

# UCLA

## UCLA Previously Published Works

### Title

Extended two-stage adaptive designs with three target responses for phase II clinical trials

### Permalink

<https://escholarship.org/uc/item/3255v0tt>

### Journal

Statistical Methods in Medical Research, 12(12)

### Authors

Kim, Seongho  
Wong, Weng Kee

### Publication Date

2018

### DOI

10.1177/0962280217709817

Peer reviewed

# Extended two-stage adaptive designs with three target responses for phase II clinical trials

Seongho Kim<sup>1</sup> and Weng Kee Wong<sup>2</sup>

Statistical Methods in Medical Research  
2018, Vol. 27(12) 3628–3642

© The Author(s) 2017

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0962280217709817

journals.sagepub.com/home/smm



## Abstract

We develop a nature-inspired stochastic population-based algorithm and call it discrete particle swarm optimization to find extended two-stage adaptive optimal designs that allow three target response rates for the drug in a phase II trial. Our proposed designs include the celebrated Simon's two-stage design and its extension that allows two target response rates to be specified for the drug. We show that discrete particle swarm optimization not only frequently outperforms greedy algorithms, which are currently used to find such designs when there are only a few parameters; it is also capable of solving design problems posed here with more parameters that greedy algorithms cannot solve. In stage 1 of our proposed designs, futility is quickly assessed and if there are sufficient responders to move to stage 2, one tests one of the three target response rates of the drug, subject to various user-specified testing error rates. Our designs are therefore more flexible and interestingly, do not necessarily require larger expected sample size requirements than two-stage adaptive designs. Using a real adaptive trial for melanoma patients, we show our proposed design requires one half fewer subjects than the implemented design in the study.

## Keywords

Adaptive design, greedy algorithm, particle swarm optimization, power, sequential design, Simon's two-stage design

## I Introduction

Phase II clinical trials concern early exploration of efficacy effects and use the most recent results from a small group of patients to make decision for the next group of patients until some prefixed termination rule is met. Group sizes can range from 20 to 120 depending on the nature and seriousness of the disease. Kramar et al.<sup>1</sup> provides a review on multistage designs for phase II clinical trials and statistical issues in cancer research and Brown et al.<sup>2</sup> gives an overview of the role of phase II trials in oncology. Increasingly, these designs employ an adaptive approach, where design decisions for the next stage are made based on cumulative responses. Chow and Chang<sup>3</sup> reviews adaptive randomization designs in clinical trials.

Simon's two-stage design was developed under a framework that if the proportion of responders in stage 1 is small, the trial is terminated; otherwise, the trial goes on to stage 2 where the cumulative response rate is now used to test whether the efficacy rate is at a higher pre-specified level. In essence, two user-selected efficacy rates are posited  $p_0$  and  $p_1$  with  $p_0 < p_1$ , and  $p_0$  is the uninspiring response rate. In stage 1, we test the null hypothesis  $H_0 : p = p_0$  and if we fail to reject the null hypothesis, we terminate the trial; otherwise, we conclude that the drug is sufficiently promising to advance to stage 2, where more responses from more patients will be used to test  $p = p_1$ . Given pre-specified Type I and II error rates for the tests at the two stages, the design questions are the number of patients required in stage 1, number of responders required in stage 1, how many additional patients are required for stage 2 and the cumulative number of responders required at the end of stage 2. The design problem is to determine the optimal combination of these four numbers, so that the expected number of patients treated with a

<sup>1</sup>Biostatistics Core, Karmanos Cancer Institute, USA

<sup>2</sup>Department of Biostatistics, UCLA School of Public Health, USA

### Corresponding author:

Seongho Kim, Biostatistics Core, Karmanos Cancer Institute School of Medicine, Wayne State University, Detroit, MI, USA.

Emails: kimse@karmanos.org; wkwong@ucla.edu

drug of low activity under the null hypothesis is minimized. Such an optimal design may not be unique and other design criteria are possible.

Lin and Shih<sup>4</sup> provided practical examples and showed that while it was relatively easy to specify the uninteresting rate  $p_0$ , the same was not true for  $p_1$ . To tackle the issue of uncertainty in targeting the hypothesis in stage 2, they proposed designs that allowed two specifications for  $p_1$  and called one a skeptical choice and the other the optimistic choice. Their design problem now has seven parameters to optimize in contrast to four in Simon's two-stage design, subject to user-specified Type I and II error constraints. The additional three parameters beyond the four parameters required for Simon's two-stage design are for testing the additional targeted efficacy rate. Both Simon and Lin and Shih employed a greedy search to find the optimal designs and the latter stated that they "did not extend the selection to more than two prefixed possible response rates mainly due to the complexity in the numerical solutions, and also because it is usually adequate for practitioners to contemplate between two (high/low) choices of the response rate  $p_1$ ."

Our work was motivated by a real problem from clients interested to conduct a single-arm two-stage phase II trial to the effect of head and neck cancer (HNC) on the incidence of obstructive sleep apnea (OSA). Our clients' main goal was to ascertain accurately the incidence rate of OSA on HNC patients in a timely manner. Because of the potential huge beneficial impact on treating HNC patients, the clients wanted a relatively simple design that could terminate the study at stage 1 if the initial response (OSA incidence) rate was poor; otherwise, stage 2 of the design used the number of responders, who experienced OSA from stage 1 to recruit more patients to more accurately assess the incidence rate of OSA. Clearly, Simon's<sup>5</sup> two-stage design proposed in 1989 and the adaptive two-stage design proposed by Lin and Shih<sup>4</sup> seemed useful.

Our clients were particularly interested to extend Lin and Shih's<sup>4</sup> approach to include an additional targeted alternative hypothesis for testing in the second stage. The main reason was the great uncertainty in the OSA incidence rate on HNC patients and the need to move a potentially promising impact further in the treatment of HNC patients expeditiously with a very good estimate of the true OSA incidence rate of HNC patients. Thus depending on the number of responders, who experience OSA in stage 1, more flexibility in specifying the incidence rate of OSA in stage 2 is required. In particular, the clients did not want to run another costly follow-up trial to more accurately evaluate the incidence rate of OSA in HNC patients. Another motivation for having such a design was also recently made aware to the second author by personal communication from a senior researcher at the National Institutes of Health. She mentioned that in oncology, there is a great tendency for researchers to be too optimistic on the efficacy rate of a new drug. Our proposed designs should help address such an issue by allowing the clinician the flexibility to perform a test at stage 2 for one of the two lower efficacy rates should the drug efficacy rate be over-specified in the first place. In what is to follow, we discuss and construct extended two-stage adaptive designs for our OSA study and two other applications. The first one is Lin and Shih's<sup>4</sup> vinorelbine, bleomycin, and gemcitabine (VBG) study for patients with recurrent or refractory Hodgkin's disease (HD) and the second one is a Phase II BREAK-2 study for melanoma patients

We were able to formulate the optimization problem quickly as an extension of that from Lin and Shih's<sup>4</sup> article, but anticipated the optimization burden for the new constrained optimization problem now with 10 parameters to optimize for the three user-selected targeted alternative hypotheses in stage 2 and subject to various user-specified Type I and II error rate constraints. Lin and Shih<sup>4</sup> acknowledged computational difficulties in their work at that time and remarked that further extension of the problem would seem to be a too challenging or an impossible computational task. In what is to follow, we present an algorithm to solve this complex constrained optimization problem and show that our extended two-stage adaptive design has advantages over both Simon's two-stage design and Lin and Shih's designs in a number of ways.

There are many variations of strategies proposed for phase II designs since Simon's two-stage design was proposed. We mention some here and refer to earlier cited articles for a comprehensive overview. Modifications and extensions of the likes of Simon's two-stage design include having two binary outcomes<sup>6</sup> or finding designs that minimize the expected sample size or minimize the expected maximum sample size, not under the null hypothesis but under the alternative hypothesis. For example, Mander and Thompson<sup>7</sup> and Mander et al.<sup>8</sup> investigated novel designs, which are optimal under the alternative hypothesis, that the tumor response rate is higher than the null hypothesis value, and also designs which allow early stopping for efficacy. Wason et al.<sup>9</sup> considered reducing sample size for phase II trials with a continuous outcome, and Kwak and Jung<sup>10</sup> proposed a two-stage adaptive optimal design for single arm trials with right-censored survival time that minimizes the expected sample size subject to Type I and II error rate specifications. Phase II designs with three stages were motivated and proposed in literature.<sup>11-13</sup> Schlesselman and Reis<sup>14</sup> noted limits and benefits of phase II trials and Zhou<sup>15</sup> gave some guidance on the choice of a design for an early phase trial.

Software for computing various types of phase II adaptive designs for such trials is available in commercial software statistical packages, like SAS, STATA, and JMP. They are also codes on webpages that generate various adaptive optimal designs for early phase trials after the error rates are specified. One such site is at <https://stattools.crab.org/Calculators/twoStage.htm>. Because none of the current software can find our proposed designs, we have developed codes to run on a web browser using the R package *shiny* (<http://shiny.rstudio.com>) to generate our proposed designs. We call our R package *ss2stagePSO* and it is freely available at <http://cansur.wayne.edu/>. We also provide an option to compare competitive designs and user-specified designs relative to the optimal designs by reporting their relative efficiencies.

In the next section, we first describe extended two-stage adaptive designs for phase II clinical trials and technical background before discussing nature-inspired stochastic population-based algorithms for finding our proposed designs. There are many of them and, as an example, we focus on particle swarm optimization (PSO), which seem to be widely used. Motivated from our recent successes with this flexible algorithm, we modified it for our application at hand, because PSO was originally designed to solve optimization problems over a continuous domain, not optimization over a domain comprising of discrete positive integer numbers, and call it discrete particle swarm optimization (DPSO). In particular, we show how DPSO solves our extended two-stage adaptive design problems and related problems effectively that a greedy algorithm cannot. We then apply our algorithm to real applications and show our proposed designs require smaller sample sizes than those that were implemented in the trial. The article is concluded with a summary

## 2 Extended two-stage adaptive designs

Suppose  $p_0$  is the maximum uninteresting response rate and there are three choices for the target response rates:  $p_1, p_2, p_3$ , where  $0 < p_0 < p_1 < p_2 < p_3 < 1$ . Data from stages 1 and 2 will be used to test one of the null hypotheses in stage 2 depending on the number of responders in stage 1. Our adaptive design assumes that a total of  $n_1$  patients are assigned at the first stage and the total number of patients and tests required in the entire trial will depend on the number of responses in the first and subsequent stages. The null hypothesis at the first stage is  $H_0: p \leq p_0$ . According to the number of responses in the first stage, the corresponding alternative hypothesis will be one of three hypotheses:  $H_{11}: p > p_1$ ,  $H_{12}: p > p_2$  and  $H_{13}: p > p_3$ . Using similar notation in Lin and Shih,<sup>4</sup> our extended two-stage adaptive design has a total of 10 parameters given by  $\theta = (s_1, r_1, q_1, n_1, s, l, r, m, q, n)$  and it operates as follows:

- Step I: Begin by recruiting  $n_1$  patients in the first stage of our design and observe the number of the responses,  $x$ , from of  $n_1$  patients.
- Step II:
  - (1) If  $x \leq s_1$ , stop the trial with failure to reject  $H_0$  (i.e.  $p \leq p_0$ )
  - (2) If  $s_1 < x \leq r_1$ , power the study at  $(1-\beta_1)$  for  $p=p_1$  and enter  $l_2 = l - n_1$  additional patients into the study. Reject the hypothesis that  $H_{11}: p > p_1$  if the total number of responses  $\leq s$  out of  $l$  patients.
  - (3) If  $r_1 < x \leq q_1$ , power the study at  $(1-\beta_2)$  for  $p=p_2$  and enter  $m_2 = m - n_1$  additional patients into the study. Reject the hypothesis that  $H_{12}: p > p_2$  if the total number of responses  $\leq r$  out of  $m$  patients.
  - (4) If  $x > q_1$ , power the study at  $(1-\beta_3)$  for  $p=p_3$  and enter  $n_2 = n - n_1$  additional patients into the study. Reject the hypothesis that  $H_{13}: p > p_3$  if the total number of responses  $\leq q$  out of  $n$  patients.

By construction, the values of  $l, m, n$  are the total number of patients required for the entire trial corresponding to the alternative hypotheses,  $H_{11}: p > p_1$ ,  $H_{12}: p > p_2$ , and  $H_{13}: p > p_3$ , respectively. The extended two-stage adaptive design has a total of 10 parameters denoted by  $\theta = (s_1, r_1, q_1, n_1, s, l, r, m, q, n)$  that we wish to optimize given error rate constraints and the stipulated three response rates. If the true response probability is  $p$ , the probability of failing to reject  $H_0$  is given by

$$G(\theta|p) = B(s_1, n_1, p) + \sum_{x=s_1+1}^{\min(r_1, s)} b(x, n_1, p)B(s-x, l_2, p) + \sum_{x=r_1+1}^{\min(q_1, r)} b(x, n_1, p)B(r-x, m_2, p) + \sum_{x=q_1+1}^{\min(q, n_1)} b(x, n_1, p)B(q-x, n_2, p) \quad (1)$$

where  $b$  and  $B$  are the probability density function and cumulative density function of a binomial distribution, respectively, with  $l = n_1 + l_2$ ,  $m = n_1 + m_2$ , and  $n = n_1 + n_2$ . It follows that when the true probability response rate

is  $p$ , the expected sample size is

$$E(N|p, \theta) = n_1 + (B(r_1, n_1, p) - B(s_1, n_1, p))l_2 + (B(q_1, n_1, p) - B(r_1, n_1, p))m_2 + (1 - B(q_1, n_1, p))n_2 \tag{2}$$

Our design problem is to find a good choice  $\hat{\theta} \in \tilde{\Theta}$ , the set containing all values of  $\theta$  that satisfies four natural error constraints

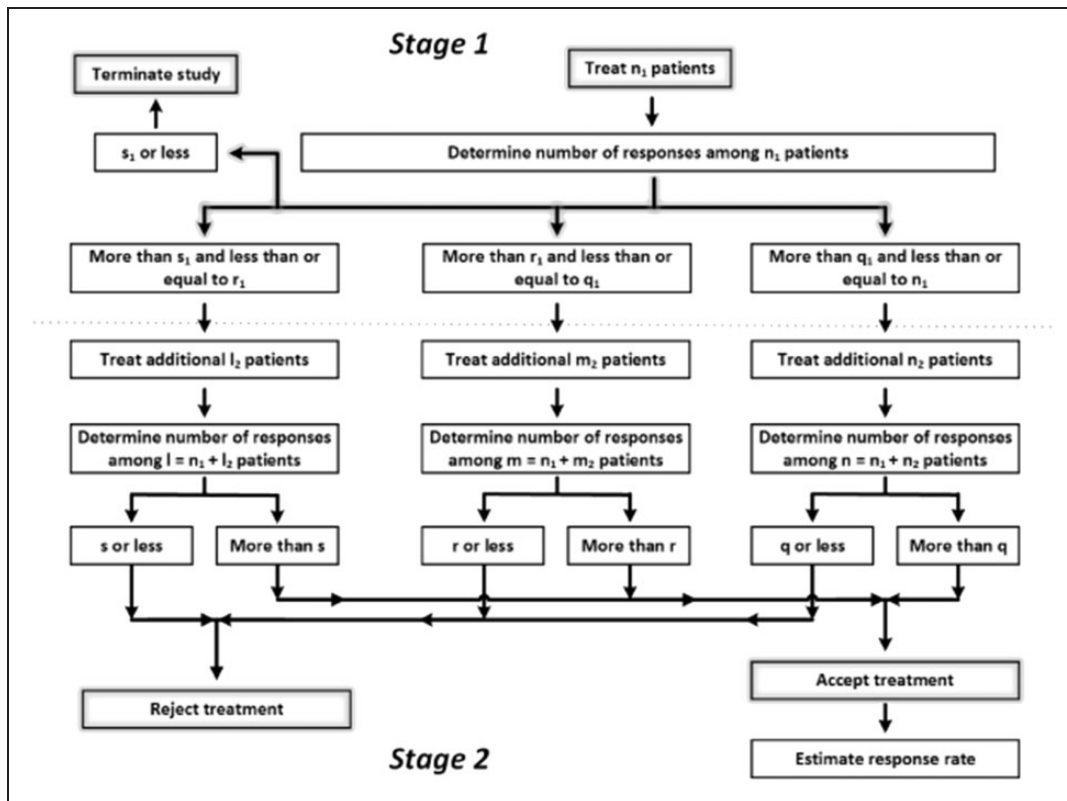
$$G(\theta|p_0) \geq 1 - \alpha, G(\theta|p_1) \leq \beta_1, G(\theta|p_2) \leq \beta_2 \text{ and } G(\theta|p_3) \leq \beta_3 \tag{3}$$

The goodness of  $\hat{\theta}$  may be determined by one of the following four optimality criteria:

- C1:  $\hat{\theta} = \arg \min_{\theta \in \tilde{\Theta}} E(N|p_0, \theta)$ ;
- C2:  $\hat{\theta} = \arg \min_{\theta \in \tilde{\Theta}} E(N|p_0, \theta)$  and  $\hat{\theta} = \arg \min_{\theta \in \tilde{\Theta}} \{\max(l, m, n)\}$ ;
- C3:  $\hat{\theta} = \arg \min_{\theta \in \tilde{\Theta}} \{\max_{i=0,1,2,3} E(N|p_i, \theta)\}$ ;
- C4:  $\hat{\theta} = \arg \min_{\theta \in \tilde{\Theta}} \{\max_{i=0,1,2,3} E(N|p_i, \theta)\}$  and  $\hat{\theta} = \arg \min_{\theta \in \tilde{\Theta}} \{\max(l, m, n)\}$ .

The optimality criteria C1–C4 are extensions of Lin and Shih’s<sup>4</sup> criteria and, if there is only one target response, criteria C1 and C2 are exactly the same as Simon’s two optimality criteria. Our proposed extended two-stage adaptive design is schematically described in Figure 1.

The above extension to three prefixed possible target responses may appear straightforward to implement, but in reality, it is computationally impractical to find a solution by the use of greedy algorithms due to its complexity. Indeed, we formulate the optimization problem and show that greedy algorithms fail to find a solution for our 10-dimensional optimization problem after running it for 30 days. To this end, we briefly review a nature-inspired



**Figure 1.** Flowchart of the extended two-stage adaptive design for Phase II clinical trials (third column) with  $s_1 < r_1 < q_1$ . Simon’s two-stage design is obtained by setting  $r_1 = n_1$  (first column) and Lin and Shih’s design is obtained by setting  $q_1 = n_1$  (second column).

stochastic population-based algorithm, PSO technique, and show that it not only provides solutions in a relatively short time but also facilitates performance comparison among different algorithms in the following sections.

### 3 DPSO

PSO, a population-based global optimization method, was introduced by Kennedy and Eberhart.<sup>16</sup> It is an evolutionary algorithm and stochastically evolves a group of particles that mimic observational behavior from nature. PSO is motivated from observing how a flock of birds move in the sky and so is a member of the class of nature-inspired stochastic population-based algorithms. This class of algorithms has been gaining lots of recognition especially in the last decade or so for its ability to solve or nearly solve hard-to-optimize high dimensional problems in the real world. Whitacre<sup>17,18</sup> documented the meteoric rise in use of such algorithms in applied fields and increasingly in academia as well. Their main appeals are that they are simple to implement, assumption-free and tend to quickly converge to the optimum or get quickly to the vicinity of the optimum solution. Additionally, they are general-purpose optimization algorithms and so are adaptable to solve different types of optimization problems after the user inputs some tuning parameters to initiate the algorithm. Genetic algorithm, simulated annealing, and PSO are examples of such algorithms.

Some recent applications of PSO to solve optimal design problems in statistics are in literature.<sup>19–22,23</sup> They tackled design problems in biomedical problems that ranged from finding D-optimal designs for several nonlinear models and mixture models with multiple constraints to finding minimax and standardized maximin type of optimal designs. The latter designs have non-differentiable optimality criteria that require a couple of nested levels of optimization and are notoriously difficult to find. Phoa et al.<sup>21</sup> applied swarm intelligence to find an optimal supersaturated design in a high dimension problem that involves judicious and repeated exchanges of columns in the design matrix to minimize correlations among the columns via the  $E(s^2)$  criterion. PSO has also successfully used for estimation in statistical problems. For example, Kim and Li<sup>24,25</sup> applied PSO to estimate parameters in nonlinear mixed-effects pharmacokinetics models, and Kim et al.<sup>26</sup> employed PSO to estimate efficacy of different lung cancer screening methods. In each of the above problems, we observed, as many others had, the flexibility of PSO and how it can be readily modified to solve an optimization problem at hand. In what is to follow, we first demonstrate yet again that PSO can be adapted to efficiently solve our adaptive design problem with a very different setup than those just mentioned above. Second, we show PSO can solve high dimensional problems with many parameters in our adaptive design problems that current greedy algorithms cannot and third, PSO can substantially outperform greedy algorithms in adaptive design problems with a small number of parameters.

Operationally, the user initiates PSO by first specifying the maximum of iterations allowed, say  $K$  and the flock size consisting of, say  $N$ , randomly generated particles for the search. Each particle represents a candidate solution to the problem and at any one time, each has a fitness value (i.e. the design criterion value). As it searches, it keeps track of its best value (personal or local best value) and communicates with the rest of the particles to determine the global best value, defined as the best personal best value among the flock up to that iteration. The particles updates these two values continuously as it moves across the search space in the direction guided by its most recent local best and the global best with a velocity determined stochastically by a combination of values of the tuning parameters, its personal best and the global best and the velocity with which it has arrived at the current position. Specifically, suppose the  $n$ th particle's position vector is  $x^n = (x_k^n)_{k=1, \dots, K}$  and its updating velocity vector is  $v^n = (v_k^n)_{k=1, \dots, K}$ ,  $n = 1, \dots, N$ . If its local best and global best are  $x_{lbest}^n$  and  $x_{gbest}$ , respectively, its velocity  $v_{k+1}^n$  and the position  $x_{k+1}^n$  at the  $(k+1)$ th iteration are determined by

$$v_{k+1}^n = w_k v_k^n + c_1 r_1 (x_{lbest}^n - x_k^n) + c_2 r_2 (x_{gbest} - x_k^n); \quad x_{k+1}^n = x_k^n + v_{k+1}^n \quad (4)$$

Here  $w_k$  is the inertia weight ( $0 \leq w_k \leq 1$ ) at the  $k$ th iteration,  $c_1$  and  $c_2$  are two positive constants called cognitive and social coefficient, and  $r_1$  and  $r_2$  are two random variates in the range  $[0, 1]$ . The lower limit values of the constants  $c_1$  and  $c_2$  determine how far particles are allowed to wander beyond the target region before pulled back, and their upper limit values determine how sudden the particles should be moved back to the target region. Following convention, we set  $c_1$  and  $c_2$  to their default values equal to 2 in our simulation studies. At the  $k$ th iteration, the inertia weight  $w_k$  is  $w_k = w_{max} - \frac{k}{K}(w_{max} - w_{min})$ , where  $w_{min}$  and  $w_{max}$  are user-defined constants satisfying  $0 \leq w_K = w_{min} \leq w_k \leq w_{max} = w_0 \leq 1$ .

During the search, the inertia weight adaptively controls the impact of the previous history of velocities on the current velocity and also influences the trade-off between global (wide-ranging) and local (nearby) exploration



abilities of the particles, as they move across the search space. A larger inertia weight facilitates global exploration (searching new areas) and a smaller inertia weight facilitates local exploration. When the inertia weight is suitably chosen, it can provide a balance between global and local exploration abilities and on average, require fewer iterations to find the global optimum.<sup>27</sup> To exploit these properties of the inertia weight, we use a dynamic inertia weight to enable PSO escape from premature convergence when it is stagnated.<sup>28</sup>

PSO was originally designed to solve optimization problems over a continuous domain, not optimization over a domain comprising of discrete positive integer numbers. Since PSO is a flexible algorithm, we modified it for our problem and call it DPSO. The task is to choose the right combination of several positive integers that would minimize the total sample size required under various hypothesis and, at the same time, meet the user-specified Type I and II error rates. The proposed DPSO has two different features from PSO: (a) the inertia weight  $w'_k$  is the nearest integer of PSO's inertia weight  $w_k$  at the  $k$ th iteration, i.e.  $w'_k = \lfloor w_k + \frac{1}{2} \rfloor = \lfloor w_{max} - \frac{k}{K}(w_{max} - w_{min}) + \frac{1}{2} \rfloor$ , where  $\lfloor x \rfloor = \arg \max_m \{m \in Z | m \leq x\}$  and  $Z$  is a set of integers, and (b) the two random sequences are generated from a discrete uniform distribution.

### 3.1 Extended two-stage adaptive design with three target responses

We consider two optimization approaches to find  $\hat{\theta}$ , the optimal value for  $\theta = (s_1, r_1, q_1, n_1, s, l, r, m, q, n)$  for our adaptive design problem. One is the greedy search and the other is DPSO proposed in this study. The greedy search adopts the strategies used in Simon's or Lin and Shih's<sup>4</sup> algorithms for the extended two-stage adaptive design. Specifically, we first calculate the required sample sizes  $n_i, i = l, m, \text{ and } n$ , for a single-stage design, where  $n_l, n_m, \text{ and } n_n$  are the required sample sizes corresponding to the target response rates  $p_1, p_2, \text{ and } p_3$ , respectively, and constrain  $l, m, \text{ and } n$  to be in the range of  $\lfloor 0.85 \cdot n_i \rfloor \leq i \leq \lfloor 1.5 \cdot n_i + 1 \rfloor$ , where  $i = l, m, \text{ and } n$ . For each value of  $l, m, \text{ and } n$ , and each of  $n_1$  in  $[1, \min(l, m, n) - 1]$ ,  $s_1$  in  $[0, n_1]$ ,  $r_1$  in  $[s_1 + 1, n_1]$ , and  $q_1$  in  $[r_1 + 1, n_1]$ , we then find the set of feasible solutions  $\tilde{\Theta}$  that satisfies the error constraints (3) for  $q$  in  $[q_1 + 1, n]$ ,  $r$  in  $[r_1 + 1, m]$ , and  $s$  in  $[s_1 + 1, l]$  and determine the optimum according to one of optimality criteria C1–C4.

To implement DPSO, we specify the lower and upper boundaries for each component of the parameter vector  $\theta = (s_1, r_1, q_1, n_1, s, l, r, m, q, n)$ . Similar to the greedy search, we use the sample sizes  $n_l, n_m, \text{ and } n_n$ , required for a single-stage design corresponding to the target response rates  $p_1, p_2, \text{ and } p_3$ , respectively and set

$$\begin{aligned} L_i &= \lfloor 0.85 \cdot n_i \rfloor, \quad i = l, m, n; \\ U_i &= \lfloor 1.5 \cdot n_i + 1 \rfloor, \quad i = l, m, n; \\ U &= \min\{U_i; i = l, m, n\}. \end{aligned} \tag{5}$$

The lower and upper boundaries for each component of  $\theta$  are then obtained as follows:

$$\begin{aligned} s_1 &\in \left[ 0, \left\lfloor \frac{U_m}{2} + 1 \right\rfloor \right]; \quad r_1 \in [0, \lfloor U \cdot p_1 + 1 \rfloor]; \quad q_1 \in [0, \lfloor U \cdot p_2 + 1 \rfloor]; \quad n_1 \in [1, U]; \\ s &\in [0, \lfloor U_l \cdot p_1 + 1 \rfloor]; \quad l \in [L_l, U_l]; \quad r \in [0, \lfloor U_m \cdot p_2 + 1 \rfloor]; \quad m \in [L_m, U_m]; \\ q &\in [0, \lfloor U_n \cdot p_3 + 1 \rfloor]; \quad n \in [L_n, U_n], \end{aligned}$$

and used in DPSO to find the optimal solution under each of the C1–C4 optimality criteria among the feasible solutions in  $\tilde{\Theta}$  that satisfy the error constraints (3).

The standard PSO is usually not sensitive to the initial values of the tuning parameters and can find an optimal solution using the default values. However, in our DPSO, a lot of iterations is required for convergence, especially when there are two or three target responses and the initial values were not among the feasible solutions in  $\tilde{\Theta}$ . For this reason, we consider two approaches (G-DPSO and D-DPSO) to find an appropriate initial set of values among the feasible solutions in  $\tilde{\Theta}$  to boost the speed of convergence of DPSO.

In the first approach, the initial value is found by a greedy search over a smaller range of values for  $n_i$  and not over the entire range, where  $i = l, m, \text{ and } n$ . Specifically, the required sample sizes  $n_i, i = l, m, \text{ and } n$ , are in the range of  $n_i - 1 \leq i \leq n_i + 1$ , where  $i = l, m, \text{ and } n$ . Within this smaller domain, we searched for an appropriate set of initial values using the same strategy as the greedy search did for the rest of the parameters. By shrinking the range of the parameters  $n_i, i = l, m, \text{ and } n$ , we potentially reduced the computational burden. We call DPSO with this first approach G-DPSO. In particular, we observed the consistency of the computation time with a short amount of time, when G-DPSO was applied to the case with one target response (see Table 1 and Supplementary Information Table S1). However, similar to the greedy search, the computation time of G-DPSO increases as the number of

**Table 1.** Various adaptive two-stage optimal designs with one target response when  $\alpha = 0.05$  and  $\beta = 0.20$ .

$p_0$	$p_1$	Optimal criteria	Method	$s_1/n_1$	$s/n$	$1 - \alpha$	$\beta$	$E(N p_0)$	$E(N p_1)$
0.05	0.20	C1	GS	0/10	3/29	0.953	0.199	17.624	26.960
			G-DPSO	0/10	3/29	0.953	0.199	17.624	26.960
		C2	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
		C3	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
		C4	GS	0/11	3/28	0.956	0.199	18.330	26.540
			G-DPSO	0/11	3/28	0.956	0.199	18.330	26.540
0.20	0.35	C1	GS	5/22	19/72	0.951	0.200	35.368	63.855
			G-DPSO	5/22	19/72	0.951	0.200	35.368	63.855
		C2	GS	3/21	15/53	0.950	0.200	41.148	51.941
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941
		C3	GS	6/31	15/53	0.950	0.198	40.436	51.983
			G-DPSO	6/31	15/53	0.950	0.198	40.436	51.983
		C4	GS	3/21	15/53	0.950	0.200	41.148	51.941
			G-DPSO	3/21	15/53	0.950	0.200	41.148	51.941
0.55	0.70	C1	GS	15/26	48/76	0.952	0.195	42.021	69.735
			G-DPSO	15/26	48/76	0.952	0.195	42.021	69.735
		C2	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO <sup>a</sup>	20/35	43/67	0.953	0.200	45.802	64.662
		C3	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO	20/35	43/67	0.953	0.200	45.802	64.662
		C4	GS	20/35	43/67	0.953	0.200	45.802	64.662
			G-DPSO <sup>b</sup>	20/35	43/67	0.953	0.200	45.802	64.662

a,b: The number of particles was increased from 10 to 20 and the population size was increased from 10,000 to 70,000.

target response  $s$  increases, especially for two or three target responses (for example, see Tables 2 and 3 and Supplementary Information Tables S2 and S3). Therefore, we devise the second approach that does not depend on the number of target responses for the cases with two or three target responses.

The second approach abbreviated as D-DPSO is used when the number of target responses is two or more. D-DPSO finds the initial set of values using an optimal set of values decided by the case with the one less number of target response. In other words, when the number of target responses is  $k$ , the initial value is decided using the sample sizes determined by the case when the number of target responses is  $k - 1$ , where  $k = 2, 3$ . By doing so, we could make the second approach not depend on the number of target responses in terms of computation time. The specific details for implementing D-DPSO to search for any one of the four types of optimal design when there are two and three target responses are as follows.

**3.1.1 The number of target responses is two**

We first obtain two sets of sample sizes,  $E_1 = \{s_1^1, n_1^1, s^1, n^1\}$  and  $E_2 = \{s_1^2, n_1^2, s^2, n^2\}$ , from the target responses  $p_1$  and  $p_2$ , respectively, based on the G-DPSO method when there is one target response. Then, using the two sets of sample sizes  $E_1$  and  $E_2$ , the initial value is set to  $\tilde{\theta}_2 = (\tilde{s}_1, \tilde{r}_1, \tilde{m}_1, \tilde{s}, \tilde{m}, \tilde{r}, \tilde{n}) = (\text{rank}_3\{s_1^1, n_1^1, s_1^2, n_1^2\}, s^1, n^1, s^2, n^2)$ , where  $\text{rank}_k S$  is the top  $k$  elements of the set  $S$  arranged in ascending order. The first three values ( $\tilde{s}_1, \tilde{r}_1, \tilde{m}_1$ ) correspond to the null response  $p_0$  (i.e. the first stage), so the sample sizes ( $s_1^1, n_1^1, s_1^2, n_1^2$ ), which correspond to the first stage of  $E_1$  and  $E_2$ , are used. Since ( $\tilde{s}, \tilde{m}$ ) and ( $\tilde{r}, \tilde{n}$ ) correspond to the target values  $p_1$  and  $p_2$  (i.e. the second stage), respectively, the values ( $s^1, n^1$ ) and ( $s^2, n^2$ ), which correspond to the second stage of  $E_1$  and  $E_2$ , are used.

**3.1.2 The number of target responses is three**

We first obtain two sets of sample sizes  $E_1 = \{s_1^1, r_1^1, m_1^1, s^1, m^1, r^1, n^1\}$  and  $E_2 = \{s_1^2, n_1^2, s^2, n^2\}$  using the D-DPSO with the target responses ( $p_1, p_2$ ) and the G-DPSO with the target response  $p_3$ , respectively. Then, the initial value



is set to  $\tilde{\theta}_3 = (\tilde{s}_1, \tilde{r}_1, \tilde{q}_1, \tilde{n}_1, \tilde{s}, \tilde{l}, \tilde{r}, \tilde{m}, \tilde{q}, \tilde{n}) = (\text{rank}_4\{s_1^1, r_1^1, m_1^1, s_1^2, n_1^2\}, s^1, m^1, r^1, n^1, s^2, n^2)$ . The first four values  $(\tilde{s}_1, \tilde{r}_1, \tilde{q}_1, \tilde{n}_1)$  correspond to the null response  $p_0$  (i.e. the first stage), so the sample sizes  $(s_1^1, r_1^1, m_1^1, s_1^2, n_1^2)$ , which correspond to the first stage of  $E_1$  and  $E_2$ , are used. Since  $(\tilde{s}, \tilde{l}, \tilde{r}, \tilde{m})$  and  $(\tilde{q}, \tilde{n})$  correspond to the target values  $(p_1, p_2)$  and  $p_3$  (i.e. the second stage), respectively, the values  $(s^1, m^1, r^1, n^1)$  and  $(s^2, n^2)$ , which correspond to the second stage of  $E_1$  and  $E_2$ , are used.

## 4 Simulation studies

We use simulation studies to evaluate the performance of extended two-stage adaptive designs. We also generate Simon's two-stage design and Lin and Shih's<sup>4</sup> adaptive two-stage design using DPSO and compare results when the number of target response is one and two, respectively. We refer to Simon's two-stage design as an adaptive two-stage design with one target response, Lin and Shih's<sup>4</sup> adaptive two-stage design as an adaptive two-stage design with two target responses, and our extended two-stage adaptive design as an adaptive two-stage design with three target responses.

Our simulation covers one set of Type I ( $\alpha$ ) and Type II ( $\beta$ ) error rates with  $(\alpha, \beta_1, \beta_2, \beta_3) = (0.05, 0.20, 0.10, 0.05)$  and three sets of target response rates, which are  $(p_0, p_1, p_2, p_3) \in \{(0.05, 0.20, 0.25, 0.30), (0.20, 0.35, 0.40, 0.45), (0.55, 0.70, 0.75, 0.80)\}$ . Note that we empirically selected these combinations for null and target response rates to reflect the general performances of each method. The optimal design under each of the four optimality criteria C1–C4 was found using two optimization methods: a greedy search (GS) and DPSO. GS looks for all possible and feasible cases to find the optimal solution, while DPSO stochastically searches for the optimal solution using equations (1) and (2). The number of particles used to optimize each design parameter was 10 and the population size was set to 10,000 in our DPSO. The total number of iterations allowed was 50. The two methods described in the previous section were applied to obtain initial values for G-DPSO and D-DPSO.

Tables 1–3 display optimal designs found from GS, G-DPSO and D-DPSO methods when there are one, two, and three target responses, respectively. Details of the computation times in minutes to evaluate computational burden for the different algorithms are available in the Supplementary Information Tables S1–S3. Note that the computation times given in Supplementary Information Tables S1 to S3 might not accurately reflect the true computational complexity of each method and we display them to give some insights on how efficient each method is. All simulations were rendered up to 30 days using a high performance Grid enabled computing system at Wayne State University. The Grid currently has the combined processing power of 4420 cores: 1200 Intel cores, 3220 AMD cores, with over 12TB of RAM and over half a petabyte of disk space.

We make two remarks before we present our results. First, to ensure a fair comparison, we implemented all the GS, G-DPSO, and D-DPSO algorithms using the statistical software R. Some of our G-DPSO results are a bit different from published results found by the GS method. This is not surprising because of the stochastic nature of the G-DPSO algorithm and the results from G- or D-DPSO depend on the algorithm convergence criterion and the tuning parameters, which for simplicity, we had set a priori to be same for all cases. Consequently, optimal designs found by the G- and D-DPSO algorithm may vary each time we ran. For instance, when  $p_0 = 0.55$  and  $p_1 = 0.70$ , our simulation settings for PSO produced different results for criteria C2 and C4 than those reported in Simon's article found by a greedy search. Some such discrepancies are indicated in Table 1 by superscripts a and b, but they can be resolved using a different set of tuning parameters for the algorithm. For example, in these instances, when the number of particles was increased from 10 to 20 and the population size was increased from 10,000 to 70,000, our DPSO algorithm converged to the same optimal design with  $s_1/n_1 = 20/35$  obtained in Simon's article in a small fraction of the time required by GS. Second, our codes repeatedly call on IMS functions for computing binomial probabilities every single time it is needed, which is unlike the codes in Simon's article where they were predefined upfront. Consequently, our CPU times may appear several factors longer than those reported in earlier work, but are in fact substantially faster than the computing time required by GS.

### 4.1 Adaptive two-stage design with one target response

When there is one target response rate, the adaptive two-stage design is the same as Simon's two-stage design. Table 1 compares the results between GS and G-DPSO when  $p_1 - p_0 = 0.15$ . As stated before, we applied D-DPSO only when the number of target response is either two or three. In all cases, both optimization methods give the exactly same optimal design with the same expected sample sizes. Table S1 in the Supplementary Information shows G-DPSO can be up to 100 times faster than GS except when  $p_0 = 0.05$ . One reason that G-DPSO is slower than GS is that the size of the solution domain when  $p_0 = 0.05$  is small and

**Table 2.** Various adaptive two-stage optimal designs with two target responses when  $\alpha = 0.05$ ,  $\beta_1 = 0.20$  and  $\beta_2 = 0.10$ .

$p_0$	$p_1$	$p_2$	Optimal criteria	Method	$s_1/r_1/n_1$	$s/m$	$r/n$	$1-\alpha$	$\beta_1$	$\beta_2$	$E(N p_0)$	$E(N p_1)$	$E(N p_2)$			
0.05	0.20	0.25	C1	GS	0/1/9	3/30	4/41	0.957	0.199	0.094	17.548	33.383	36.119			
				G-DPSO	0/1/10	3/28	3/31	0.953	0.199	0.088	17.481	27.940	29.254			
				D-DPSO	0/1/8	4/41	3/36	0.952	0.200	0.107	18.821	32.980	34.532			
			C2	GS	0/2/16	3/29	3/22	0.955	0.199	0.082	22.978	24.097	23.249			
				G-DPSO	0/2/14	3/29	3/21	0.954	0.200	0.084	21.444	23.925	22.982			
				D-DPSO	0/2/14	3/29	3/21	0.954	0.200	0.084	21.444	23.925	22.982			
			C3	GS	0/4/9	4/41	5/22	0.960	0.174	0.084	20.831	36.333	37.668			
				G-DPSO	0/5/11	3/28	6/25	0.956	0.199	0.083	18.330	26.505	27.179			
				D-DPSO	0/5/11	3/28	6/25	0.956	0.199	0.083	18.330	26.505	27.179			
			C4	GS	0/2/16	3/29	3/22	0.955	0.199	0.082	22.978	24.097	23.249			
				G-DPSO	0/2/15	3/28	3/23	0.957	0.197	0.078	21.796	24.533	24.007			
				D-DPSO	0/2/15	3/28	3/23	0.957	0.197	0.078	21.796	24.533	24.007			
0.20	0.35	0.40	C1	G-DPSO	4/9/20	17/62	10/36	0.952	0.199	0.069	35.487	53.869	53.499			
				D-DPSO	5/10/22	19/72	11/36	0.951	0.199	0.078	35.311	60.004	60.179			
			C2	G-DPSO	8/12/37	16/57	13/41	0.950	0.196	0.059	42.913	46.949	44.252			
				D-DPSO	5/10/29	16/56	11/36	0.951	0.196	0.055	43.094	46.415	42.638			
			C3	G-DPSO	6/11/27	16/58	13/43	0.950	0.200	0.063	35.833	51.406	50.886			
				D-DPSO	3/10/21	15/53	11/36	0.950	0.200	0.058	41.131	50.629	49.683			
			C4	G-DPSO	8/12/38	16/56	13/43	0.952	0.200	0.057	43.830	47.337	45.224			
				D-DPSO	3/10/21	15/53	11/36	0.950	0.200	0.058	41.131	50.629	49.683			
			0.55	0.70	0.75	C1	G-DPSO	15/20/26	48/76	27/39	0.951	0.200	0.053	41.807	63.720	61.521
							D-DPSO	15/20/26	48/76	27/39	0.951	0.200	0.053	41.807	63.720	61.521
						C2	G-DPSO	24/28/41	47/73	30/45	0.951	0.194	0.039	48.872	55.463	50.268
							D-DPSO	11/17/24	45/70	30/45	0.951	0.199	0.043	57.956	59.756	54.718
C3	G-DPSO	13/20/25				43/67	25/42	0.951	0.194	0.038	47.730	62.880	61.206			
	D-DPSO	14/20/26				43/67	23/39	0.950	0.199	0.041	45.163	59.975	56.926			
C4	G-DPSO	25/30/44				44/68	32/49	0.950	0.197	0.037	51.862	56.540	52.457			
	D-DPSO	10/16/21				43/67	28/41	0.950	0.200	0.044	51.907	60.628	57.152			

the computational time of GS can be highly dependent on the size of the solution domain. As such, the computational time required by GS can be highly variable. In contrast, the time required by G-DPSO to find the optimum does not depend on the size of the solution set once the initial values are obtained. Consequently, we observe a consistent computational time range required by DPSO for all the cases. The table also shows that the sample size requirements for optimality criteria C2–C4 are smaller than those required by criterion C1.

### 4.2 Adaptive two-stage design with two target responses

We now apply GS and DPSO (G-DPSO and D-DPSO) to generate the Lin and Shih’s<sup>4</sup> adaptive two-stage designs and Table 2 displays them. Each optimization was allowed to run up to 30 days. GS reports the optimal designs only when  $p_0 = 0.05$ , because the size of the solution area increases exponentially with the number of target responses when  $p_0$  is close to 0.5. As a result, GS could not reach out to the solutions for other cases. Interestingly, as shown in Supplementary Information Table S2, the computation time of GS when  $p_0 = 0.05$  is much longer (at least 68 times longer) than those of both G-DPSO and D-DPSO, even though the determined sample sizes are comparable with or slightly smaller than those of G-DPSO and D-DPSO. As expected, the required sample sizes of G-DPSO and D-DPSO are similar even though the computation time for G-DPSO and D-DPSO are different; the former depends on the size of the solution set but the latter does not, hence their computing times are different. For example, when  $(p_0, p_1, p_2) = (0.55, 0.70, 0.75)$ , G-DPSO takes 23 times longer than D-DPSO to determine the optimal design, but when the size of the solution set is small as when  $(p_0, p_1, p_2) = (0.05, 0.20, 0.25)$ , the computation time required by G-DPSO is shorter than that required by D-DPSO, which consistently ranges from 5.07 to 6.10 min and independent of the solution size.

**Table 3.** Various adaptive two-stage optimal designs with three target responses when  $\alpha = 0.05$ ,  $\beta_1 = 0.20$ ,  $\beta_2 = 0.10$ , and  $\beta_3 = 0.05$ .

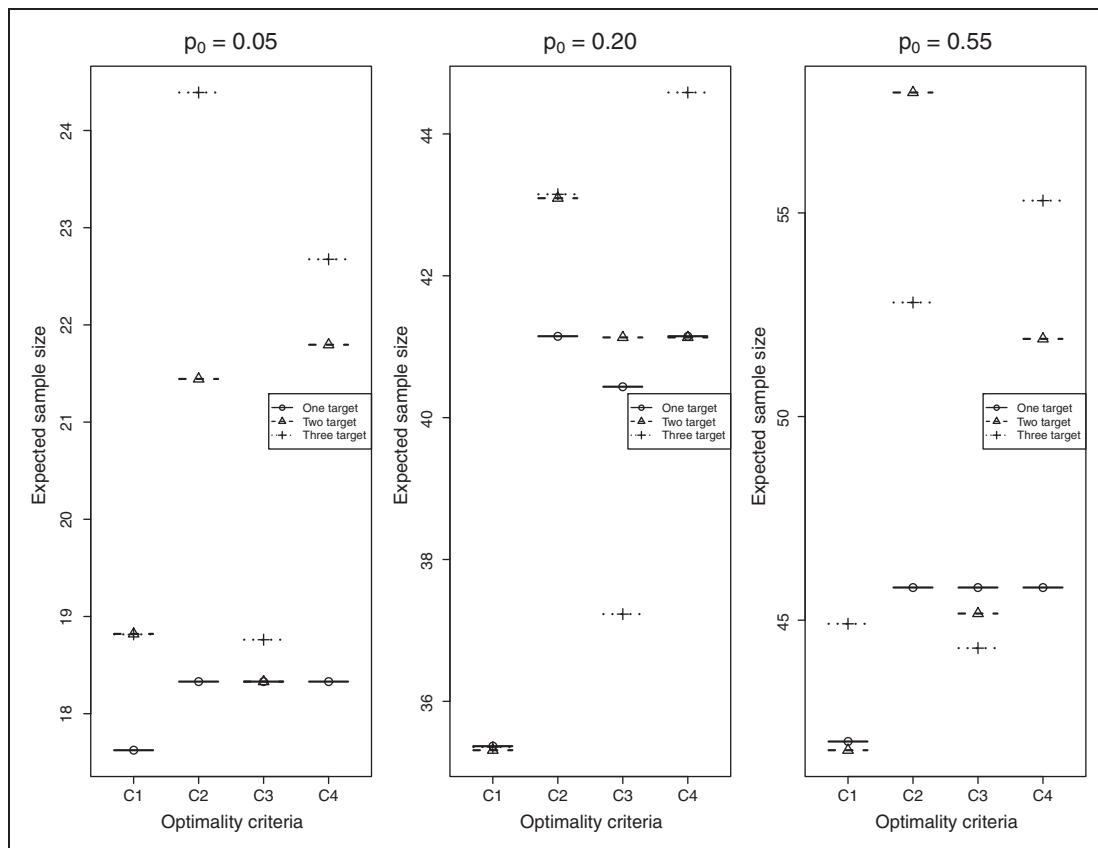
$p_0$	$p_1$	$p_2$	$p_3$	Optimal criteria	Method	$s_1/r_1/q_1/n_1$	$s/l$	$r/m$	$q/n$	$1-\alpha$	$\beta_1$	$\beta_2$	$\beta_3$	$E(N p_0)$	$E(N p_1)$	$E(N p_2)$	$E(N p_3)$		
0.05	0.20	0.25	0.30	C1	G-DPSO	0/1/4/10	3/28	3/31	5/28	0.953	0.200	0.088	0.037	17.481	27.841	29.020	29.593		
				C2	D-DPSO	0/1/2/9	4/36	3/34	3/31	3/31	0.959	0.199	0.093	0.044	18.816	30.463	31.375	31.691	
				C3	G-DPSO	0/1/2/13	3/31	3/28	3/20	3/20	0.951	0.200	0.088	0.036	21.158	23.725	22.613	21.636	
				C4	D-DPSO	0/1/2/14	4/36	3/33	3/19	3/19	0.952	0.198	0.094	0.043	24.391	24.899	22.847	21.245	
	0.20	0.35	0.40	0.45	C1	G-DPSO	5/9/10/24	17/61	10/36	11/32	0.952	0.200	0.064	0.017	36.401	48.570	45.349	40.823	
					C2	D-DPSO	5/11/13/22	19/72	12/36	14/32	14/32	0.951	0.200	0.078	0.028	35.355	62.126	63.957	61.554
					C3	G-DPSO	3/6/8/19	19/67	11/38	9/31	9/31	0.950	0.194	0.060	0.016	43.149	47.817	43.504	39.094
					C4	D-DPSO	3/8/9/17	17/62	12/36	11/28	11/28	0.951	0.196	0.059	0.014	37.889	50.724	49.244	46.353
					C1	G-DPSO	4/8/11/24	18/64	11/36	12/28	12/28	0.952	0.193	0.068	0.026	44.584	48.279	42.940	37.622
					C2	D-DPSO	15/20/21/26	48/76	29/43	23/31	23/31	0.951	0.199	0.052	0.010	41.812	63.491	60.658	51.947
					C3	G-DPSO	17/22/23/32	47/73	32/48	24/35	24/35	0.950	0.195	0.041	0.005	51.994	62.696	60.681	54.260
					C4	D-DPSO	8/13/14/17	46/72	27/39	21/31	21/31	0.951	0.195	0.048	0.009	52.800	62.503	46.623	39.493
0.55	0.70	0.75	0.80	C1	G-DPSO	15/20/21/27	46/72	27/41	24/35	0.951	0.192	0.041	0.005	44.743	59.648	54.638	46.497		
				C2	D-DPSO	12/17/18/22	47/74	26/39	24/32	24/32	0.950	0.197	0.054	0.014	44.314	63.001	60.007	52.356	
				C3	G-DPSO	17/22/23/32	47/73	32/48	24/35	24/35	0.950	0.195	0.041	0.005	51.994	62.696	46.623	39.493	
				C4	D-DPSO	7/12/14/16	45/70	26/39	19/31	19/31	0.951	0.200	0.044	0.006	55.302	60.783	56.534	50.252	

### 4.3 Adaptive two-stage design with three target responses

Table 3 displays the extended two-stage adaptive designs with three target response rates. Within 30 days run, GS always failed to determine the optimal design, but both G-DPSO and D-DPSO were able to find the optimal designs with similar results. Interestingly, when  $(p_0, p_1, p_2, p_3) = (0.55, 0.70, 0.75, 0.80)$ , the computation time of G-DPSO is almost 30 days, but that of D-DPSO is about 6.5 min, demonstrating that D-DPSO is much more time efficient than G-DPSO even in complicated situations and computation burden, see Table S3 in the Supplementary Information.

At the request of an anonymous reviewer, we also performed another simulation with  $\alpha = 0.01$ . Tables S4 to S6 in the Supplementary Information display the results. Due to the expensive computation time, we included the results only from D-DPSO when there are three target responses. The overall performances of each method with  $\alpha = 0.01$  are close to the case when  $\alpha = 0.05$ .

Figure 2 compares expected sample size requirements from the three types of two-stage designs discussed here for criteria C1–C4. We considered three testing scenarios with various hypothesized efficacy rates for the drug in the designs and investigated which method estimates the smallest sample size among the three types of two-stage designs. Interestingly, we observe that among the three types of adaptive designs, there is no clear winner that consistently requires the smallest expected sample sizes. We further compared expected sample size requirements under various alternative hypotheses shown in Figures S1 to S3 in the Supplementary Information. We observe that the adaptive designs with two or three targets are comparable similar to the null hypothesis, but the two-stage design with one target requires the largest sample sizes. Our proposed extended two-stage adaptive designs are very flexible and yet surprisingly do not necessarily require larger expected sample sizes than the other two types of adaptive designs and sometimes smaller expected sample sizes, especially for criterion 3. Even when our proposed



**Figure 2.** Expected sample sizes under the null hypothesis for the four criteria C1–C4 with 1, 2, or 3 target alternatives estimated by D-DPSO for three scenarios (from left to right): (i)  $p_0 = 0.05, p_1 = 0.20, p_2 = 0.25, p_3 = 0.30$ , (ii)  $p_0 = 0.20, p_1 = 0.35, p_2 = 0.40, p_3 = 0.45$ , and (iii)  $p_0 = 0.55, p_1 = 0.70, p_2 = 0.75, p_3 = 0.80$ . Error rates were set at  $\alpha = 0.05, \beta_1 = 0.20, \beta_2 = 0.10, \text{ and } \beta_3 = 0.05$ .

designs have larger expected sample size requirements, the additional numbers may not be large and seem manageable. Clearly, compounding the comparison problem is that the expected sample sizes also crucially depend on the various hypothesized rates in a complicated manner. The summary here is that our proposed designs offer practitioners a good alternative, especially when it is difficult to ascertain the drug efficacy rate at the onset.

## 5 Applications

### 5.1 OSA study

Our clients did not allow us to discuss the entire nature of their study and so we minimally describe the study design. This study aims to assess the effect of HNC on the incidence of OSA compared to healthy patients. The literature suggests the maximum incidence rate of snoring and sleep apnea on healthy patients is 16.5%, resulting in the null hypothesis of 16.5% (i.e.  $p_0 = 0.165$ ). There was neither historical nor preliminary data available, except that the incidence rate of OSA will be higher in HNC patients. Our clients provided an empirical range of the target response rates, which is from 24.38% to 39.00%. If Simon's two-stage design were used for sample size calculation with 80% power and 5% significance level, the required sample sizes range from 30 to 197, using the optimality and not the minimax criterion. This shows the impact of the great uncertainty of the OSA incidence rate on HNC patients has on the sample size requirements. So there is high risk of under or over powering study if one mis-specifies the response rate. Due to wide range of the target response rates, Lin and Shih's approach will not be able to cover the great uncertainty. For our application, one may use the extreme limits in the range for HNC patients as two target rates in the alternative hypotheses with the third that targets the midpoint between the extreme limits in the presumed range. Table 4 shows the D-DPSO generated sample size requirements for this scenario with three target responses ( $p_1 = 24.38$ ,  $p_2 = 31.69\%$ , and  $p_3 = 39.00\%$ ) and various error rates specifications ( $\beta_1 = 0.20$ ,  $\beta_2 = 0.15$ , and  $\beta_3 = 0.10$ ).

### 5.2 Lin and Shih's VBG study

This study is adapted from Lin and Shih's work and is to investigate the efficacy of the combination therapy of VBG on patients with recurrent or refractory HD. The null hypothesis was set at 40% based on the single agents' response rate, but the target response rates ranges from 0.5 to 0.6. So, Lin and Shih applied their design to cover this uncertainty with two target response rates,  $p_1 = 55\%$  and  $p_2 = 60\%$ , under  $\beta_1 = 0.20$  and  $\beta_2 = 0.15$  at a 5% significance level. Using our proposed design, three target responses  $p_1 = 50\%$ ,  $p_2 = 55\%$ , and  $p_3 = 60\%$  are considered with  $\beta_1 = 20\%$ ,  $\beta_2 = 15\%$ , and  $\beta_3 = 5\%$  and the optimal designs estimated are available in Table 4. The D-DPSO's required sample size in stage 1 (49 to 54) is a little larger than those of Lin and Shih's (39), because the smallest target rate in D-DPSO ( $p_1 = 0.50$ ) is smaller than that in Lin and Shih's ( $p_1 = 0.55$ ). However, the required sample sizes of both approaches in stage 2 are comparable to each other.

**Table 4.** Sample sizes from D-DPSO for the four optimality criteria for OSA, Lin and Shih's VBG, and the Phase II BREAK-2 studies.

Optimality criteria	$s_1/r_1/q_1/n_1$	$s/l$	$r/m$	$q/n$	$E(N p_0)$	$E(N p_1)$	$E(N p_2)$	$E(N p_3)$
OSA								
C1	2/8/9/21	39/188	13/55	10/39	136.92	167.28	157.91	124.73
C2 and C4	1/11/12/25	34/161	20/72	13/36	152.07	158.99	154.07	135.42
C3	1/8/9/17	37/177	12/58	10/35	144.40	166.73	167.70	153.87
Lin and Shih's VBG								
C1	20/29/35/49	83/183	33/75	39/77	101.45	157.83	153.53	129.21
C2	20/28/29/54	81/175	42/87	31/65	125.76	134.87	107.31	81.83
C3	20/31/33/51	77/168	40/76	37/68	107.64	154.27	150.26	129.11
C4	19/28/31/54	79/170	39/83	34/66	134.72	136.06	109.57	84.89
BREAK-2								
C1	4/10/11/19	26/80	14/44	12/34	51.522	72.216	65.961	58.879
C2 and C4	4/9/12/22	27/83	16/43	13/31	62.079	65.691	50.196	43.063
C3	2/7/9/13	25/77	15/44	14/39	55.526	70.032	66.463	62.186



### 5.3 Phase II BREAK-2 study

We now use a real trial from the literature and discuss advantages of our proposed extended two-stage adaptive designs. In particular, we show our proposed optimal design can reduce the sample size used in the real trial by one-half if it were implemented.

A multicenter, international, single-arm, phase II study (BREAK-2) was carried out to assess the overall response rate of dabrafenib from patients with  $BRAF^{V600E}$  mutation-positive metastatic melanoma.<sup>29</sup> The null hypothesis was set at  $p_0 = 0.25$  and the alternative hypothesis was set at  $p_1 = 0.40$ . The trial wanted to recruit at least 85 patients and the plan was to declare the treatment a success if at least 29 patients responded. The Green-Dahlberg two-stage design<sup>30</sup> was employed with an interim analysis planned after the first 30 patients with  $BRAF^{V600E}$  mutation-positive metastatic melanoma were enrolled. The expectation was that at least seven patients were needed to respond at the interim analysis to continue the study when Type I and II error rates were set at 0.05 and 0.10, respectively. The study BREAK-2 is ongoing, but not recruiting participants as of 14 December 2015 (<http://www.clinicaltrials.gov>, NCT01153763). Green and Dahlberg's two-stage design (also known as Southwest Oncology Group [SWOG] two-stage design) is similar to Simon's two-stage design but uses a simple and uniform (i.e. fixed) significance to estimate the sample sizes for each stage with 0.02 level in stage 1 and 0.055 level in stage 2, resulting in less sensitive (or more robust) to the discrepancy between the actual sample size and the planned one. For example, as stated in Green and Dahlberg,<sup>30</sup> multicenter studies are not easy to control over the number of patients accrued. In addition, it often occurs for multicenter studies that some patients who entered the study are ineligible after the accrual is suspended. To address these potential issues, BREAK-2 used Green and Dahlberg's approach and, in fact, ended up a total of 76 patients, which differ from the planned sample size of 85. Under the design parameters, using Green and Dahlberg's approach and notation, the sample sizes for each stage are  $n_1 = 45$  and  $n_2 = 40$  (i.e.  $n = 85$ ). If Simon's two-stage design is used, the sample sizes are  $n_1 = 57$  and  $n_2 = 13$  (i.e.  $n = 83$ ) with the minimax constraint and  $n_1 = 37$  and  $n_2 = 10$  (i.e.  $n = 99$ ) with the optimal constraint.

The efficacy results show that 76 patients with  $BRAF^{V600E}$  mutation-positive metastatic melanoma were enrolled and 45 patients (59%) had a confirmed response. Although its parent phase I study<sup>31</sup> (<http://www.clinicaltrials.gov>, NCT00880321) showed the same type of patients had a response rate of 50%, this phase II study chose the response rate of 40% as an alternative hypothesis by lowering the response rate of phase I study. However, based on the phase I study, it would be of benefit if the higher response rate was explored in addition to 40% because the final response rate of the phase II study was 59%. Examining the benefit of applying the adaptive two-stage design, we determined the optimal design of this phase II study with  $p_0 = 0.25$ ,  $p_1 = 0.40$ ,  $p_2 = 0.50$ , and  $p_3 = 0.55$  and with  $\alpha = 0.05$ ,  $\beta_1 = 0.15$ ,  $\beta_2 = 0.10$ , and  $\beta_3 = 0.05$  using D-DPSO.

Table 4 shows the two-stage adaptive optimal designs found from D-DPSO. The same DPSO setting of parameters was used as the simulation studies did. Comparing to the original BREAK-2 study (i.e. when  $p_1 = 0.4$ ), all the sample sizes of the adaptive design ( $s/l$  in Table 4) are comparable but smaller than that of BREAK-2 (29/85), while the sample size of the stage 1 of BREAK-2 (7/30) is larger than those of the adaptive design ( $s_1/n_1$  in Table 4), although the Type II error of the adaptive design is larger than that of BREAK-2 (i.e. 0.15 vs. 0.10). Given the response rate of 59% (i.e. the final results of BREAK-2), the probabilities of not terminating the experiment at the end of the first stage are more than 99% for all three sample sizes in the table. The probabilities of choosing the sample sizes corresponding to either  $p_2 = 0.5$  or  $p_3 = 0.55$  in the second stage are 79%, 97%, and 75%, respectively, for C1, C2, and C4, and C3, given the response rate of 59%, implying that only a half of the planned number of patients would be required for the BREAK-2 study if the adaptive design were used. Furthermore, the probabilities of choosing the sample sizes corresponding to  $p_3 = 0.55$  in the second stage are 63%, 74%, and 33%, respectively, for C1, C2, and C4, and C3, given the response rate of 59%.

## 6 Conclusions

We proposed a novel and effective nature-inspired stochastic population-based algorithm called DPSO to find extended two-stage adaptive designs. This design terminates the trial if there are too few responders in the first stage and tests 1 of 3 preselected hypothesized response rates of a drug at the second stage, where the choice depends on the number of responders in the first stage. The design controls user-selected Type I error rate in the first stage and both Type I and II error rates at the second stage. Our results show that algorithms based on a greedy search invariably failed to find extended two-stage adaptive designs and an improved version of DPSO, called D-DPSO finds the optimum. When the problem is simplified to one or two target response rates, we also showed that D-DPSO outperformed their peers by a wide margin. Additionally, we employed simulation studies



and showed all D-DPSO generated designs were able to meet pre-specified Type I and II error requirements and demonstrated how such designs can reap benefits if they were implemented in a real melanoma adaptive trial.

### Author's Note

The author Weng Kee Wong is also the corresponding author of this article.

### Acknowledgements

We are grateful to Drs. Ammar Sukari and Misako Nagasaka for allowing us to use their OSA study as a real example.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported in this article was partially supported by two grants: NSF DMS-1312603 for SK and NIH R01GM107639 for WKW. The Biostatistics Core in the Karmanos Cancer Institute at Wayne State University is supported in part by the NIH Cancer Center Support Grant P30 CA022453. The contents in this article are solely the responsibility of the authors and do not necessarily represent the official views of NIH and NSF.

### Supplemental material

Supplemental material is available for this article online.

### References

1. Kramar A, Potvin D and Hill C. Multistage designs for phase II clinical trials: statistical issues in cancer research. *Br J Cancer* 1996; **74**: 1317–1320.
2. Brown SR, Gregory WM, Twelves CJ, et al. Designing phase II trials in cancer: a systematic review and guidance. *Br J Cancer* 2011; **105**: 194–199.
3. Chow SC and Chang M. *Adaptive design methods in clinical trials*, 2nd ed. New York, NY: Chapman & Hall/CRC Press, 2011.
4. Lin Y and Shih WJ. Adaptive 2-stage designs for single-arm phase IIA cancer clinical trials. *Biometrics* 2004; **60**: 482–490.
5. Simon R. Optimal 2-stage designs for phase 2 clinical trials. *Controll Clin Trials* 1989; **10**: 1–10.
6. Kunz C and Kieser M. Optimal 2-stage designs for single-arm phase II oncology trials with two binary endpoints. *Methods Inform Med* 2011; **50**: 372–377.
7. Mander A and Thompson S. 2-Stage designs optimal under the alternative hypothesis for phase II cancer clinical trials. *Contemp Clin Trials* 2010; **31**: 572–578.
8. Mander AP, Watson JMS, Sweetig MJ, et al. Admissible 2-stage designs for phase II cancer clinical trials that incorporate the expected sample size under the alternative hypothesis. *Pharm Stat* 2012; **11**: 91–96.
9. Wason JMS, Mander AP and Eisen TG. Reducing sample sizes in 2-stage phase II cancer trials by using continuous tumor shrinkage end-points. *Eur J Cancer* 2011; **47**: 983–989.
10. Kwak M and Jung S-H. Phase II clinical trials with time-to-event endpoints: optimal two-stage designs with one-sample log-rank test. *Stat Med* 2014; **33**: 2004–2016.
11. Ensign LG, Gehan EA, Kamen DS, et al. An optimal three-stage design for phase II clinical trials. *Stat Med* 1994; **13**: 1727–1736.
12. Chen TT. Optimal three-stage designs for phase II cancer clinical trials. *Stat Med* 1997; **16**: 1–11.
13. Chen K and Shan M. Optimal and minimax three-stage designs for phase II oncology clinical trials. *Contemp Clin Trials* 2007; **28**: 32–41.
14. Schlesselman JJ and Reis IM. Phase II clinical trials in oncology: strengths and limitations of 2-stage designs. *Cancer Invest* 2006; **24**: 404–412.
15. Zhou Y. Choice of designs and doses for early phase trials. *Fundam Clin Pharmacol* 2004; **18**: 373–378.
16. Kennedy J and Eberhart R. Particle swarm optimization. *Proc IEEE Int Conf Neural Netw* 1995; **IV**: 1942–1948.

17. Whitacre JM. Recent trends indicate rapid growth of nature-inspired optimization in academia and industry. *Computing* 2011; **93**: 121–133.
18. Whitacre JM. Survival of the flexible: explaining the recent popularity of nature-inspired optimization within a rapidly evolving world. *Computing* 2011; **93**: 135–146.
19. Qiu JH, Chen RB, Wang WC, et al. Using animal instincts to design efficient biomedical studies. *Swarm Evol Comput* 2014; **18**: 1–10.
20. Chen RB, Chang SP, Wang W, et al. Minimax optimal designs via particle swarm optimization methods. *Stat Comput* 2014; **24**: 1063–1080.
21. Phoa KHF, Chen RB, Wang WC, et al. Optimizing two-level supersaturated designs using swarm intelligence techniques. *Technometrics* 2016; **58**: 43–49.
22. Wong WK, Chen RB, Huang CC, et al. A modified particle swarm optimization technique for finding optimal designs for mixture models. *PLoS One* 2015; **10**: e0124720.
23. Chen RB, Wang W, Chang SP, et al. Optimal designs for mixture models using particle swarm optimization methods. *Plos One* 2015; **10**: e0124720.
24. Kim S and Li L. A novel global search algorithm for nonlinear mixed-effects models using particle swarm Optimization. *J Pharmacokinet Pharmacodyn* 2011; **38**: 471–495.
25. Kim S and Li L. Statistical identifiability and convergence evaluation for nonlinear pharmacokinetic models with particle swarm optimization. *Comput Methods Programs Biomed* 2014; **113**: 413–432.
26. Kim S, Erwin D and Wu D. Efficacy of dual lung cancer screening by chest x-ray and sputum cytology using Johns Hopkins Lung Project Data. *Biometr Biostat* 2012; **3**: 1000139.
27. Shi YH and Eberhart RC. A modified particle swarm optimizer. In: *IEEE international conference on evolutionary computation*, Anchorage, Alaska, 4–9 May 1998.
28. Tao Z and Cai JD. A new chaotic PSO with dynamic inertia weight for economic dispatch problem. In: *2009 International conference on sustainable power generation and supply*, Nanjing, 2009, pp.1–6.
29. Ascierto PA, Minor D, Ribas A, et al. Phase II trial (BREAK-2) of the BRAF inhibitor Dabrafenib (GSK2118436) in patients with metastatic melanoma. *J Clin Oncol* 2013; **31**: 3205–3211.
30. Green SJ and Dahlberg S. Planned versus attained design in phase II clinical trials. *Stat Med* 1992; **11**: 853–862.
31. Falchook GS, Long GV, Kurzrock R, et al. Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumors: A phase I dose-escalation trial. *Lancet* 2012; **379**: 1893–1901.