HHS Public Access

Author manuscript

ACS Sustain Chem Eng. Author manuscript; available in PMC 2015 December 01.

Published in final edited form as:

ACS Sustain Chem Eng. 2013 ; 1(9): 1119–1128. doi:10.1021/sc4000412.

Two-Stage Experimental Design for Dose–Response Modeling in Toxicology Studies

Kai Wang[†], **Feng Yang**[*,†], **Dale W. Porter**[‡], and **Nianqiang Wu**[§]

[†]Department of Industrial and Management System Engineering, West Virginia University, Morgantown, West Virginia 26506, United States

[‡]National Institute for Occupational Safety and Health (NIOSH), Morgantown, Morgantown, West Virginia 26506, United States

[§]Department of Mechanical and Aerospace Engineering, West Virginia University, Morgantown, West Virginia 26506, United States

Abstract

The efficient design of experiments (i.e., selection of experimental doses and allocation of animals) is important to establishing dose–response relationships in toxicology studies. The proposed procedure for design of experiments is distinct from those in the literature because it is able to adequately accommodate the special features of the dose–response data, which include non-normality, variance heterogeneity, possibly nonlinearity of the dose–response curve, and data scarcity. The design procedure is built in a sequential two-stage paradigm that allows for a learning process. In the first stage, preliminary experiments are performed to gain information regarding the underlying dose–response curve and variance structure. In the second stage, the prior information obtained from the previous stage is utilized to guide the second-stage experiments. An optimization algorithm is developed to search for the design of experiments that will lead to dose–response models of the highest quality. To evaluate model quality (or uncertainty), which is the basis of design optimization, a bootstrapping method is employed; unlike standard statistical methods, bootstrapping is not subject to restrictive assumptions such as normality or large sample sizes. The design procedure in this paper will help to reduce the experimental cost/time in toxicology studies and alleviate the sustainability concerns regarding the tremendous new materials and chemicals.

Graphical Abstract

[*]**Corresponding Author**feng.yang@mail.wvu.edu. Phone: 304-293-9477. Fax: 304-293-4970..

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

The authors declare no competing financial interest.

Keywords

Design of experiments; Dose–response; Benchmark dose; Toxicology; Nanomaterials

INTRODUCTION

This work is primarily motivated by the rising need to assess the risks resulting from human exposure to various nanomaterials. With the rapid development of nanotechnology, nanomaterials find increasing applications in the energy,[1,2] environment,[3–5] and biomedical engineering sectors.[6,7] Workers may be directly exposed to nanomaterials during manufacturing, handling, and transport of nanomaterials, and end-users may uptake nanomaterials due to the release of nanomaterials from consumer products. Therefore, risk assessment of a large variety of nanomaterials plays a critical part to ensure the safe and sustainable development of nanotechnology.

One of the most fundamental steps in assessing the risk of a nanomaterial (or any other substance) is to understand and properly characterize its dose–response relationship.[8,9] To estimate such relationships, biological experiments need to be performed at different dose levels to observe the corresponding bioactivity responses of animals. Because of costs, ethics, or other limitations on resources or time, sample sizes are usually restricted and efficient use of available resources is critical. Thus, the design of experiments (DOE), i.e., the selection of experimental doses and the allocation of animals, plays an important role in the success of dose–response modeling.

The DOE for efficient dose–response modeling is challenging due to the following two major reasons. First, the dose–response curves may well be nonlinear,[10–12] and thus, the optimum design of experiments depends on the true underlying curves to be estimated, which is unfortunately unknown.[13,14] Second, with constrained resources, typically the amount of data allowed to be collected are not only relatively scarce and highly variable but almost certainly possess different variability across the dose range being investigated.[15] These special features of dose–response data are not adequately addressed in the existing DOE methods.

In the current practice of DOE for toxicology studies, designs are mainly selected by experimenters based on empirical experiences in a somewhat arbitrary manner.[16,17] Such

designs, which will be referred to as the traditional design, usually involve equally spaced doses on a linear or log scale and equal allocation of animals and may well lead to inefficient use of resources. Some researchers[8,18–21] recognized the nonlinearity of dose–response relationships and resorted to standard optimum design techniques[13] to approach the DOE issues. These works relied on the assumption of constant variance throughout the dose range, which typically does not hold in reality. There is also some literature by statisticians that investigated the optimum design for statistical modeling in general while taking into account the existence of variance heterogeneity. In this stream of work,[22–32] the authors usually had a particular interest in pursuing analytical solutions under various simplistic assumptions. For instance, DasGupta et al.[23] derived the optimal designs under normal prior conditions for the simple linear model when the error variance is a power function or an exponential function of the mean. These results, though theoretically appealing, are often impractical for toxicology studies where the required assumptions no longer hold.

The objective of this work is to develop a DOE procedure, which is statistically valid and practically useful, to guide the dose selection and animal allocation in biological experiments for the efficient generation of dose–response relationships. Through the following three major efforts, we intend to provide a new DOE method that adequately accommodates the nonlinear nature of dose–response curves and variance heterogeneity of toxicity data. (i) The proposed procedure is built in a two-stage paradigm. In the first stage, a batch of preliminary experiments are performed to gain information regarding the underlying dose–response curve and variance structure. In the second stage, the prior information obtained from the previous stage is utilized to guide the second batch of experiments. Compared to the traditional once-and-for-all designs, such a sequential design is known to lead to improved model estimation when a total sample size is given or to bring savings in samples if a certain model quality is desired. (ii) For the design of the second batch of experiments, a new algorithm (Algorithm 3) is developed to effectively search for the optimum design that will lead to dose–response models of the highest quality. (iii) To evaluate the model quality, which is the basis of design optimization in (ii), bootstrapping, as opposed to standard statistical methods, is employed. Bootstrapping is a computationally intensive resampling method and is not subject to restrictive assumptions such as normality or large sample sizes.

The remainder of this paper is organized as follows. The Statistical Modeling and Inference section details the modeling/inference issues involved in the estimation of dose–response curves and provides the building blocks of the two-stage DOE procedure. The procedure, which aims at achieving an efficient experimental design, is described in the Two-Stage Procedure section. In the Empirical Evaluation section, a simulation study is performed to evaluate the efficiency of the proposed DOE procedure by comparing it with the traditional design currently used in toxicology studies. The Summary section summarizes the findings of this paper and discusses future work.

STATISTICAL MODELING AND INFERENCE

In this section, we present the statistical modeling and inference methods for estimating the dose–response relationship based on a given set of experimental data. These methods provide the necessary basis for the DOE procedure in the Two-Stage Procedure section.

Heteroscedastic Dose–Response Model

Let Y be the continuous random response of a biological experiment. We represent the dose–response curve by a general model

$$\mathrm{E}[Y|x] = f(x, \boldsymbol{\theta})$$

where x denotes the dose level and $\boldsymbol{\theta}$ the $p \times 1$ vector of unknown parameters. The random variable Y is related to its mean by the relationship

$$Y = \mathrm{E}[Y|x] + \varepsilon$$

where ε is the error term. The feasible region of the dose level x is typically specified by

$$x_L \leq x \leq x_U \quad (1)$$

The lower and upper bounds, x_L ($x_L \geq 0$) and x_U ($x_U > x_L$), are specified by the experimenters based on the substance of interest.

The mean response model $f(x, \boldsymbol{\theta})$ can take any functional forms that are differentiable and monotonic. Example models include linear, polynomial, power (EPA Benchmark Dose Software[33]), and the four-parameter logistic model widely used in the literature.[34,35] In the empirical study of this paper, the following functional form is used for illustration

$$f(x, \boldsymbol{\theta}) = \theta_1 \, exp(x/\theta_2) \quad (2)$$

The error ε is assumed independent across individual observations, and to be expressed as

$$\varepsilon = \sigma h(f(x, \boldsymbol{\theta}), \boldsymbol{\gamma}) \, e \quad (3)$$

where σ is a scale parameter; e is a random error following the standard normal distribution; h describes how the standard deviation of ε depends on the mean response $f(x, \boldsymbol{\theta})$; and $\boldsymbol{\gamma}$ denotes the vector of unknown parameters involved in model h.

The formulation of the error (3) is able to model a wide range of error heterogeneity commonly encountered and has been extensively employed in the literature.[35–38] In this work, a specific functional form as given below is assumed for the variance model $h(f(x,\boldsymbol{\theta}),\boldsymbol{\gamma})$ of the empirical example in the Empirical Evaluation section

$$h(f(x, \boldsymbol{\theta}), \boldsymbol{\gamma}) = f(x, \boldsymbol{\theta})^{\gamma} \quad (4)$$

where γ is the unknown parameter. The form (eq 4) has been popularly used in toxicology and pharmacology studies.[34–36,38,39]

Estimation of Dose–Response Model and Benchmark Dose (BMD)

Given a data set $\{(x_i,y_i),i = 1,...,I\}$ with a sample size of I, how do we estimate the dose–response models in the presence of variance heterogeneity? Preliminary data analysis as that suggested by Carroll and Ruppert[36] need to first be performed to select specific functional forms for $f(x,\boldsymbol{\theta})$ and $h(f(x,\boldsymbol{\theta}),\boldsymbol{\gamma})$, which may turn out to be eqs 2 and 4, respectively, for examples. Then, the iteratively reweighted least-squares (IRLS) framework[36] is adopted to fit the models. The IRLS algorithm (Algorithm 1) is described using the notations of our dose–response modeling.

With the estimated dose–response model $f\left(x,\hat{\boldsymbol{\theta}}\right)$, the expected toxicity response can be obtained as $f\left(x,\hat{\boldsymbol{\theta}}\right)$ at any exposure level $x_0 \in [x_L,x_U]$. The model also allows for the estimation of exposures at responses of different severities; such an exposure level is referred to as benchmark dose (BMD) and is of particular interest in toxicology because it assists in setting the safety standard for the substance being investigated.

The BMD is the dose that corresponds to a specified level of adverse response called the benchmark response (BMR). The BMR can be defined as a relative change in the mean response from the control mean or as an absolute level.[33,40] Either definition can be selected based on the knowledge available regarding the substance's adverse effects, and in modeling, the BMR defined in one way can be easily converted to that defined in the other way. For illustration in this work, we let the BMR be a preselected absolute response. Given the fitted dose–response model $f\left(x,\hat{\boldsymbol{\theta}}\right)$, the BMD can be estimated as

$$\widehat{BMD}=f^{-1}\left(BMR,\hat{\boldsymbol{\theta}}\right) \quad (5)$$

where f^{-1} represents the inverse function of f.

Quantifying Estimation Uncertainty

In this part, we discuss the uncertainty quantification (or statistical inference issues) for the dose–response modeling. The following two types of uncertainties/variabilities are of primary interest.

- $Var\left[\hat{\boldsymbol{\theta}}\right]$ the variance of the estimated parameters $\hat{\boldsymbol{\theta}}$ in the fitted dose–response model $f\left(x,\hat{\boldsymbol{\theta}}\right)$.

Clearly, $Var\left[\hat{\boldsymbol{\theta}}\right]$ determines the quality of the estimated dose–response model $f\left(x,\hat{\boldsymbol{\theta}}\right)$. With the increasing of experimental data (sample size), $Var\left[\hat{\boldsymbol{\theta}}\right]$ will decrease and is directly associated with the D-optimum design criterion in the DOE literature.[13]

- $Var\left[\widehat{BMD}\right]$, the variance of the estimated BMD.

As shown in eq 5, $\widehat{BMD}$ is a function of the pre-specified BMR, which is deterministic, and the estimated parameter $\hat{\boldsymbol{\theta}}$, which is subject to uncertainties. Hence, $Var\left[\widehat{BMD}\right]$ is affected by both the BMR value and $Var\left[\hat{\boldsymbol{\theta}}\right]$. For certain BMR, $Var\left[\widehat{BMD}\right]$ directly relates to the safety standard of a substance. With $Var\left[\widehat{BMD}\right]$, a confidence interval (CI) can be formed for the BMD, and it has a clear and practically meaningful interpretation. In toxicology studies, the lower bound of the CI is referred to as BMDL and represents the lower confidence limit on the dose that would result in the required response BMR[33]

In light of the discussions above, $Var\left[\hat{\boldsymbol{\theta}}\right]$ measures the overall quality of the dose–response model, whereas $Var\left[\widehat{BMD}\right]$ not only depends on the model quality but is also associated with a particular BMR. Depending on the purpose of a study, either of these two variabilities can serve as the criterion to drive the optimal design of experiments. Either criterion can well fit into our two-stage DOE procedure.

As will be seen later in Algorithm 3 (Two-Stage Procedure section), the prerequisite of using $Var\left[\hat{\boldsymbol{\theta}}\right]$ or $Var\left[\widehat{BMD}\right]$ as a design criterion is the ability to evaluate these variabilities for a candidate design. This ability can be derived from the evaluation of $Var\left[\hat{\boldsymbol{\theta}}\right]$ or $Var\left[\widehat{BMD}\right]$ for a given sample data set, which will be discussed in the remainder of this subsection.

Bootstrap Resampling Algorithm—As mentioned in the Introduction section, standard statistical inference methods are not able to provide valid variability estimates for dose–response modeling. Hence, we use a bootstrapping resampling method to evaluate $Var\left[\hat{\boldsymbol{\theta}}\right]$ and $Var\left[\widehat{BMD}\right]$. The bootstrap is a data-based simulation method for statistical inference.[41] The basic idea of bootstrap is to use resampling (sampling from the sample, obtaining what is called bootstrap samples) to quantify the variability of the estimates of interest.[42]

Under the assumptions given in the Heteroscedastic Dose–Response Model subsection, the resampling scheme proposed by Zeng and Davidian[35] is adapted to quantify the uncertainties of the model estimation performed on the given data set $\{(x_i,y_i);i = 1,2,...,I\}$, and the resampling algorithm is described in Algorithm 2.

Zeng and Davidian[35] recommended setting the resample size to $B = 499$ based on their empirical experience. Note that M, the assumed number of design points in Input (a) of the algorithm, is allowed to be different from I, the number of real samples $\{(x_i,y_i);i = 1,2,...,I\}$ involved in Input (b). If the purpose of executing Algorithm 2 is solely to quantify the model uncertainties given a real data set $\{(x_i,y_i);i = 1,2,...,I\}$, then we set $M = I$. However, in this work, Algorithm 2 is meant to be called by Algorithm 3 (the DOE algorithm in the Two-

Stage Procedure section) to evaluate additional candidate design points, and hence, we distinct between M and I with $M \geq I$.

Quantifying the Uncertainty of Estimated Model Parameters—Given $\left\{\hat{\boldsymbol{\theta}}_b^*; \quad b=1,2,\ldots,B\right\}$ obtained from Algorithm 2, the variance-covariance matrix $Var\left[\hat{\boldsymbol{\theta}}\right]$ can be estimated as follows[36]

$$\widehat{Var}\left[\hat{\boldsymbol{\theta}}\right]=B^{-1}\sum_{b=1}^{B}\left(\hat{\boldsymbol{\theta}}_b^*-\hat{\boldsymbol{\theta}}\right)\left(\hat{\boldsymbol{\theta}}_b^*-\hat{\boldsymbol{\theta}}\right)' \quad (6)$$

where $\hat{\boldsymbol{\theta}}$ are the estimates obtained from the original data $\{(x_i,y_i);i = 1,2,...,I\}$.

Quantifying the Uncertainty of Estimated BMD—Recall the expression of BMD (eq 5). With the B fitted models $\left\{f\left(x,\hat{\boldsymbol{\theta}}_b^*\right);b=1,2,\ldots,B\right\}$ from Algorithm 2, B BMD estimates can be obtained as follows

$$\widehat{BMD}_b^*=f^{-1}\left(BMR,\hat{\boldsymbol{\theta}}_b^*\right) \quad (7)$$

for a pre-specified BMR. On the basis of $\left\{\widehat{BMD}_b^*; \quad b=1,2,\ldots,B\right\}$, $Var\left[\widehat{BMD}\right]$ is estimated as

$$\widehat{Var}\left[\widehat{BMD}\right]=B^{-1}\sum_{b=1}^{B}\left(\widehat{BMD}_b^*-\widehat{BMD}\right)^2 \quad (8)$$

where $\widehat{BMD}$ is obtained from $f\left(x,\hat{\boldsymbol{\theta}}\right)$, and the dose–response model is estimated from the original sample data $\{(\mathbf{x}_i,\mathbf{y}_i);i = 1,2,...,I\}$.

The variance estimate $Var\left[\widehat{BMD}\right]$ can be used to form the one-sided confidence interval (CI) for the BMD[33] at a selected confidence level α. The BMDL (lower bound of the one-sided CI) is given as[35]

$$BMDL=\widehat{BMD}-t_{\alpha,M-p}\sqrt{\widehat{Var}\left[\widehat{BMD}\right]} \quad (9)$$

where $t_{\alpha,M-p}$ stands for the $100\alpha^{\text{th}}$ quantile of the student t distribution with $M–p$ degrees of freedom, in which p stands for the number of fitted parameters in the dose–response model.

TWO-STAGE PROCEDURE

The major contribution of this work lies in the development of a new two-stage DOE procedure for dose–response modeling. Compared to standard once-and-for-all designs, sequential (multi-stage) designs typically lead to (i) savings in sample size if model estimates of desired quality (measured by the uncertainty/variabilty of model estimates) are to be achieved or (ii) model estimates of improved quality given a fixed total sample

size.[11,43] Because biological studies are typically performed with restricted sample sizes, the two-stage design procedure in this work is tailored to achieve models of the highest quality with a pre-specified total sample size. It is worth pointing out that the two-stage procedure can be easily adapted to a design procedure that is driven by a desired model quality and not constrained by a pre-specified total sample size; such a procedure may involve multiple (more than two) stages of experiments.

For the two-stage DOE procedure, the inputs are given as

Input 1: N, the total sample size available.

Input 2: $[x_L,x_U]$, the range of interest for the dose level, as defined in the Heteroscedastic Dose-Response Model section.

Input 3: N_I, the batch size of the experiments (i.e., the number of samples) performed at Stage I, which implies that the sample size is $N–N_I$ at Stage II.

Input 4 (optional): BMR, the pre-specified benchmark response level, which is not necessary unless the design criterion is $Var\left[\widehat{BMD}\right]$.

The outputs of the procedure include the design of experiments for the N samples; the dose–response model $f\left(x,\hat{\boldsymbol{\theta}}\right)$ and the variance model $\hat{\sigma}h\left(f\left(x,\hat{\boldsymbol{\theta}}\right),\hat{\boldsymbol{\gamma}}\right)$ both of which are estimated from the N-sample data set and allow for the BMD estimation; and the uncertainties of the estimates of interest including $\widehat{Var}\left[\hat{\boldsymbol{\theta}}\right]$ and $\widehat{Var}\left[\widehat{BMD}\right]$ if a pre-specified BMR is given.

Figure 1 provides an overview of the design procedure. In Stage I, pilot experiments with a sample size of N_I are carried following the initial design, and initial statistical modeling/inference are performed. The design augmentation in Stage II is performed utilizing the information derived from the experimental data collected in Stage I, and $N–N_I$ additional experiments are carried out following the augmented design. On the basis of the collected data from both stages, the dose–response and variance models are estimated, and related uncertainties quantified.

Clearly, the key tasks yet to be addressed in Figure 1 are the initial design in Stage I and the design augmentation is Stage II. Both tasks aim at answering the DOE questions in their designated stages, such as at what design points (i.e., dose levels) within the design space $[x_L,x_U]$ should the experiments be carried out and how many replications should be assigned to each design point?

Two alternative notations are used in this paper to represent a design. One is given as

$$\tilde{\mathscr{D}}=\begin{pmatrix} x_1 & x_2 & \cdots & x_D \\ n_1 & n_2 & \cdots & n_D \end{pmatrix} \quad (10)$$

with x_d being the d^{th} design point, n_d the number of replications at x_d and D the number of *distinct* design points. Clearly, the total sample size given by $\tilde{\mathscr{D}}$ is $\sum_{d=1}^{D} n_d$. Alternatively, a design (10) can be expressed as

$$\tilde{\mathscr{D}} = \begin{pmatrix} x_1 & x_2 & \cdots & x_M \end{pmatrix} \quad (11)$$

with $M = \sum_{d=1}^{D} n_d$ being the number of samples. Note that the array $\tilde{\mathscr{D}}$ may well include the same design points multiple times, which correspond to multiple replications. In the remainder of this paper, both eqs 10 and 11 will be used to refer to a design.

Stage I: Initial Design

At Stage I, N_I preliminary samples are to be performed. Given N, the total number of samples available, we recommend setting N_I as 1/4–1/2 of N. Guidelines for specifying N_I can also be found in Santner et al.[11] Having selected a value for N_I, how do we allocate these N_I samples? Following the notation in eq 10, the initial design is represented as

$$\tilde{\mathscr{D}}^{(I)} = \begin{pmatrix} x_1^{(I)} & x_2^{(I)} & \cdots & x_{D_0}^{(I)} \\ n_1^{(I)} & n_2^{(I)} & \cdots & n_{D_0}^{(I)} \end{pmatrix} \quad (12)$$

where D_0 is the number of distinct dose levels.

In this work, assuming that no prior information other than the dose range $[x_L, x_U]$ is available at the initial stage, we specify the initial design as follows. First, the design points $\left\{x_1^{(I)}, x_2^{(I)}, \cdots, x_{D_0}^{(I)}\right\}$ are evenly spaced on the original or log scale, as in the traditional designs, and we set $x_1^{(I)} = x_L$ and $x_{D_0}^{(I)} = x_U$, including lower and upper bounds of the dose range in the design to avoid potential extrapolation, which is considered dangerous in statistical estimation.[44] Second, D_0, the number of distinct design points, has to be at least four, which is the typical number needed to capture the trend of a dose–response curve. Third, we have $n_d^{(I)} \geq 3$ for any d to allow for the detection of variance heterogeneity and for the estimation of the variance model (eq 3). In the absence of any other concerns, we can simply set $n_1^{(I)} = n_2^{(I)} = \cdots = n_{D_0}^{(I)}$. For instance, in light of the considerations above, the initial design for the example in the Empirical Evaluation section is specified as

$$\begin{pmatrix} 0 & 7.5 & 15 & 22.5 & 30 \\ 4 & 4 & 4 & 4 & 4 \end{pmatrix} \quad (13)$$

given that $N_I = 20$ and $[x_L, x_U] = [0,30]$.

Note that the initial design is set up to enable a fair estimation of the dose–response and variance models. Following the initial design, a total of N_I samples are collected at Stage I of the procedure (Figure 1) and are denoted as $\{(x_i, y_i);\ i = 1,2,...,N_I\}$. From these N_I samples, the preliminary models are obtained by applying Algorithm 1: the fitted dose–response

model $f\left(x,\hat{\boldsymbol{\theta}}^{(I)}\right)$ and the variance model $\hat{\sigma}h\left(f\left(x,\hat{\boldsymbol{\theta}}^{(I)}\right),\hat{\boldsymbol{\gamma}}^{(I)}\right)$, both of which are used to guide the DOE in Stage II.

Stage II: Design Augmentation

The task of the Stage II design augmentation is to find out how to allocate the rest $N–N_I$ samples in the design space $[x_L,x_U]$, that is, to determine the values of $N–N_I$ design points, which are denoted as $\{x_{N_I+1},x_{N_I+2},\ldots,x_N\}$ following the notation (11).

In this work, the dosage range $[x_L,x_U]$ is divided into a relatively large number of equally spaced grid points, and only these grid points are considered as the candidate design points for the design optimization (14) for the following reason. The equally spaced grid points can be chosen in such a way that any two neighboring points are practically different. By restricting the design points to these grid points, the design optimization 14 can avoid providing design points that are numerically but not practically different.

As in eq 14, the $N–N_I$ design points are selected to minimize the design criterion subject to two constraints. First, N_I samples have already been collected. Second, these $N–N_I$ points have to be selected from the predetermined grid points within $[x_L,x_U]$. As mentioned earlier, depending on the practical need, the design criterion could be $|Var\left[\hat{\boldsymbol{\theta}}\right]|$, the determinant of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$, or $Var\left[\widehat{BMD}\right]$, the variance of $\widehat{BMD}$ for a given BMR.

$$\min_{x_{N_I+1},x_{N_I+2},\ldots,x_N} \text{Design criterion, which could be } |\text{Var}[\hat{\boldsymbol{\theta}}]| \text{ or } \text{Var}[\widehat{\text{BMD}}] \quad (14)$$

Subject to N_I samples $\{(x_i,y_i);\ i = 1,2,\ldots, N_I\}$ have already been collected;

and $\{x_{N_I+1},x_{N_I+2},\ldots,x_N\}$ can only be selected from the candidate design points,
i.e., the pre-determined evenly-spaced grid points within $[x_L,x_U]$.

Solving this optimization problem to obtain the best design is challenging. (i) The design criterion in the objective function (eq 14) represents the variability of the model estimates ($\hat{\boldsymbol{\theta}}$ or $\widehat{BMD}$) obtained from all the experimental data including those collected at first and second stages. However, at the current point of solving eq 14 and designing for second-stage experiments, only first-stage data are available, and there is no closed-form formulation for the objective function. To overcome this difficulty, in Algorithm 3, which aims at solving eq 14, a bootstrapping-based method is developed to numerically evaluate the design criterion and to provide the basis to solve the design optimization. (ii) Determining $\{x_{N_I+1},x_{N_I+2},\ldots,x_N\}$ simultaneously is difficult due to the high dimension (which is $N–N_I$) of the decision variables and the large size of the decision space. In light of this, we use a heuristics that adds one design point at a time until all the $N–N_I$ points have been found. The algorithm for solving eq 14 is described as follows assuming that $Var\left[\widehat{BMD}\right]$ is the design

criterion. The algorithm can be straightforwardly adapted to the design optimization with $|Var\left[\hat{\boldsymbol{\theta}}\right]|$ being the criterion, as will be explained later.

Algorithm 3 is initiated with the Stage I design $\tilde{\mathscr{D}}=\tilde{\mathscr{D}}^{(I)}$, and the first FOR-LOOP iterates to add one design point to the design $\tilde{\mathscr{D}}$ at a time until all the N–N_I design points have been found. At this iteration level, for a given design $\tilde{\mathscr{D}}$ that already includes N_I+k (k = 0,1,...,N–N_I–1) points, how do we determine the $(N_I+k+1)^{\text{th}}$ design point? The answer is the second FOR-LOOP, which iterates to evaluate every grid point; the best grid point is chosen as the $(N_I+k+1)^{\text{th}}$ design point, and the design D is updated accordingly. Steps (i) and (ii) are the bootstrapping-based methods developed to evaluate each grid point in terms of their additional contribution to the optimization objective, given a to-be-expanded design $\tilde{\mathscr{D}}$.

Note that Algorithm 3, which is driven by the design criterion $Var\left[\widehat{BMD}\right]$, can be easily modified to solve eq 14 with the criterion being $|Var\left[\hat{\boldsymbol{\theta}}\right]|$. The only change that needs to be made is to replace Step (ii) by evaluating $|\widehat{Var}\left[\hat{\boldsymbol{\theta}}^{(g)}\right]|$ with eq 6.

Following the additional design $\{x_{N_I+1},x_{N_I+2},...,x_N\}$ found by applying Algorithm 3, Stage II experiments will be performed for data collection. Lastly, based on all the data collected at both stages with a total sample size of N, statistical modeling and inference will be performed to obtain the final estimation for the dose–response model and the BMD and to quantify the uncertainties on those estimates of interest.

EMPIRICAL EVALUATION

The proposed two-stage design procedure has been evaluated by a simulation study, which is based on sampling through computer experiments whose outputs mimic dose–response data from real experiments. Simulation, rather than real experiments, is employed for the following reasons. First, only in a simulation-based study are the true dose–response models (i.e., the simulation models) available to evaluate the model estimates. Second, the outcome of a design procedure is random due to the randomness of responses; hence, a design procedure needs to be evaluated in a statistical manner based on the outcomes of applying it many times, which is impossible with real experiments. These advantages of simulation will become clear in the Comparison with Traditional Design subsection.

Simulation Models

In this case, the simulation data are generated using the following true dose–response and variance models

$$f\left(x,\boldsymbol{\theta}\right)=20 \quad exp\left(x/11\right) \quad (16)$$

$$h\left(f\left(x,\boldsymbol{\theta}\right),\boldsymbol{\gamma}\right)=f(x,\boldsymbol{\theta})^{0.7} \quad \text{and} \quad \sigma=0.1 \quad (17)$$

At an arbitrary dose level, say x_0, a response y_0 is simulated as

$$y_0 = f(x_0, \boldsymbol{\theta}) + \sigma h(f(x_0, \boldsymbol{\theta}), \boldsymbol{\gamma})\varepsilon = 20\,exp(x/11) + 0.1(20\,exp(x/11))^{0.7} \times \varepsilon \quad (18)$$

where ε is a random error provided by a standard normal random number generator.[45]

The simulation models (eqs 16 and 17) are derived from the real experimental data in the toxicity study of TiO_2 nanowires.[46] The dose *x* represents the TiO_2 nanowire dosage in terms of μg/mouse, and the response *y* is the BAL (bronchoalveolar lavage) PMNs measured in the units of 10^3/mouse. C57BL/6J male mice were exposed to selected dosages of TiO_2 nanowires, and at seven days post-exposure, whole lung lavage was conducted to obtain the BAL PMN counts.

The true models (eqs 16 and 17) are blind to the two-stage procedure and are only used for two purposes in this study: (i) to generate simulation data that mimic real biological data and (ii) to serve as the true benchmark to evaluate the model estimates obtained from applying the DOE procedure.

Applying the Two-Stage Procedure

For the toxicity study of TiO_2 nanowires, suppose that it is of particular interest to establish the BMD corresponding to a BMR of 177 ($\times 10^3$/mouse) for the BALPMN endpoint. (Recall that the BMR can be defined in different ways that do not affect the application of our methods, and in this paper, we treat the BMR as an absolute level.) How to specify a BMR is beyond the scope of this work, and BMR = 177 is employed for the purpose of illustration.

The two-stage DOE procedure (Figure 1) is applied to guide the experiments (that is, the simulation experiments in this empirical study) for the efficient collection of dose–response data, aiming at obtaining a good BMD estimate corresponding to the target BMR = 177. The inputs of the procedure are given as follows:

Input 1: $N = 40$, the total sample size available.

Input 2: $[x_L, x_U] = [0,30]$, the dose range of interest.

Input 3: $N_I = 20$: the batch size of the experiments (i.e., the number of samples) performed at Stage I, which implies that the sample size is $N–N_I = 20$ at Stage II.

Input 4: BMR = 177: the pre-specified benchmark response.

The criterion in the design optimization (eq 14) is set as $Var\left[\widehat{BMD}\right]$ accordingly.

Stage I

The initial design with $N_I = 20$ samples was obtained and is given in Table 1. Five evenly spaced distinct design points are selected over the dose range $[x_L, x_U] = [0,30]$ with four replications at each point. At each dose level x_i ($i = 1,2,...,20$), the response subject to random errors is generated by plugging the value of x_i into eq 18, and the response data are

given in Table 2. From the paired data {(x_i,y_i); I = 1,2,...,20}, the dose–response and variance models are estimated as follows by applying Algorithm 1

$$f\left(x,\hat{\boldsymbol{\theta}}^{(I)}\right)=20.47\,exp\,(x/11.06) \quad (19)$$

$$\hat{\sigma}h\left(f\left(x,\hat{\boldsymbol{\theta}}^{(I)}\right),\hat{\boldsymbol{\gamma}}^{(I)}\right)=0.06f\left(x,\hat{\boldsymbol{\theta}}^{(I)}\right)^{0.84} \quad (20)$$

Stage II

The augmented design was determined by applying Algorithm 3 with the following inputs. (a) The first-stage design $\tilde{\mathscr{D}}^{(I)}=\{x_i;i=1,2,\ldots,N_I\}$ with values of the design points given in Table 1. (b) The estimated dose–response model $f\left(x,\hat{\boldsymbol{\theta}}^{(I)}\right)$ and the variance model $\hat{\sigma}^{(I)}h\left(f\left(x,\hat{\boldsymbol{\theta}}^{(I)}\right),\hat{\boldsymbol{\gamma}}^{(I)}\right)$ as given in eqs 19 and 20. (c) The pool of candidate design points $\left\{x_g^{(Grid)};g=1,2,\ldots,G\right\}=\{0,0.5,1,\ldots,30\}$, which are evenly spaced grid points in [x_L,x_U] = [0,30] with a step size of 0.5. (d) The resampling size B = 499. (e) BMR = 177.

The resulting augmented design turned out to be as given in Table 3.

Following the augmented design (Table 3), simulation was performed to obtain the random responses, which are given in Table 4.

Feeding all the data collected in both stages {(x_i,y_i); I = 1,2,...,40} to Algorithm 1, the dose–response and variance models were finally estimated as

$$f\left(x,\hat{\boldsymbol{\theta}}\right)=20.44\,exp\,(x/11.07) \quad (21)$$

$$\hat{\sigma}h\left(f\left(x,\hat{\boldsymbol{\theta}}\right),\hat{\boldsymbol{\gamma}}\right)=0.09f\left(x,\hat{\boldsymbol{\theta}}\right)^{0.74} \quad (22)$$

Figure 2 plots the fitted dose–response curve (eq 21), along with the Stage I data depicted by dots and Stage II data depicted by squares.

Comparison with Traditional Design

The results presented above in the Applying the Two-Stage Procedure subsection represent the outcome of applying the two-stage procedure for one time. Because of the random nature of responses, reapplying the procedure will lead to first-stage responses different than those given in Table 2 and first-stage fitted models different than those in eqs 19 and 20. Consequently, the augmented design in Stage II will turn out differently, the final 40 sample data set will be different, and the modeling/inference results will be different.

Also, because of the randomness in responses, every time the same traditional design is applied, a different set of responses will be collected, and hence, different modeling/inference results will be obtained. The traditional design for this case is given in eq 23 and is the one used for the toxicology study in Porter et al.[46]

$$\tilde{D}^{Trad} = \begin{pmatrix} 0 & 1.875 & 7.5 & 15 & 30 \\ 8 & 8 & 8 & 8 & 8 \end{pmatrix} \quad (23)$$

The two approaches, our two-stage procedure and the traditional design, both have random outputs. The modeling/inference results are subject to the randomness of responses. Hence, these two approaches are compared in a statistical manner as follows.

Each of the two approaches was applied in this case 100 times. As a result, 100 sets of results were obtained for each approach. Because the primary goal of this study is to estimate BMD for BMR = 177, the two approaches are compared in terms of their delivered BMDL (eq 9), the lower bound of the one-sided confidence interval for the BMD.

Specifically, applying the two-stage procedure 100 times leads to 100 data sets; from each data set, the methods (Algorithms 1 and 2) in the Statistical Modeling and Inference section were performed to obtain the BMDL (eq 9). From the 100 data sets, the 100 BMDLs obtained are denoted as

$$\{BMDL_k; \quad k=1, 2, \ldots, 100\} \quad (24)$$

In the same way, applying the traditional design (eq 23) 100 times also results in 100 data sets. Performing the statistical modeling and inference on these data sets leads to 100 BMDLs, which are denoted as

$$\left\{BMDL_k^{(Trad)}; \quad k=1, 2, \ldots, 100\right\} \quad (25)$$

The true BMD can be easily calculated from the true dose−response simulation model (eq 16), and it turns out to be 24, which is represented by the horizontal line in Figure 3. The box plots are generated in Figure 3 for the BMDLs (eq 24) obtained from our two-stage procedure (left box) and the BMDLs (eq 25) obtained from the traditional design (right box). A box plot provides the basic information regarding the distribution of a data set (say, the BMDLs in eq 25), with the lower hinge being the 25th percentile, and the upper hinge being the 75th percentile of the data. For details of box plots, please refer to McGill et al.[47]

Comparing the two boxes in Figure 3, it is evident that the BMDLs resulting from our procedure are closer to the true BMD and vary over a narrower dose range. In other words, with the same sample size, collecting a data set following the traditional design leads to a BMDL that may be anywhere between the lower and upper adjacent of the right box plot, whereas collecting a data set following the two-stage procedure leads to a BMDL that falls within the much narrower and more accurate range of the left box.

SUMMARY

This work developed a two-stage experimental design procedure. Compared to the commonly used naive designs, our methods lead to dose–response models of higher quality given a limited sample size (or equivalently, leads to smaller sample size for a pre-specified model quality). The methods developed in this paper are able to reduce the experimental cost and time in toxicology studies, alleviate the rising concerns for animal ethics, and accelerate the progress toward quantifying the risk, safety, and health effects of environmental and occupational exposure to any substances (e.g., nanomaterials).

The two-stage procedure goes on the premise that no significant cross-stage variability exists, that is, there is no systematic changes from the first- to the second-stage experiments, and dosage is the dominant factor that accounts for the variability in biological responses. This assumption may well hold for experiments that are performed in the same lab by the same experimenters over a relatively short period and on animals with similar characteristics (say, weight, age, etc.). In our immediate future research, we will propose a design procedure that is able to accommodate cross-stage variability; such variability may stem from different batches of animals, different experimenters, or different research environments/laboratories. For instance, a design question to be addressed may be as follows: Given a certain substance's toxicology data obtained by some other research laboratories, how should we design our experiments so that the integrated data from multi-sources can provide the most useful information? Answers to such experimental design questions will certainly help to accelerate the toxicology assessment of various substances and to facilitate the integration of multi-source data.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation Grant CBET-1065931.

REFERENCES

1. Zhi M, Manivannan A, Meng F, Wu NQ. Highly conductive electrospun carbon nanofiber/MnO_2 coaxial nano-cables for high energy and power density supercapacitors. J. Power Sources. 2012; 208:345–353.
2. Xiang C, Li M, Zhi M, Manivannan A, Wu N. Reduced graphene oxide/titanium dioxide composites for supercapacitor electrodes: Shape and coupling effects. J. Mater. Chem. 2012; 22:19161–19167.
3. Li M, Wang Q, Shi X, Hornak LA, Wu NQ. Detection of mercury(II) by quantum dot-DNA-gold nanoparticle ensemble based nanosensor via nanometal surface energy transfer. Anal. Chem. 2011; 83:7061–7065. [PubMed: 21842845]
4. Wu NQ, Zhao M, Zheng J, Jiang C, Myers B, Chyu M, Li S, Mao SX. Porous CuO–ZnO nanocomposite for sensing electrode of high-temperature CO solid-state electrochemical sensor. Nano-technology. 2005; 16:2878–2881.
5. Wu NQ, Chen Z, Xu J, Chyu M, Mao SX. Impedance-metric Pt/YSZ/Au-Ga_2O_3 sensors for CO detection at high-temperature. Sens. Actuators, B. 2005; 110:49–53.
6. Li M, Zhang J, Suri S, Sooter LJ, Ma D, Wu NQ. Detection of adenosine triphosphate with an aptamer biosensor based on surface-enhanced Raman scattering. Anal. Chem. 2012; 84:2837–2842. [PubMed: 22380526]
7. Ren N, Li R, Chen L, Wang G, Liu D, Wang Y, Zheng L, Tang W, Yu X, Jiang H, Liu H, Wu NQ. In-situ construction of titanate/silver nanoparticle/titanate sandwich nanostructure on metallic

titanium surface for bacteriostatic and biocompatible implants. J. Mater. Chem. 2012; 22:19151–19160.

8. Edler L, Poirier K, Dourson M, Kleiner J, Mileson B, Nordmann H, Renwickf A, Slobg W, Waltonf K, Wurtzenh G. Mathematical modelling and quantitative methods. Food Chem. Toxicol. 2002; 40:283–326. [PubMed: 11893400]
9. Bretz F, Hsu J, Pinheiro J, Liu Y. Dose finding: A challenge in statistics. Biom. J. 2008; 50:480–504. [PubMed: 18663758]
10. Slob W. Dose–response modeling of continuous endpoints. Toxicol. Sci. 2002; 66:298–312. [PubMed: 11896297]
11. Sand S, von Rosen D, Eriksson P, Fredriksson A, Viberg H, Victorin K, Filipsson A. Dose–response modeling and benchmark calculations from spontaneous behavior data on mice neonatally exposed to 2,2′,4,4′,5-pentabromodiphenyl ether. Toxicol. Sci. 2004; 81:491–501. [PubMed: 15254340]
12. Oberdorster G, Oberdorster E, Oberdorster J. Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles. Environ. Health Perspect. 2005; 113:823–839. [PubMed: 16002369]
13. Atkinson, A.; Donev, A. Optimum Experimental Designs. Clarendon Press; Oxford: 1992.
14. Seber, GAF.; Wild, CJ. Nonlinear Regression. John Wiley & Sons; New York: 2003.
15. Dunn G. Optimal designs for drug, neurotransmitter and hormone receptor assays. Stat. Med. 1988; 7:805–815. [PubMed: 2841743]
16. Hamilton R, Wu N, Porter D, Buford M, Wolfarth M, Holian A. Particle length-dependent titanium dioxide nanomaterials toxicity and bioactivity. Part. Fibre Toxicol. 2009; 6:35. [PubMed: 20043844]
17. Porter D, Hubbs A, Mercer R, Wu N, Wolfarth M, Sriram K, Leonard S, Battelli L, Schwegler-Berry D, Friend S, Andrew M, Chen B, Tsuruoka S, Endo M, Castranova V. Mouse pulmonary dose– and time course–responses induced by exposure to multi-walled carbon nanotubes. Toxicology. 2010; 269:136–147. [PubMed: 19857541]
18. Kavlock RJ, Schmid JE, Setzer RW. A simulation study of the influence of study design on the estimation of benchmark doses for developmental toxicity. Risk Anal. 1996; 16:399–410. [PubMed: 8693165]
19. Krewski D, Smythe R, Fung KY. Optimal designs for estimating the effective dose in developmental toxicity experiments. Risk Anal. 2002; 22:1195–1205. [PubMed: 12530789]
20. Kuljus K, Rosen DV, Sand S, Victorin K. Comparing experimental designs for benchmark dose calculations for continuous endpoints. Risk Anal. 2006; 26:1031–1043. [PubMed: 16948695]
21. Dette H, Pepelyshev A, Wong WK. Optimal designs for dose-finding experiments in toxicity studies. Bernoulli. 2009; 15:124–145.
22. Schulz M, Endrenyi L. Design of experiments for estimating parameters with unknown heterogeneity of the error variance. Proceedings of the American Statistical Association, Computing Section. :177–181.
23. DasGupta A, Mukhopadhyay S, Studden W. Compromise designs in heteroscedastic linear models. J. Stat. Plann. Inference. 1992; 32:363–384.
24. Wong W. A unified approach to the construction of mini-max designs. Biometrika. 1992; 79:611–619.
25. Wong W. Multifactor G-optimal designs with heteroscedastic errors. J. Stat. Plann. Inference. 1994; 40:127–133.
26. Atkinson A, Cook R. D-optimum designs for heteroscedastic linear models. J. Am. Stat. Assoc. 1995; 90:429.
27. Dette H, Wong W. Bayesian D-optimal designs on a fixed number of design points for heteroscedastic polynomial models. Biometrika. 1998; 85:869–882.
28. Dette H, Studden W. Optimal designs with respect to Elfving's partial minimax criterion in polynomial regression. Ann. Inst. Stat. Math. 1998; 46:389–403.
29. Wiens D. Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. J. Am. Stat. Assoc. 1998; 93:1440–1450.

30. Brown L, Wong W. An algorithmic construction of optimal minimax designs for heteroscedastic linear models. J. Stat. Plann. Inference. 2000; 85:103–114.
31. Tack L, Goos P, Vandebroek M. Efficient Bayesian designs under heteroscedasticity. J. Stat. Plann. Inference. 2002; 104:469–483.
32. Dette H, Bretz F, Pepelyshev A, Pinheiro J. Optimal designs for dose-finding studies. J. Am. Stat. Assoc. 2008; 103:1225–1237.
33. Benchmark Dose Software, version 2.3.1. U.S. Environmental Protection Agency. 2012
34. Gelman A, Chew G, Shnaidman M. Bayesian Analysis of serial dilution assays. Biometrics. 2004; 60:407–417. [PubMed: 15180666]
35. Zeng Q, Davidian M. Bootstrap-adjusted calibration confidence intervals for immunoassay. J. Am. Stat. Assoc. 1997; 92:278–290.
36. Carroll, R.; Ruppert, D. Transformation and Weighting in Regression. Chapman and Hall; New York: 1988.
37. Davidian, M.; Giltinan, DM. Nonlinear Models for Repeated Measurement Data. Chapman & Hall/CRC; Boca Raton, FL: 1995.
38. Higgins KM, Davidian M, Chew G, Burge H. The effect of serial dilution error on calibration inference in immunoassay. Biometrics. 1998:19–32. [PubMed: 9544505]
39. Davidian M, Giltinan DM. Some general estimation methods for nonlinear mixed-effects model. J. Biopharm. Stat. 1993; 3:23–55. [PubMed: 8485535]
40. Filipsson A, Sand S, Nilsson J, Victorin K. The benchmark dose method-review of available models and recommendations for application in health risk assessment. Crit. Rev. Toxicol. 2003; 33:505–542. [PubMed: 14594105]
41. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. Chapman & Hall; New York: 1994.
42. Davison, A.; Hinkley, D. Bootstrap Methods and Their Application. Cambridge University Press; Cambridge: 1997.
43. Jennison, C.; Turnbull, BW. Group Sequential Methods: Applications to Clinical Trials. Chapman & Hall/CRC; Boca Raton, FL: 1999.
44. Freund, R.; Mohr, D.; Wilson, W. Statistical Methods. Academic Press; New York: 2010.
45. Kelton, W.; Law, A. Simulation Modeling and Analysis. McGraw Hill; Boston, MA: 2000.
46. Porter D, et al. Differential mouse pulmonary dose and time course responses to titanium dioxide nanospheres and nanobelts. Toxicol. Sci. 2013; 131:179–193. [PubMed: 22956629]
47. McGill R, Tukey J, Larsen W. Variations of box plots. Am. Stat. 1978; 32:12–16.

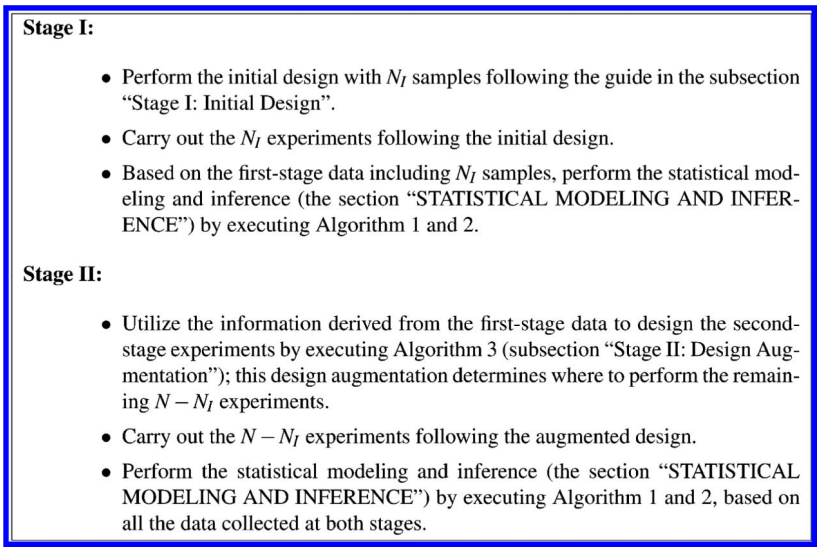

Figure 1.
Overview of the two-stage procedure.

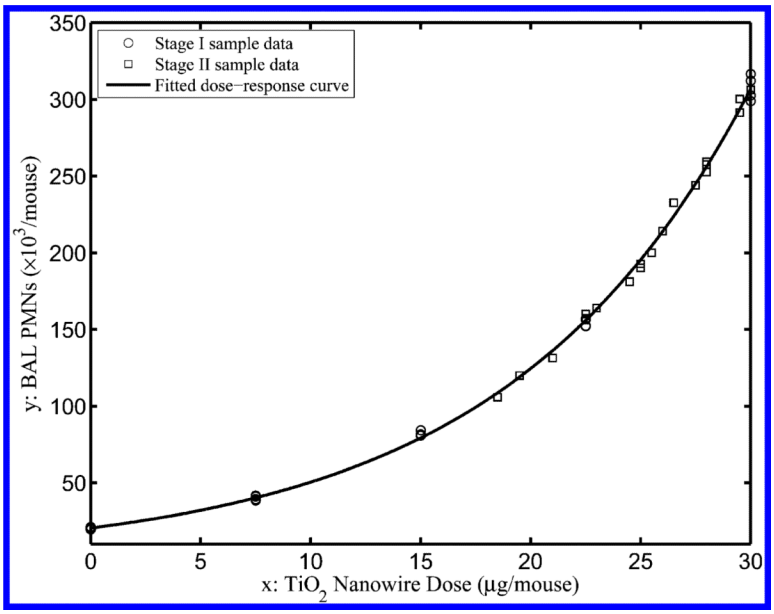

Figure 2.
Fitted dose–response model along with the simulation data.

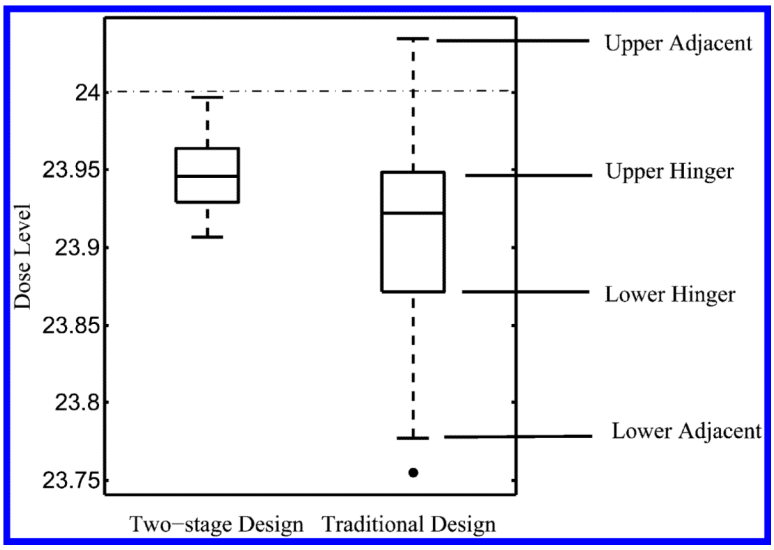

Figure 3.
Box plots for the BMDLs resulting from the two-design methods.

Table 1

Design of Experiments in Stage I

x_1	x_3	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	0	0	0	7.5	7.5	7.5	7.5	15	15

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
15	15	22.5	22.5	22.5	22.5	30	30	30	30

Table 2

Simulation Response Data in Stage I

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
21.1	19.6	20.2	19.8	41.1	38.5	41.7	39.3	84.3	80.9

y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	y_{18}	y_{19}	y_{20}
80.8	81.7	152.3	156.6	152.1	156.6	302.5	298.9	312.3	316.7

Table 3

Design of Experiments in Stage II

x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
27.5	29.5	30	29.5	28	28	23	26	25	26.5

x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}	x_{38}	x_{39}	x_{40}
28	22.5	25	24.5	22.5	22.5	21	19.5	25.5	18.5

Table 4

Simulation Response Data in Stage II

y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	y_{26}	y_{27}	y_{28}	y_{29}	y_{30}
244.0	300.3	306.4	291.4	257.6	252.7	164.0	214.2	190.2	232.6

y_{31}	y_{32}	y_{33}	y_{34}	y_{35}	y_{36}	y_{37}	y_{38}	y_{39}	y_{40}
259.4	157.5	192.7	181.1	156.2	160.0	131.3	119.9	200.2	105.8

Algorithm 1

Inputs: Data set $\{(x_i, y_i), i = 1, \ldots, I\}$

Process:

Step 0: Fit the dose-response model by performing regular (unweighted) least squares regression on the data $\{(x_i, y_i), i = 1, \ldots, I\}$, and denote the fitted model as $f\left(x, \hat{\boldsymbol{\theta}}^{(0)}\right)$; set $r = 1$.

Step 1: Fit the variance model by performing generalized least squares (GLS) regression on the squared residuals $\left\{\left(y_i - f\left(x_i, \hat{\boldsymbol{\theta}}^{(r-1)}\right)\right)^2; i = 1, 2, \ldots, I\right\}$ with the weights following $h^{-4}\left(f\left(x, \hat{\boldsymbol{\theta}}^{(r-1)}\right), \hat{\boldsymbol{\gamma}}^{(r-1)}\right)$, and denote the resulting variance model as $h\left(f\left(x, \hat{\boldsymbol{\theta}}^{(r-1)}\right), \hat{\boldsymbol{\gamma}}^{(r-1)}\right)$

Step 2: Fit the dose-response model by performing GLS regression on $\{(x_i, y_i), i = 1, \ldots, I\}$ with the weights following $h^{-2}\left(f\left(x, \hat{\boldsymbol{\theta}}^{(r-1)}\right), \hat{\boldsymbol{\gamma}}^{(r-1)}\right)$, and denote the fitted model as $f\left(x, \hat{\boldsymbol{\theta}}^{(r)}\right)$

Step 3: Set $r = r + 1$ and repeat Step 1–2 for a fixed number of times or until convergence.

Step 4: Denote $f(x, \hat{\boldsymbol{\theta}})$ and $h(f, (x, \hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\gamma}})$ and $h(f, (x, \hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\gamma}})$ as the final estimated models from the iterations above. Estimate $\hat{\sigma}$ by the usual weighted residual mean squares evaluated at those final estimates.

Outputs: The fitted dose-response model $f(x, \hat{\boldsymbol{\theta}})$ and the variance model $\hat{\sigma}h(f(x, \hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\gamma}})$

Algorithm 2

Inputs: (a) The design points at which data have been (or are assumed to have been) collected $\{x_i;\ i = 1, 2, \ldots, M\}$, with the number of design points being M; (b) the dose-response model $f(x, \hat{\boldsymbol{\theta}})$ and the variance model $\hat{\sigma}h(f(x, \hat{\boldsymbol{\theta}}), \hat{\boldsymbol{\gamma}})$, both estimated by feeding the real data $\{(x_i, y_i);\ i = 1, 2, \ldots, I\}$ to Algorithm 1; (c) the resampling size B

Process:

FOR $b = 1$ to B

FOR $i = 1$ to M

(i) set $x^*_{i,b} = x_i$;

(ii) randomly sample $e^*_{i,b}$ from the standard normal distribution $N(0,1)$;

(iii) set $y^*_{i,b} = f\left(x^*_{i,b}, \hat{\boldsymbol{\theta}}\right) + \hat{\sigma}h\left(f\left(x^*_{i,b}, \hat{\boldsymbol{\theta}}\right), \hat{\boldsymbol{\gamma}}\right)e^*_{i,b}$

END FOR

Feed $\left\{\left(x^*_{i,b}, y^*_{i,b}\right);\ i = 1, 2, \ldots, M\right\}$ to Algorithm 1, and obtain the resulting dose-response model as $f\left(x, \hat{\boldsymbol{\theta}}^*_b\right)$

END FOR

Outputs: B fitted models $\left\{f\left(x, \hat{\boldsymbol{\theta}}^*_b\right);\ b = 1, 2, \ldots, B\right\}$, from which the variabilities $\mathrm{Var}[\hat{\boldsymbol{\theta}}]$ and $\mathrm{Var}[\widehat{\mathrm{BMD}}]$ can be estimated.

Algorithm 3

Inputs: (a) $\mathrm{D}^{(I)} = \{x_i;\ i = 1, 2, \ldots, N_I\}$, the first stage design; (b) the dose-response model $f\left(x, \hat{\boldsymbol{\theta}}^{(I)}\right)$ and the variance model $\hat{\sigma}^{(I)} h\left(f, \left(x, \hat{\boldsymbol{\theta}}^{(I)}\right), \hat{\boldsymbol{\gamma}}^{(I)}\right)$, which are estimated from the first-stage samples $\{(x_i, y_i);\ i = 1, 2, \ldots, N_I\}$ by applying Algorithm 1; (c) $\left\{x_g^{(Grid)};\ g = 1, 2, \ldots, G\right\}$, the evenly spaced grid points in $[x_L, x_U]$ that serve as the pool of candidate design points; (d) the resampling size B; (e) the pre-specified BMR, which is only needed if $\mathrm{Var}[\widehat{\mathrm{BMD}}]$ is the design criterion in (14).

Process: Initialization: $\mathrm{D} = \mathrm{D}^{(I)}$

(1)FOR $k = 1$ to $N - N_I$

(2)FOR $g = 1$ to G

Evaluate the design criterion in the optimization problem (14), assuming that an experimental data set has been collected at the design $\mathrm{D} \cup \left\{x_g^{(Grid)}\right\}$:

(i) **Call** Algorithm 2 with the following three inputs: the design points $\mathrm{D} \cup \left\{x_g^{(Grid)}\right\}$; the dose-response model $f\left(x, \hat{\boldsymbol{\theta}}^{(I)}\right)$ and variance model $\hat{\sigma} h\left(f\left(x, \hat{\boldsymbol{\theta}}^{(I)}\right), \hat{\boldsymbol{\gamma}}^{(I)}\right)$; and the resampling size B.

Return B fitted models

$$\left\{f\left(x, \hat{\boldsymbol{\theta}}_b^{*(g)}\right);\ b = 1, 2, \ldots, B\right\}. \quad (15)$$

(ii) Based on the bootstrap estimates (15), evaluate the design criterion for the candidate design $\mathrm{D} \cup \left\{x_g^{(Grid)}\right\}$. Specifically, Formula (7)–(8) are used to obtain $\widehat{\mathrm{Var}}\left[\widehat{\mathrm{BMD}}^{(g)}\right]$.

(2)END FOR

Set x_{N_I+k} as the grid point that gives the minimum value of $\widehat{\mathrm{Var}}\left[\widehat{\mathrm{BMD}}^{(g)}\right]$; $g = 1, 2, \ldots, G\}$. Update $\mathrm{D} = \mathrm{D} \cup \left\{x_{N_I+k}\right\}$.

(1)END FOR

Outputs: The design D, which is equal to $\mathrm{D}^{(I)} \cup \left\{x_{N_I+1}, x_{N_I+2}, \ldots, x_N\right\}$, the union of the first-stage design $\mathrm{D}^{(I)}$ and the second-stage design points $\{x_{N_I+1}, x_{N_I+2}, \ldots, x_N\}$.