Human and Ecological Risk Assessment: Vol. 8, No. 2, pp. 305-326 (2002)

Assessing Toxicity of Mixtures: The Search for Economical Study Designs

Bertram Price,[1,*] Christopher J. Borgert,[2] Christopher S. Wells,[1] and Glenn S. Simon[3]

[1]Price Associates, Washington, DC; [2]Applied Pharmacology and Toxicology, Inc., Alachua, FL and Dept. of Physiological Sciences, University of Florida College of Veterinary Medicine, Gainesville, FL; [3]Rhodia, Inc., Research Triangle Park, NC

ABSTRACT

Toxicity screening and testing of chemical mixtures for interaction effects is a potentially onerous task due to the sheer volume of combinations that may be of interest. We propose an economical approach for assessing the interaction effects of chemical mixtures that is guided by risk-based considerations. We describe the statistical underpinnings of the approach and use examples from the published literature to illustrate concepts of local versus global mixture assessment. Our approach employs a sequential testing procedure to find the dose combinations that define the dose boundary for a specified acceptable risk level. The first test is conducted for a dose combination consisting of the acceptable doses of each individual chemical in the mixture. The outcome of this first test indicates the dose combination that should be tested next. Continuing in this manner, the boundary of dose combinations for the specified acceptable risk level can be approximated based on measurements for relatively few dose combinations. Dose combinations on one side of the boundary would have responses less than the response associated with the acceptable risk level, and dose combinations on the boundary would be acceptable levels of exposure for the mixture.

Key Words: chemical mixtures; drug interactions; chemical interactions; health risk assessment; statistical analysis; study design.

1.0. INTRODUCTION

Environmental exposure to chemicals often involves concurrent exposure to multiple chemicals. The pharmacological and toxicological behavior of chemical mixtures has long been a concern in both clinical pharmacology and environmental risk assessment, and is an increasingly active area of scientific research (Borgert *et*

* Corresponding author: Price Associates, Inc., P.O. Box 342, White Plains, NY 10605; Tel(voice):914-686-7975, Tel(fax):914-328-0561; bprice@priceassociatesinc.com
Received March 23, 2001; accepted March 24, 2001.

1080-7039/02/$.50

al. 2001). The concern is that chemicals in mixtures may interact, which means that the effect of the mixture on biological systems does not equal the sum of the effects of the individual chemicals. There is considerable interest in interactions that occur at low chemical concentrations because these may be important for establishing acceptable environmental exposure levels.

Chemical mixtures found in the environment have the potential to occur in many combinations. Any mixture of chemicals may consist of many different proportions of the component chemicals and at many different concentrations. Screening and testing chemical mixtures for synergism is a potentially onerous task due to the sheer volume of combinations that may be of interest. Identifying approaches to limit the number of tests to a practical and affordable level for screening and testing programs, therefore, is essential.

Previously, we developed five criteria for assessing the reliability of interaction studies (Borgert *et al.* 2001). Here we describe general principles, in agreement with those five criteria, that lead to economical study designs for assessing chemical mixture interactions. The scientific literature on assessing the interaction effects of mixtures is complex. It includes research on the interaction effects of chemicals in the environment, but principally deals with synergism and antagonism of drugs administered as mixtures. We limit our scope to environmental exposure to mixtures. By limiting our discussion this way, we are able to simplify the assessment process and suggest general economies in testing that achieve specified levels of statistical reliability.

Statistical principles constitute an essential component of assessing interaction effects for chemical mixtures. Data used to differentiate mixtures that act synergistically (antagonistically) from those that display no interactive effects have an inherent statistical character. Determining whether or not a mixture is synergistic (antagonistic), therefore, requires the application of statistical tests and confidence intervals. Statistical experimental design also is applicable in some circumstances to guide the economical development of data for assessing mixture interactions.

This article consists of four additional sections. In Section 2, we discuss design and analysis concepts that underlie a mixture assessment study. In Section 3, we discuss examples from the mixtures literature, focusing on opportunities to identify and apply economical designs. Section 4 contains our recommended strategies for mixture assessment study designs. In Section 5, we provide our conclusions.

2.0 STATISTICAL DESIGN AND ANALYSIS APPLIED TO MIXTURES

2.1 Background

Our starting point is to provide a clear and operationally meaningful statement of interaction for mixtures. We consider the status quo to be "no interaction" and will use the terminology "no interaction hypothesis." We define the no interaction using one of two models. The choice between these two models is the first of three dichotomies that could affect the development of study designs for assessing interaction. To avoid unnecessary complication in explaining the concept, we present the definitions for mixtures of two chemicals. Also, our subsequent discussion applies equally to chemicals measured in dose units or concentration units. Through-

out this report, we use the term "dose" to mean dose or concentration depending on the context.

Loewe Additivity (Loewe and Muischnek 1926) — Loewe Additivity means that one chemical behaves like a dilution of the second chemical, and *vice versa*. Mathematically, $a/A^* + b/B^* = 1$, where a/A^* and b/B^* are fractions of equieffective doses of chemicals A and B. Stated differently, if doses A* and B* of chemicals A and B result in the response R*, a mixture of the two chemicals is classified as additive for the response R* if every dose combination (a, b) satisfying $a/A^* + b/B^* = 1$ produces the response, R*.

Bliss Independence (Bliss 1939) — Independent effects of A and B in combination (E_{A+B}) are calculated from the individual effects (E_A and E_B) as $E_{A+B} = E_A+E_B-(E_A x E_B)$. If E is the probability of an effect (*e.g.*, mortality probability), the Bliss Independence model expresses probabilistic independence between the effects of the two chemicals. Independence implies that the relative effect of one chemical in the presence of a second chemical corresponds to the effect of the first chemical alone (Poch *et al.* 1995b; Poch *et al.* 1995a; Unkelbach 1992). Bliss independence also takes the form $E_{A+B} = E_A x E_B$ where the response is the complement of the probability described above (*e.g.*, survival instead of mortality).

Guidance for choosing between Loewe additivity and Bliss independence for the "no interaction hypothesis" is not well established. Ideally, the decision should be based, in part, on the anticipated toxicological outcomes resulting from the chemicals in the mixture, however, support based on toxicology has not been adequately developed for either model. From an empirical perspective an assessment could be made for both models subject to subsequent interpretation. Greco *et al.* (1995) points out a weakness in the Bliss model with an example involving a rapidly decreasing dose response curve (DRC) where application of Bliss independence would lead to the conclusion that a chemical is synergistic with itself. Berenbaum (1989) shows that Loewe additivity and Bliss independence are identical for exponential DRCs. Where the DRCs are linear, Berenbaum (1989) shows that the response expected under Loewe additivity for a dose concentration is the sum of the responses for the doses individually.[1] The USEPA (1999) cites the primary criterion for choosing between dose (Loewe additivity) or response (Bliss independence) addition as the no interaction approach is the functional similarity between the chemicals in the mixture. Greco *et al.* (1995) and Kortenkamp and Altenberger (1998) contain comprehensive discussions comparing these models. In the scientific literature on mixtures it appears that the Loewe model is preferred wherever a formal model of no interaction is used, however, adaptations of the Bliss model are used as a model of no interaction in some mixtures of carcinogens.

We use Loewe additivity as our no interaction hypothesis in this report. We, therefore, refer to "no interaction" simply as "additivity." For two chemicals, A_1 and A_2, all mixtures (a_1, a_2) on the Loewe additivity line $a_1/A_1^* + a_2/A_2^* = 1$ are expected to produce the same response as A_1^* and A_2^*, the equieffective doses of chemicals

[1] In the scientific literature on mixtures, the circumstance where the response expected under additivity for a dose combination is the sum of the responses for the doses individually is referred to as "response addition," "effect addition," or "effect summation."

A_1 and A_2 individually. The generalization of the Loewe additivity line to K>2 chemicals is the additivity plane:[2]

$$\sum_{i=1}^{k}\left(ai / A_i^*\right) = 1 \qquad (1)$$

where $\{a_i/A_i^*\}$ are fractions of equieffective doses of chemicals $\{A_i\}$.

For any specific dose combination, a deviation of a measured response from the response predicted by the additivity model is an indication of interaction.[3] If the DRCs for the component chemicals of the mixture have positive slopes, synergy would be inferred if the measured response resulting from the mixture were larger than the response expected based on additivity. Antagonism would be inferred if the measured response were less than the response expected based on additivity. If the DRCs have negative slopes, synergy would be inferred if the measured response resulting from the mixture were smaller than the response expected based on additivity, and antagonism would be inferred if the measured response were larger than the response expected.[4]

The second dichotomy that could affect the development of study designs for assessing additivity concerns the goal of the study — either to determine if the mixture is globally additive or if the mixture is additive for a specified limited set of dose combinations.[5] In some studies of mixtures, certain dose combinations exhibit additivity, others exhibit synergy, and others exhibit antagonism (Kortenkamp and Altenburger 1998; Levasseur *et al.* 1997). It would appear, therefore, that local analysis generally would be more meaningful than global analysis.

The third dichotomy is also a "goals" issue. Is the study intended as an exploratory analysis of additivity or as a confirmatory analysis? The exploratory mode would assist a researcher in identifying additional dose combinations for testing following a pilot investigation. The confirmatory mode applies where the test results would be used to reach a conclusion, with a reasonable level of statistical certainty, concerning additivity for specific dose combinations or a specific range of dose combinations. The number of dose combinations to be evaluated and their specification may depend on whether the goal is exploratory or confirmatory.

Our approach for evaluating the additive hypothesis is to compare measured responses to responses that would be expected under additivity for specific chemical dose combinations. Measured responses exhibit statistical variability, which is a

---

[2] The additivity plane is a simplex (*i.e.*, a plane in k-1 dimensional space with k vertices inside the dose hyper-cube).

[3] The deviation must be statistically different than zero. Statistical analysis of differences between observed responses and those expected under the no interaction hypothesis is discussed later in this report.

[4] In the remainder of this report, the reader may assume that DRCs have positive slopes, unless there is an explicit statement to the contrary. We use this assumption to simplify the presentation, without any loss of generality.

[5] We use Loewe additivity throughout this report; therefore, we do not discuss the first dichotomy further.

consequence of both biological variability and measurement error.[6] Expected response values calculated under the assumption of additivity also exhibit statistical variability because these calculations are based on DRC parameter estimates derived from measured responses for individual chemicals. Therefore, we apply statistical methods to ensure that conclusions concerning additivity properly account for statistical variability.

A study design for assessing additivity of chemical mixtures must be tailored to assessment objectives (*e.g.*, global versus local, exploratory versus confirmatory, etc.) A design must also answer three general questions. How many different dose combinations should be tested? Which dose combinations should be tested? How many replicates of each selected dose combination should be tested? The answers to these questions depend on, in addition to the assessment objectives, the models used to represent the data and the statistical methods used to evaluate the results. We provide general guidance for answering these questions. Specific answers depend on the characteristics of the particular assessment and require consideration of both toxicological and statistical issues.

2.2 Assessing Additivity

We begin by describing how to determine whether or not a mixture of two chemicals is additive for a specific dose combination. Let g_1 and g_2 be the DRCs for chemicals 1 and 2, respectively (*i.e.*, $g_1(x_1) = R_1$ for chemical 1; $g_2(x_2) = R_2$ for chemical 2). We test for additivity at any specific dose combination (x_1, x_2), by measuring the response, $R(x_1, x_2)$, and comparing it to the expected response if the chemicals were additive. The expected additive response is defined as the solution, $R_{ADD}(x_1, x_2)$, to:

$$\{ x_1/[g_1^{-1}(R)] + x_2/[g_2^{-1}(R)]\} = 1 \qquad (2)$$

(Gennings *et al.* 1990; Greco *et al.* 1995). ($g_i^{-1}(\cdot)$ is the inverse function of the DRC, $g_i(\cdot\text{-})$.) If $R(x_1, x_2)$ equals $R_{ADD}(x_1, x_2)$, the mixture at dose combination (x_1, x_2) would be classified as additive. If $R(x_1, x_2)$ exceeds $R_{ADD}(x_1, x_2)$, the mixture at (x_1, x_2) would be classified as synergistic. If $R(x_1, x_2)$ is less than $R_{ADD}(x_1, x_2)$, the mixture (x_1, x_2) would be classified as antagonistic. The decision leading to the classification — additive, synergistic, or antagonistic — must be based on a statistical test of the difference between $R(x_1, x_2)$ and $R_{ADD}(x_1, x_2)$. Statistical tests are discussed later. This procedure allows an additivity decision to be made for any dose combination regardless of the response.

Another approach to evaluating additivity results from focusing on a particular response, for example a response expected in a specified percentage of the study population or a response equal to a specified percentage of the control response. For example, if the specified percentage were 50%, we would measure the response at one or more dose combinations that, under the additivity hypothesis, would be expected to produce a 50% response. Let the ED50s (*i.e.*, doses corresponding to

[6] Measurement error is statistical variation associated with the measurement process. It reflects random fluctuations in replicate measurements of a given experimental unit or separate, but similar, experimental units.

50% responses) for chemicals 1 and 2 be X_1^* and X_2^*, respectively (*i.e.*, $g_1(X_1^*) = 0.50$ and $g_2(X_2^*) = 0.50$). If the chemicals are additive for the 50% response outcome, all dose combinations (x_1, x_2) that satisfy the additivity equation, $x_1/X_1^* + x_2/X_2^* = 1$, would be expected to produce a 50% response. Additivity of the mixture for the 50% response outcome could be tested for any specific dose combination (x_1, x_2) satisfying the additivity equation by measuring the response, $R(x_1, x_2)$, for that combination and comparing it to 50%.

2.2.1 Dose-Response Modeling: Estimation and Design

The first step in evaluating additivity for a mixture is to identify and characterize a DRC for each chemical in the mixture. Each DRC characterization must be adequate for the intended application. This means selecting doses for each chemical individually that will: (1) be sufficient to identify the form of the DRC; (2) cover a sufficient range of responses for the additivity analysis; and (3) produce DRC parameter estimates with sufficient precision for the additivity analysis.

The DRC estimation problem is not unique to additivity assessment. The biostatistical literature contains many references describing approaches for designing a study to estimate DRCs. Generally, DRC identification and estimation involves: (2) selection of effect endpoints and dosage duration; (3) preliminary testing to evaluate the full range of effects; (3) additional testing to estimate a DRC structural model; and (4) final testing to obtain optimal (high precision) estimates of the DRC parameters. We are not providing detailed guidance on DRC identification and estimation here. Instead, we reference the rich literature on this subject, including USEPA (1996), Gart *et al.* (1986), Gennings (1995), Greco *et al.* (1995), Kissin *et al.* (1987), and Tallarida (1992).

2.2.2 Local Additivity

For a local assessment of additivity, we assume that a range of interest has been established for doses of both chemicals. The range for chemical 1 is all doses less than or equal to X_1; the range for chemical 2 is all doses less than or equal to X_2. A graphical representation of the broadest region for evaluating additivity, therefore, may be viewed as the x_1-x_2 plane with vertices at $(0, 0)$, $(0, X_2)$, $(X_1, 0)$, and (X_1, X_2). We proceed with one of two possible approaches.

In the first approach, we may select combinations in the region of interest, measure the response for each combination, and statistically compare the measured response to the response expected for that combination under additivity. For a particular dose combination (x_1, x_2), denote the measured responses by $\{R_j(x_1, x_2)\}$[7], and denote the response expected under the additivity hypothesis be $R_{ADD}(x_1, x_2)$. (Recall that $R_{ADD}(x_1, x_2)$ is the value of R that satisfies Equation 2.). If the average of the replicate responses, $Ave\{R_j(x_1, x_2)\}$, is statistically equal to (greater than, less than) $R_{ADD}(x_1, x_2)$, the mixture at dose combination (x_1, x_2) would be classified as additive (synergistic, antagonistic).

For the second approach to evaluating local additivity, we begin by identifying a particular response of interest, R*. This response may be, for example, an outcome

[7] We anticipate replicate measurements at the dose combination (x_1, x_2). The subscript, j, on R represents replicates.

expected in a specified percentage of the population under study (*e.g.*, a 50% response) or some specified percentage of the maximum response. Doses of the two chemicals associated with the selected response, denoted by $X_1^* = g_1^{-1}(R^*)$ and $X_2^* = g_2^{-1}(R^*)$, are used to define the additivity line:

$$x_1/ X_1^* + x_2/ X_2^* = 1 \qquad (3)$$

Under the additivity assumption, all dose combinations (x_1, x_2) that satisfy EQ (3) would produce a response equal to R*. To test for additivity, first select dose combinations that satisfy the additivity equation and measure their responses. For any particular dose combination, (x_1, x_2), if the estimated response, $Ave\{R_j(x_1, x_2)\}$, were statistically equal to R*, the mixture would be classified as additive for that dose combination. If $Ave\{R_j(x_1, x_2)\}$were statistically greater (less) than R*, the mixture at (x_1, x_2) would be classified as synergistic (antagonistic).

For the two situations described above, the additivity decision is determined from the outcome of a statistical test based on the difference $[Ave\{R_j(x_1, x_2)\} - R_{ADD}(x_1, x_2)]$, or $[Ave\{R_j(x_1, x_2)\}- R^*]$.

Assuming that the estimated responses, $Ave\{R_j\}$and R_{ADD}, are approximated by the normal distribution, the statistical test would be conducted using the ratio of the difference to its standard error. The ratio, ignoring its algebraic sign, would be compared, for example, to the 97.5^{th} percentile of the standard normal distribution (*i.e.*, 1.96) for statistical test with a 5% significance level. If the unsigned ratio were smaller than 1.96, the null hypothesis of additivity would be accepted. If the ratio were larger than 1.96, additivity would be rejected and a classification of synergism or antagonism, respectively, would be given depending on whether the numerator of the ratio were positive or negative.[8]

We suggest in the discussion above that multiple measurements (replicates) of the response for a particular dose combination may be required to assess additivity. How many replicates are required? The answer comes from standard statistical hypothesis testing theory. We have described a statistical test with a significance level or Type I error rate of 5%.[9] That is to say, if the rule for determining additivity described above is employed, the probability of falsely classifying the dose combination as synergistic or antagonistic is at most 0.05 (5%). This 5% Type I error rate applies regardless of the number of replicate measurements used to calculate $Ave\{R_j(x_1, x_2)\}$. There is, however, a Type II error and a Type II error rate associated with this test – the probability of falsely classifying the dose combination as additive. The magnitudes of the Type II error and error rate are used to determine the number of replicate measurements.

It is customary to work with the complement of the Type II error rate, namely the power of the statistical test. In words, we are concerned about the probability that synergy (antagonism) will be correctly detected. The power of the test depends on

[8] If the estimated responses are not normally distributed, alternative statistical methods may be applied. A discussion of these alternatives is beyond the scope of the current report.

[9] Other terminology for the Type I error rate and significance level include α-level, and p-level.

two parameters: (1) the number of replications that are used to form Ave$\{R_j (x_1, x_2)\}$; and (2) the magnitude of a synergistic (antagonistic) effect that has been identified as important to detect. The magnitude of this latter parameter should be determined foremost from toxicological considerations rather than statistical considerations. For example, in a particular situation it may be determined that only synergistic responses that are at least twice the response expected under additivity are biologically important.

Determining the number of replications is part of the planning stage of a mixture assessment study. In a local assessment of additivity using the methods described earlier in this section, the plan should include a statement of the magnitude of synergistic (antagonistic) response that is considered important and the power of the statistical test for additivity (*i.e.*, the minimum probability that is acceptable for detecting the specified effect). These two parameters then can be used in standard statistical calculations to determine the required number of replicate measurements of the response for each dose combination.[10]

2.2.3 Global Additivity

Global additivity assessment means conducting an analysis that results in a conclusion of additivity, synergy, or antagonism that applies to all dose combinations. At the outset, one must be circumspect about claims of global additivity, synergy, or antagonism, because mixtures typically exhibit all three characteristics for different dose combinations.

There are two general approaches to assessing global additivity. The approach usually presented in the literature employs a mathematical model to describe the relationship between dose combinations and responses. A general model, referred to as a response surface model (RSM), typically employs linear terms for the individual chemicals and interaction terms for all combinations of the chemicals (Myers and Montgomery 1995). If the DRCs for individual chemicals are linear or can be made linear by transforming the response variable with an appropriate function, $g(\cdot)$, the RSM for a mixture of k chemicals takes the form:

$$g[R(x)] = \beta_0 + \sum\beta_j x_j + \sum\sum\beta_{ij} x_i x_j \qquad (4)$$

where the double summation on i and j is for i<j, each from 1 to K, and $x = (x_1, x_2, \ldots, x_K)$ is the vector of doses. β_0 represents the control response, which may or may not be zero (Gennings 1995). $\sum\beta_j x_j$ are the additive terms. The remaining terms in EQ(4) are the interaction terms. $g(\cdot)$ represents a monotonic mathematical function that is used to transform the response so it can be adequately represented by a linear function of dose for the individual chemicals.

[10] By omission of $R_{ADD}(x_1, x_2)$ and R* from the discussion of power, we tacitly assume that these quantities are known with certainty (*i.e.*, are not subject to statistical variation). This is not true for $R_{ADD}(x_1, x_2)$. Therefore, determining the number of replications for $R(x_1, x_2)$ will not entirely determine the power of the statistical test. The impact of the statistical variation in R_{ADD} (x_1, x_2) can be taken into account for determining the number of replicate response measurements, but that discussion is beyond the scope of this article.

Response surface models are fit to test data for selected dose combinations. A statistical test is used to assess the hypothesis that all the interaction parameters in the response model are equal to zero. An outcome indicating that the interaction parameters were zero would support an additivity conclusion. If the interaction parameters are statistically significant (*i.e.*, statistically different from zero), the results would indicate synergism, antagonism, or possibly both. To determine if a particular dose combination were synergistic or antagonistic, the model, including the interaction terms, would be used to predict the response for that dose combination. If the predicted response were statistically larger than the response excluding the interaction terms, the dose combination would be classified as synergistic. If the predicted response were smaller than the response predicted without interaction terms, the dose combination would be classified as antagonistic. In some cases, all response predictions using the interaction terms may lead to a classification of synergy and in other cases all response predictions may lead to a classification of antagonism. This type of uniformity, however, should not be expected.

Greco *et al.* (1995) describe another response surface modeling approach to assessing global additivity using their "flagship" model:

$$1 = x_1/[g_1^{-1}(R)] + x_2/[g_2^{-1}(R)] + \alpha x_1 x_2/\{[g_1^{-1}(R)]\ [g_2^{-1}(R)]\}^{1/2} \qquad (5)$$

α is the additivity parameter and $\{g_i(\cdot)\}$ are DRCs. If α were positive, the mixture at the dose combination (x_1, x_2) would be classified as synergistic. If α were negative, (x_1, x_2) would be classified as antagonistic. If α were zero, (x_1, x_2) would be classified as additive (Greco *et al.* 1995).

A second approach to assessing global additivity involves testing individual dose combinations that were selected to be representative of all combinations of interest. The test for an individual dose combination is based on the statistical comparison of $R\ (x_1, x_2)$ to $R_{ADD}\ (x_1, x_2)$. Global additivity (synergy, antagonism) would be asserted if the tests for all combinations indicated additivity (synergy, antagonism). In this second approach, the statistical test would be similar to the test described for local additivity, but with appropriate modification to account for the higher probability of finding a statistically significant result when multiple tests are conducted.[11]

Experimental design is critical for both approaches to global assessment because the potential number of dose combinations that could be tested is uncountable. In this circumstance, the judicious selection of dose combinations for testing is required to obtain the necessary information at an affordable cost. The designs that are discussed in the mixtures testing literature include full and fractional factorial designs, response surface designs, ray designs, and optimality designs (Gart *et al.* 1986; Greco *et al.* 1995).

A full factorial design includes all dose combinations of the doses used to study each chemical in the mixture separately. For a two-chemical mixture with testing planned at 4 dose levels for each chemical, a full factorial design would require $4 \times 4 = 16$ dose combinations to be tested. Each dose level of the first chemical would be tested in conjunction with each dose level of the second chemical. A full factorial

[11] We are referring to the statistical inference problem of multiple comparisons. At a minimum, the Bonferroni approach should be employed. See, for example, Miller (1985).

design allows evaluation of the additive contribution of the chemicals to the response, as well as all interaction contributions. If one of the dose levels for each chemical were set to zero, the data would include four points for estimation of each DRC and one point for estimation of the control response.

The statistical literature contains an abundance of articles and books on the application of factorial designs and fractional factorial designs. The majority of this literature addresses 2^k designs,[12] which means a mixture of k chemicals each tested at 2 dose levels. The number 2^k is a count of all combinations that must be tested in this full factorial design. Other forms of the full factorial design are possible. For example, the full factorial design consisting of 16 dose combinations described earlier is referred to with shorthand notation as a 4^2 design. Gennings and Schwartz (1998) analyze data from a 5^3 design. Including replication at each combination, where k is large or even moderately large, the total number of tests also would be large. Experimental design theory provides a method for reducing the size of the experiment through use of fractional designs, (*e.g.*, a half fraction, $2^{(k-1)}$, or more generally a p^{th} fraction, $2^{(k-p)}$). The fraction of dose combinations is selected in a prescribed fashion to preserve the integrity of statistical tests of the lower-order interaction terms in the response model.

We do not believe that 2^k designs or their fractions provide sufficient information for analyzing global additivity. Unless the DRCs are linear or each DRC can be transformed to a linear dose-response form using the same transformations, the limitation of two dose levels, a high and a low, is unlikely to lead to sufficient data to make a global additivity determination.

Ray designs also may be used to test for global additivity. Using a mixture of two chemicals as an example, a ray is a straight line emanating from the (0,0) dose combination. All dose combinations on the line have the same proportions (by weight) of chemical 1 and chemical 2. Dose combinations increase in concentration moving out on the ray from the (0,0) dose combination. A globally representative set of dose combinations may be obtained by selecting a few different rays, selecting dose combinations along each ray, and measuring the responses for the selected combinations. The dose combinations defined by a particular ray may be viewed as concentrations of a unique chemical. The concentration for this unique chemical increases in proportion to the distance on the ray from the (0, 0) dose. We view the response function along a ray as a DRC for the mixture with fixed proportions of each chemical. Therefore, the guidance used for selecting doses for estimating DRCs for individual chemicals is applicable to any mixture represented by a ray.

3.0 EXAMPLES: STATISTICAL DESIGN AND ANALYSIS IN PUBLISHED MIXTURE ASSESSMENT STUDIES

We have selected three published studies of mixtures as examples to further discuss some of the issues in designing studies to assess additivity. These three studies were selected from among many in the available literature because the authors of these studies incorporate statistical testing to assess additivity and they discuss experimental designs.

[12] See, for example, Box *et al.* (1978).

Kissin *et al.* (1987)

Kissen *et al.* study the loss of righting reflex in rats resulting from exposure to mixtures of morphine with thiopental and fentanyl with thiopental. The authors assess additivity using a response of 50% of the exposed rats as their measure of effect. For each pair of chemicals, the interaction between the agents is determined in two steps. First, the dose effect curves for all three individual chemicals are obtained and the ED50 values for loss of the righting reflex are calculated. Second, an isobolographic analysis is used to assess the type of drug interaction at the 50% response level.

The isobolographic analysis used by Kissen *et al.* is illustrated in Figure 1. For each pair of chemicals, the authors establish the additivity plane based on the individual ED50s. Then, Kissen *et al.* select three rays, named B, C, and D, from the (0, 0) dose to the additivity plane. Recall that the dose combinations (x1, x2) on a particular ray are different concentrations of a mixture consisting of constant proportions of morphine and thiopental (fentanyl and thiopental). The weight ratios (m) that define these constant proportions for mixtures of morphine and thiopental are m_B=(1:0.7), m_C=(1:3.5), and m_D=(1:17). Likewise, the weight ratios that define these constant proportions for mixtures of fentanyl and thiopental are m_B=(1:0.0005), m_C=(1:0.0025), and m_D=(1:0.0125).

The authors measured responses at a number of dose combinations on each ray in order to approximate the concentration corresponding to the 50% response. The concentrations corresponding to the 50% response along rays B, C, and D are the ED50s for the mixtures, labeled β, γ, and δ, respectively, in Figure 1. The joined line connecting β, γ, δ and the ED50s is the ED50 isobol. Kissen *et al.* assess additivity by comparing the doses at points β, γ, and δ to the doses expected to produce a 50%

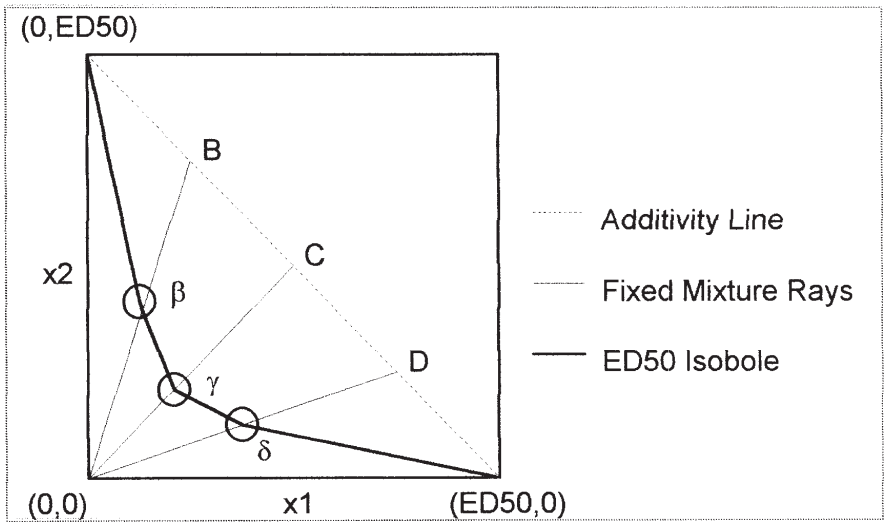


Figure 1. Testing Strategy used by Kissin *et al.* 1987. An isobolographic illustration of the scheme used by Kissin *et al.* (1987) to test binary dose combinations of hypnotic agents. The scales of both axes are normalized by each chemical's respective ED50. The isobole is not necessarily symmetric.

response under additivity. Recall that the dose expected to produce a 50% response under additivity is the dose combination at the intersection of the ray and the additivity plane. Kissen *et al.* find the doses for the 50% responses along the rays to be statistically smaller than the doses corresponding to additivity. They conclude, therefore, that the mixture is synergistic.

Kissen *et al.* applied statistical methods similar to those derived by Tallarida (1992) for comparing a mixture dose with the equieffective additive dose for the same mixture. The authors provide no explanation of the logic behind their choice of rays or the dose concentrations they chose to measure along each ray. It is apparent, however, that this study could have been successfully conducted with significantly fewer dose combinations. The study requires response measurements at multiple doses of the individual chemicals in order to obtain a sufficiently precise estimate of the ED50 for each chemical. But, as described below, response measurements at only one dose combination on each ray were necessary to evaluate interaction for mixtures of the individual chemicals.

The ED50s for the individual chemicals determine dose combinations that produce a 50% response if the mixture is additive (*i.e.*, the dose combinations on the additivity line). To test for additivity relative to a 50% response, Kissen *et al.* could have limited their study to measuring the responses for the dose combinations at the intersection of the rays and the additivity line. Additivity, then, would be assessed for each dose combination by testing for a statistically significant difference between the measured response for the dose combination on the additivity line and the expected additive response, which is 50%. If the measured response at any dose combination on the additivity line were statistically larger than a 50% response, that mixture would be classified as synergistic. If all three responses (*i.e.*, one measurement on each ray) were statistically larger than 50%, the conclusion would be the same as the conclusion reached by Kissen *et al.* based on the estimation and testing of ED50s along each of four rays.

The economy in testing suggested above requires only one assumption — that the DRCs for the mixtures represented by the rays are monotonic (nondecreasing). With this assumption, if the response at any dose combination on the additivity line is greater than the 50% response, a lower dose along the ray defining that mixture would result in a 50% response. The statistical test to compare doses and the statistical test to compare responses are logically equivalent. We accept the assumption of monotonicity of DRCs along the rays as reasonable in all circumstances. Therefore, testing only dose combinations on the additivity line is justified and the savings associated with this approach can be substantial.

Gennings (1995)

Gennings used data from a 2^5 factorial design for studying the effect of polycyclic aromatic hydrocarbons to demonstrate an economical additivity assessment approach as an alternative to response surface modeling. Gennings shows the fit of a full response surface model including 10 interaction parameters involving two chemicals, 10 interaction terms involving three chemicals, five interaction terms involving four chemicals, and one interaction term for all five chemicals. Gennings does not conduct a statistical test of the hypothesis that all interaction terms are

simultaneously equal to zero. This statistical test would constitute an assessment of global additivity. Gennings reports the p-values for testing individual interaction terms. These tests exhibit a haphazard pattern of statistical significance, which supports Gennings (and our) view that local rather than global assessment is the more meaningful objective.

Gennings does not rely on the response surface model for assessing additivity. Instead she uses estimates of the DRCs for the individual chemicals to calculate the response expected under additivity for any of the dose combinations that were tested. Gennings employs a transformation of the response variable (square root transformation) to obtain DRCs that are linear. Additivity is determined by statistically comparing the response expected under additivity to the measured response. Due to linearity of the DRCs, the response expected under additivity is the sum of the responses predicted for the doses of the individual chemicals from their estimated DRCs.

Gennings uses 4 to 6 doses for each chemical to estimate the DRCs. The spacing of these doses allows for testing of the linearity assumption for the DRCs. If the DRCs were known in advance to be linear when the square root transformation is applied to the response, allocating the doses for DRC estimation equally between the minimum and maximum doses would produce DRC estimates with maximum precision. Savings would be realized by measuring and testing the responses only for those dose combinations of interest rather than measuring the responses at 32 arbitrary dose combinations.

Taylor *et al.* (1995)

Taylor *et al.* study combinations of direct acting environmental mutagens that produce genetic reversions in single strains of salmonella. The source of these mutations (reversions) is believed to be either from genetic base deletions or complex frameshift mutations. Three chemicals studied in pairs by standard Ames plate-incorporation assays for these reversions include an organic extract of diesel exhaust (DE), a single chemical of diesel exhaust 1-nitropyrene (1NP), and chlorinated furanone (MX). These three chemicals act directly, meaning that they do not require S9 protein intermediates to induce mutations in salmonella.

The analysis is a local assessment because specific dose combinations were studied for each pair of chemicals. The authors also discuss the application of the results for a global assessment. Based on preliminary experiments for the individual DRCs, limitations were placed on selecting dose combinations for testing. The dose selection criteria were: (1) produce at least twice the background reversions; (2) keep to the linear portion of the DRCs; and (3) reversions should not exceed the upper capacity for the analytical technique when any two agents were combined. Thirteen dose combinations of the three chemicals were tested in triplicate (MX and 1NP — 4 dose combinations; MX and DE — 5 dose combinations; and DE and 1NP — 4 dose combinations).

Taylor *et al.* did not fit DRCs to the data for individual chemicals, but instead used only information about the DRCs consisting of response data at the doses in the study. The authors applied the following model to test for additivity:

$$\mu((D_A, D_B) = \mu_0 + \mu(D_A) + \mu(D_B) + \delta(D_A, D_B), \quad (6)$$

where D_A and D_B are doses of chemicals A and B, respectively, $\mu((D_A, D_B)$ is the combination response, μ_0 is the expected number of revertants in the control plates, $\mu(D_A)$ and $\mu(D_B)$ are the expected number of revertants based on application of the individual chemicals, and $\delta(D_A, D_B)$ is the interaction term (*i.e.*, the deviation from a purely additive response). In this assessment, the predicted additive response is the sum of the responses for each chemical individually plus the control level. That is, $R_{ADD}(D_A, D_B) = \mu_0 + \mu(D_A) + \mu(D_B)$, which is consistent with Loewe additivity where the DRCs are approximately linear.

Taylor *et al.* applied Generalized Least Squares to estimate the parameters in the model for each pair of chemicals and applied the Wald test as the statistical test of the hypothesis that all interaction terms were equal to zero (Rotnitzky and Jewell 1990). If the response to the exposure from the dose combinations were purely additive, all of the δ terms would be equal to zero.

Where the Wald test indicated a departure from additivity, Taylor *et al.* interpreted the result by inspecting the magnitude and algebraic sign of the difference between the measured response at (D_A, D_B) and $R_{ADD}(D_A, D_B)$. If the majority of the differences for a particular pair of chemicals were greater than zero (less than zero), Taylor *et al.* classified the mixture as synergistic (antagonistic). Only (MX + DE) was shown to be nonadditive by the Wald test, and by further consideration, synergistic.

4.0 STRATEGIES FOR MIXTURE ASSESSMENT STUDY DESIGNS

Our objective is to find ways of economizing in studies for assessing the additivity of mixtures of chemicals. We emphasize the statistical nature of additivity assessment data and note that a study design to assess additivity must determine the number of dose combinations to be tested, which combinations should be tested, and the number of replicate measurements for each dose combination. We prefer a "local" assessment approach rather than a "global" approach, because some dose combinations of a given mixture may exhibit additivity while other combinations may exhibit either synergy or antagonism. Therefore, any effort to establish a global additivity conclusion would, in general, be difficult to interpret. We provide the following three cases of a local assessment approach.

First, for a local assessment of additivity where the dose combinations of interest are specified in advance, only one of the three design questions must be answered: how many replicate measurements are required for each combination? The answer is determined by specifying the power of the statistical test and the magnitude of difference that is toxicologically meaningful between the response for the dose combinations and the expected additive response. The requirements on power should be more stringent in a confirmatory assessment than in an exploratory assessment leading to a requirement for a larger number of replicate measurements in a confirmatory assessment. This type of local assessment, where the dose combinations of interest are explicit, can be accomplished with estimates of the DRCs for the individual chemicals in the mixture and the measured responses for the combinations of interest. With this approach, it is not necessary to develop data that

would support the estimation of a complex response function for the dose combinations. The total number of measurements needed to estimate individual DRCs and the responses for the dose combinations of interest would be substantially less than the number of measurements needed to fit a response model that is intended to represent all dose combinations.

For the second of the three cases, we consider an assessment of additivity that is limited to a specified response (*e.g.*, a 50% response). In this case, all three of the design questions must be answered; however, the dose combinations selected for measurement would be restricted to the additivity plane for the specified response. The total number of dose combinations selected on the additivity plane usually would be determined by resource constraints (*i.e.*, time and costs for conducting the measurements) and secondarily by the goals of the test. Absent prior information about the anticipated toxicity of dose combinations, the locations of the dose combinations should divide the additivity plane into equal segments (Figure 2). The number of replicate measurements for each dose combination would be determined through consideration of the power of the statistical test and the magnitude of difference to be detected, as described previously.

For the third case, we focus on risk-based objectives for environmental exposure. The ultimate objective of an analysis of exposure to chemicals in the environment is to establish an exposure level that is not anticipated to cause significant adverse effects. In simpler terms, we seek an exposure level associated with acceptable health and environ-

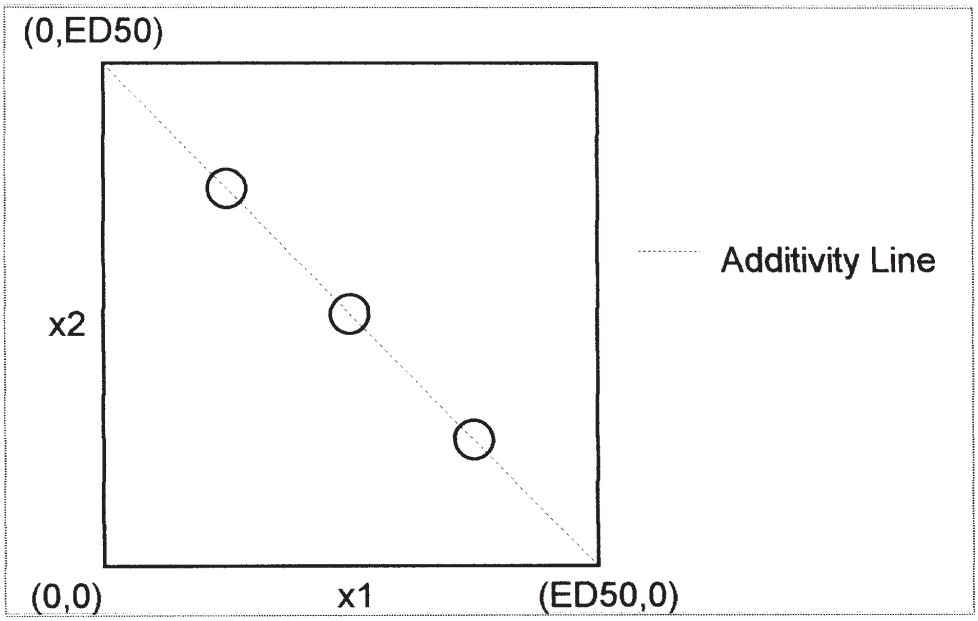


Figure 2. **Testing strategy for a specified effect level.** Recommended dose combinations for testing the additivity (no-interaction) hypothesis are equally spaced along the ED50 additivity line when the additivity assessment is limited to the 50% effect level. Dose combinations selected from the additivity line follow a regular geometric division of the additivity plane into equal parts. The scales of both axes are normalized by each chemical's respective ED50.

mental risk limits. We subsequently refer to such a level simply as an acceptable level (AL). An AL for exposure to an individual chemical usually is derived from a specified low percentile of the DRC, a benchmark dose (BMD) (Crump 1995; USEPA 1996), a "no observed adverse effect level" (NOAEL), or, for carcinogens, a dose that results in an acceptable cancer risk level (*e.g.*, a risk that does not exceed $1x10^{-6}$).

Where our interest is ALs, rather than focusing on a general assessment of mixture additivity, a more pertinent question is how to characterize the dose combinations that are ALs. As an example, consider a mixture of two chemicals and assume that BMDs have been established for each. A BMD for an individual chemical is determined from a benchmark response (BMR) using the DRC to find the dose corresponding to the BMR. The additivity line would be:

$$x_1/BMD_1 + x_2/BMD_2 = 1, \tag{7}$$

which defines all dose combinations (x_1, x_2) that, if the mixture were additive, would result in a response equal to the BMR. For mixtures of two chemicals, the dose combinations of interest would be a plane with the dose combinations (0, 0) (0, BMD_2), (BMD_1, 0), and (BMD_1, BMD_2) as vertices (Figure 3).

We seek those dose combinations that produce a response no greater than the BMR. We can investigate this question for any specific dose combination using the methods described above and build a description of the "AL" dose combinations (*i.e.*, those combinations with responses no greater than BMR). We suggest a sequential procedure, testing first at the mixture point (BMD_1, BMD_2).[13] This test is referred to as Test 1 in Figure 3. If the response at (BMD_1, BMD_2) is no greater than BMR, it is reasonable to expect that the response for dose combinations along the line segments (0, BMD_2) to (BMD_1, BMD_2) and (BMD_1, 0) to (BMD_1, BMD_2) would be no greater than BMR. One verification test for the dose combinations at the central points of these segments may be appropriate in this case. If Test 1 and the optional verification tests indicate responses no greater than BMR, we would conclude that every dose combination in the plane produces a response no greater than BMR. All of these dose combinations, therefore, would be considered to be in the acceptable exposure range. The mixture would be classified as "AL" dose combinations.[14] In this case the mixture may be additive, or it may be antagonistic, but if obtaining information about an AL is the principal objective of the analysis, additional tests necessary to determine antagonism would not be necessary.

If the response at Test 1 is greater than BMR, additional testing along the fixed mixture ray from (0, 0) to (BMD_1, BMD_2) would follow. We would search for a dose combination on the fixed mixture ray from (0, 0) to (BMD_1, BMD_2) with response equal to BMR. The next test in the search would be conducted at the intersection of that ray and the additivity line (Test 2 in Figure 3). The subsequent testing sequence depends on the outcome of Test 2.

---

[13] The assessment logic described here is inherently sequential, but practical laboratory testing constraints dictate that the tests would be prepared and conducted concurrently.

[14] Our conclusion in this case is based on the non-controversial assumption that the DRCs for the individual chemicals and for mixtures of the chemicals are non-decreasing functions of dose.

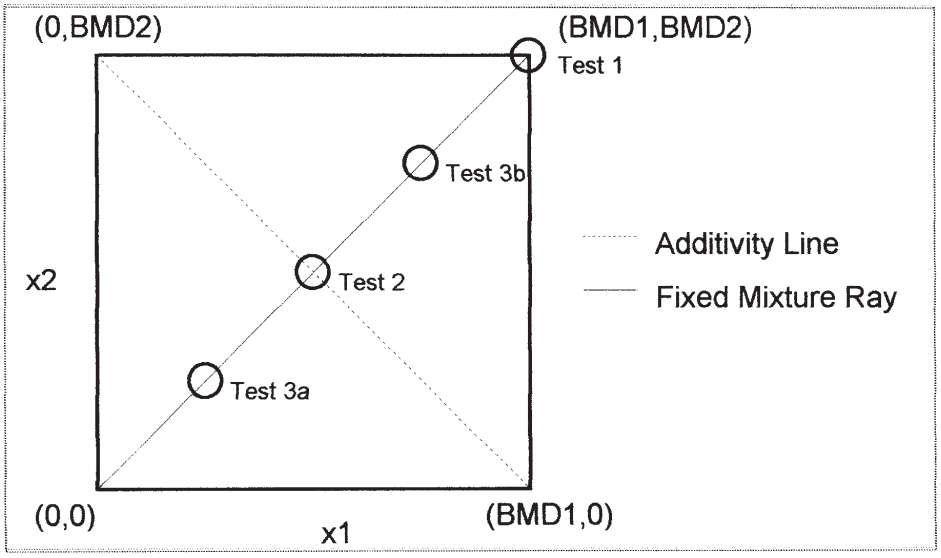
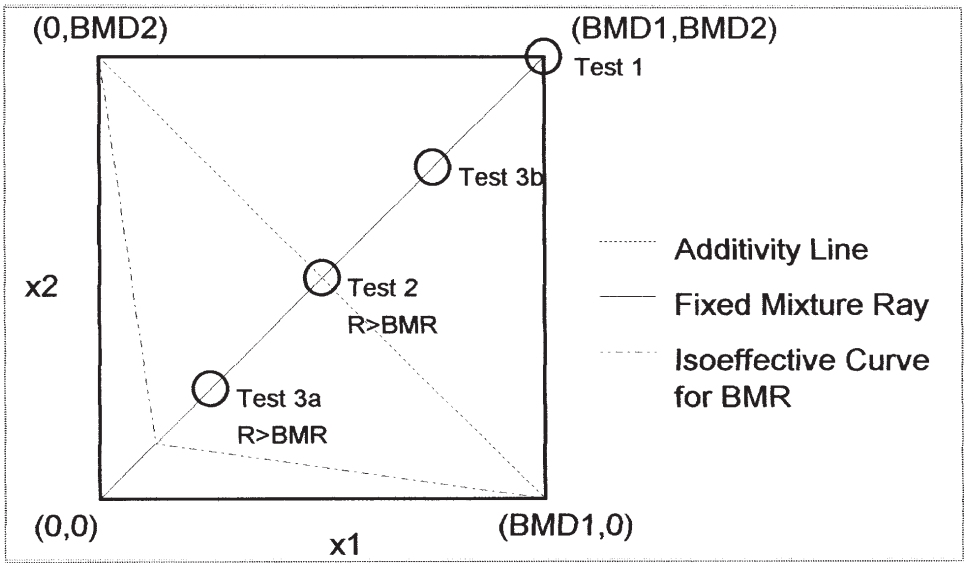


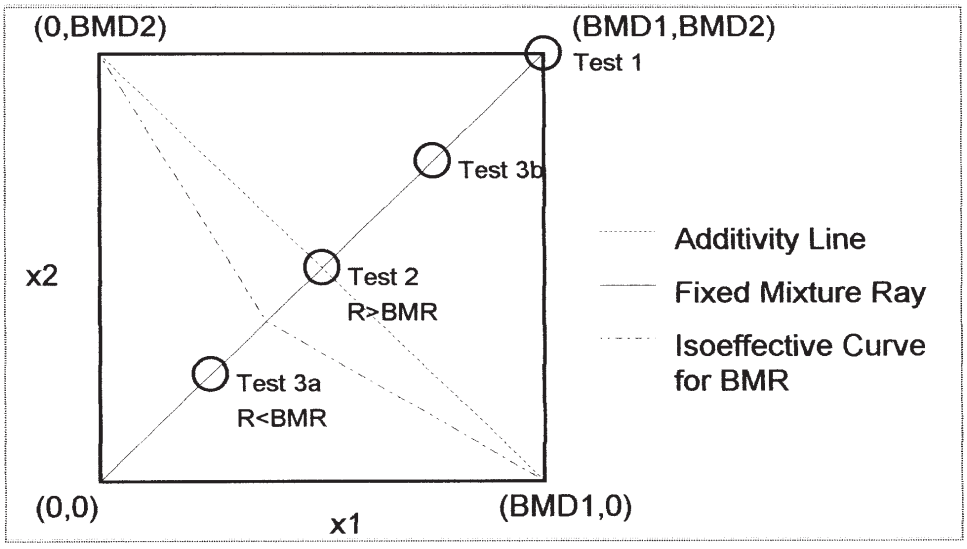
Figure 3. Testing locations to approximate the isoeffective curve for a mixture of two chemicals. When the goal is to find an acceptable level of exposure for a mixture, the BMDs of the individual chemicals may be used in an economic testing scheme where the initial dose combination selected for testing is a mixture based on the BMDs (Test 1). If Test 1 and the optional verification tests (Tests 2, 3a, 3b) indicate responses no greater than a BMR, we would conclude that every dose combination in the plane produces a response no greater than BMR. If the response at Test 1 is greater than BMR, additional testing along the fixed mixture ray from (0, 0) to (BMD_1, BMD_2) would follow.

(a) If the estimated response for Test 2 were greater than BMR, the next test should be conducted on the ray half way between the additivity plane and (0, 0) — Test 3a in Figure 3. If the result of Test 3a were greater than BMR, another test could be conducted half way between Test 3a and (0, 0). If the result of Test 3a were less than the BMR, the next test would be conducted half way between Test 3a and the additivity line. This approach, testing at the midpoint between the previous two tests, could be continued until the response is statistically equal to the BMR. The BMR isoeffective dose combination curve then could be approximated by line segments from (0, BMD_2) to the point on the ray where the response is equal to BMR and from this point to (BMD_1, 0). We suggest, however, that testing may be curtailed after Test 3a and the isoeffective curve approximated as shown in Figures 4(a) and 4(b). Dose combinations below the line segments would have responses less than BMR. Doses on the line segments would be interpreted as ALs.

(b) If the estimated response for Test 2 were less than BMR, the next test would be conducted on the ray half way between the additivity line and Test 1 — *i.e.*, Test 3b in Figure 3. If the result of Test 3b were smaller than BMR, another test would be conducted half way between this dose combination and Test 1. If the result of Test 3b were larger than the BMR, the next test would be conducted half way between Test 3b and the additivity line. This approach, testing at the midpoint



A



B

Figure 4. Sequential testing strategy to approximate the isoeffective curve for a mixture of two chemicals. Testing at the midpoint between the previous two tests could be continued until the response is statistically equal to the BMR. Alternatively, if the result of Test 3a were less than the BMR, testing may be curtailed after Test 3a and the isoeffective curve approximated as shown in Figures 4(a) and 4(b). If the result of Test 3b were larger than the BMR testing may be curtailed after Test 3b and the isoeffective curve approximated as shown in Figures 4(c) and 4(d). Dose combinations below the line segments would have responses less than BMR. Doses on the line segments would be interpreted as acceptable levels.

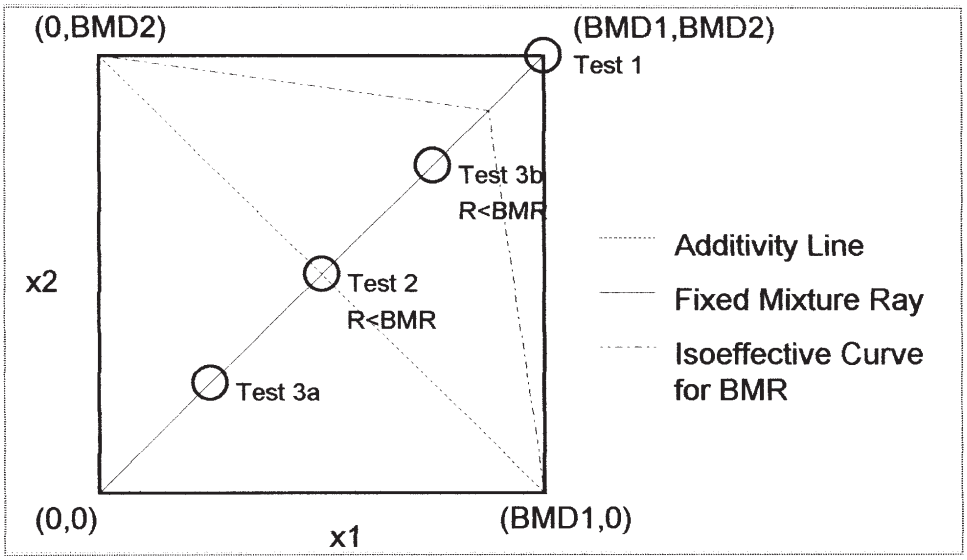


Figure 4C

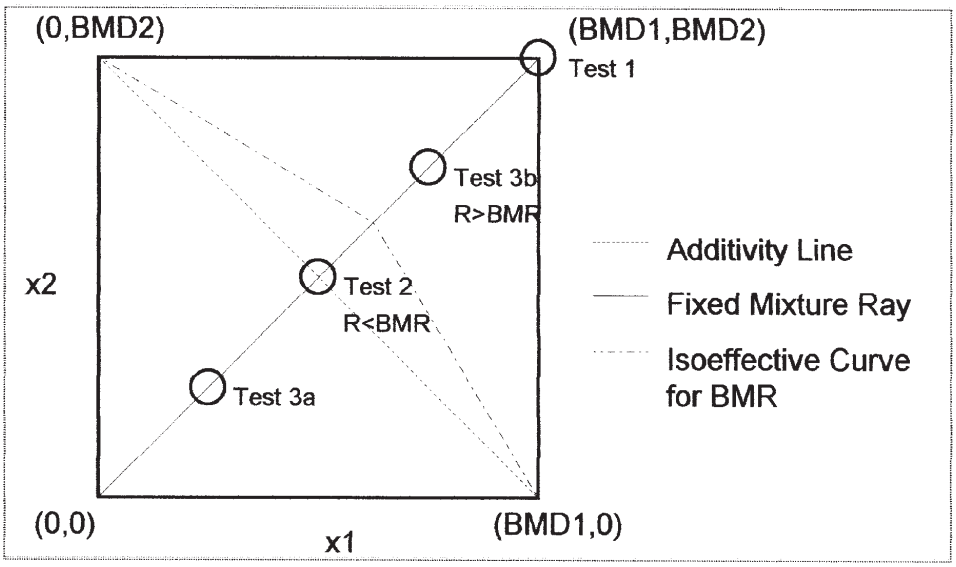


Figure 4D

between the previous two tests, could be continued until the response is statistically equal to the BMR. The BMR isoeffective dose combination curve then could be approximated by line segments from (0, BMD_2) to the point on the ray where the response is equal to BMR and from this point to (BMD_1, 0). As suggested above, testing may be curtailed after Test 3b and the isoeffective curve approximated as shown in Figures 4(c) and 4(d). Dose combinations below the line segments would have responses less than BMR. Doses on the line segments would be interpreted as ALs.

The approach described above also applies where the maximum acceptable response corresponds to the no observed adverse effect level (NOAEL). Since a NOAEL is determined from the doses in a dose-response study, dose combinations can be evaluated relative to the NOAEL without a full description of the DRCs for the individual chemicals in the mixture. For two chemicals, the additivity line based on NOAELs would be:

$$x_1/NOAEL_1 + x_2/NOAEL_2 = 1 \qquad (8)$$

Testing would proceed as described above. Upon completion of this process, a table could be constructed showing dose combinations on the isoeffective curve consistent with the NOAEL. These dose combinations are themselves ALs or may be transformed by application of uncertainty factors into ALs for the mixture consisting of the chemicals.

In some circumstances, sequential testing may not be practical. However, proceeding sequentially is not essential. For example, Tests 1, 2, 3a, and 3b may be conducted simultaneously and analyzed as needed to determine the approximate isoeffective dose combinations.

The procedure described above for mixtures of two chemicals is a conceptual foundation for analyzing mixtures where information about an AL is the objective. The procedure may be generalized to address mixtures of three chemicals or more. A description and examples of the generalization to three or more chemicals is beyond the scope of this article.

5.0 CONCLUSIONS

Exposure to chemicals in the environment involves concurrent exposure to more than one chemical. Screening and testing mixtures for additivity versus synergy or antagonism is a potentially onerous task due to the sheer volume of combinations that may be of interest. We have addressed the statistical nature of the additivity assessment data and differentiate between global and local assessments. We have identified a few general approaches for limiting the number of tests to practical and economically feasible levels. Our principal purpose is to provide guidance for designs to assess the effects of environmental exposures to mixtures. We note that the greatest efficiency in designs for assessing mixtures is achieved by having a specific well-defined objective, and for environmental exposures the objective is to control risk. In situations where an acceptable risk level has been translated into an acceptable level for environmental exposure to the component chemicals in a mixture, we suggest a sequential testing procedure that limits the number of dose

combinations that needs to be evaluated. This sequential procedure leads to an approximation of the isoeffective boundary for dose combinations. Dose combinations on one side of the boundary would have responses less than the acceptable risk level. Doses on the boundary would be interpreted as acceptable levels for the mixture.

ACKNOWLEDGMENTS

This work was partially funded by the American Chemistry Council and RhÙne-Poulenc. The views expressed herein are those of the authors and not necessarily those of the American Chemistry Council or RhÙne-Poulenc.

NOTE FROM HERA'S EDITORS IN CHIEF

This manuscript was prepared and submitted to *the Journal of Human and Ecological Risk Assessment* (HERA) under the journal's Author-Directed Peer Review System (HERA 6(1) 2000). The following peer reviewers, approved by HERA's Managing Editor, reviewed drafts of this manuscript and approved the authors' revisions: Douglas J. Crawford Brown, Ph.D. Director, Environmental Education Programs; Professor, Department of Sciences and Engineering. The Carolina Environmental Program, University of North Carolina, Chapel Hill, NC. William R. Greco, Ph.D. Department of Biomathematics, Roswell Park Cancer Institute, Buffalo, NY. Allan S. Susten, Ph.D., D.A.B.T. Assistant Director for Science. Division of Health Assessment and Consultation, Agency for Toxic Substances and Disease Registry, Atlanta, GA. HERA's Managing Editor approved the final manuscript, as submitted for publication.

REFERENCES

Berenbaum MC. 1989. What is synergy? [published erratum appears in Pharmacol Rev (1990) 41(3):422]. Pharmacol Rev 41:93-141

Bliss CI. 1939. The toxicity of poisons applied jointly. Ann Appl Biol 26:585-615

Borgert CJ, Price B, Wells C, *et al.* 2001. Evaluating chemical interaction studies for mixture risk assessment. Human Ecol Risk Assess 7(2):259-306

Box EP, Hunter JS, and Hunter WG. 1978. Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. Wiley & Sons, NY, NY, USA

Crump K. 1995. Calculation of benchmark doses from continuous data. Risk Anal 15(1):79-89

Gart J, Krewski D, Lee P, *et al.* 1986. Statistical Methods in Cancer Research Volume III - The Design and Analysis of Long-term Animal Experiments. International Agency for Research on Cancer (IARC), Lyon, France

Gennings C. 1995. An efficient experimental design for detecting departure from additivity in mixtures of many chemicals. Toxicology 105:189-97

Gennings C, Carter WHJr, Campbell ED, *et al.* 1990. Isobolographic characterization of drug interactions incorporating biological variability. J Pharmacol Exp Ther 252(1):208-17

Gennings C and Schwartz P. 1998. Combination threshold models with design optimization along fixed-ratio rays. J Agricultural, Biol, and Environ Stat 3(1):1-16

Greco WR, Bravo G, and Parsons JC. 1995. The search for synergy: a critical review from a response surface perspective. Pharmacol Rev 47:331-85

Kissin I, Mason JO, and Bradley ELJr. 1987. Morphine and fentanyl hypnotic interactions with thiopental. Anesthesiology 67:331-5

Kortenkamp A and Altenburger R. 1998. Synergisms with mixtures of xenoestrogens: A reevaluation using the method of isoboles. Science Total Environ 221**:**59-73

Levasseur LM, Greco WR, Rustum YM, *et al.* 1997. Combined action of paclitaxel and cisplatin against wildtype and resistant human ovarian carcinoma cells. Cancer Chemother Pharmacol 40:495-505

Loewe S and Muischnek H. 1926. Effect of combinations: mathematical basis of problem. Arch Exp Pathol Pharmakol 114:313-26

Miller R. 1995. Simultaneous Statistical Inference. Springer-Verlag, NY, NY, USA

Myers R and Montgomery D. 1995. Response Surface Methodology: Process and Product Optimization Using Designed Experiments. Wiley & Sons, NY, NY, USA

Poch G, Reiffenstein RJ, and Baer HP. 1995a. Quantitative estimation of potentiation and antagonism by dose ratios corrected for slopes of dose-response curves deviating from one. J Pharmacol Toxicol Methods 33(4)**:**197-204

Poch G, Reiffenstein RJ, Kock P, *et al.* 1995b. Uniform characterization of potentiation in simple and complex situations when agents bind to different molecular sites. Can J Physiol Pharmacol 73:1574-81

Rotnitzsky A and Jewell N. 1990. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. Biometrika 7:485-97

Tallarida RJ. 1992. Statistical analysis of drug combinations for synergism [published erratum appears in Pain (1993) 53(3):365] [see comments]. Pain 49:93-7

Taylor MS, Setzer RW, and DeMarini DM. 1995. Examination of the additivity assumption using the spiral and standard Salmonella assays to evaluate binary combinations of mutagens. Mutat Res 335:1-14

Unkelbach HD. 1992. What does the term "non-interactive" mean? Arch Complex Environ Studies 4:29-34

USEPA (U.S. Environmental Protection Agency). 1996. Benchmark Dose Technical Guidance Document. EPA/600/P-96/002A. External review Draft. Washington, DC, USA. August 9

USEPA (U.S. Environmental Protection Agency). 1999. Guidance for Conducting Health Risk Assessment of Chemical Mixtures. NCEA-C-0148. External Scientific Peer Review Draft. Washington, DC, USA. April