Published in final edited form as:

Risk Anal. 2012 October ; 32(10): 1784–1797. doi:10.1111/j.1539-6924.2012.01834.x.

Optimal Design for the Precise Estimation of an Interaction Threshold: The Impact of Exposure to a Mixture of 18 Polyhalogenated Aromatic Hydrocarbons

Sharon D. Yeatts^{1,*}, Chris Gennings^{2,3}, and Kevin M. Crofton⁴

¹Division of Biostatistics and Epidemiology, Department of Medicine, Medical University of South Carolina, Charleston, SC, USA

²Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA

³Solveritas, L.L.C, Richmond, VA, USA

⁴Neurotoxicology Division, National Health and Environmental Effects Research Laboratory, Office of Research and Development, U.S. EPA, Research Triangle Park, NC, USA.

Abstract

Traditional additivity models provide little flexibility in modeling the dose–response relationships of the single agents in a mixture. While the flexible single chemical required (FSCR) methods allow greater flexibility, its implicit nature is an obstacle in the formation of the parameter covariance matrix, which forms the basis for many statistical optimality design criteria. The goal of this effort is to develop a method for constructing the parameter covariance matrix for the FSCR models, so that (local) alphabetic optimality criteria can be applied. Data from Crofton *et al.* are provided as motivation; in an experiment designed to determine the effect of 18 polyhalogenated aromatic hydrocarbons on serum total thyroxine (T₄), the interaction among the chemicals was statistically significant. Gennings *et al.* fit the FSCR interaction threshold model to the data. The resulting estimate of the interaction threshold was positive and within the observed dose region, providing evidence of a dose-dependent interaction. However, the corresponding likelihood-ratio-based confidence interval was wide and included zero. In order to more precisely estimate the location of the interaction threshold, supplemental data are required. Using the available data as the first stage, the Ds-optimal second-stage design criterion was applied to minimize the variance of the hypothesized interaction threshold. Practical concerns associated with the resulting design are discussed and addressed using the penalized optimality criterion. Results demonstrate that the penalized Ds-optimal second-stage design can be used to more precisely define the interaction threshold while maintaining the characteristics deemed important in practice.

© 2012 Society for Risk Analysis

* Address correspondence to Sharon D. Yeatts, Division of Biostatistics and Epidemiology, Medical University of South Carolina, 135 Cannon St., Suite 303, P.O. Box 250835, Charleston, SC 29425, USA; tel: (843) 792-0178; fax: (843) 876-1126; yeatts@musc.edu..

DISCLAIMER

This manuscript has been reviewed by the National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Keywords

Dose-dependent interaction; implicit differentiation; implicit models; optimal design; synergy

1. INTRODUCTION

Humans are exposed to mixtures of chemicals on a daily basis, and the number of chemicals involved may be quite large.^(1,2) Dose–response and toxicity information, however, often exist for the single chemicals in a mixture, but not for the mixture itself.⁽³⁾ Clearly, the study of all possible mixture combinations is not possible. Teuschler *et al.*⁽²⁾ state that “the mixture components that are tested and their relative proportions in the mixture should reflect those seen in environmental samples.” Following the work of Gennings *et al.*⁽⁴⁾ and Casey *et al.*,⁽⁵⁾ the ray design is a useful method for focusing experimental resources on a particular ratio of chemicals.

When a mixture is under investigation for regulatory purposes, the analysis is often based on the assumption of additivity among the chemicals in the mixture.⁽⁶⁾ However, the effects of complex mixtures can be additive or interactive (i.e., synergistic or antagonistic⁽⁷⁾). The basic concept of an interaction is that the presence of one chemical alters the biological activity of another. Statistically, this translates to a change in the slope of the dose–response relationship for one chemical due to the presence of the other chemical.^(2,8) If the slope of the dose–response relationship is not affected by the presence of another chemical, then the relationship is said to be additive.

The effects of chemical mixtures on a biological endpoint is often complex. Indeed, according to Gennings *et al.*,⁽⁴⁾ “mechanisms of . . . chemical interaction are rarely uniform across the entire dose–response range and instead tend to be highly dose-dependent.” In the regulatory process, the U.S. Environmental Protection Agency supports an assumption of low-dose additivity,⁽⁶⁾ this assumption is supported by Carpy *et al.*,⁽⁹⁾ whose literature review demonstrated that interactions are rare at low exposure levels, suggesting that the occurrence of interactions is generally in a high exposure region, and Könnemann and Pieters,⁽¹⁰⁾ who stated that interactions “may not be very relevant in the low-dose region.” This assumption of low-dose additivity and high-dose interaction implies the existence of an interaction threshold, a boundary dividing the region of additivity from the interaction region. In 2002, an expert panel, commissioned by the Society of Toxicology to suggest directions for future research in the risk assessment of chemical mixtures, stressed the importance of using low-dose exposure in mixtures experiments in an effort to demonstrate additivity at low exposure levels.⁽²⁾ As a result, the estimation of an interaction threshold is an important aspect of the analysis of complex chemical mixtures.

Methods for predicting additivity typically rely on a single model that includes terms for each of the active single chemicals. The additivity model is assumed to adequately represent the single chemical data, as it is often used to describe the dose–response relationships of the single chemicals. Often, such a model implies that certain features of the model, a plateau parameter for instance, are similar across all of the single chemicals;⁽⁵⁾ however, this may not be the case. To allow more flexibility in adequately modeling the dose–response

relationships of the single chemicals, more flexible approaches have been developed.^(4,11) The flexible single chemical required (FSCR) method, detailed by Gennings *et al.*,⁽⁴⁾ allows for the estimation of different range and dose threshold parameters for each of the single chemicals and the fixed-ratio mixture ray. Additivity is then implicitly defined according to Berenbaum's interaction index, which can be represented using the parameters associated with the single chemical dose–response relationships. Gennings *et al.* demonstrated the FSCR additivity model for a mixture of six chemicals⁽⁴⁾ and the FSCR interaction threshold model for a mixture of 18 polyhalogenated aromatic hydro-carbons (PHAHs);⁽¹²⁾ associated inference was based on a likelihood-ratio test of additivity. Hass *et al.*⁽¹³⁾ followed the approach described in Altenberger *et al.*,⁽¹¹⁾ but included bootstrap confidence intervals to compare the additivity curve to the observed mixture data.

Statistical optimality criteria can be used to determine the optimal experimental design for such nonlinear models in order to achieve the most statistically efficient allocation of resources, in terms of both the location of the design points and the proportion of subjects allocated to each point. These criteria can be applied with a focus on the complete vector of model parameters, or some subset thereof, in order to obtain precise parameter estimates or achieve a target statistical power. Such considerations may be particularly important for complex nonlinear models such as those in the FSCR family, wherein an equal allocation of subjects among design points evenly spaced throughout the dose range may not be optimal, or worse, may not support the fit of the model. A commonly used design criterion is the D-optimality criterion, which is based on the parameter covariance matrix. These implicit models, however, cannot be written in such a way that the mean is an explicit function of the data. As a result, we cannot explicitly differentiate with respect to the model parameters in order to construct the parameter covariance matrix. The goal of this article is to develop a method for constructing the parameter covariance matrix, so that we can determine the optimal experimental design for supporting the FSCR interaction threshold model.

2. MOTIVATING EXAMPLE

Crofton *et al.*⁽¹⁴⁾ conducted a study of a mixture of 18 PHAHs. The primary objective of the study was to determine if additivity adequately described the outcome for the mixture of 18 PHAHs, which were combined according to a fixed-ratio mixture ray. A secondary goal of the analysis was to determine whether the departure from additivity, if significant, was dose-dependent. The hypothesis of additivity was tested using the FSCR methodology detailed by Gennings *et al.*^(4,12) and reviewed here for demonstration of the model's implicit nature.

Recall the situation where we have single chemical data available for each of the c chemicals in the mixture, as well as mixture data taken along the fixed-ratio ray of interest, for a total of $c + 1$ rays. Without loss of generality, let the scaled FSCR additivity model be described as follows:

$$\mu = \begin{cases} \begin{cases} 100, & x < \delta_i \\ \alpha_i + (100 - \alpha_i) \exp \left\{ \frac{\beta_i}{SF_i} (x - \delta_i) \right\}, & x \geq \delta_i \end{cases}, & i=1, \dots, c, \\ \mu_{(add)} \epsilon^{t_{(add)}} (\mu_{(add)}) = \left[\sum_{i=1}^c \frac{a_i}{\frac{SF_i}{\beta_{i(add)}} \ln \left(\frac{\mu_{(add)} - \alpha_{i(add)}}{(100 - \alpha_{i(add)})} \right) + \delta_{i(add)}} \right]^{-1}, & i=c+1. \end{cases} \quad (1)$$

where $i = 1, \dots, c$ represents the ray for each of the c single chemicals, and $c + 1$ represents the fixed-ratio mixture ray under consideration. Here, x represents the dose of the i th chemical for $i = 1, \dots, c$ or the total dose of the mixture for $i = c + 1$. Along the i th ray, the parameter β_i is associated with the slope, δ_i represents the unknown threshold, and α_i represents the unknown minimum response. The model for the mixture ray ($i = c + 1$) defines an additive relationship among the single chemicals. The subscript (add) indicates that these parameters are estimated under this constraint of additivity, using both single chemical and fixed-ratio mixture ray data. As written, Equation (1) specifies a maximum response of 100; the model can easily be modified to include an unknown range parameter, γ_i , such that $\alpha_i + \gamma_i$ represents the maximum response.

The scaling factor, represented by SF_i , is determined separately for each of the c single chemicals and the mixture and is included to improve the properties of the parameter covariance matrix if necessary. If the dose scales are similar across the c chemicals, such that the scaling factor is unnecessary, let SF_i equal one for all i . In this case, the scaled FSCR additivity model given in Equation (1) reduces to the FSCR additivity model given in Gennings *et al.*,⁽⁴⁾ and β_i represents the slope of the i th chemical. Scaling the slope parameter does not affect the prediction of the model; it reduces the range of the variance associated with the slope parameters.

Along the fixed-ratio mixture ray, let Δ be the interaction threshold parameter, and let μ_{Δ} be the mean response at the interaction threshold as predicted under additivity. Then, the scaled FSCR interaction threshold model is written as follows:

$$\mu = \begin{cases} \begin{cases} 100, & x < \delta_i \\ \alpha_i + (100 - \alpha_i) \exp \left\{ \frac{\beta_i}{SF_i} (x - \delta_i) \right\}, & x \geq \delta_i \end{cases}, & i=1, \dots, c, \\ \mu_{(add)} \epsilon^{t_{(add)}} (\mu_{(add)}) = \left[\sum_{i=1}^c \frac{a_i}{\frac{SF_i}{\beta_{i(add)}} \ln \left(\frac{\mu_{(add)} - \alpha_{i(add)}}{(100 - \alpha_{i(add)})} \right) + \delta_{i(add)}} \right]^{-1}, & t < \Delta \\ \alpha_{mix} + (\mu_{\Delta} - \alpha_{mix}) \exp \{ \beta_{mix} (t - \Delta) \}, & t \geq \Delta \end{cases}, \quad i=c+1. \end{cases} \quad (2)$$

In this parameterization, for total doses of the mixture less than Δ , the model describes an additive relationship among the chemicals in the mixture. For total doses larger than Δ , the model allows for interaction.

In Equations (1) and (2) above, the implicit nature of the additivity region is evident. This region of the model cannot be written in such a way that the mean is an explicit function of

the dose and model parameters. Instead, the mean response under the assumption of additivity is determined implicitly via the interaction index.

In our motivating example, the design of the experiment was sufficient to support the complicated FSCR interaction threshold model, which is presented in Fig. 1. It was shown that significant departure from additivity exists along the ray. However, although the estimate of the interaction threshold was positive, the quasi-likelihood ratio-based confidence interval was wide and included zero. Therefore, there is insufficient evidence to support the conclusion that the interaction is dose-dependent.

Since the model adequately describes the dose–response relationship of the mixture, there are two possible explanations for the width of the confidence interval. One explanation is that the true interaction threshold is zero, indicating that the interaction along the fixed-ratio mixture ray is not dose-dependent. The other explanation is more problematic: the experiment performed may have been insufficient to provide reasonable power for detection, or to allow the precise estimation, of the interaction threshold. If this is the case, then the second goal has not been met; the data do not permit an adequate answer to the question at hand.

In the analysis of this mixture, we are without a true answer to the question of the dose-dependent nature of the interaction. We can say that we don't have enough evidence to conclude that the interaction is dose-dependent. There is some evidence, however, since the estimate of the interaction threshold is positive. Therefore, we are interested in increasing the precision associated with the estimate of the interaction threshold by adding design points to the original experiment, which we consider the first-stage data. Our intention is to design a second-stage experiment, which is characterized by the location and allocation of these additional design points.

To begin, we developed a method for constructing the parameter covariance matrix associated with the FSCR models via implicit differentiation. The parameter covariance matrix was then used to calculate the statistical optimality criterion for design development. Of particular interest is the situation in which single chemical data, used to define additivity along the ray, are already available. Statistical optimality criteria can then be employed to find the optimal mixture design for supporting the FSCR interaction threshold model. Alternatively, it may be the case that mixture data are already available, as in the motivating example, but do not permit an adequate answer to the question at hand. In this case, the available mixture data may be considered a first-stage experiment, and statistical optimality criteria can be employed to find the optimal second-stage design for addressing the primary hypothesis.

3. METHODS

3.1. Estimation and Inference Associated with FSCR Models

Let $\phi = [\theta]$ represent the $p \times 1$ vector of model parameters, where θ represents the $(p - 1) \times 1$ vector of model parameters excluding τ , such that $E\{y\} = \mu(\phi)$ and $\text{var}\{y\} = \tau \mathbf{V}$.

Following Gennings *et al.*,⁽¹²⁾ a Nelder–Mead direct search algorithm containing an embedded

bisection algorithm is used in order to estimate the model parameters. The quasi-likelihood ratio test can be used to compare two nested models, as described in Gennings *et al.*,⁽¹²⁾ and quasi-likelihood ratio confidence intervals can be determined. Although these methods allow us to test for interaction among the chemicals in the mixture ray, they do not allow us to implement traditional local optimality criteria in the design stage. For design purposes, a method for computation of the covariance matrix of the parameter estimates is preferable; such a method would also allow us to compute Wald-type test statistics and confidence intervals.

Let the matrix $\mathbf{F}$ represent the matrix of derivatives of the mean response with respect to the model parameters, such that $\mathbf{F} = [\boldsymbol{\mu}' \boldsymbol{\phi}']$ is an $n \times p$ matrix. Then, the asymptotic covariance matrix⁽¹⁵⁾ is

$$\text{var}\{\hat{\boldsymbol{\phi}}\} = \tau(\mathbf{F}'\mathbf{V}^{-1}\mathbf{F})^{-1}. \quad (3)$$

The variance associated with the interaction threshold is given by the (p, p) th element of the parameter covariance matrix, $\text{var}\{\hat{\boldsymbol{\phi}}\}$.

3.2. Implicit Differentiation

The computation of the covariance matrix requires that the first derivatives of the model with respect to the model parameters, $\mathbf{F}$, be specified. In the case of the single chemicals and mixture doses beyond the interaction threshold, where the model is explicitly stated, these derivatives can be found in the usual way. The additivity region of the model, however, is an implicit function, and the derivatives with respect to the model parameters are not readily available in this region.

We employed implicit differentiation to construct the parameter covariance matrix. Following Finney *et al.*,⁽¹⁶⁾ we can take the derivative of both sides of the implicit function with respect to each of the model parameters. The result of this implicit differentiation is that we arrive at equations that can be solved to obtain the derivative of the mean with respect to the model parameters. The resulting derivatives are used to construct the estimated parameter covariance matrix for the scaled FSCR interaction threshold model. In the following subsection, we demonstrate the process for a mixture of two chemicals.

3.2.1. Implicit Differentiation of the Scaled FSCR Additivity Model for a Mixture of $c = 2$ Chemicals—Without loss of generality, consider the situation where $c = 2$, so that, for $t < \tau$, the additivity region of the model given in Equation (1) is written as follows:

$$\mu_{add} = \mu + t(\mu) \left[\sum_{i=1}^2 \frac{a_i}{\frac{SF_i}{\beta_i} \ln\left(\frac{\mu - \alpha_i}{(100 - \alpha_i)}\right) + \delta_i} \right]^{-1}.$$

Let $\xi_1 = \frac{SF_1}{\beta_1} \ln\left(\frac{\mu - \alpha_1}{(100 - \alpha_1)}\right) + \delta_1$ and $\xi_2 = \frac{SF_2}{\beta_2} \ln\left(\frac{\mu - \alpha_2}{(100 - \alpha_2)}\right) + \delta_2$. The additivity region of the model can be rewritten in the following manner:

$$\mu_{add} = \left\{ \mu - \epsilon - t(\mu) = \left[\frac{a_1}{\xi_1} + \frac{a_2}{\xi_2} \right]^{-1} = \left[a_1 \xi_1^{-1} + a_2 \xi_2^{-1} \right]^{-1} \right\}. \quad (4)$$

To find the derivative of the mean with respect to α_1 , we implicitly differentiate both sides of Equation (4) with respect to α_1 , as follows:

$$\begin{aligned} \frac{\partial t(\mu)}{\partial \alpha_1} &= \frac{\partial \left[a_1 \xi_1^{-1} + a_2 \xi_2^{-1} \right]^{-1}}{\partial \alpha_1} \\ 0 &= - \left[a_1 \xi_1^{-1} + a_2 \xi_2^{-1} \right]^{-2} \times \left[\left(-a_1 \xi_1^{-2} \right) \frac{\partial \xi_1}{\partial \alpha_1} - a_2 \xi_2^{-2} \frac{\partial \xi_2}{\partial \alpha_1} \right]. \end{aligned} \quad (5)$$

To complete this equation, we can differentiate ξ_1 and ξ_2 with respect to α_1 in the following manner:

$$\frac{\partial \xi_1}{\partial \alpha_1} = \frac{SF_1}{\beta_1} \left(\frac{100 - \alpha_1}{\mu - \alpha_1} \right) * \left[(\mu - \alpha_1) \left((100 - \alpha_1)^2 \right) + \left((100 - \alpha_1)^{-1} \right) \left(\frac{\partial \mu_{add}}{\partial \alpha_1} - 1 \right) \right].$$

As defined, ξ_i is a function of only α_i , β_i , and δ_i . Because ξ_i represents the $ED_i(\mu)$, the μ contained within ξ_i is associated only with the i th chemical. Therefore, $\frac{\partial \xi_2}{\partial \alpha_1} = \frac{\partial \xi_2}{\partial \beta_1} = \frac{\partial \xi_2}{\partial \delta_1} = 0$.

We can substitute these derivatives into (5) and solve the resulting equation for $\frac{\partial \mu_{add}}{\partial \alpha_1}$. The process can be similarly carried out to find $\frac{\partial \mu_{add}}{\partial \alpha_2}$, $\frac{\partial \mu_{add}}{\partial \beta_1}$, $\frac{\partial \mu_{add}}{\partial \beta_2}$, $\frac{\partial \mu_{add}}{\partial \delta_1}$, and $\frac{\partial \mu_{add}}{\partial \delta_2}$.

4. APPLICATION

4.1. Data Source

Data used in this analysis were from Crofton *et al.*⁽¹⁴⁾ Long-Evans rats were orally dosed with varying concentrations of 18 different PHAHs (two dibenzodioxins, four dibenzofurans, and 12 polychlorinated biphenyls (PCBs), including dioxin-like and nondioxin-like PCBs) for four consecutive days. Serial dilutions of a custom-synthesized mixture were also tested; the mixture was constructed such that the ratio of chemicals was based on environmental concentrations. Serum total T4 was measured 24 hours after the last dose via radioimmunoassay. A full description of the study can be found in Crofton *et al.*⁽¹⁴⁾

A thorough analysis of the mixture of 18 PHAHs was presented in Gennings *et al.*⁽¹²⁾ As described by Gennings *et al.*, the minimum response parameters were combined to arrive at a more parsimonious model. The scaled FSCR full model was fit to the single chemicals and the fixed-ratio mixture ray simultaneously. Nonunity scaling factors were used for each single chemical with a dose range larger than 1,000 $\mu\text{g/kg}$. Due to the inclusion of the scaling factors, the analysis on which our design work is based differs slightly from that presented in Gennings *et al.*; the associated model fit and prediction, however, are consistent. The empirical dose–response relationship is described in Fig. 1. The predicted interaction threshold is 105.3 $\mu\text{g/kg}$. Recall that the objective is to find the optimal second-stage design for the precise estimation of the interaction threshold. It should be noted that we have both single chemical and mixture data to support the hypothesized model

parameters. The estimates of the model parameters resulting from the analysis of the first-stage data serve as the hypothesized parameter values for the design work.

4.2. Determination of the Optimal Design

Let n_s be the number of observations available from the single chemical experiments, and let $n_{mix_original}$ be the number of observations available from the first-stage mixture experiment. Neither the single chemical data nor the mixture data from the first-stage experiment will be affected by our design work. Rather, we are interested in constructing an optimal design to supplement the data already available. Let n_{mix_design} be the number of observations to be included in the second-stage design, such that $N = n_s + n_{mix_original} + n_{mix_design}$ is the total number of observations to be used in the final analysis.

The Nelder–Mead Simplex Algorithm was used to find the optimal design by optimizing the appropriate design criterion. The algorithm searches to find the optimal location of the design points, as well as the proportion of observations allocated to each design point. The first dose point in each design is always zero, though the proportion of subjects included at the control dose iterates within the search algorithm. The location and allocation of observations that optimize the design criterion is the optimal m point design. The design is

represented by $\begin{Bmatrix} \mathbf{t} \\ \mathbf{q} \end{Bmatrix}$, where $\mathbf{t}$ is the corresponding vector of design points, and $\mathbf{q}$ is the vector of corresponding sample allocations.

In the case where the optimal m point design required fewer than m dose groups, as a result of overlapping dose locations or negligible allocations, we reduced our grid of starting values to an $m-1$ point design. The number of design points was reduced until the design could be reduced no further, either because all remaining design points were required or because a smaller design was considered unacceptable. Issues with the practicality of the optimal design were addressed using the penalized optimality criterion,⁽¹⁷⁾ which uses desirability functions to penalize impractical experimental designs.

4.2.1. The Ds-optimal Design—The Ds-optimal design is the design that minimizes the determinant of some subset of the covariance matrix related to a parameter set of particular interest. In this case, we are interested in focusing the second-stage experiment on the estimation of the interaction threshold. As a result, the Ds-optimal second-stage design is the second-stage design that minimizes the variance of the hypothesized interaction threshold,

$$var \left\{ \hat{\Delta} \right\}.$$

4.2.2. The Penalized Ds-optimal Design—As described in Parker and Gennings,⁽¹⁷⁾ a penalty function is added to the design criterion to penalize experimental designs, based on the investigator's predefined experimental design preferences. The penalized optimal design is the design that jointly minimizes the variance of the hypothesized interaction threshold and the penalty function. Let $j = 1, \dots, k$ represent the k experimental design preferences under consideration. Each characteristic p_j is transformed into its desirability, d_j , where $0 \leq d_j \leq 1$. In this setting, $d = 0$ indicates that characteristic p_j is highly undesirable, and $d_j = 1$

indicates that characteristic p_j is highly desirable. The overall desirability of the k combined characteristics is represented by

$$D(\mathbf{t}, \mathbf{q}) = (d_1 d_2 \cdots d_k)^{1/k}, \quad (6)$$

and the penalty function is represented by $1 - D(\mathbf{t}, \mathbf{q})$. Let $h(\mathbf{t}, \mathbf{q}, f(\phi), m, N)$ be the value of an alphabetic optimality criterion, such as the Ds-optimality criterion previously discussed. In this case, $h(\mathbf{t}, \mathbf{q}, f(\phi), m, N)$ is the variance of the hypothesized interaction threshold.

4.2.3. Experimental Design Preferences—The penalized optimality criterion⁽¹⁷⁾ is based on desirability functions, which are used to capture the investigator's experimental design preferences. A logistic function was used to create the desirability functions, though other functions can be used to achieve the appropriate shape.⁽¹⁷⁾ Fig. 2 contains plots of these desirability functions. Preferences for the number of doses, the number of animals per dose, and dose spacing, were chosen as follows:

1. Number of Subjects per Dose Group:

The number of subjects per dose group can be thought of as a target approach; the allocation of between 10 and 18 subjects per dose group is fairly standard in the toxicological literature. This target region is captured in the desirability function described by the solid curve. For experimental designs with more than four design points, this target range is appropriate, since our design work is based on a second-stage sample size of 100 subjects. For experimental designs with four or fewer design points, it is acceptable to have more than 18 subjects per dose group. This expanded target region is captured in the desirability function described by the dashed curve. In addition, it is important to note that the target region may depend on the importance of the particular chemical or mixture under study. For important compounds, the overall sample size may be increased, so it may be acceptable to increase the number of subjects allowed at each design point.

2. Dose Spacing (Percent Difference between Doses):

Due to accuracy restrictions in the serial dilution method used to obtain dosing solutions,⁽¹⁴⁾ the acceptable spacing between dose groups depends on dose region. We describe dose spacing as the percent difference between two consecutive dose groups. A smaller percentage is acceptable for doses in the high-dose region than in the low-dose region. For instance, a 10% difference is more reliably detected between doses of 500 and 550 $\mu\text{g/kg}$ than between doses of 5 and 5.5 $\mu\text{g/kg}$. After careful consideration, the dose space was divided into three dose regions, and a desirability function was constructed for each region. The low-dose region requires a large percentage difference between doses, and the percentage difference required decreases as the dose increases. The desirability function for doses less than 100 $\mu\text{g/kg}$ specifies that each dose in this region must be at least 30% larger than the dose below, as shown in the solid curve. The desirability function for doses between 100 and

500 $\mu\text{g/kg}$ is a parallel shift to the left, and the desirability function for doses larger than 500 $\mu\text{g/kg}$ is a further shift to the left.

3. Dose Region:

Teuschler *et al.*⁽²⁾ commented on the importance of low-dose studies, since the dose–response relationship evident in the high-dose region may not be applicable for low doses. As such, it may be of interest to restrict the design search to doses within the low-dose region, or within some particular region of interest. In general, this desirability function represents a constraint on the dose region. In the mixture of 18 PHAHs, the FSCR analysis of the first-stage experiment indicated that the interaction threshold is no larger than 544 $\mu\text{g/kg}$. Therefore, the desirability function representing dose region restricts the placement of dose groups larger than 550 $\mu\text{g/kg}$.

4. Percentage of the Sample below the Interaction Threshold:

If the nonlinear model of interest contains one or more threshold parameters, used to define two or more regions of the model, it may be of interest to place dose groups in each region. In the mixture of 18 PHAHs, we would like to be able to use the data from the second stage to validate our analysis of the first-stage experiment. After the second-stage experiment has been conducted, we are able to compare the second-stage data to the model resulting from the analysis of the first-stage data. Ideally, we are looking for the predicted model to adequately describe the data from both stages. It is important, then, that the second-stage design allocates enough of the sample below the threshold for us to inspect model fit in the additivity region. The same is true for the interaction region. The desirability function for the percentage of the sample below the interaction threshold describes a target region of 30–45%. As the number of subjects included in the second-stage experiment increases, this desirability curve may shift to the left. For experiments with a large sample size, we may require a smaller percentage of the sample allocated to a particular dose region.

5. Percentage of Design Points below the Interaction Threshold:

This design preference is somewhat related to that described in⁽⁴⁾ above the percentage of the sample below the interaction threshold. If we are interested in allocating a certain percentage of the sample below the interaction region, for validation purposes, we may want to ensure that a certain percentage of the second-stage design points are located below the interaction threshold. This is to guard against a situation in which 35% of the second-stage sample is allocated below the interaction threshold but to the same design point. In an experiment with less than 10 groups, a roughly equal distribution of design points above and below the interaction threshold is appropriate. This gives us a number of validation points in the additivity region. The desirability function for the percentage of design points below the interaction threshold describes a target region of 40–55%. As the number of dose groups increases, however, this curve will likely shift to the left. For experiments with a large number of

design points, we would require a smaller percentage of those design points to be located in a particular dose region.

The sample size and the number of design points to use in the experiment are questions of power. The user specifies these values in the process, rather than controlling them through desirability functions. The minimum alphabetic optimality criterion depends on both the sample size available and the number of design points specified. Because these values are prespecified and fixed for a given design, they do not contribute to the overall desirability of the design.

5. RESULTS

5.1. The Ds-optimal Design

The Ds-optimal m point designs are given in Table I, where $\mathbf{t}$ represents the vector of total dose locations, and $\mathbf{q}$ represents the vector of corresponding proportions. As can be seen in the table, there is a tendency for dose groups to cluster at the hypothesized interaction threshold. These design points can be collapsed into one dose group at 105 $\mu\text{g/kg}$, as illustrated in Table I. Rounding doses to the hundredths place, however, it is often the case that the design points actually fall just above or just below the hypothesized interaction threshold. For the Ds-optimal eight-point design, for instance, we see that one such design point is located just below the interaction threshold, whereas the other two are located just above. In keeping with this distinction, when we round the design points to the nearest integer, the total dose just below the interaction threshold can be rounded down, whereas the total dose just above the interaction threshold can be rounded up. This design is also described in Table I and is denoted with an asterisk (*). This allows us to investigate the importance of their location relative to the interaction threshold, which is an important distinction. Combining dose groups without regard to their location relative to the interaction threshold results in a dramatic increase in the variance of the interaction threshold. Thus, the Ds-optimal second-stage design requires data both above and below the hypothesized interaction threshold in order to precisely estimate its location.

The Ds-optimal second-stage design is the design that minimizes the variance of the interaction threshold. We can see from Table I, then, that the Ds-optimal design is the Ds-optimal four-point design, which is described by

$$\left\{ \begin{array}{cccc} 0 & 98.3 & 104.5 & 105.5 \\ <0.001 & 0.009 & 0.335 & 0.656 \end{array} \right\}.$$

This design has associated variance of the hypothesized interaction threshold equal to 3,195.6 and is presented in Fig. 3. The dashed line represents the additivity model, and the hypothesized mixture model is represented by the solid line. The dots indicate data available from the first stage of the experiment, and the triangles represent the location of the design points indicated by the Ds-optimal second-stage design.

Though this design may be statistically optimal, there are some practical problems with its implementation. Since less than 0.1% of the sample is allocated to the control dose, the Ds-

optimal second-stage design does not require the use of a control group, which is a problem. A toxicologist may not be willing to conduct an experiment that does not include a control group. Practically speaking, the entire sample is allocated to the third and fourth design points. These doses are very close together, so that the design reduces to a single design point. It is unlikely, for instance, that a toxicologist will be able to distinguish a dose of 104.5 $\mu\text{g/kg}$ from a dose of 105.5 $\mu\text{g/kg}$. Many of the designs given in Table I share these practical problems.

In the next sections, we apply the penalized optimality criterion, developed by Parker and Gennings,⁽¹⁷⁾ and the desirability functions described in Section 4.2.3, to arrive at a second-stage design that is more likely to be carried out in practice. Parker⁽¹⁸⁾ admits that “although absolute optimality properties are lost, the penalized optimal design is optimal in accordance with design preferences set by the investigator and the alphabetic optimality criterion.” This trade-off is reasonable in situations such as this, where certain characteristics of the statistically optimal design may not be considered reasonable by scientific investigators.

5.2. The Penalized Optimal Design

The penalized optimal second-stage designs are given in Table II, where $\mathbf{t}$ represents the vector of total dose locations and $\mathbf{q}$ represents the vector of corresponding proportions. The Ds-optimal second-stage designs are also provided for comparison. As can be seen in Table II, the penalized optimal four-point design results in the highest desirability and the lowest variance of the penalized optimal second-stage designs. Therefore, the penalized optimal four-point design, which is described in Fig. 4, is the experimental design most practically suited for second-stage work to define the interaction threshold. The dashed line represents the additivity model, and the hypothesized mixture model is represented by the solid line. The dots indicate data available from the first stage of the experiment, and the triangles represent the location of the design points indicated by the penalized optimal second-stage design. The desirability associated with the penalized optimal second-stage design is 0.957, which is a sizeable increase over the desirability associated with the Ds-optimal second-stage design, <0.001 . In addition, the variance of the hypothesized interaction threshold associated with the penalized optimal second-stage design is a mere 7.4% larger than the variance associated with the Ds-optimal second-stage design, which is a fairly minimal increase.

The penalized optimal second-stage design, presented in Fig. 4, clearly addresses the practical issues of the Ds-optimal design. Approximately 15% of the sample is allocated to the control dose. The remaining three design points are scattered around the hypothesized interaction threshold. Their proximity to each other is controlled by the desirability functions, as is the number of subjects per design point, the location of the design points, the percentage of the sample below the interaction threshold, and the percentage of the design points below the interaction threshold. Thus, the resulting penalized optimal design has the characteristics deemed important in practice, as quantified by the desirability functions, while allowing only a minimal increase in the variance of the interaction threshold.

6. DISCUSSION

This article presents a methodology for constructing the parameter covariance matrix for implicit mixtures models, such as the FSCR interaction threshold model. This methodology allows for the development of statistically optimal designs focused on the estimation and/or statistical testing of the interaction threshold. The approach was applied to a mixture of 18 PHAHs in order to determine the Ds- and penalized Ds-optimal second-stage designs. The penalized optimal design criterion represents a compromise between a statistically optimal design, which may be impractical or even impossible to implement, and a design preferred by the investigator, which may add little additional information regarding the hypothesis of interest.

As described in Section 4.2, our application of this methodology determined the optimal second-stage design in order to supplement the single chemical and mixture data already available; the final analysis, in this case, would combine the first-stage data (both the single chemical and the mixture data) with the second-stage mixture data in order to estimate model parameters, conduct hypothesis testing, and calculate confidence intervals. Analysis of the second-stage data alone would result in suboptimal characteristics, since the first-stage data were considered fixed in the second-stage design determination. Alternatively, one could determine the second-stage design using only the parameter estimates derived from the first-stage data to inform the second-stage design determination, in which case the single chemical data and the second-stage data would be combined for final analysis.

It should be noted that, while analysis of mixtures data often combines data across experiments (i.e., single chemical and mixture experiments are often conducted sequentially), this approach assumes that experimental conditions remain constant across experiments, which may not be the case. In order to determine the applicability of combining data across experiments, positive controls could be included in the experiment for validation of this assumption but excluded from the analysis. Alternatively, the analysis might incorporate varying scale parameters for each stage in order to reflect potential changes in experimental conditions; this approach would allow for a test of coincidence to determine whether such a modification is supported by the data.

The proposed methodology details a procedure for constructing the parameter covariance matrix associated with implicit models such as the FSCR. The method was applied to existing data in order to illustrate the determination of Ds- and penalized Ds-optimal designs, and there are some limitations which should be discussed. We have not addressed the power associated with the designs proposed. Both designs are optimal in that they achieve minimum variance; the penalized Ds-optimal design achieves minimum variance given the design preferences of the investigator. However, the power associated with the design is impacted by the variance, as well as the sample size. For the 100 subjects used in this example, the proposed optimal designs fall just short of significance regarding the interaction threshold under the hypothesized conditions. The number of subjects required to achieve a target power for a given design can be determined using the methods described in Casey *et al.*⁽⁵⁾ As previously stated, it is important to consider carefully the research question when determining the experimental design, as well as the sample size. As is the

case in traditional factorial experiments, a test for interaction requires a larger sample size than a test for the associated main effects.

Limitations pertaining to the data itself are described in Crofton *et al.*⁽¹⁴⁾ In particular, the failure of the FSCR model to detect interaction in the low-dose region may be due to low statistical power, given the impact of the experimental design and sample sizes on statistical power. In addition, extrapolation from these acute dosing studies may not be applicable to chronic exposures, or to other animal models, given the uncertainty regarding variation between species.

The method and application presented here represent a useful tool for determining an experimental design suitable for precisely estimating the location of the interaction threshold. Given the assumption of low-dose additivity and high-dose interaction supported by the literature^(9,10,6), estimation of the interaction threshold remains an important goal in the risk assessment of chemical mixtures.⁽²⁾

Acknowledgments

The research presented in this article was supported in part by the National Institute of Environmental Health Sciences, National Institutes of Health grant T32 007334 (SDY). The authors thank the anonymous referees for their helpful comments on the earlier version of this manuscript.

REFERENCES

1. Monosson E. Chemical mixtures: Considering the evolution of toxicology and chemical assessment. *Environmental Health Perspectives*. 2005; 113(4):383–390. [PubMed: 15811826]
2. Teuschler L, Klaunig J, Carney E, Chambers J, Conolly R, Gennings C, Giesy J, Hertzberg R, Klaassen C, Kodell R, Paustenbach D, Yang R. Support of science-based decisions concerning the evaluation of the toxicology of mixtures: A new beginning. *Regulatory Toxicology and Pharmacology*. 2002; 36(1):34–39. [PubMed: 12383716]
3. Teuschler LK, Hertzberg RC. Current and future risk assessment guidelines, policy, and methods development for chemical mixtures. *Toxicology*. 1995; 105(2–3):137–144. [PubMed: 8571352]
4. Gennings C, Carter WH Jr, Carney EW, Charles GD, Gollapudi BB, Carchman RA. A novel flexible approach for evaluating fixed ratio mixtures of full and partial agonists. *Toxicological Sciences*. 2004; 80(1):134–150. [PubMed: 15084752]
5. Casey M, Gennings C, Carter WH Jr, Moser VC, Simmons JE. Power and sample size calculations for linear hypotheses associated with mixtures of many components using fixed-ratio ray designs. *Environmental and Ecological Statistics*. 2006; 13:11–23.
6. US EPA. Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures.. In: OoRaD National Center for Environmental Assessment. , editor. *Series Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures*. Washington, DC: 2000.
7. Berenbaum MC. The expected effect of a combination of agents: The general solution. *Journal of Theoretical Biology*. 1985; 114(3):413–431. [PubMed: 4021503]
8. Gennings C, Carter WH Jr, Carchman RA, Teuschler LK, Simmons JE, Carney EW. A unifying concept for assessing toxicological interactions: Changes in slope. *Toxicological Sciences*. 2005; 88(2):287–297. [PubMed: 16081521]
9. Carpy SA, Kobel W, Doe J. Health risk of low-dose pesticides mixtures: A review of the 1985–1998 literature on combination toxicology and health risk assessment. *Journal of Toxicology and Environmental Health, Part B: Critical Reviews*. 2000; 3(1):1–25.
10. Könnemann WH, Pieters MN. Confusion of concepts in mixture toxicology. *Food and Chemical Toxicology*. 1996; 34(11–12):1025–1031. [PubMed: 9119311]

11. Altenberger R, Backhaus T, Boedeker W, Faust M, Scholze M, Grimme LH. Predictability of the toxicity of multiple chemical mixtures to vibrio fischeri: Mixtures composed of similarly acting chemicals. *Environmental Toxicology and Chemistry*. 2000; 19(9):2341–2347.
12. Gennings C, Carter WH Jr, Carchman RA, DeVito MJ, Simmons JE, Crofton KM. The impact of exposure to a mixture of eighteen PHAHs on thyroid function: Estimation of an interaction threshold. *Journal of Agricultural, Biological and Environmental Statistics*. 2007; 12(1):96–111.
13. Hass U, Scholze M, Christiansen S, Dalgaard M, Vinggaard AM, Axelstad M, Metzdorff SB, Kortenkamp A. Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. *Environmental Health Perspectives*. 2007; 115(Suppl 1):122–128. [PubMed: 18174960]
14. Crofton KM, Craft ES, Hedge JM, Gennings C, Simmons JE, Carchman RA, Carter WH Jr, DeVito MJ. Thyroid-hormone-disrupting chemicals: Evidence for dose-dependent additivity or synergism. *Environmental Health Perspectives*. 2005; 113(11):1549–1554. [PubMed: 16263510]
15. McCullagh, P.; Nelder, JA. *Generalized Linear Models*. 2nd ed.. Chapman and Hall; New York: 1989.
16. Finney T. *Calculus*. Massachusetts: Addison-Wesley Publishing Co. 1994
17. Parker SM, Gennings C. Penalized optimal experimental designs for nonlinear models. *Journal of Agricultural, Biological and Environmental Statistics*. 2008; 13:334–354.
18. Parker, SM. *Solutions to reduce problems associated with experimental designs for nonlinear models: Conditional analyses and penalized optimal designs*. Virginia Commonwealth University; Richmond, VA: 2005.

Fig. 1.
Predicted T4 versus total dose of the mixture, adapted from Gennings *et al.*⁽¹²⁾

Fig. 2.
Desirability functions for the penalized optimal second-stage design.

Fig. 3.

The Ds-optimal second-stage design for the estimation of the interaction threshold. The vertical dotted line in (B) represents the location of the hypothesized interaction threshold. The triangles represent the location of the second-stage design points.

Fig. 4.

The penalized optimal second-stage design for the estimation of the interaction threshold.

The vertical dotted line in (B) represents the location of the hypothesized interaction threshold. The triangles represent the location of the second-stage design points.

Table I

Ds-Optimal Second-stage Designs for the Estimation of the Interaction Threshold

Number of Points		Second-stage Design: $\begin{Bmatrix} t \\ q \end{Bmatrix}$	Variance of the Interaction Threshold
8	Actual	$\begin{Bmatrix} 0 & 38.9 & 105.3 < & 105.3 > & 105.3 > & 105.9 & 116.6 & 578.0 \\ 0.093 & 0.125 & 0.113 & 0.112 & 0.185 & 0.222 & 0.105 & 0.045 \end{Bmatrix}$	3,413.8
	Practical	$\begin{Bmatrix} 0 & 39 & 105 & 106 & 117 & 578 \\ 0.093 & 0.125 & 0.410 & 0.222 & 0.105 & 0.045 \end{Bmatrix}$	4,113.0
	*	$\begin{Bmatrix} 0 & 39 & 105 & 106 & 117 & 578 \\ 0.093 & 0.125 & 0.113 & 0.519 & 0.105 & 0.045 \end{Bmatrix}$	3,438.0
7	Actual	$\begin{Bmatrix} 0 & 37.0 & 105.0 & 105.3 > & 105.3 > & 105.3 > & 113.4 \\ 0.001 & 0.156 & 0.155 & 0.250 & 0.165 & 0.157 & 0.116 \end{Bmatrix}$	3,281.5
	Practical	$\begin{Bmatrix} 0 & 37 & 105 & 113 \\ 0.001 & 0.156 & 0.727 & 0.116 \end{Bmatrix}$	7,041.9
	*	$\begin{Bmatrix} 0 & 37 & 106 & 113 \\ 0.001 & 0.156 & 0.727 & 0.116 \end{Bmatrix}$	3,587.8
6	Actual	$\begin{Bmatrix} 0 & 4.0 & 104.5 & 105.3 > & 105.4 & 109.7 \\ 0.020 & 0.001 & 0.291 & 0.095 & 0.319 & 0.274 \end{Bmatrix}$	3,228.7
	Practical	$\begin{Bmatrix} 0 & 105 & 110 \\ 0.020 & 0.705 & 0.274 \end{Bmatrix}$	4,458.6
	*	$\begin{Bmatrix} 0 & 105 & 106 & 110 \\ 0.020 & 0.291 & 0.414 & 0.274 \end{Bmatrix}$	3,232.3
5	Actual	$\begin{Bmatrix} 0 & 96.7 & 99.9 & 105.5 & 106.8 \\ < 0.001 & 0.205 & 0.164 & 0.326 & 0.305 \end{Bmatrix}$	3,213.9
	Practical	$\begin{Bmatrix} 0 & 97 & 100 & 106 & 107 \\ < 0.001 & 0.205 & 0.164 & 0.326 & 0.305 \end{Bmatrix}$	3,216.6
4	Actual	$\begin{Bmatrix} 0 & 98.3 & 104.5 & 105.5 \\ < 0.001 & 0.009 & 0.335 & 0.656 \end{Bmatrix}$	3,195.6
	Practical	$\begin{Bmatrix} 0 & 105 & 106 \\ < 0.001 & 0.335 & 0.656 \end{Bmatrix}$	3,200.1
3	Actual	$\begin{Bmatrix} 0 & 105.2 & 105.3 > \\ 0.120 & 0.255 & 0.625 \end{Bmatrix}$	3,238.5
	Practical	$\begin{Bmatrix} 0 & 105 \\ 0.120 & 0.880 \end{Bmatrix}$	16,593.8
	*	$\begin{Bmatrix} 0 & 105 & 106 \\ 0.12 & 0.255 & 0.625 \end{Bmatrix}$	3,245.0
2	Actual	$\begin{Bmatrix} 0 & 105.3 > \\ 0.329 & 0.671 \end{Bmatrix}$	3,428.72
	Practical	$\begin{Bmatrix} 0 & 105 \\ 0.329 & 0.671 \end{Bmatrix}$	42,304.7
	*	$\begin{Bmatrix} 0 & 106 \\ 0.329 & 0.671 \end{Bmatrix}$	3,434.6

Notes: The “Actual” design is the actual result of the Nelder–Mead Simplex Algorithm over a grid of starting values. In the “Practical” design, design points have been collapsed and/or eliminated as indicated by the “Actual” design.

t represents the vector of total dose locations and q represents the vector of corresponding proportions.

* denotes that design points have been collapsed in keeping with their location relative to the interaction threshold.

Table II

Optimal Second-Stage Designs for the Estimation of the Interaction Threshold

Number of Points	Optimality Criterion	Second-Stage Design: $\begin{Bmatrix} \mathbf{t} \\ \mathbf{q} \end{Bmatrix}$	Variance of the Interaction Threshold Desirability	
8	Ds ($\lambda = 0$)	$\begin{Bmatrix} 0 & 38.9 & 105.3 < & 105.3 > & 105.3 > & 105.9 & 116.6 & 578.0 \\ 0.093 & 0.125 & 0.113 & 0.112 & 0.185 & 0.222 & 0.105 & 0.045 \end{Bmatrix}$	3,413.8	0.128
	Penalized ($\lambda = 0.25$)	$\begin{Bmatrix} 0 & 65.9 & 85.6 & 106.7 & 123.5 & 152.9 & 181.4 & 503.5 \\ 0.076 & 0.115 & 0.145 & 0.166 & 0.184 & 0.145 & 0.095 & 0.075 \end{Bmatrix}$	3,809.1	0.901
6	Ds ($\lambda = 0$)	$\begin{Bmatrix} 0 & 4.0 & 104.5 & 105.3 > & 105.4 & 109.7 \\ 0.020 & 0.001 & 0.291 & 0.095 & 0.319 & 0.274 \end{Bmatrix}$	3,228.7	0.003
	Penalized ($\lambda = 0.75$)	$\begin{Bmatrix} 0 & 64.7 & 84.7 & 107.5 & 142.4 & 173.3 \\ 0.110 & 0.175 & 0.167 & 0.195 & 0.195 & 0.157 \end{Bmatrix}$	3,784.3	0.933
4	Ds ($\lambda = 0$)	$\begin{Bmatrix} 0 & 98.3 & 104.5 & 105.5 \\ < 0.001 & 0.009 & 0.335 & 0.656 \end{Bmatrix}$	3,195.6	<0.001
	Penalized ($\lambda = 0.75$)	$\begin{Bmatrix} 0 & 78.7 & 105.3 & 127.8 \\ 0.153 & 0.250 & 0.313 & 0.285 \end{Bmatrix}$	3,431.3	0.957

Note: $\mathbf{t}$ represents the vector of total dose locations, $\mathbf{q}$ represents the vector of corresponding proportions, and λ is the user-defined scaling constant.