D-Optimal Experimental Designs to Test for Departure from Additivity in a Fixed-Ratio Mixture Ray

Todd Coffey,* Chris Gennings,*†† Jane Ellen Simmons,† and David W. Herr††

*Department of Biostatistics, Virginia Commonwealth University, Richmond, Virginia 23298; †National Health and Environmental Effects Research Laboratory, ORD, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711

Received May 17, 2005; accepted September 7, 2005

Traditionally, factorial designs for evaluating interactions among chemicals in a mixture may be prohibitive when the number of chemicals is large. Using a mixture of chemicals with a fixed ratio (mixture ray) results in an economical design that allows estimation of additivity or nonadditive interaction for a mixture of interest. This methodology is extended easily to a mixture with a large number of chemicals. Optimal experimental conditions can be chosen that result in increased power to detect departures from additivity. Although these designs are used widely for linear models, optimal designs for nonlinear threshold models are less well known. In the present work, the use of D-optimal designs is demonstrated for nonlinear threshold models applied to a fixed-ratio mixture ray. For a fixed sample size, this design criterion selects the experimental doses and number of subjects per dose level that result in minimum variance of the model parameters and thus increased power to detect departures from additivity. An optimal design is illustrated for a 2:1 ratio (chlorpyrifos:carbaryl) mixture experiment. For this example, and in general, the optimal designs for the nonlinear threshold model depend on prior specification of the slope and dose threshold parameters. Use of a D-optimal criterion produces experimental designs with increased power, whereas standard nonoptimal designs with equally spaced dose groups may result in low power if the active range or threshold is missed.

Key Words: additivity; nonadditivity; nonlinear threshold models; optimal designs.

INTRODUCTION

Humans are exposed to mixtures of chemicals in their daily lives. Concern about exposure to multiple pesticides and the unknown sensitivity of the young were among the factors that led to the passage of the Food Quality Protection Act (FQPA, 1996). This law requires the U.S. Environmental Protection Agency (EPA) to consider exposure to combinations of pesticides that may have the same mechanism of action when setting regulatory standards. The EPA has recommended default approaches for use in component-based health risk assessments of chemical mixtures. These methods include dose-additive models for compounds with a common mode of action, and response-additive models for chemicals with different modes of actions (USEPA, 2000). Within these two general categories of models of additivity, different experimental designs may be used to test the null hypothesis of no interaction among the chemicals in the mixture. For example, an approach based on isobole methodology can be used to examine different ratios of the chemicals in a mixture (mixing ratio) that are predicted to produce a single constant level of effect (e.g., ED50) (Berenbaum, 1988; Gennings et al., 1990; Gessner, 1988). An extension of this design is the response surface approach that examines a wide region of the mixture space involving both different mixing ratios and levels of effect to see if a planar response surface results (Grecco et al., 1995; Myers et al., 1989; Solana et al., 1987). Such traditional experimental designs for evaluating interactions among chemicals of a mixture are feasible when the number of chemicals is small. However, when there are many chemicals in the mixture, these designs may be prohibitive because of the requirement of a large number of dose groups. For example, a full-factorial design requires $n^k$ treatment groups, where $n$ is the number of dose levels per chemical and $k$ is the number of chemicals in the mixture. If three chemicals were tested at five dose levels each, a total of 125 treatment groups would be required. Each of these treatment groups should have replicates, further increasing the experimental costs and efforts.

Efficient, less than full-factorial experimental designs make evaluation of large numbers of chemicals in a mixture economically and technically feasible. As an example, Groten et al. (1997) used a fractional $2^9$ factorial design in a study of nine chemicals that required the assumption that higher order interactions were negligible and that a linear approximation is
optimal experimental conditions that result in increased power for a departure from additivity that requires relatively few dose responses from doses with nonbackground responses. The Single Chemical Required (SCR) methodology (Casey et al., 2004; Gennings et al., 2002) allows the estimation of additivity or interaction for a mixture with any number of chemicals. This method requires that single chemical dose–response curves are available for each component of the mixture. By use of a fixed-ratio mixture ray with chemicals in proportions that are relevant to the scientific question of interest, this approach results in a test for a departure from additivity that requires relatively few dose groups compared to full-factorial experimental designs.

Using the SCR methodology, the investigators can choose optimal experimental conditions that result in increased power to detect departures from additivity. The power of a test can be defined as the probability of observing a departure from additivity, provided that one exists (e.g., Neter et al., 1996, pp. 1052–1060), and it depends on the amount of departure specified in the alternative hypothesis. In the present work, a methodology is described for the development of experimental designs that require much fewer dose groups than traditional full-factorial designs, yet these designs have high power to detect departures from additivity. They are based on a D-optimality criterion that minimizes the generalized variance of parameters from the mixture model based on specified values of the model parameters. For a fixed sample size, this kind of design selects the experimental doses of the mixture, and number of observations per dose level, that result in minimum variance of the model parameters, thereby increasing the power to detect departures from additivity. The scenario is that single chemical dose–response data are available and an additive model is estimated. The objective is to determine total doses and sample size allocation to be used for a fixed mixing ratio of the chemicals that satisfies a specified design optimality criterion. The example in this article uses a mixture of two chemicals. However, the methods presented permit extension to any number of chemicals, provided that suitable single chemical dose–response data are available for each chemical.

The optimal experimental designs presented can vary as a function of the expected relationship between the dose of the mixture and the biological response of interest. The example presented examines changes in cholinesterase (ChE) activity in erythrocytes produced by a 2:1 mixture of chlorpyrifos:carbaryl. Using an additive model defined by the SCR methodology, we provide examples of optimal designs for a greater than additive, less than additive, and mixed (less than additive in one dose region and greater than additive in another dose region) interactions. As with any experimental paradigm, judgment on the part of the investigators as to the effect size they wish to detect is required in designing the experiments. For these examples, a 33% change in the ED_{20} (i.e., a 20% decrease from the vehicle control) for erythrocyte ChE activity was chosen.

**MATERIALS AND METHODS**

**Experimental methods.** The experimental details related to data collection are reported in a related manuscript (Gordon et al., 2005). A time-course (0.5, 1, 2, 4, 8, 24 h) for inhibition of ChE activity in erythrocytes of adult male Long-Evans rats was determined after treatment with a 2:1 mixture of chlorpyrifos:carbaryl using administered total mixture dosages of 0, 28, and 35 mg/kg. The time-course study indicated that near-maximal levels of ChE inhibition occurred 4 h after treatment. Based on this result, single chemical dose–response curves for ChE inhibition produced by chlorpyrifos or carbaryl were determined 4 h after treatment. All samples were prepared rapidly, frozen on dry ice, and stored at −80°C (Hunter and Padilla, 1999) until they were assayed for cholinesterase activity. A slightly modified radiometric assay was used to quantify total cholinesterase activity (Johnson and Russell, 1975) at 26°C. All samples were run in triplicate, with a counting efficiency of approximately 57%. The average activity was converted first to nanomoles of acetylcholine hydrolyzed per minute per gram, and then to proportion of control (as described below).

**Statistical methods.** The SCR method has been described previously (Casey et al., 2004; Gennings et al., 2002). This methodology permits the analysis of mixtures while maintaining feasible experimental designs, does not require the constraint of parallel dose–response curves for the chemicals in the mixture, and adequately accounts for biological variability when testing the null hypothesis of additivity. Briefly, dose–response curves for individual chemicals are determined. Based on the single chemical dose–response curves, an additivity model is developed. Experimental mixture data are generated using mixtures at fixed mixing ratios, and then modeled to produce the mixture curve. The results of this mixture model are then compared to the additivity model in terms of total dose of the mixture. The statistical comparison of the two models based on prediction along the specified fixed-ratio ray is a test of additivity for the fixed ratio of the mixture of chemicals. Differences between the mixture and additivity models are considered to be a demonstration of departure from additivity. Details of the SCR method are provided in the Appendix.

The definition of additivity (i.e., zero interaction) we use is based on Berenbaum (1985) and is based on the classical isobologram for the combination of two chemicals (e.g., Loewe, 1953; Loewe and Muischnek, 1926). That is, in a combination of c chemicals, let E_i represent the dose/ concentration of the ith component alone that yields a fixed response, y_0, and let x_i represent the dose/concentration of the ith component in combination with the c agents that yields the same response. According to this definition of additivity, if the substances combine with zero interaction, then

\[ \sum_{i=1}^{c} x_i E_i = 1. \]  

(1)

If the left-hand side of Equation 1, termed the interaction index, is less than 1, then a greater than additive effect can be claimed. If the left-hand side of Equation 1 is greater than 1, then a less than additive interaction can be claimed. The additivity models used in the SCR approach are algebraically equivalent to Berenbaum’s definition of additivity (Equation 1) (Carter et al., 1988; Gennings et al., 1998; Meadows et al., 1998). The additivity models used in the SCR approach are algebraically equivalent to Berenbaum’s definition of additivity (Equation 1) (Carter et al., 1988; Gennings et al., 1998; Meadows et al., 1998).
et al., 2002). This definition of additivity is a general form for what is commonly known as dose-addition.

The SCR methodology requires single chemical dose–response data. With these data, optimal experimental conditions for the mixture experiment can be chosen to increase the power of detecting departure from additivity. Although there are many statistical criteria for an optimal experimental design, it is common to minimize the variability associated with the model parameters. Designs based on this criterion are known as D-optimal designs (e.g., Myers and Montgomery, 2002). The procedure minimizes the generalized variance, defined as the determinant of the variance-covariance matrix, of the model parameters. The smaller standard errors from this procedure result in increased power compared to standard, nonoptimal designs, because, for a fixed effect size, smaller standard errors result in more evidence against the null hypothesis. With the SCR methodology, for example, using the models described in the appendix (i.e., the nonlinear exponential threshold model), the hypothesis of additivity is parameterized in terms of the slope and threshold parameters of the statistical model, and is written in Equation 2:

\[ H_0 : \theta_{mix} = \theta_{add} \text{ and } \delta_{mix} = \delta_{add} \text{ vs. } H_1 : \theta_{mix} \neq \theta_{add} \text{ or } \delta_{mix} \neq \delta_{add} \]  

If a threshold model is not used, Equation 2 is altered to involve only the slope parameters.

The null hypothesis is evaluated using a Wald test. The D-optimal criterion produces an optimal design by minimizing the generalized variance of these parameters. The variance-covariance matrix of these parameters depends on the location of the dose groups for the mixture, as well as the allocation of the total sample size to each dose group. Details of the D-optimality criterion and the dependence of the variance-covariance matrix on the location of the dose groups and allocation of the total sample size are provided in the Appendix.

Using the SCR methodology, the predicted model under additivity is estimated from the single chemical data, and thus the slope (θ_{add}) and threshold (δ_{add}) parameters under additivity are estimated based on observed data. However, because the parameters associated with the mixture model (θ_{mix} and δ_{mix}) are unknown, they must be specified. Therefore, the D-optimality criterion minimizes the variance according to the assumed values of these parameters. Thus, different specifications of the slope and threshold parameters of the mixture model will result in different optimal designs. To find the D-optimal design, a numerical algorithm is needed. We implemented the Nelder-Mead simplex routine (Nelder and Mead, 1965) to minimize the generalized variance of the model parameters. The computation of this direct-search algorithm was performed using SAS version 8.2 (SAS, 1999), and this article presents an example to illustrate the application of this methodology. The dose–response relationship for chlorpyrifos and carbaryl (as prototypical organophosphate and carbamate pesticides) in a 2:1 mixing ratio was determined. The biological response of interest was the ChE activity in erythrocytes, expressed as a proportion of the vehicle control group’s activity. Expressing the data in this manner facilitated convergence in the statistical model.

RESULTS

Both chlorpyrifos and carbaryl inhibited ChE activity in erythrocytes 4 h after the animals were treated (Fig. 1). The reader is referred to Gordon and colleagues (2005) for a detailed description of this portion of the study and its results. The statistics for the single chemical dose–response curves are provided in Table 1. As explained in the Appendix, the single chemical dose–response curves have common maximum levels of ChE inhibition (H) and range of decrease in ChE activity (γ). A common threshold (δ_{add}) parameter, but distinct slope parameters (β_{1}, β_{2}), result in the estimation of different dose thresholds for carbaryl (δ_1) and chlorpyrifos (δ_2).

Using the SCR methodology described in the Appendix, an additivity model was created by modeling the single chemical data using Equation 3. The modeling step was implemented using the Nelder-Mead simplex algorithm in the SAS procedure NLP (SAS, 1999). The resultant slope of the additivity model is presented in Table 1.

The goal of this work was to design mixture experiments that would minimize the generalized variance of the slope and threshold parameters of the statistical model in Equation 5. After finding the optimal design, the power for detecting departure from additivity is reported.

Because the optimal designs depend on the predicted outcome of the experimental results from the mixture experiment, three scenarios of interest were chosen. First, where carbaryl and chlorpyrifos have a greater than additive interaction; second, in a mixed relationship interaction that is less than additive in the low-dose region and greater than additive at higher doses. Finally, we seek a less than additive interaction.

FIG. 1. Single chemical dose–response curves for inhibition of erythrocyte cholinesterase activity 4 h after treatment with carbaryl or chlorpyrifos. The dose–response functions used to generate the additivity model are indicated, as are the individual data points (○). See Gordon et al., 2005 for experimental details.
In each scenario, the shape of the curve for the mixture data was chosen to represent one possibility of a biologically meaningful effect. In the reported examples, an effect size of a 33% shift in the ED$_{20}$ (i.e., a 20% decrease in ChE activity from the vehicle control) was modeled by selecting appropriate parameter values for the slope and threshold. The threshold and slope parameters were allowed to change by a varying percentage for each scenario, but their values had to satisfy the particular example (i.e., for the greater than additive scenario, the mixture curve was greater than additive for the entire active range; for the less than additive scenario the mixture curve was less than additive for the entire active range; for the additivity scenario, the additivity curve was less than additive for the entire active range and subsequently greater than additive).

For all designs, a total of 50 animals was assumed. This number was chosen to correspond to a mixture experiment that allocated 10 subjects to five different dose groups.

**Greater than Additive Interaction Scenario**

For all scenarios, the ED$_{20}$ for the additive model is 1.54 mg/kg and the threshold is 0.95 mg/kg (see Table 1 for details). For the greater than additive interaction scenario, a steeper slope was chosen for the mixture curve compared to the additivity curve. The threshold was slightly reduced so that the greater than additive interaction resulted in a 33% shift in the ED$_{20}$. It was assumed that the mixture and additivity curves reached the same maximal effect, and thus both curves were equivalent in the high-dose region.

The shift in the greater than additive response curve shown in Figure 2 results in an ED$_{20}$ of 1.03 mg/kg (a 33% decrease from additivity), and under this scenario, the additivity ED$_{20}$ of 1.54 mg/kg became the ED$_{43}$. For this example, the threshold for the mixture is 0.74 mg/kg, which is a 22% decrease from the threshold under additivity.

As shown in Table 2, an equally spaced five-point design in five-unit intervals with 10 observations per dose group has only 8% power to detect a departure from additivity. The low power of this design is due to the placing of the dose groups in locations that do not define the threshold and active range of the mixture curve (Fig. 2). The same equal allocation for dose groups spaced in 4- and 3-unit intervals has about 16% and 39% power, respectively. However, the optimal five-point design (Design 4) has 64% power to detect a departure from additivity, whereas the optimal four-point design (Design 5) has 74% power. The optimal designs achieve such an increase in power because the treatment group locations and allocation

<table>
<thead>
<tr>
<th>Design</th>
<th>Design points</th>
<th>Number of observations</th>
<th>Power</th>
<th>Generalized variance$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0, 5, 10, 15, 20</td>
<td>10, 10, 10, 10</td>
<td>0.08</td>
<td>72720.0</td>
</tr>
<tr>
<td>2</td>
<td>0, 4, 8, 12, 16</td>
<td>10, 10, 10, 10</td>
<td>0.16</td>
<td>272.1</td>
</tr>
<tr>
<td>3</td>
<td>0, 3, 6, 9, 12</td>
<td>10, 10, 10, 10</td>
<td>0.39</td>
<td>1.50</td>
</tr>
<tr>
<td>4</td>
<td>0, 0.74, 2.1, 14.0, 17.9</td>
<td>8, 13, 15, 7, 7</td>
<td>0.64</td>
<td>0.00054</td>
</tr>
<tr>
<td>5</td>
<td>0, 0.74, 2.46, 16.4</td>
<td>7, 11, 26, 6</td>
<td>0.74</td>
<td>0.00044</td>
</tr>
</tbody>
</table>

$^a$Variability proportional to the area of the confidence region of the slope and threshold parameters.
of observations to each dose group are done in a way that minimizes the variance of the model parameters. In addition to the vehicle control, a dose group is located at 0.74 mg/kg, which is the threshold for this scenario. The dose groups at 2.1 and 2.46 are useful in estimating the slope, and the higher dose groups provide additional support for the slope. It is also important to note that 72% of the observations are allocated to the low-dose region for the optimal five-point design and 88% for the optimal four-point design. To reach 80% power, the optimal 4-point design needs 17 additional animals, or 67 total animals. By contrast, if Design 1 were implemented with 1000 animals, it would still only have 39% power.

**Mixed Interaction Scenario**

The curve for the mixed interaction scenario was chosen to have a larger threshold and steeper slope than the additivity curve. This resulted in a less than additive interaction in the low-dose region and a gradual shift to a greater than additive effect as the dose level increased. Both curves were predicted to be equivalent in the high-dose region.

As stated previously, the ED20 under additivity is 1.54 mg/kg, and the threshold is 0.95 mg/kg. The shift in the mixed interaction curve shown in Figure 3 results in an ED20 of 2.05 mg/kg (33% increase), and the additivity threshold of 1.54 mg/kg is associated with the control response. For this scenario, the threshold for the mixed interaction curve became 1.75 mg/kg, which is an 82% increase from that under additivity.

For this mixed interaction scenario, a five-point design with 10 observations per dose group, equally spaced in 5-unit intervals, has 10% power to detect a departure from additivity (Table 3). The same allocation pattern for designs with 4- and 3-unit intervals has about the same power. The low power for these designs results from missing the threshold and active range of the mixture curve (Fig. 3). However, the optimal five-point design (Design 4) has 78% power, and the optimal four-point design (Design 5) has 84% power to detect a departure from additivity. As with the greater than additive scenario, the optimal designs achieve such a striking increase in power because the dose groups are located to effectively estimate the threshold and slope parameters and because a high percentage of the observations are allocated to the low-dose region, where the change in the mixture curve is expected. For this mixed interaction scenario, both optimal designs locate a dose group at the threshold (1.75 mg/kg) and at 3.2 mg/kg, which allows estimation of the slope parameter. The higher dose groups provide additional support for the slope parameter. The optimal 5-point design needs only 5 additional animals, or 55 total animals, to reach 80% power.

**Less than Additive Interaction Scenario**

The mixture curve of the less than additive interaction scenario was chosen to have a larger threshold than the additivity curve, but the slopes of the two curves was similar. As with the other scenarios, both curves were equivalent in the high-dose region. This scenario also resulted in a 33% shift in the ED20 but differed from the other situations because the modeled mixture curve was similar to the additivity model curve for doses greater than the ED20. Figure 4 shows that the two curves are very similar for doses in the upper end of the active range, as well as the high-dose region.

The shift in the less than additive response curve shown in Figure 4 results in an ED20 of 2.05 mg/kg (33% increase), and under this scenario, the additivity threshold of 1.54 mg/kg has the same response as control groups. The threshold for the less than additive model is 1.57 mg/kg, which is a 64% increase from that under additivity (0.95 mg/kg).

### Table 3: Design Table for the Mixed-Interaction Scenario

| Design | Design points | Number of observations | Power | Generalized variance
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0, 5, 10, 15, 20</td>
<td>10, 10, 10, 10, 10</td>
<td>0.10</td>
<td>773.7</td>
</tr>
<tr>
<td>2</td>
<td>0, 4, 8, 12, 16</td>
<td>10, 10, 10, 10, 10</td>
<td>0.14</td>
<td>5.02</td>
</tr>
<tr>
<td>3</td>
<td>0, 3, 6, 9, 12</td>
<td>10, 10, 10, 10, 10</td>
<td>0.08</td>
<td>0.061</td>
</tr>
<tr>
<td>4</td>
<td>0, 1.75, 3.2, 11.6, 13.4</td>
<td>9, 11, 15, 7, 8</td>
<td>0.78</td>
<td>0.00068</td>
</tr>
<tr>
<td>5</td>
<td>0, 1.75, 3.2, 12.0</td>
<td>9, 15, 19, 7</td>
<td>0.84</td>
<td>0.00056</td>
</tr>
</tbody>
</table>

*Variability proportional to the area of the confidence region of the slope and threshold parameters.
Under the less than additive interaction scenario (Table 4), a five-point design with 10 observations per group equally spaced in 5-unit intervals has only 5% power to detect a departure from additivity. A similarly spaced design with 4- and 3-unit intervals has about the same power. These equally spaced designs miss the threshold as well as the majority of the active range of the mixture curve. The optimal five-point design (Design 4) has 26% power to detect a departure from additivity, whereas the optimal four-point design (Design 5) has 27% power. The optimal designs again increase the power of detecting a departure from additivity by locating dose groups at the predicted threshold (1.57 mg/kg) and at doses that will be effective at estimating the slope. It is again important to note that 86% of the observations are allocated to the low-dose region for the two optimal designs. The power of the optimal designs under this scenario is much lower than that under the greater than additive or mixed interaction scenarios. This result is a consequence of the specific choice of a less than additive interaction and the observed single chemical data that produced the additivity model. For this particular case, the curves were mostly coincident for doses greater than the shifted ED$_{20}$ (2.05 mg/kg). The similarity between the curves resulted in small differences between the model parameters. Smaller changes are more difficult to detect, and this led to the decreased power. However, in general, the D-optimal design criterion does not systematically result in designs that have lower power for detecting less than additive effects than greater than additive interactions. As an example of the difficulty of detecting a difference in this specific scenario, using 1000 animals with Design 1 has only 5% power. Although a 33% shift in the ED$_{20}$ was achieved, this example demonstrates that, for a nonlinear model, the optimal design depends not only on the type of interaction and effect size but also on the shape of the predicted mixture curve.

**DISCUSSION**

Use of statistical methods for experimental design can increase the power to detect biological effects without requiring a large increase in sample size. The utility of efficient experimental designs has been examined for studies designed to determine a benchmark dose (Kavlock et al., 1996; Slob et al., 2005; Weller et al., 1995). In this article, we demonstrate how statistical methods can determine optimal experimental designs for investigating nonadditive interactions in mixtures studies. The need for efficient experimental designs for mixtures research has been emphasized, as traditional factorial designs can require expenditure of large amounts of time and money (Cassee et al., 1998; Groten et al., 2001; Simmons, 1995). Even when using experimental mixture procedures such as the ray design (Gennings and Schwartz, 1998; Meadows et al., 2002), this manuscript illustrates that traditional equally spaced dose levels with equal allocation of animals to each group may have very low statistical power to detect departures from additivity. In such cases, it may be advisable to conserve valuable resources and clearly state that an assumption of additivity is required. Alternatively, optimal statistical design techniques may be used to increase the power to detect deviations from additivity.

The procedures demonstrated in this article involve several steps. First, single chemical dose–response curves are generated. Second, the responses from these experiments are modeled, and the coefficients are used to construct the additivity model. The third step is to design the mixture experiment, collect the data, and see if the results indicate a deviation from dose-additivity.

The D-optimal designs for nonlinear models that are presented depend on the assumed values of the model parameters. This requires specification of the predicted type of interaction (greater than additive, less than additive, mixed interaction).
interactions), the shape of the mixture dose–response curve, and the magnitude of the change (effect size) to detect while designing the mixtures experiment. As with any power calculation, the investigator can choose to maintain a constant sample size and calculate the resultant power or calculate the required sample size to produce the desired statistical power (Casey et al., 2005a, 2005b; Cohen, 1992; Meadows-Shropshire et al., 2005; Muller and Benignus, 1992; Muller et al., 1992). For the purposes of this demonstration, a 33% change in the ED20 for inhibition of erythrocyte ChE activity was chosen. The ED20 was chosen as this represents a low-level effect, yet is within the detection range for most ChE assays. Pilot experiments are a useful means to gain insight into directions of nonadditivity. Such pilot studies can be designed as a part of on-going experiments in a “leapfrog design” (Muller et al., 1983). After determining a direction of nonadditivity using a small sample size, the investigator can construct the D-optimal experimental design. Two-stage designs can be implemented, which use the data from the pilot study and the data from the larger optimally designed experiment, to determine departures from additivity.

In the threshold models we examined, the D-optimal designs contained observations at the expected threshold. To achieve the desired statistical power, the groups near the threshold may require larger numbers of observations than the groups near the ends of the response range. Designs with equally spaced dose groups over the mixture dose–response function may have low statistical power owing to inadequate definition of the threshold or slope parameters of the model. A relatively minor adjustment to the location of a dose group can result in a large change in power, especially when the slope of the mixture curve is steep. Thus, the investigator should assure that the spacing of dose groups effectively estimates both the threshold and the slope of the mixture response curve.

As a general observation, the four-point designs had greater power than the five-point designs. This occurred because the dose groups of the four-point designs more efficiently estimate the two parameters (θmix and δmix), by putting a greater percentage of observations at critical design points. For example, in the greater than additive scenario (see Table 2), the optimal four-point design allocates 37 observations to the active range, whereas only 28 are allocated in the optimal five-point design. A similar conclusion exists for the mixed interaction scenario shown in Table 3. However, the five-point optimal designs are more robust to misspecification of the threshold and slope parameters than the four-point optimal designs. This suggests that when the investigator does not have prior information regarding these parameter estimates (from pilot studies), the five-point design may be a desirable alternative to the four-point optimal design.

Several other observations about these D-optimal designs can be clarified. Low statistical power can be a result of similarity between the mixture and additivity curves. This is because the hypothesis of additivity is evaluated with a test of coincidence of the two curves. In general, curves that are farther apart are more likely to have greater power to detect departure from additivity. This was observed in the optimal designs for the less than additive scenario, which had less power than the optimal designs for the mixed or greater than additive scenarios. This finding was related to the similarity of the mixture and additivity curves for doses greater than the ED20. When a 60% increase in the ED20 was allowed, the difference between the curves was increased and the optimal designs resulted in 52–54% power. This is in agreement with other sorts of power analysis, where larger effect sizes result in greater statistical power.

In summary, use of statistical methods for experimental design can increase the power to detect effects without requiring a large increase in sample size. The method described in this article uses a D-optimality criterion to minimize the variance of model parameters associated with a chemical mixtures ray design. The reduction in the model variance is accomplished by placing treatment groups for the mixture experiment at points that efficiently estimate the model parameters. Often these locations are regions where the mixture model is anticipated to be altered (e.g., change in ED20, change in response slope) from the additivity model by exposure to the chemical mixture. Unequal allocation of observations may also be necessary to optimize the parameter estimation. Pilot studies with the mixture of interest can greatly assist in developing the D-optimal experimental design. Although the present work created designs for continuous responses, the D-optimal criterion can be used for any type of regression model, including those for binary responses. These models are also applicable when the dose metric is internal dosage (e.g., blood levels of compound), which may change the single chemical dose–response curves, as the theory of construction of the additivity model and the D-optimal designs is not altered. In times of declining resources, assurance of proper experimental design and adequate power to detect deviations from dose-additivity is of paramount importance. A properly designed experiment can increase efficiency, decrease cost, and increase confidence that the results are scientifically defendable.

**APPENDIX**

**General Framework**

Assume a study of c chemicals with dose–response data available for each chemical. In addition, dose–response data of r fixed-ratios of mixtures of the c chemicals are available for a total of R = c + r rays. Single chemical and mixture rays are defined by their fixed ratios denoted by [a1:a2:⋯:ac] such that

\[ \sum_{i=1}^{c} a_i = 1; \]
where, for the $ith$ single chemical, $a_i = 1$ and $a_b = 0$ for $b \neq i$ for $i = 1, \ldots, c$. Let $y_{ijk}$ be the response from the $kth$ observation in the $ith$ dose of the $ith$ chemical; $i = 1, \ldots, c + r, j = 1, \ldots, d; k = 1, \ldots, n_{ij}$ with mean $\mu_{ijk}$. The variance of $Y$ is assumed to be a function of the mean, i.e., $\tau \text{V}(\mu)$, and the method of maximum quasi-likelihood is used for parameter estimation (e.g., McCullagh and Nelder, 1989, chapter 9). A special case of this general form for the variance is that of a common variance, which results in a least-squares estimation approach to estimate model parameters. In the example, we assume the variance of $Y$ increases with the mean, i.e., $\tau \mu$.

**Testing for Interaction Using a Fixed-Ratio Ray Design: SCR Approach**

It is of interest to determine whether the components in a specified mixing ratio of chemicals interact. Interaction is detected in the SCR approach by comparison of the results of mixture experiments to the corresponding additivity model. It has been previously demonstrated (Carter et al., 1988; Gennings, 2002) that generalized linear models, and nonlinear models which can be linearized (i.e., $h(\mu) = \sum_{i=1}^{c} \beta_i x_i$), have planar contours and therefore agree with the definition of additivity in Equation 1 (see Gennings et al., 2005 for discussion). Further, these models can be adjusted to be in the form of a threshold additivity model (Schwartz et al., 1995). Specifically, for increasing dose–response relationships,

$$h(\mu) = \begin{cases} 
\beta_0, & \sum_{i=1}^{c} \beta_i x_i < \delta_{add} \\
\beta_0 + \sum_{i=1}^{c} \beta_i x_i - \delta_{add}, & \sum_{i=1}^{c} \beta_i x_i \geq \delta_{add} 
\end{cases}$$

and the inequalities are switched for decreasing curves. If additional nonlinear parameters are necessary, as denoted by the vector $\omega$, the left-hand side of the model statement is written as $h(\mu; \omega)$. Consideration of threshold additivity models permits inference regarding low-dose regions that are associated with background response. It is important to note that a threshold model may be used when it is of interest to estimate the join point between a background response region and a region of dose responsiveness. It should not be used as proof of the existence of a biological threshold. Although the SCR approach is applicable to general forms of nonlinear and threshold models (e.g., Gennings et al., 2002), we describe the method via the models used to produce the optimal designs.

To illustrate the approach, we consider a nonlinear exponential threshold model where $\omega = [\alpha, \gamma]$ and

$$g(\mu; \omega) = \ln \left( \frac{\mu - \alpha}{\gamma} \right).$$

Using single chemical data, a threshold additivity model for decreasing dose–response curves is given by

$$\mu_{add} = \begin{cases} 
\alpha + \gamma \sum_{i=1}^{c} \beta_i x_i > \delta_{add} \\
\alpha + \gamma \exp \left( \sum_{i=1}^{c} \beta_i x_i - \delta_{add} \right) - \sum_{i=1}^{c} \beta_i x_i \leq \delta_{add} 
\end{cases}$$

where $x_i$ is the dose of the $ith$ chemical, $\beta_i$ is the unknown slope parameter of the $ith$ chemical in the mixture of $c$ components $\alpha$ is a maximum effect parameter, i.e., the minimum response plateau of the dose–response curve $\alpha + \gamma$ is the magnitude of the control response, and $\delta_{add}$ is an unknown parameter associated with the threshold.

The parameter estimates for Equation 3 are found by solving an iterative algorithm specific to nonlinear models. Although many routines and software packages are available, we used the Nelder-Mead simplex algorithm in the SAS procedure NLP (SAS, 1999). The dose threshold for the $ith$ chemical is given by

$$\delta_{add}^* = \frac{\delta_{add}}{\beta_i}, i = 1, \ldots, c.$$

Thus, the model assumes that the dose–response curves for each of the chemicals in the mixture ranges between $\alpha$ and $\alpha + \gamma$; the dose–response curves are allowed to have different slope parameters ($\beta_i$); and the thresholds, $\delta_{add}^*$, are somewhat restricted to be jointly related to $\delta_{add}$. The last statement follows because $\delta_{add}$ is a common parameter for all of the $i$ chemicals (i.e., $\beta_i \delta_{add}^* = \delta_{add}$). When none of the dose thresholds are estimated within the experimental region, the model in Equation 3 is overparameterized and is replaced by the corresponding smooth model: here,

$$\mu_{add} = \alpha + \gamma \exp \left( \sum_{i=1}^{c} \beta_i x_i \right).$$

The adequacy of the model fit is validated with a goodness-of-fit test.

We are interested in detecting and characterizing an interaction among the $c$ chemicals in the mixture for the fixed mixing ratio of the chemicals. Following Gennings et al. (2002), define $t$ as the total dose of the mixture

$$t = \sum_{i=1}^{c} x_i.$$

Along a given ray with total dose $t$, the amount of the $ith$ chemical is $x_i = a_i t$, where $a_i$ is the ratio of the $ith$ chemical in the mixture for $i = 1, \ldots, c$, and

$$\sum_{i=1}^{c} a_i = 1.$$
The slope associated with this mixture under the assumption of additivity is given by

\[ \theta_{add} = \sum_{i=1}^{c} \beta_i a_i \]

Thus, the dose–response curve of the mixture in terms of total dose for the fixed mixing ratios under the assumption of additivity is given by

\[ \mu_{add} = \begin{cases} \alpha + \gamma & \theta_{add} > \delta_{add} \\ \alpha + \gamma \exp(\theta_{add} - \delta_{add}) & \theta_{add} \leq \delta_{add} \end{cases} \]

which is a re-parameterization of the threshold additivity model given in Equation 3. When the model in Equation 3 is replaced by the corresponding smooth model, Equation 4 becomes

\[ \mu_{add} = \alpha + \gamma \exp(\theta_{add}) \]

If the dose threshold is estimated outside the experimental region, the mixture data along the specified mixture ray is fit to a similarly parameterized mixture model of the form

\[ \mu_{mix} = \begin{cases} \alpha + \gamma & \theta_{mix} > \delta_{mix} \\ \alpha + \gamma \exp(\theta_{mix} - \delta_{mix}) & \theta_{mix} \leq \delta_{mix} \end{cases} \]

or \( \mu_{mix} = \alpha + \gamma \exp(\theta_{mix}) \). The hypothesis of additivity along the specified ratio of the chemicals is a hypothesis of coincidence between the additivity model in Equation 4 and the mixture model given in Equation 5, i.e., for the threshold models, \( H_0 : \theta_{mix} = \theta_{add} \) and \( \delta_{mix} = \delta_{add} \) versus \( H_1 : \theta_{mix} \neq \theta_{add} \) or \( \delta_{mix} \neq \delta_{add} \).

If smooth models are used, the hypothesis of additivity is \( H_0 : \theta_{mix} = \theta_{add} \) versus \( H_1 : \theta_{mix} \neq \theta_{add} \). An F-test is used to test this hypothesis of additivity (e.g., Casey et al., 2004).

### D-Optimal Design Criterion

Following Casey et al. (2005), let \( \Gamma = \begin{bmatrix} \alpha \gamma \beta \delta_{add} \delta_{mix} \end{bmatrix}^T \) represent the \( p \times 1 \) vector of model parameters from Equations 3 and 5,

\[ c = \begin{bmatrix} 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix} \]

represent the linear contrast matrix that corresponds to the test of additivity in Equation 2, and \( N \) represent the total sample size for the mixture experiment. When the variance of \( \bar{Y} \) is assumed to be a function of the mean, i.e., \( \text{Var}(\bar{Y}) = \tau \text{V}(\mu) \), then the generalized Wald statistic for testing the null hypothesis in Equation 2 is

\[ W = \left( c\hat{\Gamma} \right)^T \left( c(\hat{D}^T \hat{V}^{-1} \hat{D})^{-1} c \right)^{-1} \left( c\hat{\Gamma} \right) \]

where \( df \) represents the number of rows in \( c \) (here, \( df = 2 \)). For the vector of mixture dose groups \( t = [t_1, t_2, \ldots, t_d] \) and the proportion of subjects allocated to each group, \( q = \left[ n_1, n_2, \ldots, n_d \right] / N \), the D-optimal design criterion minimizes the determinant of the variance–covariance matrix,

\[ \Omega = \left\{ c(\hat{D}^T \hat{V}^{-1} \hat{D})^{-1} c \right\} \]

written mathematically as

\[ \min_{t} \left| c(\hat{D}^T \hat{V}^{-1} \hat{D})^{-1} c \right| . \]

Here, \( V \) is a \( N \times N \) diagonal matrix with \( V(\mu) \) on the diagonal and \( D \) is a derivative matrix:

\[ D = \begin{bmatrix} \partial \mu_{11}(\Gamma; t_1) / \partial \Gamma^T \\ \vdots \\ \partial \mu_{1n_1}(\Gamma; t_1) / \partial \Gamma^T \\ \partial \mu_{21}(\Gamma; t_2) / \partial \Gamma^T \\ \vdots \\ \partial \mu_{2n_2}(\Gamma; t_2) / \partial \Gamma^T \\ \vdots \\ \partial \mu_{d1}(\Gamma; t_d) / \partial \Gamma^T \\ \vdots \\ \partial \mu_{dn_d}(\Gamma; t_d) / \partial \Gamma^T \end{bmatrix} \]

This form for \( D \) shows how \( \Omega \) depends on the \( d \) dose group locations and the allocation \( (n_j) \) to each dose group.

### ACKNOWLEDGMENTS

This research was supported in part by T32 ES007334–01A1 (National Institute of Environmental Health Sciences [NIEHS], National Institutes of Health) and CR-828-11401 (U.S. Environmental Protection Agency). The authors thank Drs. Woodrow Setzer and Vernon Benignus for helpful discussions and critique of the manuscript. We also thank Ms. Jamie Graff and Mr. Matthew McMurray for excellent technical support in collecting the cholinesterase data.

### REFERENCES


