))\pi_3(x) & 0 & \pi_3(x) \end{bmatrix} \\ &\quad \begin{bmatrix} 1 & x & 0 & 0 \\ 0 & 0 & 1 & x \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Since $y^T \sim \text{Multinomial}(1, \pi(x)^T)$, the likelihood function at dose x is

$$L(\theta; x) = \prod_{j=1}^3 \pi_j^{y_j}(x),$$

and so the log-likelihood is $l(\theta; x) = y^T \log \pi(x)$. By the chain rule, the score vector at dose x is

$$\begin{aligned} \frac{\partial l(\theta; x)}{\partial \theta} &= \frac{\partial l(\theta; x)}{\partial \log \pi(x)} \frac{\partial \log \pi(x)}{\partial \pi(x)} \frac{\partial \pi(x)}{\partial \theta} \\ &= G(x)^T \begin{bmatrix} \pi_1^{-1}(x) & 0 & 0 \\ 0 & \pi_2^{-1}(x) & 0 \\ 0 & 0 & \pi_3^{-1}(x) \end{bmatrix} y, \end{aligned}$$

and the unit information matrix $I(\delta_x, \theta)$ for the observation y at dose x is given by

$$\begin{aligned} I(\delta_x, \theta) &= E \left(\frac{\partial l(\theta; x)}{\partial \theta} \right) \left(\frac{\partial l(\theta; x)}{\partial \theta^T} \right) \\ &= G^T(x) \begin{bmatrix} \pi_1^{-1}(x) & 0 & 0 \\ 0 & \pi_2^{-1}(x) & 0 \\ 0 & 0 & \pi_3^{-1}(x) \end{bmatrix} (Eyy^T) \\ &\quad \begin{bmatrix} \pi_1^{-1}(x) & 0 & 0 \\ 0 & \pi_2^{-1}(x) & 0 \\ 0 & 0 & \pi_3^{-1}(x) \end{bmatrix} G(x). \end{aligned}$$

Since $y^T \sim \text{Multinomial}(1, \pi(x)^T)$, we have

$$Eyy^T = EyEy^T + \text{Var}(y) = \begin{bmatrix} \pi_1(x) & 0 & 0 \\ 0 & \pi_2(x) & 0 \\ 0 & 0 & \pi_3(x) \end{bmatrix},$$

and so,

$$\begin{aligned} I(\delta_x, \theta) &= G^T(x) \begin{bmatrix} \pi_1^{-1}(x) & 0 & 0 \\ 0 & \pi_2^{-1}(x) & 0 \\ 0 & 0 & \pi_3^{-1}(x) \end{bmatrix} G(x) \\ &= \begin{bmatrix} 1 & x & 0 & 0 \\ 0 & 0 & 1 & x \\ 0 & 0 & 0 & 0 \end{bmatrix}^T \begin{bmatrix} (\pi_1(x) + \pi_2(x))\pi_3(x) & 0 & 0 \\ 0 & \frac{\pi_1(x)\pi_2(x)}{\pi_1(x) + \pi_2(x)} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= (\pi_1(x) + \pi_2(x))\pi_3(x) \begin{bmatrix} 1 & x & 0 & 0 \\ x & x^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ &\quad + \frac{\pi_1(x)\pi_2(x)}{\pi_1(x) + \pi_2(x)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x \\ 0 & 0 & x & x^2 \end{bmatrix}. \end{aligned}$$

Locally c -optimal designs for estimating the MED and MTD of the CR model with $b_1 \neq b_2$ can be similarly derived. Recalling that MED is the dose which maximizes π_2 the probability of having efficacy without toxicity. We derive MED by solving $\frac{d\pi_2(x)}{dx} = 0$. After a few steps of simplification, we get MED as the implicit solution of the following equation

$$g(x, \theta) = b_2(1 + \exp(-a_1 - b_1x)) - b_1(1 + \exp(a_2 + b_2x)) = 0.$$

To find the locally c -optimal design for estimating the MED , we need to get its gradient ∇_{MED} first. We follow [6]'s way to handle such equation $g(x, \theta)$ without an explicit solution for MED as follows.

By the implicit function theorem, if the function $g(x, \theta)$ has continuous first derivatives, and $MED(\theta)$ is continuous, then

$$\frac{\partial g(x, \theta)}{\partial \theta} \Big|_{MED} = \frac{\partial g(x, \theta)}{\partial x} \Big|_{MED} \frac{\partial x}{\partial \theta} \Big|_{MED}.$$

It follows that

$$\begin{aligned} \nabla_{MED}^T &= \frac{\partial x}{\partial \theta} \Big|_{MED} \\ &= \left[\frac{\partial g(x, \theta)}{\partial x} \Big|_{MED} \right]^{-1} \frac{\partial g(x, \theta)}{\partial \theta} \Big|_{MED} \end{aligned}$$

where

$$\frac{\partial g(x, \theta)}{\partial x} \Big|_{MED} = -b_1b_2(\exp(-a_1 - b_1MED) + \exp(a_2 + b_2MED)),$$

and

$$\begin{aligned} \frac{\partial g(x, \theta)}{\partial \theta} \Big|_{MED} &= \\ &= \begin{bmatrix} -b_2 \exp(-a_1 - b_1MED) \\ -b_2MED \exp(-a_1 - b_1MED) - (1 + \exp(a_2 + b_2MED)) \\ -b_1 \exp(a_2 + b_2MED) \\ 1 + \exp(-a_1 - b_1MED) - b_1MED \exp(a_2 + b_2MED) \end{bmatrix}^T. \end{aligned}$$

Thus, the locally c -optimal design for estimating the MED maximizes

$$-\nabla_{MED}^T I^{-1}(\xi, \theta) \nabla_{MED}.$$

In the CR model, since

$$\log(\pi_3(x)/(1 - \pi_3(x))) = a_1 + b_1x,$$

a simple calculation shows $MTD = \frac{1}{b_1}(\logit(\rho) - a_1)$. Similarly, the locally c -optimal design for finding MTD is to maximize

$$-\nabla_{MTD}^T I^{-1}(\xi, \theta) \nabla_{MTD},$$

where $\nabla_{MTD}^T = \frac{\partial MTD}{\partial \theta} = (-\frac{1}{b_1}, \frac{a_1 - \logit(\rho)}{b_1^2}, 0, 0)$. In the traditional dose finding designs, MTD is usually estimated using a 3+3 design or set the MTD level as the dose level at which two of six patients experienced toxicity, see [20] for example. Therefore in what is to follow, we set $\rho = 0.3$, but other values can be directly used as well.

6. EQUIVALENCE THEOREMS FOR LOCALLY SINGLE-OBJECTIVE OPTIMAL DESIGNS FOR THE CR MODEL

In what is to follow, we treat a design ξ as an approximate design, i.e. a probability measure on a given dose space. If N is the total number of subjects in the study and ξ has support at x_1, \dots, x_k with corresponding weights w_1, \dots, w_k and $w_1 + \dots + w_k = 1$, then the implemented design is to take Nw_i at $x_i, i = 1, \dots, k$ after rounding each Nw_i to a positive integer $[Nw_i]$ and subject to $[Nw_1] + \dots + [Nw_k] = N$. In particular, this means that for a K -point approximate design with weights w_i at dose $x_i, i = 1, 2, \dots, K$, the normalized information matrix for the design ξ is

$$I(\xi, \theta) = \sum_{i=1}^K w_i I(\delta_{x_i}, \theta)$$

[26] proposed working with approximate designs after noting that if the criteria are all formulated as convex or concave functional of the design information matrix, they are easier to construct, study and check for their optimality. The latter is accomplished via an equivalence theorem for each convex or concave functional. They all have the form but are different for different criteria. Recalling the directional derivative of $\Psi(I(\xi, \theta)) = \log|I(\xi, \theta)|$ at ξ in the direction of δ_x is

$$\psi(x, \xi, \theta) = \text{trace}[I(\delta_x, \theta)I^{-1}(\xi, \theta)] - P$$

and P is the number of parameters in the mean function. For example, if $b_1 \neq b_2$ in the CR model, $P = 4$. Applying standard convex analysis results for the D -optimality criterion, we conclude the following statements are equivalent ([26]):

1. The design ξ^* maximizes $\Psi(I(\xi, \theta))$;
2. The design ξ^* minimizes $\max_{x \in \mathcal{X}} \psi(x, \xi, \theta)$;
3. $\max_{x \in \mathcal{X}} \psi(x, \xi^*, \theta) = 0$, and it achieves its maximum at the support points of the design.

This is the equivalence theorem for D -optimality that allows us to use (3) to check optimality of any design ξ in practice. Each convex or concave functional has a unique equivalence theorem but they all have the same form.

In the same paper, [26] also showed the equivalence theorem to verify the optimality of a locally c -optimal designs for estimating MED or MTD . Both can be subsumed under the criterion $\Psi(I(\xi, \theta)) = \text{trace}[AI^{-1}(\xi, \theta)]$ when A is chosen appropriately. For estimating the former, the directional derivative of $\Psi(I(\xi, \theta)) = -\nabla_{MED}^T I^{-1}(\xi, \theta) \nabla_{MED} = -\text{trace}[\nabla_{MED} \nabla_{MED}^T I^{-1}(\xi, \theta)]$ at ξ in the direction of δ_x is

$$\begin{aligned} \psi(x, \xi, \theta) &= \text{trace}[I(\delta_x, \theta) I^{-1}(\xi, \theta) (\nabla_{MED} \nabla_{MED}^T) I^{-1}(\xi, \theta)] \\ &\quad - \text{trace}[(\nabla_{MED} \nabla_{MED}^T) I^{-1}(\xi, \theta)]. \end{aligned}$$

When ξ is optimal, $\psi(x, \xi, \theta)$ is bounded above by 0 with equality at the support points. Similarly we can derive the equivalence theorem for locally c -optimal design to estimate other quantities such as MTD . Because equivalence theorems confirm the optimality or non-optimality of the design, these theorems are generally referred to also as checking conditions and the function $\psi(x, \xi, \theta)$ is referred as the sensitivity function of the design ξ .

6.1 Equivalence of Compound and Constrained Optimal Designs

In practice, there are usually not one, but a couple of objectives of interest and it is desirable to incorporate them at the design stage. There are two typical approaches to construct a multiple-objective design: the compound and constrained optimal design ([13]). Because different objectives $\Psi_i(I(\xi, \theta))$ ($i = 1, \dots, m$) may have different magnitude in their values, we define each $\Psi_i(I(\xi, \theta))$ in terms of the design efficiency $e_i(\xi)$ as did in [13]. For example, the D-efficiency for an arbitrary design ξ is defined by $e_1(\xi) = \left(\frac{|I(\xi, \theta)|}{|I(\xi_D, \theta)|}\right)^{1/P}$, and its c -efficiency is $e_2(\xi) = \frac{\nabla_c^T I^{-1}(\xi_c, \theta) \nabla_c}{\nabla_c^T I^{-1}(\xi, \theta) \nabla_c}$, where ξ_D is the locally D-optimal design and ξ_c is the locally c -optimal design, ∇_c is the gradient of $c(\theta)$ with respect to θ . To find a multiple-objective compound optimal designs, we consider a concave functional of these efficiencies and maximize each of them.

Suppose there are two competing objectives Ψ_1 and Ψ_2 of interest. Denote Δ_s the set of designs ξ that maximize $\Psi_2(I(\xi, \theta))$ subject to the constraint that $\Psi_1(I(\xi, \theta)) \geq s$, where s is a user-selected constant. The constrained optimal design is defined by

$$\xi_s = \arg \max_{\xi \in \Delta_s} \Psi_1(I(\xi, \theta)),$$

where the maximum is taken over the entire Δ_s .

For any user-selected constant $\lambda \in [0, 1]$, a two-objective compound optimal design finds a design ξ_λ that maximizes the weighted average of two concave functionals

$$\Psi(\xi|\lambda) = \lambda \Psi_1(I(\xi, \theta)) + (1 - \lambda) \Psi_2(I(\xi, \theta)).$$

The constrained optimal design is harder to find than the compound optimal design. The latter is easier to determine because a concave combination of concave functionals is still concave for a fixed λ , and the equivalence theorem is just a weighted means of the checking conditions for the optimal design for each objective. However, the weighting parameter λ may be hard to interpret in practice. For example, if we set $\lambda = 0.5$, does it mean we put are equally interested in the two objectives $\Psi_1(\xi)$ and $\Psi_2(\xi)$?

For linear models, [13] pointed out that for a two-objective compound optimal design ξ_λ that maximizes $\lambda \Psi_1(I(\xi, \theta)) + (1 -$

$\lambda) \Psi_2(I(\xi, \theta))$, it is equivalent that ξ_λ maximizes $\Psi_2(I(\xi, \theta))$ subject to $\Psi_1(I(\xi, \theta)) \geq \Psi_1(I(\xi_\lambda, \theta))$ which is the primary objective. To find a multiple-objective optimal design, we first formulate the optimal design problem as a constrained design ξ_s , and then determine the value of λ such that the compound optimal design ξ_λ is equivalent to the constrained optimal design ξ_s .

Extensions to constructing optimal designs for nonlinear models under 3 or more objective criteria using the Lagrange's multiplier method were considered in [11]. For instance, if we have a 3 design criteria, we set $\lambda = (\lambda_1, \lambda_2, \lambda_3)^T$, each λ_i is in $[0, 1]$ and $\sum_{i=1}^3 \lambda_i = 1$. The 3 objective compound optimal design ξ_λ is then found by maximizing

$$\Psi(\xi|\lambda) = \lambda_1 \Psi_1(I(\xi, \theta)) + \lambda_2 \Psi_2(I(\xi, \theta)) + \lambda_3 \Psi_3(I(\xi, \theta)).$$

This problem is equivalent that ξ_λ maximizes $\Psi_3(I(\xi, \theta))$ subject to $\Psi_1(I(\xi, \theta)) \geq \Psi_1(I(\xi_\lambda, \theta))$ and $\Psi_2(I(\xi, \theta)) \geq \Psi_2(I(\xi_\lambda, \theta))$.

7. THREE-OBJECTIVE LOCALLY OPTIMAL DESIGN FOR THE CR MODEL VIA PSO

In [16], only MED was of interest and the authors focused on finding locally c -optimal design for estimating the MED . This study extends their work to construct a design that efficiently estimates MED and MTD simultaneously. Other parameters such as b_1 and b_2 determine the slopes of three possible curves and so it is important to make sure the generated optimal design provides good estimates for all parameters in the CR model. Accordingly, in addition to estimating MED and MTD , we incorporate D-optimality as well. For a given λ , we construct a three-objective compound optimal design by maximizing the weighted sum of three concave functionals

$$\Psi(\xi|\lambda) = \lambda_1 \Psi_1(I(\xi, \theta)) + \lambda_2 \Psi_2(I(\xi, \theta)) + \lambda_3 \Psi_3(I(\xi, \theta))$$

where $\lambda = (\lambda_1, \lambda_2, \lambda_3)^T$, λ_i 's $\in [0, 1]$ and $\sum_{i=1}^3 \lambda_i = 1$. Here $\Psi_1(I(\xi, \theta))$ is the log c -efficiency of the design ξ for estimating the MTD

$$\begin{aligned} \Psi_1(I(\xi, \theta)) &= \log\left(\frac{\nabla_{MTD}^T I^{-1}(\xi_{MTD}, \theta) \nabla_{MTD}}{\nabla_{MTD}^T I^{-1}(\xi, \theta) \nabla_{MTD}}\right) \\ &= \log(e_1), \end{aligned}$$

$\Psi_2(I(\xi, \theta))$ is log c -efficiency of the design ξ for estimating the MED

$$\begin{aligned} \Psi_2(I(\xi, \theta)) &= \log\left(\frac{\nabla_{MED}^T I^{-1}(\xi_{MED}, \theta) \nabla_{MED}}{\nabla_{MED}^T I^{-1}(\xi, \theta) \nabla_{MED}}\right) \\ &= \log(e_2), \end{aligned}$$

and $\Psi_3(I(\xi, \theta))$ is log D-efficiency of the design ξ for estimating all parameters in the model

$$\begin{aligned} \Psi_3(I(\xi, \theta)) &= \log\left(\frac{|I(\xi, \theta)|}{|I(\xi_D, \theta)|}\right)^{\frac{1}{P}} \\ &= \log(e_3). \end{aligned}$$

The checking condition for a three-objective compound optimal design is just a weighted sum of the checking conditions for the optimal design for each objective, i.e. we have ξ is optimal for the three-objective problem if and only if for all $x \in \chi$

$$\lambda_1 \frac{\text{trace}[I(\delta_x, \theta) I^{-1}(\xi, \theta) (\nabla_{MTD} \nabla_{MTD}^T) I^{-1}(\xi, \theta)]}{\text{trace}[(\nabla_{MTD} \nabla_{MTD}^T) I^{-1}(\xi, \theta)]}$$

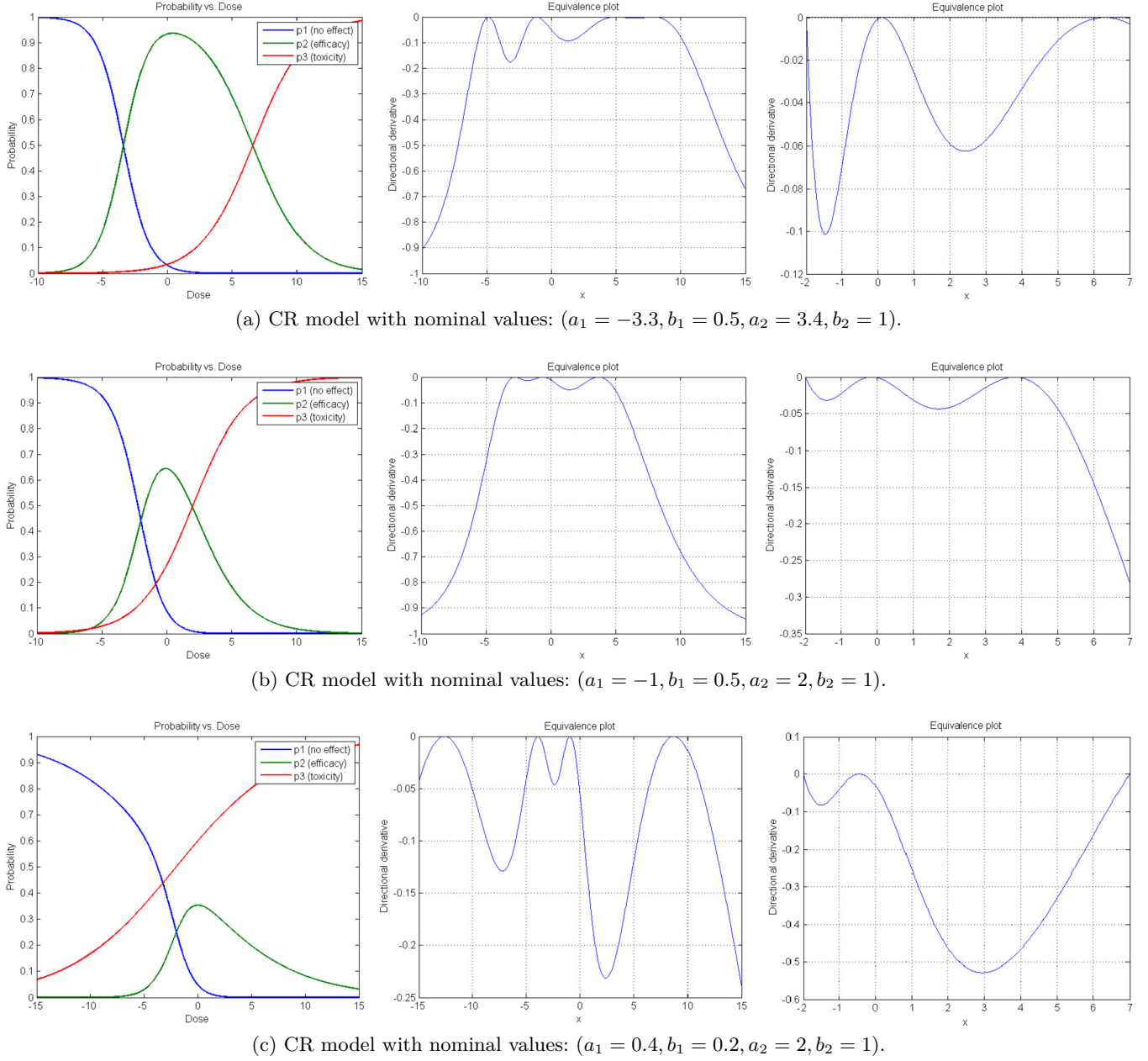


Figure 1: Mean Responses from the CR model with corresponding nominal values; directional derivatives of the compound criterion evaluated at the PSO-generated multiple-objective design on the unrestricted design space (middle); and directional derivatives of the compound criterion evaluated at the PSO-generated multiple-objective design on the design space $[-2,7]$ (right).

$$\begin{aligned}
 & + \lambda_2 \frac{\text{trace}[I(\delta_x, \theta)I^{-1}(\xi, \theta)(\nabla_{MED} \nabla_{MED}^T)I^{-1}(\xi, \theta)]}{\text{trace}[(\nabla_{MED} \nabla_{MED}^T)I^{-1}(\xi, \theta)]} \\
 & + \lambda_3 \frac{\text{trace}[I(\delta_x, \theta)I^{-1}(\xi, \theta)]}{P} \leq 1,
 \end{aligned}$$

with equality at the support points.

We apply standard PSO to find the locally c -optimal design for estimating the LD_{30} in the CR model with nominal values $a_1 = -3.3$ and $b_1 = 0.5$. Here and elsewhere, we use 100 particles and 1000 iterations to find the locally optimal design. The PSO-

generated design is supported at a single dose at 4.905, which can be directly verified by checking $LD_{30} = \frac{1}{0.5}(\text{logit}(0.3) + 3.3) = 4.9054$. Similarly, we use PSO to find different types of locally compound optimal designs using various nominal values for the CR model and note that all generated designs have four or fewer support points. The three sets of nominal values we have are selected to generate mean response curves that we expect, i.e. the maximum efficacy varies from 0.3 to 0.9, see Figures 1(a), (b) and (c). In each figure, probability curves are shown in the left plot, and the

Table 1. Three-objective compound optimal designs for estimating the *MTD*, *MED* and all parameters with $\rho = 0.3$, $\lambda_1 =$ and $\lambda_2 = 1/3$ on the unrestricted designs interval.

(a_1, b_1, a_2, b_2)	$x_1(w_1)$	$x_2(w_2)$	$x_3(w_3)$	$x_4(w_4)$
(-3.3, 0.5, 3.4, 1)	-4.875 (0.102)	-1.139 (0.464)	5.016 (0.322)	7.874 (0.112)
(-1, 0.5, 2, 1)	-2.790 (0.202)	-0.637 (0.513)	3.683 (0.284)	-
(0.4, 0.2, 2, 1)	-12.610 (0.366)	-3.918 (0.158)	-0.942 (0.470)	8.727 (0.006)

Table 2. Three-objective compound optimal designs for estimating the *MTD*, *MED* and all parameters with $\rho = 0.3$, $\lambda_1 = \lambda_2 = 1/3$ on $\chi = [-2, 7]$.

(a_1, b_1, a_2, b_2)	$x_1(w_1)$	$x_2(w_2)$	$x_3(w_3)$
(-3.3, 0.5, 3.4, 1)	-2.000 (0.152)	0.1045 (0.502)	6.328 (0.345)
(-1, 0.5, 2, 1)	-2.000 (0.330)	-0.156 (0.403)	3.820 (0.267)
(0.4, 0.2, 2, 1)	-2.000 (0.356)	-0.438 (0.319)	7.000 (0.325)

middle one is the equivalence plot for compound optimal designs in the unrestricted dose space. Table 1 shows selected compound optimal designs with equal weights λ_i 's=1/3 for all 3 criteria on the unrestricted designs space. We observe that generated optimal designs are supported at three or four points, and the weights at these support points are unequal.

Invariance of the optimal designs to linear transformation of the design is a desirable property for an optimal design. *D*-optimal designs for linear models are examples of optimal designs with such a property, but *A*-optimal designs, that minimize the average variance of all the estimated parameters, do not have such a property, i.e. an *A*-optimal design on a new interval cannot be deduced from the *A*-optimal design on the origin interval. Interestingly, some algorithm may work on a design space but not on another design space, especially when the latter has are many interacting factors and each factor is defined on intervals with very different widths; see [52] for details and examples when they found locally *D*-optimal designs for a logistic model with many factors and all pairwise interactions are included.

To further investigate the influence of dose space on the compound optimal design, we verified the algorithm was still able to find compound optimal designs on an unrestricted space that may not be symmetric about 0 for the 3 sets of nominal values for the model parameters. For example, Table 2 displays the PSO-generated compound optimal designs on the asymmetric dose interval $[-2, 7]$ for all the three cases. Interestingly, the previous four point optimal designs become three points, with one or two points located on the boundaries of the restricted design space. We also observe that the middle support point is not the same as the ones in optimal designs on unrestricted dose space. The plots on the right hand side of Figures 1(a), (b) and (c) display the corresponding sensitivity functions and visually confirm that all the PSO-generated three-objective designs are optimal among all designs on $[-2, 7]$.

In our PSO implementation, for each fixed λ in $[0, 1]$, we use the default parameter settings in the standard PSO to find the three-objective compound optimal design ξ_λ for the CR model. Particles that wander outside of the dose space are pulled back

to its nearest boundary $[lb, ub]$:

$$z_i^{(t)} = \begin{cases} ub & \text{if } z_i^{(t)} > ub \\ lb & \text{if } z_i^{(t)} < lb \\ z_i^{(t)} & \text{otherwise} \end{cases}$$

where $z_i^{(t)}$ is the particle i 's position at time t . The rationale for our strategy is that optimal designs frequently have support points at the boundary of the dose space, see [29] for example. Results from a small numerical experiment not included here suggest that generally, PSO with boundary repair outperforms PSO with random repair in searching for locally *D*- and *c*-optimal designs. Using equivalence theorems, we were also able to verify that PSO-generated locally designs under the *D*-optimal and *c*-optimality criteria for the CR model were optimal for the different sets of nominal values in [16].

8. DIFFERENT EFFICIENCIES OF THE COMPOUND OPTIMAL DESIGN ξ_λ

Can we find a three-objective optimal design ξ for the CR model such that its *D*-efficiency $e_3(\xi)$ is maximized subject to both its *c*-efficiencies $e_1(\xi)$, $e_2(\xi)$ for *MTD* and *MED* are equal to or greater than 0.9? Such a design may or may not exist depending on how competitive the criteria are and also the efficiency requirements in the constraints. If such a design exists, what values for the λ_i 's should be chosen in the compound optimal design?

To answer the question we need to investigate the impact of different combinations of λ_1 and λ_2 on the efficiencies of the compound optimal design ξ_λ relative to the three criteria. We first discretize λ_1 and λ_2 using a grid size of 0.05, and find the corresponding compound optimal design ξ_λ for each pair of λ_1 and λ_2 combination subject to $\lambda_1 + \lambda_2 \leq 1$. The efficiencies $e_i(\xi_\lambda)$ under the i th criterion are displayed in Figure 2, 3 and 4 for different nominal values sets. In these figures, the upper left plot shows the *c*-efficiencies of $e_1(\xi_\lambda)$ as measured by the locally *c*-optimal design ξ_{MTD} for *MTD*; the upper right plot shows *c*-efficiencies of $e_2(\xi_\lambda)$ relative to the locally *c*-optimal design ξ_{MED} for *MED*; the bottom left plot shows *D*-efficiencies of $e_3(\xi_\lambda)$; bottom right shows minimal efficiencies of all three efficiencies.

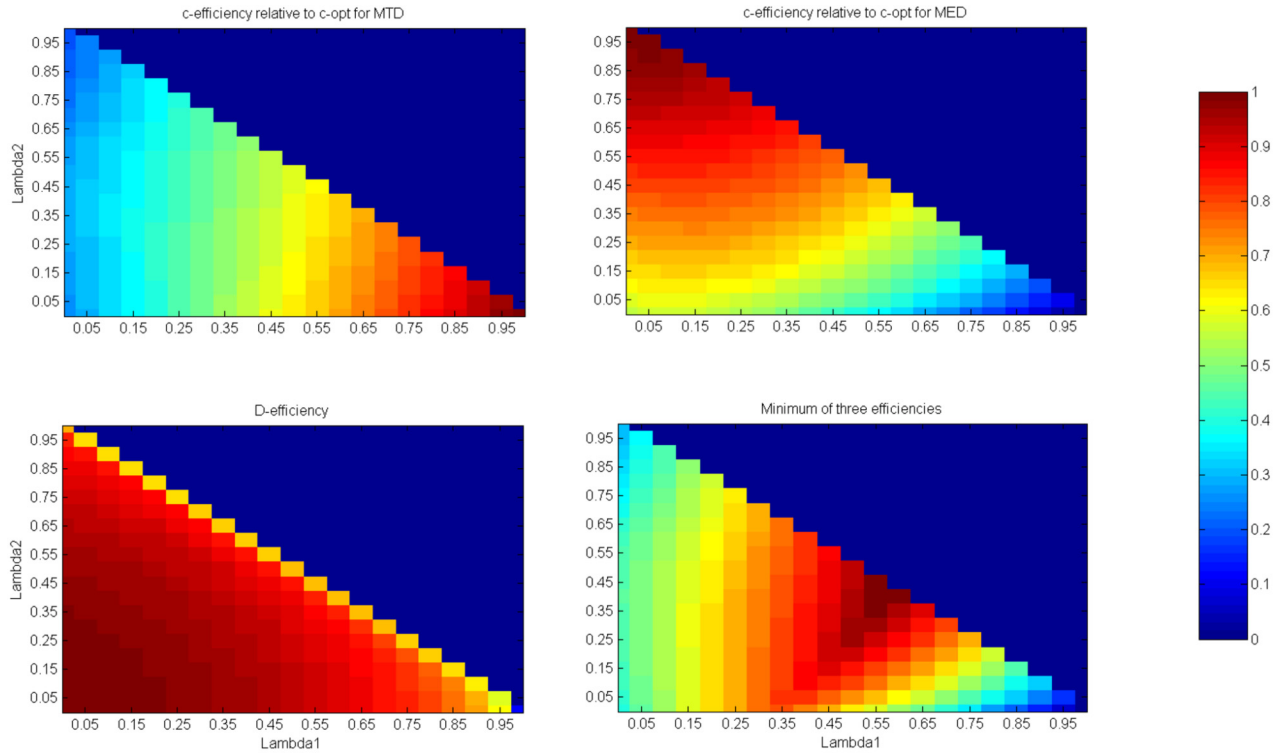


Figure 2: Different efficiency plots of compound optimal designs for the CR model with nominal values $a_1 = -3.3$, $b_1 = 0.5$, $a_2 = 3.4$ and $b_2 = 1$.

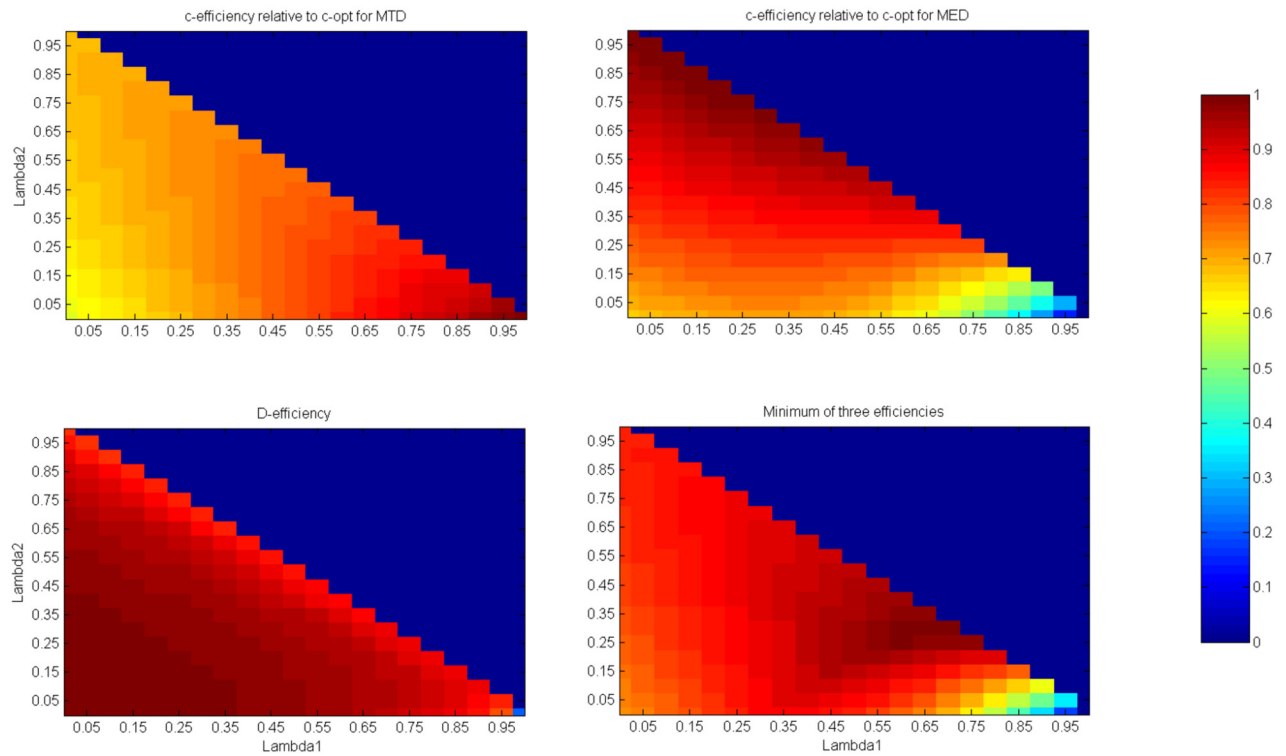


Figure 3: Different efficiency plots of compound optimal designs for the CR model with nominal values: $a_1 = -1$, $b_1 = 0.5$, $a_2 = 2$ and $b_2 = 1$.

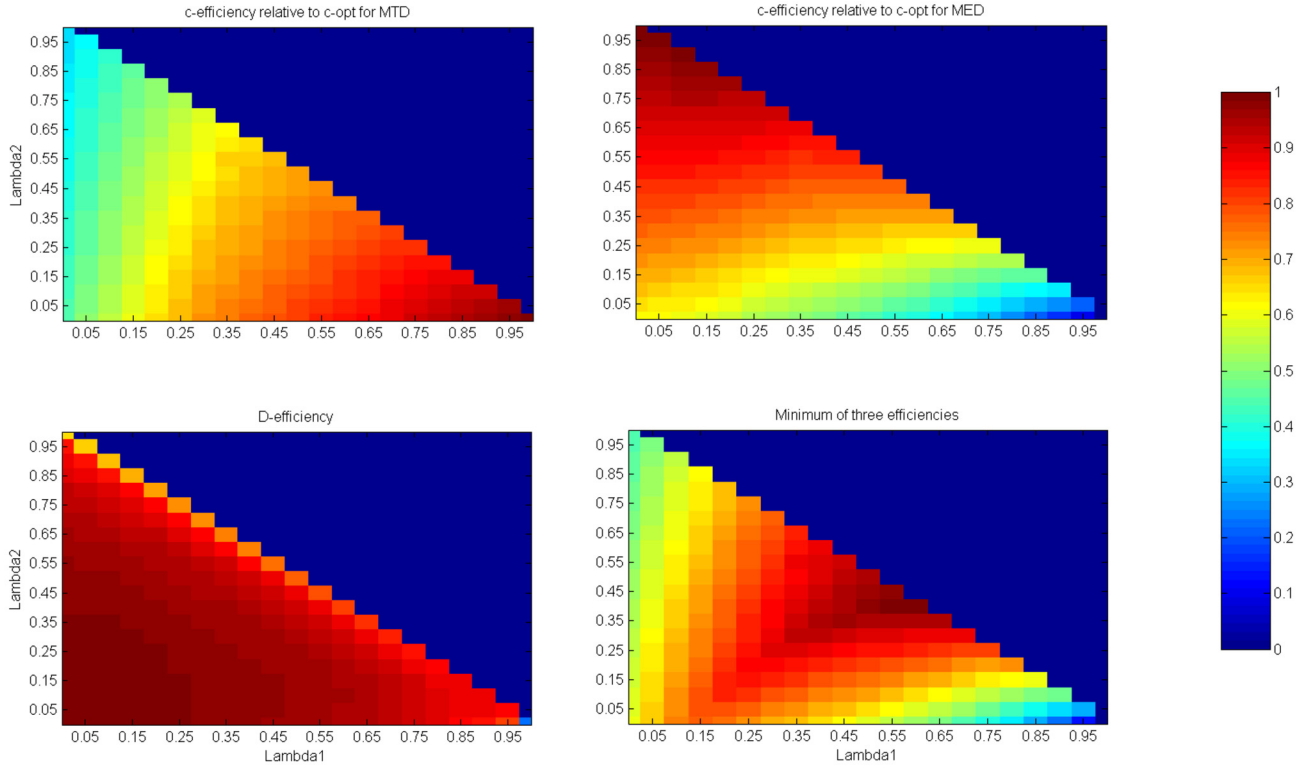


Figure 4: Different efficiency plots of compound optimal designs for the CR model with nominal values: $a_1 = 0.4$, $b_1 = 0.2$, $a_2 = 2$ and $b_2 = 1$).

These efficiency plots provide an answer to the question posed at the beginning of the section. The answer is no, at least for the three sets of nominal values we show here; there is no compound optimal design ξ_λ that maximizes D -optimality with subject to $e_1(\xi_\lambda) \geq 0.9$ and $e_2(\xi_\lambda) \geq 0.9$ no matter what combination of λ_i 's we choose. By observing the efficiency plots of the three criteria under different nominal values sets, we find $e_1(\xi_\lambda)$ and $e_2(\xi_\lambda)$ seem to be very competitive with each other; see Figures 2, 3 and 4). As expected, when one criterion gains in value, the other loses, and vice versa. Competitive criteria will result in a substantial loss for a small gain in the other, and vice versa.

For each set of the nominal values, Table 3 displays the combination of λ_i 's that produces the maximum of the minimum efficiencies. For example, $\lambda_1 = 0.55$, $\lambda_2 = 0.35$, $\lambda_3 = 0.1$ is the weights combination that produces the maximum of the minimum efficiency as 0.63 for all three criteria. That means with this choice for λ , we can find a compound optimal design ξ_λ that maximizes the precision for estimating all parameters and at the same time has at least 63% efficiency for estimating the MED and MTD .

In addition, we observe that for all the three nominal values sets, the compound optimal designs maintain high D -efficiency for most of the combinations of λ_i 's. This indicates for those nominal values that consistently produce high D -efficiency for different criteria weight λ_i 's, we may reduce the three-objective compound optimal design to two-objective for estimating the MTD and MED only so that the optimal design problem is simplified.

We close this section by noting that the above approach of finding locally optimal designs can be directly extended to finding Bayesian optimal designs with one or multiple objectives. Other methods for finding Bayesian optimal designs for Phase I trials and do not require that each objective be a convex function of the information matrix are available. Some seminal work and notable work in this direction are [42, 40, 56], among many others.

As an application, we revisit the clinical trial in [56] to identify the biologically optimal dose for a biologic agent to boost haemoglobin (HbL) levels for patients whose HbL were below the normal range. After the drug administration, the outcomes of patients could be categorized as 'no response', implying the drug had no effect, 'success', implying that the HbL levels were back to

Table 3. ξ_D , ξ_{MTD} and ξ_{MED} -efficiencies of ξ_λ for selected λ vectors.

(a_1, b_1, a_2, b_2)	λ_1	λ_2	λ_3	$e_1(\xi_\lambda)$	$e_2(\xi_\lambda)$	$e_3(\xi_\lambda)$
$(-3.3, 0.5, 3.4, 1)$	0.55	0.35	0.1	0.63	0.63	0.87
$(-1, 0.5, 2, 1)$	0.7	0.3	0	0.82	0.85	0.88
$(0.4, 0.2, 2, 1)$	0.6	0.4	0	0.75	0.75	0.81

the normal range or ‘toxicity’, meaning that patients were over-stimulated and their HbL were raised beyond the normal range. The probabilities of these event sum to unity and the authors also chose the continuation ration model because of its flexibility and particular usefulness for modelling the three probabilities realistically. The probabilities have unknown parameters and so the information matrix of a design also has unknown parameters in them. Consequently, the design criterion which is formulated as a convex function of the information matrix also contains unknown parameters. Like, [42, 40], [56] also used a Bayesian approach and so they specified prior densities for the unknown parameters and averaged them out in the design criterion before finding a design to optimize the criterion. Locally optimal designs used in this paper can be viewed the same except that the prior densities are all one point priors. Thus depending on the situation, the practitioner may prefer a Bayesian optimal design or a locally optimal design. In the latter case, a single best guess for each parameter is needed and the prior elicitation process may be easier for the practitioner to do so. Frequently best guesses come from prior experience with a similar drug or from a pilot study to estimate the model parameters; see [10] for implementation details of a locally optimal design.

9. COMPARISONS OF ALGORITHMS

When tackling design or any optimization problems, especially high dimensional ones, where the solution is unknown, it is always not clear which metaheuristic algorithm to use. Solely relying on results from a single metaheuristic algorithms is risky since they do not guarantee convergence to the global optimum. It is thus important to run several metaheuristic algorithms and observe whether they all produce similar solution. If it is the case, the confidence in our solution is increased.

We now briefly compare performance of PSO with two other algorithms. We choose to compare CPU time required by PSO to find optimal designs with two of its competitors: Differential evolution (DE) and Cocktail algorithm (CA). Some of the models we used in the comparisons are listed in Tables 4 and 5. The compartment model is taken from [5] and it has 3 parameters and some of the interesting features to estimate were area under the

Table 4. CPU time required to find PSO and DE-generated locally D-optimal designs. The numbers in parentheses are flocks size and the maximum number of iterations allowed.

Model	Algorithm	CPU time
a) Compartmental model	PSO(20,200)	0.4
	DE(20,200)	2.9
b) Logistic quadratic model	PSO(100,200)	1.4
	DE(80,400)	57.4
c) Hill model	PSO(20,200)	0.3
	DE(20,200)	1.4
d) Bivariable linear model	PSO(80,300)	1.2
	DE(60,200)	11.0
e) CR model ($b_1 = b_2$)	PSO(20,200)	0.1
	DE(20,200)	1.4
e) CR model ($b_1 \neq b_2$)	PSO(100,300)	2.5
	DE(40,300)	6.5

Table 5. CPU time required by PSO and DE to find locally c-optimal designs. The numbers in parentheses are flocks size and the maximum number of iterations allowed.

Model	Algorithm	CPU time
f) t_{max}	PSO(60,100)	4.5
	DE(20,200)	2.3
g) AUC	PSO(40,1000)	2.8
	DE(40,200)	4.6
h) MED in CR($b_1 \neq b_2$)	PSO(80,200)	1.2
	DE(40,200)	7.4

curve (AUC), the time to maximum (t_{max}) and the model parameters. The Hill model is the 4-parameter logistic model and the bivariable linear model has two additive factors. All have normally and independently distributed errors with constant variance. In the following tables, the numbers in parentheses are flocks size and the maximum number of iterations allowed for PSO and the single number in parentheses represents the number of points in the discretized dose interval

DE is a metaheuristic algorithm like PSO, which is a general purpose optimization algorithm. We use it to find locally D-optimal design for models (a) to (e) in Table 4, and locally c-optimal designs for estimating the three quantities (f) to (h) in Table 5. In this section, we worked with known optimal designs and the generated designs by PSO, DE and CA are considered to be optimal when all their design points and their weights are the same as the known optimal designs rounded to the 4 decimal places.

CA is a deterministic algorithm, which means that the same input results in the same answer, which in our case, results in the same optimal design. CA requires the design space to be discretized into a finite number of dose levels to approximate the continuous space. This approach can limit the ability of CA to find the optimal design because the discretized points may not correspond to a design point of the optimal design. However, our experience is that the CA-generated designs can usually still be highly D-efficient (with D-efficiency greater than 90%). In addition, we observe from Table 6 that the number of support points in the PSO-generated designs tend to be smaller than those generated by CA. Thus the simpler PSO-generated designs are attractive and also likely less costly to implement because of the fewer design points.

Cocktail algorithm (CA) is a deterministic algorithm which always produces the same output given one input. For PSO, we still use the same tuning parameters in section 5.2.1 such that the algorithm produces the locally D-optimal at 90% chance. Our empirical experience with CA is that although its generated designs can be very efficient even when the grid size n is as small as 10, they may have many more support points than the true optimal design does. For example, we compare the locally D-optimal designs for bivariable linear model d) generated by PSO and CA in Table 6. The numbers in the parenthesis after PSO are flock size and maximum iteration number; the numbers in the parenthesis after CA are the grid size used in different optimization problems. PSO is able to find the true optimal design supported at 6 points with D-efficiency 1. Both of the CA generated designs have D-efficiency greater than 0.99, but they

Table 6. *PSO and CA generated locally D-optimal designs for bivariable linear model in (d).*

Algorithm	CPU time	D_{eff}	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
PSO(80,300)	1.2	1	-1	-1	0	0	1	1		
weight			0	1	0	1	0	1		
CA (10)	0.1	0.997	-1	-1	-0.1111	-0.1111	0.1111	0.1111	1	1
weight			0	1	0	1	0	1	0	1
CA (1000)	94.6	1	-1	-1	-0.0101	-0.0101	0.0101	0.0101	1	1
weight			0	1	0	1	0	1	0	1
weight			0.1875	0.1875	0.6250	0.6250	0.6250	0.6250	0.1875	0.1875

Table 7. *Comparisons of the locally D-optimal design of the CR model with constant or non-constant slopes on $\chi = [-10, 10]$ between PSO and CA. The numbers in parentheses are flocks size and the maximum number of iterations allowed for PSO and the single number in parentheses represents the number of points in the discretized dose interval.*

Model	Algorithm	# of support points	CPU time
CR model with $b_1 = b_2 = 1$	PSO(20,200)	3	0.1
	CA(100)	5	1.7
	CA(1000)	5	20.8
	CA(10000)	4	71.3
CR model with $b_1 = 0.5, b_2 = 1$	PSO(100,500)	4	1.5
	CA(100)	5	2.9
	CA(1000)	6	14.3
	CA(10000)	5	122.7

have two more support points near $(0, 0)^T$ and $(0, 1)^T$ due to the discretized dose space. In biomedical researches such as a dose response study, practitioners usually prefer a simple design with fewer support points to a complicated design with a lot of points on the condition that both designs provide almost the same efficiency. Therefore in addition to the average CPU time, we use the number of support points in the generated design as another comparison criterion to measure the accuracy of CA generated design.

We observe from Tables 4-7, that generally CPU times required by PSO are shorter than that required by DE or CA. In a couple of cases, we observe that the opposite results are true but when PSO outperforms the two algorithms, the shorter CPU times are substantially shorter than either of the two algorithms.

10. CONCLUSIONS

Our work generalizes the work of [16] and [57] to find the three-objective design for the CR model. We show PSO is a powerful and flexible tool to find a three-objective compound optimal design for simultaneously estimating *MED* that produces maximal efficacy, the maximum tolerated dose (*MTD*) and all parameters in the CR model with non-constant slope $b_1 \neq b_2$. We were able to find the compound optimal designs under different combinations of weights λ , and investigate the impact of λ on efficiencies relative to different criteria. Such technique enables researchers and practitioners to construct the desired compound optimal design through appropriate weights combination of three optimal criteria in a more flexible and informative way.

We used simulation studies and investigated efficiencies of the PSO-generated designs when we applied two repair mechanisms

to bring out of area points back to the dose space: one employs a random repair mechanism and brings outlying points back to the dose space and the other exploits a common property of optimal designs and bring them back to the boundary of the dose space. The latter option greatly expedites PSO in searching for locally D- and c-optimal designs for a variety models, including the continuation-ratio model. Additionally, we compared performance of PSO with boundary repair to two other popular algorithms: the Cocktail algorithm (CA) in the statistics literature [55] and Differential Evolution (DE) algorithm in the engineering literature [38]. In the simulations using the same set of models, we find PSO outperforms DE in terms of faster speed to converge to locally D- and c-optimal designs. In comparisons between PSO and CA, we find CA usually has difficulty in allocating weights for adjacent points when the grid size is increased, and CA generated designs do not maintain a consistent structure when the grid size or dose space changes. In contrast, PSO consistently finds the locally D-optimal designs regardless what design space is chosen. Since we set PSO to round all the design points and weights to 4 decimal places, such precision enables us explore the structure of optimal designs at a microscopic level.

The techniques developed here are broadly applicable to find types of optimal designs for other nonlinear models. For example, standardized maximin optimal designs were found for estimating parameters in Michaelis-types of models that study how substrate concentration affects reaction with the enzyme in [9]. Maximin optimal designs maximizes the minimal efficiency under an uncertain situation; in [9], it was over mis-specification of values for the nominal parameters in the nonlinear model. When the estimated parameters vary substantially in values, a more

appropriate criterion is to use a standardized maximin optimality criterion and the standardized maximin optimal design is found by solving a multi-nested optimization problem.

The PSO codes to find designs reported in the paper can be directly amended to find different types of multiple-objective optimal designs for other models. For example, the codes were recently amended to find dual-objective Bayesian optimal designs for estimating a percentile and model parameters simultaneously in a Beta regression for modeling proportion of malformed purple sea urchins in the embryo after they were treated with a toxic agent [12]. The Beta regression has 4 parameters and the implemented code allows a discrete prior distribution be specified up to 5 plausible sets of nominal values. More information is available at the interactive website at https://elviscuihan.shinyapps.io/Dc_optimal_design/, where references, input instructions and explanations are available. Readers who are interested in the PSO codes can send their request to the second author.

10.1 Hybrid Metaheuristic Algorithms

When a metaheuristic algorithm does not produce a solution near to the optimum, it is common to use one of its variants. A variant is just one of many modifications of the algorithm that improves the original version in various ways. The variant can have improved convergence property, for a more targeted application, have better self-finding tuning parameters or the like. Popular metaheuristic algorithms, like PSO, probably have a couple of dozens of variants now. Examples of PSO variants are [14, 48, 8, 39] and [51], to name a few.

Another option is to hybridize the algorithm with one or more carefully selected metaheuristic algorithms so that the hybridized version combines the strengths of each algorithm. A common guiding principle is that the hybridized algorithm performs better than each of the individual algorithm, see [7, 17, 32] for details, an overview of hybridization methods and their varied applications.

As an example, a hybridized metaheuristic algorithm was used to extend the capabilities of the celebrated Simon 2-stage design for a Phase II trial [35]. Instead of a single paired set of the null and alternative hypotheses, [27] proposed a framework to allow for a single null hypothesis and 3 possible alternative hypotheses where only one will be tested in the second stage and the choice depends on results in the first stage. Earlier papers had noted, see for example [30], that it is relatively easy to specify the null hypothesis in practice, but not so for specifying a single alternative in an informed way. This is because in an early phase trial, it can be problematic to specify the effectiveness of the drug being tested for the particular medical condition. An over or under specification of its efficacy rate can affect the power of the test and in the decision making process. To this end, [30] allowed two alternative hypotheses and depending on results from Phase I, one of the alternative hypotheses will be tested. They noted that the optimization problem searches for 5 integers, appropriately ordered, to solve the complicated nonlinear problem with three different error constraints. They used a greedy search (i.e. searched over all possibilities) and they remarked that they exhausted the computing power at that time.

Expanding on this framework and to further reduce the difficulty or uncertainty of specifying the alternative hypothesis, [27] allows for 3 alternative hypotheses, and only will be tested based on results from the Phase I results. The greedy approach

did not work after weeks of computing time and metaheuristics was used for the search. PSO was first used to solve the discrete optimization problem, and it did not work even after changing the flock size and values of the tuning parameters multiple times. Omitting details, [27] was able to hybridize PSO with appropriate algorithms and found a solution to a 10 integer optimization problem, appropriately ordered, that met the 5 nonlinear error constraints and optimized the two criteria in Simon's original paper. While there is no theory to verify its optimality, it was verified that the numerical results agreed with those of [35, 30] when one or two alternative hypotheses were omitted in our framework.

Hybridized metaheuristics was also used to find an optimal discrimination design to ascertain the most plausible nonlinear model among several nonlinear models [9]. There were 5 possible models postulated by [37] with an uni-variate continuous outcome that may have various error distributional assumptions. They considered different types of discrimination criteria and applied a hybrid metaheuristic algorithm to generate an optimal design in a toxicology experiment.

Not surprisingly, nature-inspired algorithms have been promptly applied to better understand various aspects of COVID-19. Various such algorithms, such as PSO, DE and Imperialist Competitive Algorithm (ICA) have been used to tackle the control and spread of COVID-19. ICA is based on human behavior and proposed by [3]. For example, [34] implemented ICA to predict trends in the COVID-19 pandemic in Hungary, [15] used DE to monitor spread of the COVID-19 virus in Italy, [23] applied PSO to estimate model parameters in SEIR models commonly used in epidemiology to track an epidemic, [31] applied PSO and used real time data to estimate and predict death rates caused by COVID-19. Additionally, [36] used DE to classify COVID-19 patients from chest CT images and most recently, [25] proposed a COVID-19 optimizer algorithm specifically for modeling and controlling the coronavirus distribution process and one its objectives is to minimize the number of infected countries to slow down the spread. The authors also showed their algorithm outperformed PSO and GA by 53% and 59%, respectively, and newer created metaheuristic algorithms, like the Volcano Eruption Algorithm and the Gray Wolf optimizer, by 15% and 37%, respectively.

Pareto Optimization (PO) is a common approach to solve optimization problems with multiple objectives. [1] applied PO to tackle problems posed by COVID-19, which can infect many people quickly resulting in huge requests of medical care at varying levels. Coping with how, when and where to admit COVID-19 patients to various hospitals is a complex multi-objective optimization problem. For instance, to decrease the in-bed time, save lives and resources, the choice of the most suitable hospital for the patient to be admitted has to be balanced by expected admission time, hospital readiness and severity of the COVID-19 patient. These are multi-objective optimization problems and the author showed their strategy using data from 254 patients in Saudi Arabia outperformed the lexicographic multi-objective optimization method. Recently, evolutionary algorithms have made remarkable progress in tackling many types of multi-objective optimization problems [44, 46, 43, 45] and we expect metaheuristic algorithms will make important contributions to solve COVID-19 multi-objective optimization problems and beyond. Likewise, machine learning has made many advances in tackling COVID-19 problems [21]. Research in metaheuristic algorithms is very active

[51] in various ways, including from the mathematical front, see for example, [28, 47]. Further integration of metaheuristics and machine learning techniques should enable us to tackle pandemic and complicated optimization problems more effectively.

ACKNOWLEDGEMENTS

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Wong is also grateful for the Yushan Scholarship Award from the Ministry of Education in Taiwan and the hospitality of the Department of Statistics at the National Cheng Kung University in Tainan, where this manuscript was completed.

FUNDING

The research of Wong reported in this publication was partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639.

Accepted 24 June 2023

REFERENCES

- [1] ABDELAZIZ, A. M., ALARABI, L., BASALAMAH, S. and HENDAWI, A. (2021). A Multi-objective optimization method for hospital admission problem – a case study on COVID-19 patients. *Algorithms* **14**(2) 38. <https://doi.org/10.3390/a14020038>.
- [2] AGRETI, A. (1990) *Categorical Data Analysis*. Wiley. [MR1044993](https://doi.org/10.1002/978-1-4419-1644-0_9)
- [3] ATASHPAZ-GARGARI, C. E. ; LUCAS (2007). Imperialist Competitive Algorithm: An algorithm for optimization inspired by imperialistic competition. *IEEE Congress on Evolutionary Computation* **7** 4661–4666.
- [4] ATKINSON, A. C. (1992). A segmented algorithm for simulated annealing. *Statistics & Computing* **2** 221–230.
- [5] ATKINSON, A. C. and DONEV, A. N. (1992) *Optimum Experimental Designs*. OXFORD University Press.
- [6] ATKINSON, A. C. and HAINES, L. M. (1996). Designs for non-linear and generalized linear models. *In Handbook of Statistics* **13** 437–475. [https://doi.org/10.1016/S0169-7161\(96\)13016-9](https://doi.org/10.1016/S0169-7161(96)13016-9). [MR1492576](https://doi.org/10.1016/S0169-7161(96)13016-9)
- [7] BLUM, C., PUCHINGER, J., RAIDL, G. R. and ROLI, A. (2011). Hybrid metaheuristics in combinatorial optimization: A survey. *Applied Soft Computing* **11** 4135–4151. https://doi.org/10.1007/978-1-4419-1644-0_9. [MR2766839](https://doi.org/10.1007/978-1-4419-1644-0_9)
- [8] BONYADI, M. R. and MICHALEWICZ, Z. (2014). A locally convergent rotationally invariant particle swarm optimization algorithm. *Swarm Intelligence* **8**(3) 159–198.
- [9] CHEN, R. -B., CHEN, P. -Y., HSU, C. -L. and WONG, W. K. (2020). Hybrid algorithms for generating optimal designs for discriminating multiple nonlinear models under various error distributional assumptions. *PLOS ONE* **15**(10). <https://doi.org/10.1371/journal.pone.0239864>.
- [10] CHERNOFF, H. (1953). Locally Optimal Designs for Estimating Parameters. *Annals of Mathematical Statistics* **24**(4) 586–602. <https://doi.org/10.1214/aoms/1177728915>. [MR0058932](https://doi.org/10.1214/aoms/1177728915)
- [11] CLYDE, M. and CHALONER, K. (1996). The Equivalence of Constrained and Weighted Designs in Multiple Objective Design Problems. *Journal of the American Statistician Association* **91**(43) 1236–1244. <https://doi.org/10.2307/2291742>. [MR1424621](https://doi.org/10.2307/2291742)
- [12] COLLINS, M. D., CUI, E., H., S. W. HYUN and WONG, W. K. (2021). A Model-based Approach to Designing Developmental Toxicology Experiments using Sea Urchin Embryos. *Archives of Toxicology* **96** 919–932.
- [13] COOK, R. D. and WONG, W. K. (1994). On the Equivalence of Constrained and Compound Optimal Designs. *Journal of the American Statistician Association* **89**(426) 687–692. [MR1294092](https://doi.org/10.1080/01621459408838648)
- [14] EBERHART, R. and KENNEDY, J. (1995). A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on* 39–43. IEEE.
- [15] FALCO, I. D., CIOPPA, A. D., SCAFURI, U. and TARANTINO, E. (2020). Coronavirus COVID-19 spreading in Italy: optimizing an epidemiological model with dynamic social distancing through Differential Evolution. *arXiv preprint arXiv:2004.00553v3*.
- [16] FAN, S. K. and CHALONER, K. (2004). Optimal designs and limiting optimal designs for a trinomial response. *Journal of Statistical Planning and Inference* **126** 347–360. <https://doi.org/10.1016/j.jspi.2003.08.004>. [MR2090863](https://doi.org/10.1016/j.jspi.2003.08.004)
- [17] FAN, S. Q. K. S., LIANG, Y. -C. and ZAHARA, E. (2004). Hybrid simplex search and particle swarm optimization for the global optimization of multimodal functions. *Engineering Optimization* **36**(4) 401–418.
- [18] FEDOROV, V. V. and WU, Y. (2007). Dose finding designs for continuous responses and binary utility. *Journal of Biopharmaceutical Statistics* **17** 1085–1096. <https://doi.org/10.1080/10543400701645132>. [MR2414563](https://doi.org/10.1080/10543400701645132)
- [19] FEDOROV, V. (1972). *Theory of optimal design*. Academic, New York. [MR0403103](https://doi.org/10.1080/00036817208839648)
- [20] FUMOLEAU, P., TRIGO, J. M., ISAMBERT, N., SIMOND, D., GUPTA, S. and CAMPONE, M. (2013). Phase I dose-finding study of cabazitaxel administered weekly in patients with advanced solid tumours. *BMC Cancer* **13** 1471–2407.
- [21] GOMES, D. C. D. S. and SERRA, G. L. D. O. (2021). Machine learning model for computational tracking forecasting the COVID-19 dynamic propagation. *IEEE Journal of Biomedical and Health Informatics* **25**(3) 515–622.
- [22] HAINES, L. M. (1987). The application of the annealing algorithm to the construction of exact optimal designs for linear regression models. *Technometrics* **29**(4) 439–447. [MR2637962](https://doi.org/10.1080/001418019808838648)
- [23] HE, S., PENG, Y. and SUN, K. (2020). SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics* **101** 10–100071107102005743.
- [24] HEISE, M. A. and MYERS, R. H. (1996). Optimal designs for bivariate logistic regression. *Biometrics* **52** 613–624.
- [25] HOSSEINI, E., GHAFOOR, K. Z., SADIQ, A. S., GUZANI, M. and EMROUZNEJAD, A. (2020). COVID-19 Optimizer algorithm, modeling and controlling of coronavirus distribution process. *IEEE Journal of Biomedical and Health Informatics* **24**(10) 2765–2775.
- [26] KIEFER, J. (1974). General Equivalence Theory for optimum designs (Approximate Theory). *The Annals of Statistics* **2** 849–879. [MR0356386](https://doi.org/10.1214/aoms/1177728915)
- [27] KIM, S. and WONG, W. K. (2018). Extended two-stage adaptive designs with three target responses for phase II clinical trials. *Statistical methods in Medical Research* **27**(12) 3628–3642. <https://doi.org/10.1177/0962280217709817>. [MR3878648](https://doi.org/10.1177/0962280217709817)
- [28] LAI, T. L., CHOI, K. P., TONG, T. X. and WONG, W. K. (2021). A Statistical Approach to Adaptive Parameter Tuning in Nature-inspired Optimization and Optimal Sequential Design of Dose-Finding Trials. *Statistica Sinica*. In press. <https://doi.org/10.5705/ss.20. MR4338089>
- [29] LI, G. and MAJUMDAR, D. (2009). Some results on D-optimal designs for nonlinear models with applications. *Biometrika* **96**(2) 487–493. <https://doi.org/10.1093/biomet/asp004>. [MR2507158](https://doi.org/10.1093/biomet/asp004)
- [30] LIN, Y. and SHIH, W. J. (2004). Adaptive two-stage designs for single-arm phase IIA cancer clinical trials. *Biometrics* **60** 482–490. <https://doi.org/10.1111/j.0006-341X.2004.00193.x>. [MR2067004](https://doi.org/10.1111/j.0006-341X.2004.00193.x)
- [31] MAKADE, R. G., CHAKRABARTI, S. and JAMIL, B. (2020). Real time estimation and prediction of the mortality caused due

- to COVID-19 using particle swarm optimization and finding the most influential parameter. *Infectious Disease Modelling* **5** 772–782.
- [32] MOHAMMADI, M., NASTARAN, M. and SAHEBGHARANI, A. (2016). Development, application, and comparison of hybrid metaheuristics for urban land-use allocation optimization: Tabu search, genetic, GRASP, and simulated annealing algorithms. *Computers, Environment and Urban Systems* **60** 23–36.
- [33] MONTEPIEDRA, G., MYERS, D. and B., Y. A. (1998). Application of genetic algorithms to the construction of exact D-optimal designs. *Journal of Applied Statistics* **25** 817–826.
- [34] PINTER, G., FELDE, I., MOSAVI, A., GHAMISI, P. and GLOAGUEN, R. (2020). COVID-19 Pandemic prediction for Hungary; a hybrid machine learning approach. *MDPI Mathematics* **8** 890.
- [35] SIMON, R. (1989). Optimal Two-Stage Designs for Phase II Clinical Trials. *Controlled Clinical Trials* **10** 1–10.
- [36] SINGH, D., KUMAR, V., VAISHALI and Kaur, M. (2020). Classification of COVID-19 patients from chest CT images using multi-objective Differential Evolution-based convolutional neural networks. *European Journal of Clinical Microbiology and Infectious Diseases* **39** 1379–21389.
- [37] SLOB, W. (2002). Dose-response modeling of continuous endpoints. *Toxicological Sciences* **66** 298–312.
- [38] STORN, R. and PRICE, K. (1997). Differential Evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* **11** 341–359. <https://doi.org/10.1023/A:1008202821328>. MR1479553
- [39] SUN, J., XU, W. and FENG, B. (2004). A global search strategy of quantum-behaved particle swarm optimization **1** 111–116.
- [40] THALL, P. F. and COOK, D. J. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics* **60**(3) 684–693. <https://doi.org/10.1111/j.0006-341X.2004.00218.x>. MR2089444
- [41] THALL, P. F. and RUSSELL, K. E. (1998). A Strategy for Dose-Finding and Safety Monitoring Based on Efficacy and Adverse Outcomes in Phase I/II Clinical Trials. *Biometrics* **54** 251–264.
- [42] THALL, P. F., COOK, D. J. and ESTEY, E. H. (2006). Adaptive dose selection using efficacy-toxicity trade-offs: Illustrations and practical considerations. *Journal of Biopharmaceutical Statistics* **16**(5) 623–638. <https://doi.org/10.1080/10543400600860394>. MR2252311
- [43] TIAN, Y., YANG, S. and ZHANG, X. (2020). An Evolutionary Multiobjective Optimization Based Fuzzy Method for Overlapping Community Detection. *IEEE Transactions on Fuzzy Systems* **28**(11) 2841–2855.
- [44] TIAN, Y., ZHANG, X., WANG, C. and JIN, Y. (2018). An Evolutionary Algorithm for Large-Scale Sparse Multiobjective Optimization Problems. *IEEE Transactions on Evolutionary Computation* **24**(2) 380–393.
- [45] TIAN, Y., SU, X., SU, Y. and ZHANG, X. (2020). EMODMI: A Multi-Objective Optimization Based Method to Identify Disease Modules. *IEEE Transactions on Emerging Topics in Computational Intelligence*.
- [46] TIAN, Y., LIU, R., ZHANG, X., MA, H., TAN, K. C. and JIN, Y. (2020). A Multi-Population Evolutionary Algorithm for Solving Large-Scale Multi-Modal Multi-Objective Optimization Problems. *IEEE Transactions on Evolutionary Computation*. MR4361281
- [47] TONG, T. X., CHOI, K. P., LAI, T. L. and WONG, W. K. (2021). Stability Bounds and Almost Sure Convergence of Improved Particle Swarm Optimization Methods. *Research in Mathematical Sciences*. <https://doi.org/10.1007/s40687-020-00241-4>. MR4257863
- [48] VAN DEN BERGH, F. and ENGELBRECHT, A. P. (2002). A new locally convergent particle swarm optimiser. In *Systems, Man and Cybernetics, 2002 IEEE International Conference on* **3** 6. IEEE.
- [49] WHITACRE, J. M. (2011). Recent trends indicate rapid growth of nature-inspired optimization in academia and industry. *Computing* **93** 121–133. <https://doi.org/10.1007/s00607-011-0154-z>. MR2860177
- [50] WHITACRE, J. M. (2011). Survival of the flexible: explaining the recent popularity of nature-inspired optimization within a rapidly evolving world. *Computing* **93** 135–146. <https://doi.org/10.1007/s00607-011-0156-x>. MR2860178
- [51] WU, M., HU, W., HU, J. and YEN, G. G. (2021). Adaptive Multiobjective Particle Swarm Optimization Based on Evolutionary State Estimation. *IEEE Transactions on Cybernetics* **51**(7) 3738–3751.
- [52] XU, W. W., K., W., TAN, K. C. and XU, J. X. (2019). Finding High-Dimensional D-Optimal Designs for Logistic Models via Differential Evolution. *IEEE Access* **7** 7133–7146.
- [53] YANG, X. S. (2010). Engineering Optimization: An Introduction with Metaheuristic Applications. In *Particle Swarm Optimization* John Wiley and Sons.
- [54] YEAP, B. Y., CATALANO, P. J., RYAN, L. M. and DAVIDIAN, M. (2003). Robust two-stage approach to repeated measurements analysis of chronic ozone exposure in rats. *Journal of Agricultural, Biological, and Environmental Statistics* **8**(4) 438–454.
- [55] YU, Y. (2011). D-optimal designs via a cocktail algorithm. *Stat Comput* **21** 475–481. <https://doi.org/10.1007/s11222-010-9183-2>. MR2826686
- [56] ZHANG, W., SARGENT, D. J. and MANDREKAR, S. (2006). An adaptive dose-finding design incorporating both toxicity and efficacy. *Statistics in Medicine* **25** 2365–2383. <https://doi.org/10.1002/sim.2325>. MR2240943
- [57] ZHU, W. and WONG, W. K. (2000). Multi-objective Designs in A Dose-response Experiment. *Journal of Biopharmaceutical Statistics*, **10:1** **10** 1–14.
- [58] ZOCCHI, S. S. and ATKINSON, A. C. (1999). Optimum Experimental Designs for Multinomial Logistic Models. *Biometrics* **55** 437–444. <https://doi.org/10.1111/j.0006-341X.1999.00437.x>. MR1705110

Jiaheng Qiu. Day One Biopharmaceuticals, Brisbane, California, USA.

Weng Kee Wong. University of California, Los Angeles, California, USA. E-mail address: wkwong@ucla.edu