Human and Ecological Risk Assessment: Vol. 7, No. 2, pp. 259-306 (2001)

Evaluating Chemical Interaction Studies for Mixture Risk Assessment

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

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

ABSTRACT

We describe a set of criteria to evaluate the quality of data and interpretations in chemical interaction studies. These criteria reflect the consensus of the literature on interaction analysis developed over decades of research in pharmacology, toxicology, and biometry; address common pitfalls in published interaction studies; and can be easily applied to common methods of interaction analysis. The criteria apply broadly to interaction data for drugs, pesticides, industrial chemicals, food additives, and natural products and are intended to assist risk assessors who must evaluate interaction studies for use in component-based mixture risk assessments. The criteria may also assist researchers interested in conducting interaction studies to inform mixture risk assessment. The criteria are also intended to serve larger scientific goals, including increasing the repeatability of results obtained in chemical interaction studies, enhancing the reliability of conclusions drawn from interaction data, providing greater consistency of interpretations among various analysts, and decreasing uncertainty in using interaction data in risk assessments. We describe the basis for each criterion and demonstrate their utility by using them to evaluate interaction studies from the recent toxicological and pharmacological literature, which serve as examples of different types of data sets that the risk assessor may encounter.

Key Words: chemical mixtures, drug interactions, chemical interactions, health risk assessment, data evaluation.

---

[1] Corresponding author: Christopher J. Borgert, Ph.D., 238 Turkey Creek Alachua, Florida 32615; Tel(off): (904) 462-1266; Tel(fax): (904) 462-1267; cjborgert@APTHARMATOX.com

1080-7039/01/$.50

I. INTRODUCTION AND BACKGROUND

For a number of years, the scientific and regulatory communities have expressed concern that risk assessments for synthetic chemicals fail to adequately address the toxicity of chemical mixtures found in products, food, and the environment. Recently, this concern has intensified. The Food Quality Protection Act of 1996 emphasized chemical mixtures by requiring assessments of cumulative risks (Fed. Reg. 1998a, b). In 1997, the Presidential Commission on Risk Assessment and Risk Management (CRARM 1997) recommended moving beyond risk assessments for individual chemicals to a broader focus on the toxicity of chemical mixtures. In 1998, the U.S. Environmental Protection Agency's (USEPA) Endocrine Disruptor Screening and Testing Advisory Committee (EDSTAC) recommended screening six specific chemical mixtures (EDSTAC 1998). The USEPA has for many years focused attention on chemical mixtures in the risk assessment process (USEPA 1989a) including the recommendation of Toxic Equivalency Factors for dioxins and dibenzofurans (USEPA 1986). More recently, the USEPA supplemented its 1986 mixture risk guidelines (USEPA 1986) and Technical Support Document (USEPA 1990) with a draft Guidance Document on health risk assessment of chemical mixtures (USEPA 1999). The USEPA's approach to risk assessment for chemical mixtures begins with an evaluation of data quality to determine the adequacy of (1) exposure data for the mixture; (2) information on health effects of the mixture; and (3) information on interactions between components of the mixture. In this paper, our interest is in evaluating studies that provide the third type of information, interactions between mixture components.

In this paper we describe a set of criteria to evaluate the quality of data and interpretations in toxicological interaction studies. These criteria may assist risk assessors in identifying studies that can be used in component-based mixture risk assessments as well as those studies that are less useful due to inadequacies in design or interpretation (a component-based approach is used when obtaining data on the whole mixture is impractical). The criteria also may be used in applying a weight-of-evidence approach to interaction studies that reach contradictory conclusions regarding whether or what types of interactions occur.

To be meaningful, criteria for evaluating interaction studies should be based on accepted methods of analysis, address common pitfalls, and be easily applicable to methods typically used in interaction studies. Kortenkamp and Altenburger noted that methods for analyzing interactions have been developed over the last 100 years in the fields of pharmacology, toxicology, ecotoxicology, epidemiology, chemotherapy, and biometry (Kortenkamp and Altenburger 1998). Indeed, the scientific literature contains many publications describing methods for conducting interaction studies. Some examples that provide methodological background include: (Berenbaum 1978, 1981, 1984, 1985, 1989; Carter *et al.* 1979; Carter *et al.* 1984; Carter and Wampler 1986; Caudle and Williams 1993; Chou and Talalay 1984; Gennings 1995, 1996; Gennings *et al.* 1990; Gessner 1995; Greco *et al.* 1995; Greco *et al.* 1990; Loewe and Muischnek 1926; Machado and Robinson 1994; Martinez-Irujo *et al.* 1996; Poch 1980a,b; Poch *et al.*1990a; Poch and Holzmann 1980; Poch and Londong 1985, 1993; Poch and Pancheva 1995; Poch *et al.* 1995a.; Poch *et al.* 1990; Poch *et al.*1990b; Scaramellini *et al.* 1997; Suhnel 1990, 1996;

Tallarida 1992; Tallarida *et al.* 1989; Tallarida and Raffa 1996). The scientific literature also contains many recent studies that claim to have demonstrated interactions between chemicals, some examples of which include: (Abdelghani *et al.* 1997; Abou-Donia *et al.* 1985; Arnold *et al.* 1996a, 1996b, 1997a,b; Banks and Soliman 1997; Baynes *et al.* 1997; Ben-Shlomo *et al.* 1993; Bianchi *et al.* 1994; Bianchi-Santamaria *et al.* 1997; Bouakka *et al.* 1988; Brown 1997; Christ *et al.*1990; Cohen *et al.* 1972; Cohen and Murphy 1971; Dent and Orrock 1997; Dorian *et al.* 1983; Fletcher *et al.* 1995; Garfield and Bukusoglu 1996; Groten *et al.* 1996; Holzmann *et al.* 1992; Howe *et al.* 1998; Hu *et al.* 1987; Jaworska *et al.* 1997; Kissin *et al.* 1990, 1992; Kissin *et al.* 1987; Kissin *et al.* 1989, 1991; Kreitzer and Spann 1973; Lau and Wang 1996; Le *et al.* 1997; Levasseur *et al.* 1997; Livingston and Dethlefsen 1979; Lu *et al.* 1997; Marinovich *et al.* 1994; Marinovich *et al.*, 1996; Martin *et al.* 1997; Mayer and Doolittle 1995; McCain *et al.* 1997; McLachlan 1997; Misslin *et al.* 1988; Monette *et al.* 1997; Naguib and Abdulatif 1993; Piatti *et al.* 1994; Porter *et al.* 1993; Proudfoot *et al.* 1997; Qureshi *et al.* 1992; Rajamoorthi *et al.* 1997; Ramakrishna and Ramachandran 1978; Recker and Kier 1997; Redai *et al.* 1995; Ribas *et al.* 1997; Riekkinen *et al.* 1997; Sanders *et al.*1993; Shou *et al.* 1994; Siegmund *et al.* 1992; Siller *et al.* 1997; Soto *et al.* 1994; Stinecipher and Shah 1997; Sumpter and Jobling 1995; Takahashi *et al.* 1987; Taylor *et al.* 1995; Thompson and Moerschbaecher 1982; Tverskoy *et al.* 1989; Vale *et al.* 1997; van Birgelen *et al.* 1994a,b, 1996; van Steveninck *et al.* 1993; Verma *et al.* 1997; Verschoyle *et al.* 1982; Vonier *et al.* 1996; Wessinger and Balster 1987; Yu *et al.* 1997).

Meaningful criteria should also serve larger scientific goals. The criteria should help researchers generate data that are more interpretable, interpretations that are more consistent with the data, and hypotheses that are more testable. Application of the criteria should also promote consistency of interpretation among various analysts viewing the same data. The criteria should provide a basis for better understanding differences between studies and methods of analysis, and should also facilitate understanding the differences between various analysts with respect to their assumptions and conclusions. Application of the criteria should ultimately lead to better risk assessments because the relative strengths of published studies may be understood in the context of the goals of the risk assessment, and because uncertainties can be better quantified according to a replicable standard of measure. The value of the criteria presented here could be judged by the degree to which they further such goals.

The following set of criteria for evaluating interaction studies are based on our evaluation of the literature and reanalysis of the data reported in a number of published studies:

Criterion 1. Dose response curves (DRCs) for the mixture components should be adequately characterized.

Criterion 2. An appropriate "no-interaction" hypothesis should be explicitly stated and used as the basis for assessing synergy and antagonism.

Criterion 3. Combinations of mixture components should be assessed across a sufficient range to support the goals of the study.

Criterion 4. Formal statistical tests should be used to distinguish whether the response produced by a dose combination is different (larger or smaller) from that predicted by the "no-interaction" hypothesis.

Criterion 5. Interactions should be assessed at relevant levels of biological organization.

The five criteria proposed here do not attempt to address every issue pertaining to the use of interaction data in risk assessment. The criteria assume that assays and assumptions employed in interaction studies also satisfy standards of minimal epistemic status (*i.e.*, issues pertaining to the quality of the data themselves as opposed to inferences drawn from the data). For example, the criteria do not stipulate but rather assume that assays used in interaction studies be valid for measuring the intended biological endpoint. The criteria do, however, address issues of data quality regarding statistical tests specific to interaction analysis (Criterion 4). Several of the criteria address the quality of inferences drawn from interaction experiments rather than the quality of the data themselves (Criteria 1-3, 5). Nonetheless, the criteria leave unapprised many of the scientific judgements that must be applied when choosing among various biological models or choosing between models of noninteraction for performing a study. Our intent here is to provide criteria that are specific to interaction analysis and that will assist the evaluation and comparison of interaction studies, but that do not bias the assessor toward specific approaches.

Satisfying these criteria may not render the results of an interaction study suitable for every purpose or every risk assessment. Nevertheless, the literature appears to express a consensus that these criteria are the minimum standards for testing chemical interactions and reliably interpreting the dose response data. As such, these criteria can be applied broadly to interaction studies for drugs, pesticides, industrial chemicals, food additives, and natural products. The degree to which a particular study fulfills each criterion must be judged in the context of the goals of the study and the interpretations drawn from the data. Whether or not the results of a study should be used in risk assessment must be judged by the relevance of the study to the exposure scenario of interest, the degree to which the study satisfies the *set* of five criteria, and the level of uncertainty acceptable in the risk estimate.

The remainder of this article is divided into three sections. Section II describes the foundation of the criteria in detail, using formulae and examples where needed. Section III provides examples of how the criteria can be applied to studies from the scientific literature and can help identify common deficiencies. Section III is divided into four subsections: (A) identifying studies that satisfy all five criteria; (B) using the criteria to weight contradictory data; (C) using the criteria to uncover inadequacies in studies; and (D) using the criteria as a design tool to improve study reliability. In Section IV, we discuss how the criteria advance larger scientific goals and offer a few concluding remarks**.**

II. EXPLANATION AND DISCUSSION OF THE CRITERIA

Before describing the criteria in detail, a brief discussion of key terminology is in order. "No-interaction hypothesis" refers to the testable proposition that the effect

of chemicals in combination conforms to an appropriate non-interaction model. "Non-interaction model" refers to a mathematical function for predicting the level of response produced by any dose combination assuming that the individual components do not interact in producing a biological response. "Dose combination" refers to simultaneous treatment with specific amounts of two different agents, and a "fixed mixture" refers to a constant ratio of the mixture components across a range of concentrations. "Combination response" refers to the observed response produced by the dose combination, and a "non-interactive response" is a combination response that conforms to a non-interaction model. "Synergism" or "antagonism" refer to combination responses greater than or less than the responses predicted by the non-interaction model.

Criterion 1. *Dose response curves (DRCs) for the mixture components should be adequately characterized.*

The need to characterize the dose response curves (DRCs) of the individual chemical components of a mixture is stressed widely in the literature (Berenbaum 1981, 1989; Calabrese 1991; Groten *et al.* 1996; Merlin 1994; Suhnel 1990). Without adequate DRC characterization for the individual components, it is not possible to determine whether a biological effect of a mixture is due to interactions between the components. As such, Criterion 1 assists in evaluating the quality of inferences drawn from an interaction study rather than the quality of the data themselves.

Ideally, dose response characterization for the individual components should be sufficient to determine slope, inflection points, maximal and minimal response levels,[1] and should include the dose range corresponding to the dose combinations to be tested. Various dose response models may be employed to represent observed data, including the probit or logit. Greco and co-workers suggest the three-parameter Hill model (Greco *et al.* 1995; Hill 1910), which has the ability to accommodate a broad range of dose response data, but more recently have used the four-parameter Hill model, which includes an additional background parameter for residual effects of inhibitory agents at very high concentrations (Faessel *et al.* 1998; Levasseur *et al.* 1997). Mathematical transformations of the response measurement that result in a linear relationship between transformed response and dose greatly simplify the interaction analysis. In some dose ranges, simple linear interpolation may replace more complex mathematical dose response models for characterizing the dose response relationship used to assess interactive effects.

The significance of this criterion can be illustrated by considering a theoretical interaction experiment using two doses of a single chemical rather than doses of different chemicals. The dose response relationship for this single chemical is sigmoidal as depicted by the bold sigmoid-shaped curve in Figure 1. Rather than characterizing the DRC through its full range, this theoretical interaction experiment employs only two doses of the chemical, D1 and D2. The measured effects at

[1] The response measurement may be continuous (*i.e.*, the concentration of another chemical), binary (*i.e.,* the rate of sex reversal in individuals), or otherwise discrete (*i.e.*, ordered categories for severity), so long as the observations yield some quantitative result.

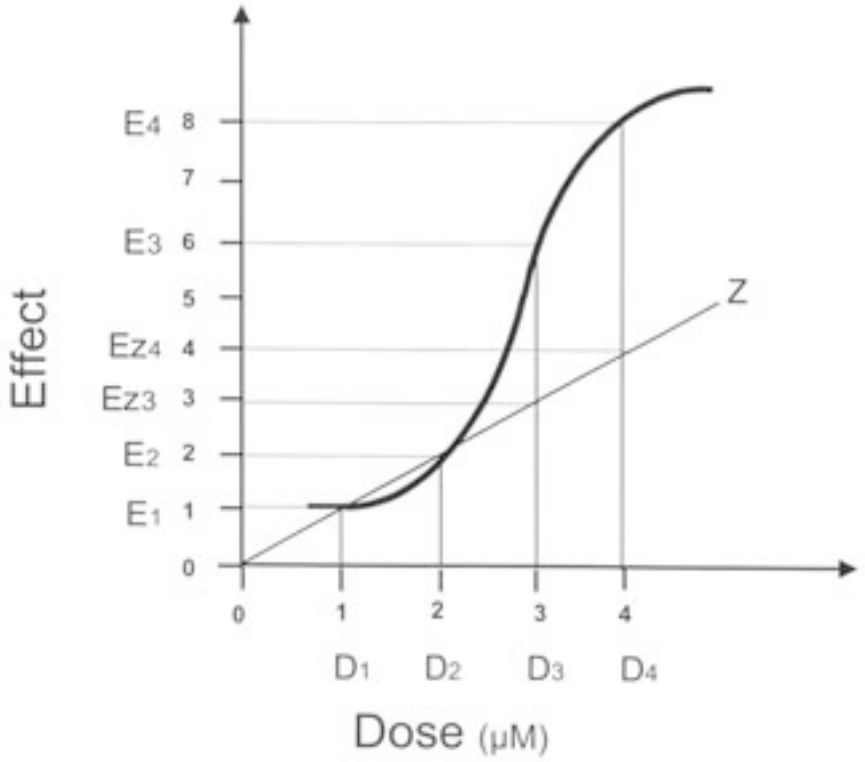


Figure 1. Illustration of how inadequate dose-response characterization could lead to the erroneous conclusion that two doses of a single chemical synergize. The measured effects at D1 and D2 are denoted by E1 and E2 respectively, and the measured effect at zero dose (control) is zero. Although E1 and E2 lie on the actual dose-response curve of the chemical, in the absence of other data points along the true curve, they define an artificially linear dose-response relationship (line Z). The predicted effects of combined doses D1 + D2 (D3) and D2 + D2 (D4) are EZ3 and EZ4, respectively. However, because the true dose-response curve is sigmoidal, the observed effects produced by the combinations D1 + D2 and D2 + D2 will be E3 and E4, respectively, rather than EZ3 and EZ4 as predicted by line Z.

D1 and D2 are denoted by E1 and E2, respectively, and the measured effect at zero dose (control) is zero. Although E1 and E2 lie on the actual DRC of the chemical, in the absence of other data points along the true curve, they define an artificially linear dose response relationship (line Z in Figure 1). Based on the dose response relationship described by line Z, the predicted effects of dose combinations D1 + D2 and D2 + D2 are EZ3 and EZ4, respectively.

Because the true DRC of this chemical is sigmoidal, however, the observed effects produced by the combinations D1 + D2 and D2 + D2 will be E3 and E4, respectively, rather than EZ3 and EZ4 as predicted by line Z (Figure 1). Comparison of the observed versus predicted effects would lead to the erroneous conclusion that both dose combinations D1 + D2 and D2 + D2 for this single chemical were synergistic (*i.e.* the responses E3 and E4 were greater than expected) when in fact, they are precisely as would have been expected (*i.e.*, dose additive) if the full DRC had been characterized. Inadequate dose response characterization could therefore lead one to conclude that two doses of a single chemical synergize when mixed. Without adequate DRC characterization, erroneous conclusions might be compounded greatly for mixtures of chemicals with differently shaped DRCs.

Criterion 2. *An appropriate "no-interaction" hypothesis should be explicitly stated and used as the basis for assessing synergy and antagonism.*

Interactions are inferred when a mixture of chemicals produces a biological response greater or less than expected based on mathematical concepts of non-interaction (Berenbaum, 1981, 1989; Boyd *et al.* 1990; Greco *et al.* 1995; Loewe and Muischnek 1926; Machado and Robinson 1994; Suhnel 1990; Unkelbach 1992). Thus, a "no-interaction hypothesis" must be defined that predicts the level of response expected if chemicals do not interact. Use of this concept allows the interactive effects of agents in combination to be assessed according to classical modes of hypothesis testing. A combination response that does not conform to the chosen no-interaction hypothesis is an indication of interaction, *i.e.*, the measured response deviates from the response predicted by the non-interaction model. Synergy is inferred if the measured response resulting from the mixture exceeds the response predicted under the no-interaction hypothesis; antagonism is inferred if the measured response is less than predicted under the no-interaction hypothesis. Since conclusions regarding chemical interactions must be based on testing an appropriate no-interaction hypothesis, Criterion 2 is used to evaluate the quality of inferences drawn from a study rather than the quality of data obtained.

Two models of non-interaction have been well developed in the pharmacological and toxicological literature and are appropriate as the basis for a no-interaction hypothesis.[2] Loewe additivity (Loewe and Muischnek 1926) is based on the concept that an agent cannot interact with itself. Loewe additivity predicts that two non-interacting compounds will behave as dilutions of one another when combined. Expressed mathematically, $A/A^* + B/B^* = 1$, where A/A^* and B/B^* are fractions of

[2] Greco *et al.* (1995) provide a comprehensive discussion of the debate over the best reference model for "no-interaction". Greco *et al.* (1995) propose Loewe additivity as the first choice and Bliss independence as an alternative.

isoeffective (equally effective) doses of A and B. Loewe additivity is commonly referred to as "dose addition" or "concentration addition".

Loewe additivity would accurately predict additive responses E3 and E4 for dose combinations D1 + D2 and D2 + D2 in Figure 1, regardless of the form of the DRCs. When the DRCs are linear, a special case of Loewe additivity occurs in which the predicted additive response is simply the sum of the responses. The outcome for this special case is sometimes referred to as "effect" or "response" addition.

The second model of non-interaction is Bliss independence (Bliss 1939). Bliss independence is often used as the general definition of "response" additivity, and is equivalent to Loewe additivity only under the special case of DRC linearity. Independent effects of A and B in combination (E_{A+B}) are calculated from the individual effects (E_A and E_B). The Bliss independence model expresses probabilistic independence between the two compounds. Expressed mathematically, $E_{A+B} = E_A+E_B-(E_A x E_B)$, where E is the probability of an effect (*e.g.*, the probability of mortality). Independence implies functional independence between two chemicals such that the incremental effect of one compound is unchanged in the presence of a second (Berenbaum 1981; Poch *et al.* 1990a,b, 1995b; Unkelbach 1992). In other words, the body responds to A as if B were not present. Bliss independence has an alternate form for the complement of the effect. For example, if the effect is mortality, the complement is survival. Whereas Bliss independence for mortality is described as $E_{A+B} = E_A+E_B-(E_A x E_B)$, Bliss independence for survival is $E(_{A+B})' = E_A' x E_B' = (1-E_A)(1-E_B)$.

Bliss independence can be illustrated by considering the action of two anti-tumor agents. Suppose that a dose of agent A produces 30% cell kill, while a dose of agent B produces 20% cell kill. For cell kill, independence predicts: 20% + 30% – (20% × 30%) = 44%. A response greater or less than 44% would indicate an interaction. Conversely, a dose of agent A results in 70% cell survival, while a dose of agent B results in an 80% cell survival. For cell survival, the expected combination response $E(_{A+B})$' is 70% × 80% = 56%.

The choice of a non-interaction model can influence the interpretation of a study. Greco *et al.* (1995) describes a sham experiment involving reasonable and adequately characterized DRCs where assuming Bliss independence leads to the paradoxical conclusion that a drug is synergistic with itself. If the DRCs are exponential (*e.g.*, $1-e^{-\beta d}$), then Loewe additivity and Bliss independence are equivalent (Berenbaum 1989). For other DRC functional forms, conclusions based on Loewe additivity and Bliss independence, generally, would not be the same (Greco *et al.* 1995; Poch *et al.* 1990b).

Currently, there are no compelling criteria for choosing between the two non-interaction models when designing an interaction study. Simultaneously discerning the "correct" non-interaction model and deviations from that model is a conundrum for which there would appear to be no general solution based upon real data. One approach for choosing the non-interaction model that might be considered is to select the non-interaction model based on the biological mechanism of action of the chemicals. However, considerable mechanistic detail might be required, as demonstrated by Jackson (1993). Jackson showed that differences in the Vmax of a single enzyme can change the interaction of two inhibitors from synergy to antagonism if one inhibitor interacts with the active site of the enzyme and the other limits

substrate availability at the active site. Few toxicological mechanisms are known in such detail. In fact, it is the paucity of information about mechanisms that leads us to an empirical approach for assessing interactive effects. If mechanisms were understood well enough to differentiate between the two non-interaction models, that same understanding could be used to predict the expected outcome for a mixture (Berenbaum 1981; Berenbaum 1989). Then, an interaction experiment would be a test of mechanistic understanding rather than an empirical test of a no-interaction hypothesis.

If the goal of a study is test interactions rather than to probe mechanistic understanding, then the study cannot be deemed either more or less valid based simply on the no-interaction hypothesis tested. The critical feature we are concerned with in this criterion is that an appropriately defined no-interaction hypothesis is tested and the results of the study interpreted with respect to that particular model. The mixtures literature favors Loewe additivity as the no-interaction model, in part due to paradoxes such as the one described by Greco (see above). In many cases, it may be informative to compare combination responses by both models of non-interaction (Kodell and Pounds 1991; Poch *et al.* 1990a,b).

It is important to recognize that the literature is inconsistent in terminology for the two main types of interactions. Although attempts have been made to standardize nomenclature (Greco *et al.* 1992, 1995; Henschler *et al.* 1996; Unkelbach and Wolf 1984; USEPA 1990, 1999), authors frequently differ in their interpretations and often neglect to define the terms they use. Some authors sub-classify the terms antagonism or synergism (Calabrese 1991; Poch and Londong 1985), while others use several words synonymously. For example, potentiation often refers to a form of synergy in which one chemical does not produce a measurable response at any dose, but sometimes the term is used simply to mean "greater than additive," *i.e.*, as a general synonym for synergy. Although we recognize that finer classification of terms may be useful for some purposes, we refer only to the two major categories of interactions — synergism and antagonism — because the criteria also apply to their subcategories.

Criterion 3: *Combinations of mixture components should be assessed across a sufficient range to support the goals of the study.*

To understand the necessity for this criterion, it is important to appreciate that the pharmacological and toxicological characteristics of a mixture are dependent upon the identity of the chemical components, the concentration of the mixture, and the concentration ratio of the components within the mixture. For a given set of chemicals, different mixture concentrations and component ratios can exhibit different types of interactions and biological effects (Berenbaum 1981). Even a fixed ratio of two chemicals can produce different types of interactions when tested across a broad range of mixture concentrations. Thus, to completely characterize interactions of a mixture requires testing numerous combinations of all mixture components across their full dose response ranges. Only when the goals of the study and the interpretations are more narrowly constrained is it sufficient to test fewer dose ratios across a narrower dose response range. Since it is likely that some dose combinations will interact and others will not (Berenbaum 1981, 1989; Greco *et al.*

1992, 1995; Poch *et al.* 1990b), it can be difficult to make global statements about interactions that are applicable to all dose combinations of a mixture. Because of this limitation, it is important to guard against over-interpreting data from studies where the design could, at best, support only narrowly constrained conclusions. Criterion 3 is thus used to evaluate the quality of inferences drawn from a study rather than the quality of data obtained.

The goals of an interaction study should be used to determine the number of different ratios of the individual components and the dose range across which a mixture should be tested, *i.e.*, the dose combination(s) of interest. It is possible for the goals of an interaction study to be more narrowly constrained than providing global characterizations of interactions. For example, Howe *et al.* (1998) assessed a single dose combination of the herbicides alachlor and atrazine found in a formulated product and appropriately constrained their interaction interpretation to this fixed mixture. Howe's study constitutes a local assessment of interactions that may be meaningful for certain occupational risk assessments. For other types of environmental risk assessments, different dose combinations of alachlor and atrazine may need to be tested because dose combinations found in the environment may differ greatly from the starting material. Not only might chemicals degrade at different rates and have different transport characteristics in the environment, but also interactions that occur when concentrations of the individual chemicals alone are sufficient to produce biological effects may not occur at lower mixture concentrations.

A number of different approaches can be used to study chemical interactions of mixtures (for reviews see Berenbaum 1989; Cassee *et al.* 1998; Greco *et al.* 1995). A full-factorial design tests a full complement of component ratios across the full dose response range of the mixture. Fractional factorial designs reduce the number of tests to a specified subset of mixture combinations while maintaining a substantial proportion of the information that would be produced with a full factorial design. Ray designs test fixed mixtures, *i.e.*, a constant ratio of the mixture components, across a range of concentrations. Berenbaum has described an approach for testing interactions between any number of agents by assessing fixed mixtures comprised of fractions of specific effect concentrations, such as the EC50 concentration (Berenbaum 1978). Whatever design is employed in a mixture study, it is important that the data set generated is sufficient to allow conclusions that satisfy the goals of the study. Thus, numerous approaches are candidates to fulfill our third criterion providing that the study design and the methods used to evaluate the data support the goals of the study and the interpretations drawn from it.

Criterion 4. *A formal statistical test should be used to distinguish whether the response produced by a dose combination is different from that predicted by the "no-interaction" hypothesis.*

Researchers often fail to appropriately test whether the observed response differs from the predicted "no-interaction" response (Berenbaum 1989; USEPA 1990). Some researchers evaluate only whether responses differ statistically from controls and whether dose combination responses differ statistically from individual constituent responses. Such comparisons do not adequately test whether the data can

support an interaction conclusion. In order to support a conclusion, departures from non-interaction that occur within the biological response range of the assay must be detectable at an acceptable level of statistical confidence. Therefore, our criteria for evaluating the quality of interaction studies include the explicit requirement for a statistical test appropriate for interaction analysis. Since the statistical test is an integral component of interaction data, Criterion 4 addresses an issue of data quality and is needed to fulfill a standard of minimal epistemic status for an interaction study.

Statistical analysis may be applied to data from toxicity tests of mixtures in either an exploratory or confirmatory mode. An exploratory analysis can be applied in a pilot investigation to assist the researcher in identifying potential interactions and additional dose combinations to test for confirmation. A confirmatory analysis uses combination responses to establish, with a reasonable level of statistical certainty, an interaction conclusion. A variety of statistical methods have been used to infer whether mixture components interact (*e.g.*, simple t-tests, linear models including ANOVA and multiple regression, and logistic regression).[3] In a confirmatory analysis, sufficient replication is important to assure an acceptable level of statistical power (*i.e.*, the probability of detecting a departure from the "no-interaction" hypothesis if, in fact, the "no interaction" hypothesis is false). Recording statistical power in a research report is particularly important where the result indicates acceptance of the "no-interaction" hypothesis. Statistical power is also important for interpreting the magnitude of the interaction that can be observed, a factor that may influence how the data are used in risk assessment.

The statistical model and assumptions underlying the analysis should not introduce a bias that favors one of the possible outcomes, either interaction or non-interaction. To avoid bias, it is essential to apply statistical methods and statistical tests in a manner that is consistent with a clearly specified no-interaction hypothesis and appropriate model of non-interactive effects. A clear statement defining the response expected if the mixture components do not interact (Criterion 2) underlies all analyses and conclusions of synergism, antagonism, or no-interaction. Without a clearly stated no-interaction hypothesis, the results of any statistical test cannot be interpreted. Statistical confirmation is necessary for inferences concerning both single dose combinations as well as those that are interpreted as global (*i.e.*, for all dose combinations).

Criterion 5. *Interactions should be assessed at relevant levels of biological organization.*

It is critical to select the proper level or levels of biological organization at which interactions are assessed so that the results can be interpreted meaningfully. Interaction experiments conducted at the level of the whole organism or population can be difficult to interpret without information from underlying levels of biological organization. On the other hand, numerous interactions may be detected at the molecular, biochemical, or cellular level that never manifest change in the organism

[3] For an in-depth discussion of the strengths, weaknesses and appropriate use of various approaches see Carter *et al.* 1988; Gessner 1988; and Greco *et al.* 1995.

or in populations of organisms. Because such phenomena have been observed repeatedly in the search for drug interactions to improve antibiotic and cancer chemotherapy, pharmacological synergism and antagonsim are not considered to be synonymous with clinical advantage and disadvantage (Berenbaum 1988; Merlin 1994; Wampler *et al.* 1992). McInnes and Brodie (1988) have discussed the pitfalls of failing to adequately assess interactions at relevant levels of biological organization. Criterion 5 is intended to help avoid such pitfalls and to assist in evaluating the quality of interpretations drawn from interaction data.

Although predicting effects on organisms or populations of organisms is often the ultimate goal of interaction analyses, interactive effects of chemicals may not be discernable at higher levels of biological organization for a number of reasons. One possible reason is an inability to fully assess DRCs (Criterion 1) or to test a sufficient range of dose combinations (Criterion 3) in particular types of assays at the whole-organism or population level. A population or community may also be exposed to other unmeasured stressors, or may have high background responses that lower the statistical power of detecting changes in the effect of concern. Another possibility is that a high level of biological organization may respond identically to qualitatively different types of interactions. Complex homeostatic feedback mechanisms and counterbalancing physiological systems not only present the opportunity for interactive effects in whole organisms and populations, but these may also confuse the interaction (or lack thereof) underlying a response.

There are also many reasons that interactions at lower levels of biological organization might never manifest change in an organism or population. Pharmacokinetic behavior *in vivo* may preclude pharmacodynamic interactions that are detectable *in vitro*. Interactions at the molecular level may occur in organisms only at concentrations above those that produce toxicity unrelated to the interaction. Chemicals may trigger adaptive or compensatory responses in organisms that abrogate interactions at the cellular or biochemical level.

The fifth criterion is different conceptually and practically from either the obvious requirement to corroborate *in vitro* results with *in vivo* effects or the implication that an understanding of mechanism is essential for characterizing interactions. This criterion emphasizes that experiments and observations at multiple levels of biological organization may be necessary to properly characterize interactions that occur in complex biological systems. Each level of biological organization may have limitations for assessing interactions, and such limitations should be minimized by selecting assays at a level (or levels) relevant to the goals of the study. Failure to appreciate limitations in the level of biological organization utilized can lead to errors of interpretation and characterization common in the interaction literature. Carefully discerning the limitations of various approaches can help focus investigative research and toxicity assessments on interactions that have significance to human and ecological health.

Because this criterion may not be readily understood apart from its application, we extend our discussion here by considering theoretical interaction experiments using two classic therapeutic agents, propranolol and epinephrine, as an example.[4]

[4] A later discussion in this review (see subsection entitled "Synergistic Pheromone Analogs") also underscores the importance of assessing interactions at relevant levels of biological organization and will assist the reader in understanding and applying the fifth criterion.

Epinephrine, one of the endogenous autonomic neurotransmitter hormones, is a universal α-β-adrenergic receptor agonist and is also used clinically as a pressor agent. Propranolol is a nonselective β-adrenergic receptor antagonist and the first of such drugs to gain wide clinical use, mainly in the treatment of hypertension. A surprising response has been seen in some patients on propranolol that receive epinephrine during minor surgical procedures. Rather than propranolol blunting the pressor action of epinephrine, some patients have experienced an acute hypertensive crisis (Dzubow 1986; Whelan 1987).

Interaction experiments performed at suborganismal levels of biological organization might not predict synergism between propranolol and epinephrine. Competitive binding studies and adrenergic functional assays performed in a tissue preparation that contains predominantly α-adrenergic receptors, such as skin, would be expected to show no interaction between propranolol and epinephrine. If the same types of experiments were performed in cardiac muscle, a tissue that contains predominantly β1-adrenergic receptors, propranolol would be expected to displace radiolabeled epinephrine competitively, have no β1-stimulatory activity alone, but to antagonize β1-stimulation by epinephrine. Thus, depending on the tissue and assay, either non-interaction or antagonism might be found.

The reported hypertensive crisis appears to be dose dependent and is thought to be due to unopposed α-pressor activity of epinephrine in the presence of the β-receptor blockade by propranolol (Dzubow 1986). Epinephrine normally has a bimodal effect on vascular tone, constricting some vessels through α-stimulation and relaxing others through β-stimulation. In the presence of propranolol, only vasoconstriction occurs, which can produce the abnormally large increase in blood pressure. Experiments in whole animals would be necessary to detect a synergism of epinephrine-induced blood pressure elevation because it is the integrated response of different receptor systems and different tissues to these drugs that produces the precipitous rise in blood pressure. On the other hand, the clinical observation of hypertensive crisis in patients receiving combinations of epinephrine and propranolol is initially surprising and gives little clue as to the type of interaction that produces the response.

III. EVALUATING INTERACTION STUDIES USING THE FIVE CRITERIA

In the following section, we attempt to demonstrate the utility of the five criteria for evaluating interaction studies by applying them to a series of publications from the recent toxicological literature. The goal of this section is to show how the risk assessor can use the criteria to evaluate interaction studies when conducting mixture risk assessments. To meet this goal, we apply the criteria to three different types of data sets that the risk assessor may encounter when evaluating interaction studies. We first discuss three studies that differ in goals and in methodology, yet that all fulfill our criteria. These three studies are conducted with sufficient rigor to be used in lieu of default assumptions in a component-based approach to assessing mixture effects. Next, we discuss an example in which the criteria are used to weight two studies that come to contradictory conclusions regarding the interaction between chemicals. Third, we apply the criteria to studies that lack sufficient rigor to be used in lieu of a default additivity assumption for the toxicity of chemicals in mixture. We

conclude this section by comparing our conclusions with those of others and discussing how the five criteria can be used to improve the reliability of interaction studies in toxicology. A summary of these evaluations is presented in Table 1.

A. Studies That Satisfy All Five Criteria

1. Kissin *et al.* (1987)

Kissin and co-workers demonstrated convincingly that the loss of righting reflex in rats administered thiopental is synergized by the opiate agonists morphine and fentanyl. The loss of righting reflex in rats has been used extensively in pharmacology to test the hypnotic effects of drugs (Erickson *et al.* 1980; Garfield and Bukusoglu 1996; Gessner 1995; Hu *et al.* 1987; Kissin *et al.* 1989, 1990, 1991,1992; Kurtz *et al.* 1996; Tverskoy *et al.* 1989), and is used to predict sedative-hypnotic potency in humans. Although the exact mechanisms that produce loss of the righting reflex are not fully characterized, the results of the assay are directly interpretable with respect to the sedative-hypnotic potency of the dose of drug administered. Thus, Kissin and co-workers tested the interaction at a relevant level of biological organization (Criterion 5).

DRCs were characterized at five dose levels for each drug individually (Criterion 1) and for three different fixed mixtures of the agents (*e.g.,* morphine alone; thiopental alone; thiopental plus morphine at ratios of 1:0.7, 1:3.5, 1:17) (Criterion 3). A similar scheme was used to test combination responses of thiopental and fentanyl. The drugs were tested individually at doses equally spaced across a range from one that blocked the righting reflex in none of the animals to one that blocked in all of the animals (fulfills Criterion 1). Results of the individual dose response experiments were used to determine the dose combinations to be tested (fulfills Criterion 3). A probit procedure was used to determine ED50 values from the DRC of each fixed mixture. ED50 additivity isoboles and ED50 dose combinations for each fixed mixture were plotted on a single isobologram. Since isobolographic techniques are based on Loewe-additivity, the requirements of Criterion 2 are met (see Figures 2 and 3 for an explanation of isobolographic methods).

The distance between the ED50 dose for a fixed mixture and a corresponding reference point on the additivity isobole was used to perform a statistical test (Criterion 4) of the no-interaction hypothesis. The reference point on the additivity isobole was determined by its intersection with a line connecting the origin and the ED50 dose of the fixed mixture. The standard error of this distance was computed by the method of propagation of error (Ku 1966). Error estimates from a fixed mixture ED50 point as well as from single-drug ED50 points were used. The ratio of the measured distance to its standard error was subjected to an approximate *t*-test under the null hypothesis of Loewe additivity. Differences from the predicted additive response were found for three ED50 dose combinations of morphine-thiopental ($p < 0.001$). Two ED50 dose combinations of fentanyl-thiopental were also different from the response expected based on Loewe additivity ($p < 0.01$), and one dose combination was additive (non-significant at $p < 0.5$). Criterion 4 is thus fulfilled by Kissin *et al.* (1987).

The concise study by Kissin and colleagues (Kissin *et al.* 1987) demonstrates very clearly and simply how the five criteria are addressed and fulfilled by an interaction

Table 1. Evaluation summary for studies critiqued in Section III.

Study	Criterion 1 (C1) DRC Characterization	Criterion 2 (C2) No-Interaction Hypothesis	Criterion 3 (C3) Dose Combinations	Criterion 4 (C4) Statistical Test	Criterion 5 (C5) Biological Organization	Summary
Kissin *et al.* (1987)	Tested 5 doses of each agent equally spaced from 0% to 100% response. Fully satisfies C1.	Loewe Additivity (isobolographic analysis). Fully satisfies C2.	Tested 3 dose combinations at constant dose of thiopental. Satisfies C3; could have tested additional dose combinations containing various doses of thiopental.	Determined ED_{50} values by Probit analysis; departure from additivity by propagation of error formulae and approximate t-test. Fully satisfies C4.	Measured loss of righting reflex in laboratory animals, a standard test of sedative hypnotic activity that allows interpretation of drug interactions. Fully satisfies C5.	Fully Satisfies C1, 2, 4, 5. Satisfies C3.
Levasseur *et al.* (1997)	Tested 7 doses across a 10^6-fold concentration range that produced cell survival from 10% to 100%. Fully satisfies C1.	Loewe Additivity (Universal response surface approach and isobolographic analysis). Fully satisfies C2.	Tested 5 fixed mixtures across a 10^6-fold concentration range. Fully satisfies C3.	Used equations for Hill model and Loewe additivity to derive an interaction equation. Applied statistical test to interaction parameter. Fully satisfies C4.	Used cell survival to define role of interactions in clinical efficacy of combined anti-tumor drug treatment. Fully satisfies C5.	Fully Satisfies C1-5. Used non-interaction equation that allowed estimation of the magnitude of interactions.
Taylor *et al.* (1995)	Used 7 doses to characterize DRC and 4 or 5 doses to identify linear range of the DRC between minimum and maximum response levels. Fully satisfies C1.	Loewe Additivity. Fully satisfies C2.	Tested 4 to 5 dose combinations for each pair. Fully satisfies C3.	Applied the Wald statistical test to departure from additivity. Distinguished synergy from antagonism by sign (+, -) of the interaction term. Fully satisfies C4.	Assessed interactions at cellular level in a validated mutagenicity test. Fully satisfies C5 for intended purpose (mutagenic potency testing).	Fully Satisfies C1-5.

Table 1 (continued)

Study	Criterion 1 (C1) DRC Characterization	Criterion 2 (C2) No-Interaction Hypothesis	Criterion 3 (C3) Dose Combinations	Criterion 4 (C4) Statistical Test	Criterion 5 (C5) Biological Organization	Summary
Barger *et al.* (1990)	Did not characterize concentration response relationships for pheromone or the analogue. Fails C1.	Did not define a no-interaction hypothesis or test a non-interaction model. Fails C2.	Tested only one dose combination of the agents. Fails C2.	Did not apply a statistical test to compare observed with expected results. Fails C4.	Tested the population level, but could not discern synergy from antagonism under field conditions. Fails C2.	Fails C1-5.
Mayer and Dodemide (1995)	Tested 4 concentrations of the pheromone and 4 of one analogue. Identified maximal and minimal response levels. Satisfies C1.	Response Addition defined, but applied inappropriate equation. Fails C2.	Tested only 2 dose combinations, but of an ineffective analogue with sub-effective and median effective concentrations of the pheromone. Satisfies C3.	Applied a statistical test to compare observed with expected results, but tested inappropriate non-interaction model. Partially satisfies C4.	Tested dose combinations at both the levels of the individual neuron and the whole organism. Fully satisfies C5.	Satisfies C1, 3, 5. Fails C2. Partially satisfies C4. Reanalysis against Loewe Additivity agrees with reported synergism, but needs statistical confirmation.
Arnold *et al.* (1997)	Characterized DRCs of estradiol-17β, estriol and estrone but may not have identified the maximal response levels. Partially satisfies C1.	Response Addition implied, but used questionable equation. Partially satisfies C2.	Tested 12 (full-factorial) dose combinations of the agents *in vitro*, probably within the response range corresponding to the individual DRCs. Satisfies C3.	Did not apply a statistical test to compare observed with expected results. Fails C4.	Tested molecular level in one species; whole animal in another species, but there is no established relationship between the two models. Fails C5.	Satisfies C3; Partially satisfies C1, 2. Fails C4, 5. Failed to report antagonistic combinations, but broadly concluded synergy.
Bergeron *et al.* (1994)	Tested only 2 concentrations of chemicals. Fails C1.	Did not define a no-interaction hypothesis or test a non-interaction model. Fails C2.	Tested only two dose combinations of the agents at 1:1 ratio, but relationship to individual DRCs unclear. Partially satisfies C2.	Did not apply a statistical test to compare observed with expected results. Fails C4.	Used whole eggs and embryos to determine sex ratio, but relevance of (fixed-temperature model is questionable. Fails C5.	Partially Satisfies C3. Fails C1, 2, 4, 5. Conclusions require several assumptions, including statistical significance.

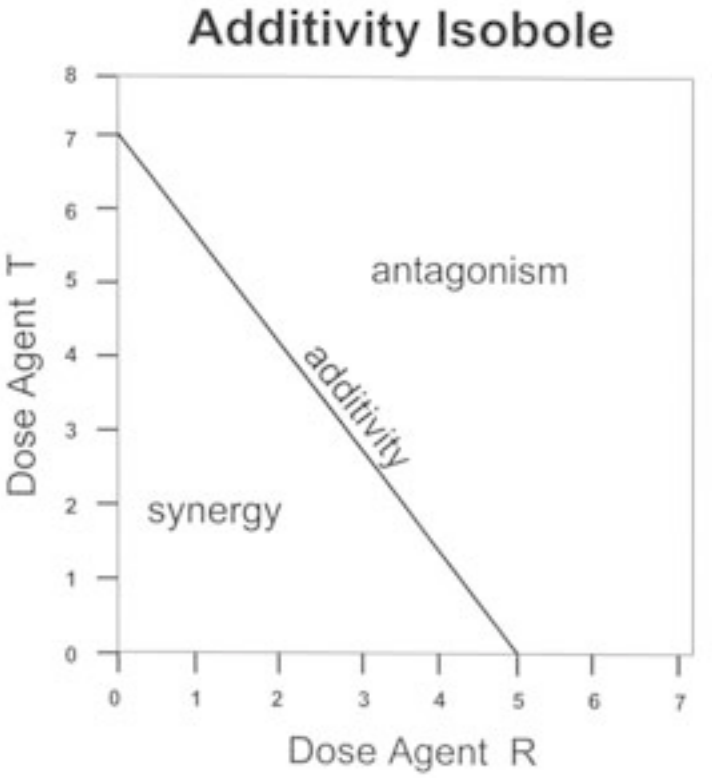


Figure 2. Illustration of an isobologram showing the additivity isobole for two theoretical chemicals, R and T. An isobologram is a dose versus dose plot for two chemicals. Doses of R and T that produce the same level of effect in a particular assay, *i.e.*, that are isoeffective, are plotted on the axes and connected by a line of Loewe additivity, called the "additivity isobole".

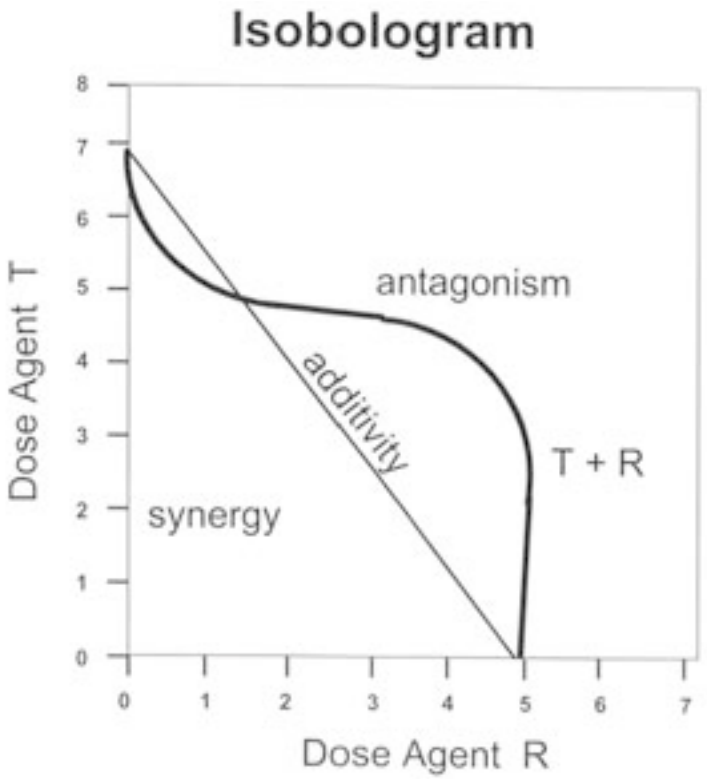


Figure 3. Illustration of an observed isobologram for theoretical chemicals R and T. Combination doses of R and T isoeffective with individual doses are plotted according to their dose coordinates. Combination doses that conform to the no-interaction hypothesis, *i.e.,* are Loewe additive, fall along the additivity isobole. Synergistic combinations of R and T fall below and antagonistic combinations above the additivity isobole.

study (see Table 1). The Kissin study is also an example of using an assay with direct relevance to the whole organism to test chemical interactions. Interaction studies are often intended to have relevance to either whole organisms or to populations of organisms. Other levels of biological organization, however, may also be of interest. Sometimes it is the molecular level of biological organization that is relevant because the goals of the study are to understand fundamental modes of toxicity at the molecular level, such as mutagenicity. In other instances, a molecular or cellular level of biological organization is relevant because it can provide information about the mechanism underlying effects observed in organisms or populations. The next study discussed is an example in which the cellular level is the relevant level of biological organization to test interactions between chemotherapeutic agents.

2. Levasseur *et al.* (1997)

Synergistic cytotoxicity is of particular interest to pharmacologists working on the development of chemotherapeutic regimens. Synergistic cytotoxicity presents the potential for increased efficacy against disease without increasing toxicity to the patient (Greco *et al.* 1996). The combination of paclitaxel and cisplatin shows good clinical efficacy against ovarian carcinoma. The mechanism of this efficacy is not fully understood, but its elucidation could provide clues as to how the therapeutic regimen might be optimized. Levasseur and co-workers studied the growth inhibitory effects of combinations of paclitaxel and cisplatin in various drug-sensitive and drug-resistant human carcinoma cell lines in order to examine a potential cellular basis for the combination efficacy of these agents.

The study was designed to address the following specific questions regarding potential interactions between the agents. Might synergy at the cellular level contribute to the clinical efficacy of paclitaxel and cisplatin? What is the shape of the interaction response surface for paclitaxel and cisplatin for a set of human cell lines and drug exposure schedules? Does drug resistance modify the cellular response to the combination of paclitaxel and cisplatin? These clearly stated goals provided a rational basis for the study design and the type of interaction analysis used.

Table 1 summarizes how the Levasseur study fully satisfies all five criteria, but several features of the study are worth discussing in more detail. Levasseur *et al.* (1997) used a statistical method that not only satisfied Criterion 4 but also allowed them to quantify the magnitude of an interaction, a value that could be important for many risk assessment applications. To accomplish this, they fitted an interaction equation to the growth inhibition data for each dose combination with a nonlinear regression technique that enabled the estimation of an interaction parameter α along with its standard error. The no-interaction hypothesis (Loewe additivity) would be rejected if the 95% confidence interval for the interaction term α did not include zero. The magnitude of α is directly related to the intensity of the interaction and to the degree of bowing of the isobole. Positive values of α indicate synergy, whereas negative values of α indicate antagonism.

Although the Levasseur study on paclitaxel and cisplatin was conducted *in vitro* with no corroborating results *in vivo*, we consider it to have been conducted at a relevant level of biological organization (Criterion 5). In this instance, *in vitro* results

alone are readily interpretable because the goal of the study was to investigate potential cellular mechanisms underlying the enhanced clinical efficacy of combination therapy in cancer patients. The authors discuss that explanations based on pharmacokinetic alterations, cell cycle dependent toxicity, or pharmacodynamic mechanisms cannot account for the efficacy of combination treatment. Therefore, it was reasonable to investigate interactive effects at the cellular level in an attempt to better explain the clinical efficacy and hopefully to optimize the therapeutic regimen.

Levasseur and colleagues found qualitatively different types of interactions at different dose combinations of paclitaxel and cisplatin (Levasseur *et al.* 1997). For example, in A121 cells, only one treatment schedule produced a consistent interaction — antagonism — for all effect levels at all dose combinations tested (Figure 8a in the published study). Other treatment schedules produced synergy, antagonism or no-interaction at a particular effect level depending upon the binary dose combination (Figure 8b and Figures 9a-c in the published study). For A121 cells treated for 24 hours with paclitaxel prior to 4 hours of treatment with cisplatin, dose combinations yielded unambiguous synergy only for the higher effect levels. Isoboles for the IC50 and below showed complex response surfaces with localized regions of synergy and antagonism. The dose combination-dependence of these observed interactions underscores the necessity of assessing more than one ratio of the mixture components — in this case, five fixed mixtures were used — over a range of concentrations (Criterion 3).

Overall, the authors characterized their results as generally supporting the no-interaction hypothesis. It is important to understand that such a global characterization is possible and defensible only if a sufficient number of dose combinations are tested. Studies that do not incorporate such rigor cannot reliably support generalized statements about interactions between chemicals. The rigor of testing numerous dose combinations of paclitaxel and cisplatin at the cellular level allowed Levasseur and colleagues to draw reliable conclusions regarding potential mechanisms underlying the good clinical efficacy of combination therapy. Given the lack of consistent synergism at the cellular level, the authors attributed the clinical efficacy of combination therapy to the non-overlapping spectrum of organ toxicity of the two agents. A high dosage of both drugs can be attained, and thus a high level of anti-tumor activity, without increasing organ toxicity in the patient.

Another interesting feature is that a high degree of experimental variability was noted for interactions between paclitaxel and cisplatin. Ten experiments were conducted using each drug combined with itself to serve as a negative control for the no-interaction hypothesis. For some of these "sham" dose combinations, inconsistent isoboles were produced, but these were attributable to dilution errors or to increased sensitivity of cells pre-treated with a first dose of the chemical at earlier time points in some treatment schedules. Otherwise, these negative control experiments confirmed the no-interaction hypothesis. In dose combination experiments with both drugs, however, the shapes of the isoboles often varied from experiment to experiment for some treatment schedules, despite the fact that the same techniques have been used to produce consistent isoboles for combinations of other anti-tumor drugs (Faessel *et al.* 1998; Gaumont *et al.* 1992; Greco *et al.* 1990). The variability in this set of experiments suggests that close attention must be paid to

inter-experimental variability in interaction studies. Indeed, the authors use the variability in this set of results as an illustration of why caution should be exercised when interpreting interesting but unreplicated interactions (Levasseur *et al.* 1997). They stress that for interaction studies, it is extremely important to replicate experiments a sufficient number of times to be certain that the interactive effects are observed consistently.

3. Taylor *et al.* (1995)

Our third example that satisfies all five criteria is a study by Taylor and co-workers that tested the assumption of additivity for binary combinations of mutagens. The literature prior to this study was contradictory as to the mutagenicity of binary mixtures, with superadditivity (synergy), additivity, and subadditivity (antagonism) being reported (Albertini and Gocke 1992; Crebelli *et al.* 1991; Hass *et al.* 1987; Hermann 1981; Kaden *et al.* 1979; Kawalek and Andrews 1981; Ogawa *et al.* 1985, 1987; Salamone *et al.* 1979; Somani *et al.* 1981; Thornton-Manning *et al.* 1989; Whong *et al.* 1989; Yoshida *et al.* 1979). These complex responses were dependent on the dose combination tested. The majority of interactions reported in these studies were non-additive responses, with superadditivity (synergy) being reported most frequently. In addition, few studies tested interactions between two mutagens. Instead, interactions between a mutagen and a comutagen were tested (Taylor *et al.* 1995). A comutagen is a non-mutagenic chemical that enhances or inhibits the activity of a mutagen. Taylor and co-workers sought to extend the body of literature on the mutagenicity of binary combinations to include binary combinations of known mutagens and to evaluate relevant mutagenic environmental mixtures. They tested mutagenicity with and without metabolic activation with a standard S9 liver fraction.

Taylor and co-authors (Taylor *et al.* 1995) use the term "mixture" to refer to an entity found in the environment that consists of many different chemicals. They did not attempt to assess interactions between the components of complex mixtures, but instead tested mixtures as single entities and in binary combination with single chemicals. Single chemicals were selected on the basis of representing chemical classes known to be responsible for much of the mutagenic activity of the complex environmental mixtures. To test interactions between direct-acting mutagens, an organic extract of diesel exhaust was tested in the absence of S9, as was 1-nitropyrene, which contributes a significant fraction of the mutagenic activity of diesel exhaust. A chlorinated furanone found in some chlorinated drinking water supplies was also tested without metabolic activation. To test interactions between indirect-acting mutagens, an organic extract of combusted polyethylene and one of its constituents, benzo(a)pyrene, were tested with S9 metabolic activation, as was 4-aminobiphenyl, a compound found in cigarette smoke. By testing these single compounds alone and in binary combinations with the complex mixture, Taylor *et al.* obtained information about potential interactions between the single mutagenic components and the rest of the mixtures.

In preliminary experiments, Taylor *et al.* (1995) used seven doses to define the DRC for each agent individually (meets Criterion 1). Based on the resulting individual curves, doses were selected for the binary studies that fulfilled three conditions. The doses selected produced at least a twofold increase in mutant yield when

used individually. The doses selected were within the linear region of the DRC. The sum of the response for any two doses of different agents would not exceed 300 revertants per plate. Four or five doses were selected for testing individually and in combination. These conditions meet the requirements of Criterion 3 that different ratios of the mixture components be assessed across a sufficient concentration range. These conditions also form a defensible basis for selecting a model of non-interaction for testing the no-interaction hypothesis (satisfies Criterion 2). A statistical analysis of responses for dose combinations in the linear range of the DRCs was employed that is equivalent to testing the Loewe non-interaction hypothesis.

Models were fitted to the mutagenic responses at all doses of a given pair of chemicals, both alone and in combination. The model's fit included an interaction term that quantified the deviation from a purely additive response, thus allowing a statistical test to be applied to the non-interaction model (Taylor *et al.* 1995). The no-interaction hypothesis (Criterion 2) for the data was tested using the Wald test statistic (meets Criterion 4). The direction of departure from non-interaction, *i.e.,* synergy versus antagonism, was determined according to whether the value of the interaction term was less than or greater than zero.

Mutagenicity is a toxicological effect that requires evaluation at the molecular level of biological organization. This is because the expression of mutation at the level of the whole organism is sometimes ambiguous as to the underlying toxic effect. The standard method for assessing mutagenic potential is the *Salmonella typphimurium* (Ames) assay. Although there are ways to assess mutagenicity using whole organisms, the Ames assay uses fewer resources and time. The Ames assay has been used extensively for more than 4 decades and the limitations of its predictivity and relevance for mutagenicity and carcinogenicity in mammals has been well characterized (Ashby and Tennant 1988; Tennant *et al.* 1987). Because the *Salmonella* assay is required for safety assessment under several regulatory programs (*e.g.,* those pursuant to FIFRA, FFDCA), it is a reasonable goal to understand how mixtures behave in this assay. Because of those factors and its stated goals, the Taylor *et al.* study meets the criterion for assessment of interactions at a relevant level of biological organization (Criterion 5).

According to the literature on mutagenic activity of mixtures published prior to the Taylor study, non-interaction (additivity) has generally been observed at lower doses of mutagens, whereas the interactions observed were generally at higher doses. Synergy is reported to occur more frequently than antagonism. In agreement with previous literature, Taylor and co-workers found that combination responses were generally additive at low doses and interactive at high doses. Rather than a uniform interaction, they observed regions of local synergy and antagonism in the combination responses. In contrast to previous reports, however, they observed antagonism more frequently than synergism at higher doses, and this was true especially for combinations of indirect-acting mutagens. One possibility for this difference might be that other studies used different concepts of non-interaction for the no-interaction hypothesis. Other possibilities, of course, include the fact that Taylor and co-workers tested binary combinations of mutagens rather than mutagens in combination with comutagens, as did most of the prior studies. Taylor *et al.* conclude that collectively, the literature supports additivity as a reasonable

assumption for binary combinations of relatively low doses of mutagens, but that additivity is less applicable to binary combinations of mutagens at high doses.

It is important to realize that Taylor and co-workers were able to draw meaningful conclusions about the utility of these assays to test the effects of mixtures against those of single compounds because the study design was appropriate for the goals of the study. Had the study not met the Criteria 1 through 4, it would have been impossible to determine if either assay actually detected interactions between known mutagenic agents in mixtures. The relevance and limitations of the *Salmonella* assay for predicting mutagenicity have been established, hence the biological level used to assess mutagenicity was relevant (satisfies Criterion 5). Meeting these conditions also allowed the comparison of two forms of the assay for their utility in studying interactions.

The preceding examples (Kissin *et al.* 1987; Levasseur *et al.* 1997; Taylor *et al.* 1995) illustrate that the five criteria provide a means of consistently evaluating studies that differ in goals, study designs, and methods of analyzing interactive effects. Despite such differences, the criteria can be applied equally to each study and help reveal that all three were conducted by established methods sufficiently thorough to yield reliable results. The results and conclusions of these studies are of sufficient quality to be used in lieu of default assumptions regarding the expected toxic effect of chemical mixtures. However, before applying the results of an interaction study to a particular risk assessment, a risk assessor should also consider the relevance of the conditions and doses used in a study to the exposure scenarios in question. Interactions that occur at high doses may not be manifested at the low environmental contaminant levels that many risk assessments must consider (Jonker *et al.* 1990, 1993; Jonker *et al.* 1996; Seed *et al.* 1995).

B. Using the Criteria to Weight the Quality of Contradictory Data

Risk assessors may often be faced with published studies that reach contradictory conclusions regarding the type of chemical interaction exhibited by two or more chemicals for a particular response. The following two studies investigate interactions between insect pheromones and structurally related synthetic analogs. An evaluation of these two studies illustrates how the criteria can assist in weighting contradictory results in order to choose the best model for combined action of the chemicals when present in mixtures (see Table 1). A comparison of these studies also illustrates why the fifth criterion is essential to reliable interaction analysis.

Synergistic Pheromone Analogs

Burger *et al.* (1990)

Mayer and Doolittle (1995)

Female moths emit chemical sex signals, called sex pheromones, which are detected by olfactory specialist neurons housed within characteristic, often sexually dimorphic, sensilla located on the antenna of male moths. These antennal neurons are specialized to detect primarily one constituent of the female signal at low airborne concentrations. At elevated concentrations, however, the neurons will

respond to other compounds. Stimulation of the specialized neuron by female sex pheromone induces male moths to fly upwind in search of the emitting female. A male navigates toward a female by continuously redirecting his flight path — right or left — according to the antennal specialized neuron that is more highly stimulated. When both specialized neurons are maximally stimulated, the male stops his upwind flight in anticipation of finding the female nearby.

In 1990, Burger and co-workers (Burger *et al.* 1990) reported that 7-vinyldecyl acetate (7-VD), an analog of the female pheromone of false codling moths, *Crytophlebia leucotreta* (Meyrick), reduced field trap captures of males when incorporated into otherwise effective sex pheromone bait. These researchers found that while 370 males were captured at traps containing only a virgin female moth, no males were captured at traps that included both a female moth and a dispenser containing 2 μL of 7-VD. The authors concluded that 7-VD was a pheromone inhibitor.

Because it is technically unfeasible to measure levels of airborne sex pheromones in the field, Burger and co-workers did not determine the levels of the natural pheromone or synthetic analog emitted from traps. Neither did they measure concentrations that reached male moths at distances away from traps. It was thus impossible for Burger and colleagues to compare the responses produced by specific doses of individual compounds with those of specific combinations. Although somewhat implied by their report, Burger and co-authors did not use the term "antagonism" to characterize the effect of the analog, but instead refer to "inhibitory" pheromonal effects.

Based on the structural relationship of 7-VD to one of the two false codling moth pheromone components, Mayer and Doolittle (1995) undertook the synthesis of analogs that would have a structurally similar relationship to the sex pheromone (Z)-7-dodecenyl acetate (Z7-12:Ac) of the cabbage looper *Trichoplusia ni* (Hübner). They compared both the behavioral and neurophysiological effects of the analogs 6-vinlyldecyl acetate (6-VD) and 10-vinyltetradecyl acetate (10-VD) alone and binary mixtures with Z7-12:Ac. Action potentials were recorded from specialist neurons within individual antennal sensilla exposed to carefully controlled airborne concentrations of the compounds and mixtures. Binary mixtures of Z7-12:Ac and the analogs produced a response greater than the sum of the responses produced by the individual stimuli alone. In contrast with the inhibitory effect on male trapping observed in the field by Burger and co-workers, neither of the analogs inhibited the neurostimulatory effect of the pheromone component Z7-12:Ac. Normally effective concentrations of Z7-12:Ac in combination with analogs produced supra-maximal stimulation of the specialist neuron.

Mayer and Doolittle extended their electrophysiological studies by conducting combination studies on live male cabbage loopers in a wind tunnel. Consistent with the results of the electrophysiological studies, concentrations of pheromone that were incapable of eliciting an upwind flight and mating response in males when tested alone produced a marked increase in upwind flight and mating behavior when present in combination with the analogues. These results confirm that in addition to producing a synergistic electrophysiological response at the specialized neuron, the particular pheromone analogs tested were able to elicit synergistic responses of the relevant behavioral endpoints in individual male moths. In combination with optimally effective levels of pheromone, flight response was inhibited by the analogs.

The corroboration of results at the level of the neuron with behavioral responses elicited in the organism not only confirm that the synergism has molecular, physiological, and organismal relevance, but may also help to explain why Burger and coworkers observed an inhibition of the pheromone behavioral response in the field. Rather than a pharmacological antagonism at the molecular level, inhibition of the pheromone behavioral response in the wind tunnel and in the field would appear to be due to synergistic overstimulation of the specialized receptor neuron. At supramaximal stimulation of specialized neurons, males may not be induced to fly toward the emitting female because they perceive that the female is already in close proximity due to the intensity of the pheromone stimulus. If flight is initiated, successful navigation may be impossible because the difference in stimulation level between the two antennae necessary for directed flight would be overcome at distances far from the emitting female. The physiological and behavioral consequences of supra-maximal stimulation of the specialist neuron are consistent with these explanations.

Despite testing the effects of 7-VD on the population response of interest, Burger and co-workers (Burger *et al.* 1990) were unable to interpret the interaction of the analog with natural pheromone. The response observed in the field could have occurred by either antagonism or synergism of pheromone action (Criterion 5). Burger and colleagues were unable to test the individual agents (Criterion 1) or combinations of the agents (Criterion 3) across an appropriate dose range in field population studies, a limitation that further constrained their ability to distinguish the type of interaction responsible for the effect. They neither defined a no-interaction hypothesis (Criterion 2) nor tested observed results statistically against those predicted by a model of non-interaction (Criterion 4). Thus, the Burger *et al.* (1990) study does not satisfy any of the criteria proposed here (Table 1 summarizes the evaluation).

In their own combination experiments with pheromones and analogs, Mayer and Doolittle (1995) did not test a broad range of dose combinations (Criterion 3). They did, however, use estimated DRCs to choose effective and sub-effective concentrations of Z7-12:Ac, the natural pheromone. One of the analogs (6-VD) was tested at four airborne concentrations ranging from 2×10^{-12} *M* to 1×10^{-10} *M*, and the second analog (10-VD) at two concentrations, 3×10^{-13} and 3×10^{-12} *M*. These concentrations of the analogs were tested alone or in combination with two concentrations of Z7-12:Ac (1.9×10^{-13} and 2.1×10^{-11} *M*), the lower representing a subeffective concentration and the higher an effective concentration. This satisfies Criteria 1 and 3 because the goals of the study were to investigate the possibility that inhibition of pheromone activity was a likely explanation for the effect of the pheromone analog on male moths in the field, as reported by Burger.

Mayer and Doolittle also specifically defined synergism, and in so doing, defined a no-interaction hypothesis (Criterion 2) based on effect summation. The authors state:

> *Synergism, in the sense that we use it for these studies, is the resultant response of the specialist neuron to the cooperative action of two or more discrete chemicals that is greater than the sum of the responses to the individual compounds.*

There is debate regarding whether Loewe additivity, Bliss independence, or models based on slopes of DRCs are the most appropriate for receptor-mediated, graded

responses (Poch *et al.* 1995b; Scaramellini *et al.* 1997). However, Mayer and Doolittle's no-interaction hypothesis — "the sum of the responses to the individual compounds" — is not equivalent to Loewe additivity because the dose response curves are not linear. Neither is it a statement of Bliss Independence, but the authors have provided no other appropriate non-interaction model (fails Criterion 2).

We estimated the expected responses under Loewe additivity and our results agree with the authors' conclusions (see Tables 1 and 2). We note, however, that our results cannot confirm the study's conclusions because they are not based on a statistical test between the predicted and observed responses, which would require data on the individual measurements.

We also inspected the data of Mayer and Doolittle (1995) by the isobolographic technique (see Figures 2 and 3 for a discussion of isobolograms), which is based on the Loewe additivity model of non-interaction. Figures 4a through 4e show isobolograms at various response levels estimated from a point-to-point linear interpolation of the dose-response data for electrophysiological responses of Z7-12:Ac, 6-VD and their binary combinations. These isobolograms show that estimated isoeffective dose combinations fall below the line of additivity estimated from the dose-response data for the agents alone. Numbers printed beside the points indicate combination index values (Berenbaum 1978, 1981). The combination index is a calculation of the degree of departure from the line of additivity, which is the straight line between isoeffective doses of the individual agents plotted on either axis of the graph. Mathematically, the combination index is defined as $CI = d_a/D_a + d_b/D_b$, where D indicate individual doses of agents a and b isoeffective with dose combinations of the combination ab. If the combination index is significantly less than 1, the dose combination is synergistic; if the combination index is significantly greater than 1, the dose combination is antagonistic. These plots provide a visual estimate of the degree of synergy suggested by the published data of Mayer and Doolittle (1995). Although lacking here, a statistical test should be applied to determine whether the confidence intervals for dose combinations overlap the confidence interval around the line of additivity (Gennings *et al.* 1990).

Mayer and Doolittle (1995) used a probit analysis to distinguish differences between treated and control groups in their behavioral studies in the wind tunnel. They did not apply a statistical analysis to confirm differences between observed and expected effects in electrophysiological studies (fails Criterion 4), although they do report standard errors for their measurements. They should have made statistical comparisons between the measured response and the estimated expected response (Loewe additivity) taking account of all statistical variation in the measurements and estimates.

Application of the criteria (see Table 1) shows that the Mayer and Doolittle (1995) study is a more reliable test of chemical interactions than the study by Burger and colleagues (Burger *et al.* 1990). Mayer and Doolittle satisfied Criteria 1, 3, and 5, whereas Burger and co-workers did not satisfy any of the five criteria. This analysis illustrates how the five criteria can be used to weight studies that come to different conclusions regarding chemical interactions, or to weight the evidence for interactive toxicity reported for different chemical mixtures. Use of these criteria to weight

Table 2. Interaction analysis for Z7-12:Ac and 6-VD at the HS(a) antennal sex pheromone specialist receptor neuron.

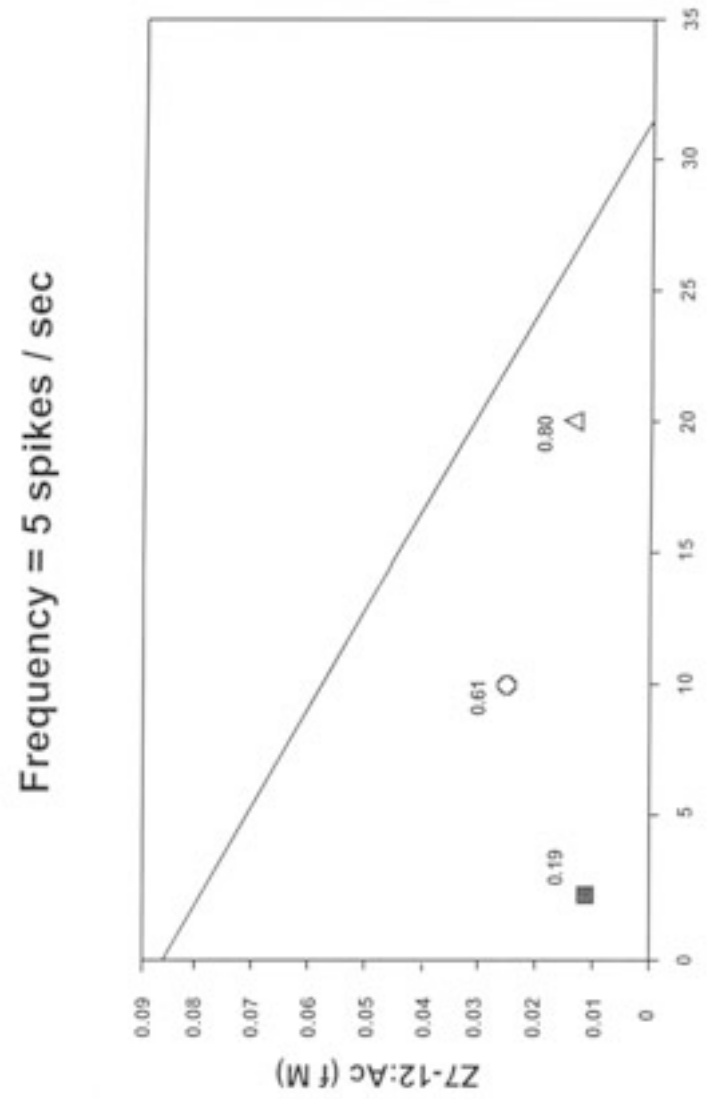
Airborne Concentration[a] (M)		Frequency of Action Potentials in HS(a) Neurons (spikes/sec)		
Z7-12:Ac	6-VD	Observed Frequency[b]	Predicted Loewe Additive Frequency[c,d]	Ratio: Observed Frequency to Predicted Loewe Additive Frequency[d]
0	0	0.3 (9)	-	-
0	$\sim 2 \times 10^{-12}$	4.4 ± 3.9 (8)	-	-
0	$\sim 1 \times 10^{-11}$	0.6 ± 0.3 (8)	-	-
0	$\sim 2 \times 10^{-11}$	1.9 ± 0.7 (8)	-	-
0	$\sim 1 \times 10^{-10}$	23.6 ± 8.9 (8)	-	-
1.9×10^{-13}	0	8.7 ± 2.6 (8)	-	-
1.9×10^{-13}	$\sim 2 \times 10^{-12}$	17.1 ± 4.7 (8)	8.7	2.0
1.9×10^{-13}	$\sim 1 \times 10^{-11}$	35.1 ± 8.7 (7)	8.8	4.0
1.9×10^{-13}	$\sim 2 \times 10^{-11}$	42.7 ± 13.6 (6)	8.9	4.8
1.9×10^{-13}	$\sim 1 \times 10^{-10}$	73.5 ± 37.1 (5)	24.1	3.0
2.1×10^{-11}	0	40.7 ± 10.9 (5)	-	-
2.1×10^{-11}	$\sim 2 \times 10^{-12}$	na	41.1	-
2.1×10^{-11}	$\sim 1 \times 10^{-11}$	na	42.1	-
2.1×10^{-11}	$\sim 2 \times 10^{-11}$	na	44.8	-
2.1×10^{-11}	$\sim 1 \times 10^{-10}$	na	63.0	-

[a] Concentrations as reported in Table 1 of Mayer and Doolittle, 1995.

[b] Frequency ± standard error as reported in Table 1 of Mayer and Doolittle, 1995. The number of neurons measured is indicated in parentheses.

[c] Predicted Loewe Additive Frequencies satisfy 1=da/Da +db/Db, the equation that defines Loewe additivity.

[d] A statistical test has not been applied because the original data were unavailable at the time of analysis.



4a

Figures 4a-e. Isobolograms showing synergistic combinations of Z7-12:Ac and 6-VD at various response levels in an electrophysiological assay. Isoeffective doses of the individual agents and their combinations were estimated using JMP Statistical Discovery Software, SAS Institute, version 3.2 by a point-to-point linear interpolation of the dose-response data of Mayer and Doolittle (1995). Numbers printed beside the points indicate combination index values. The combination index is a calculation of the degree of departure from the additivity isobole.

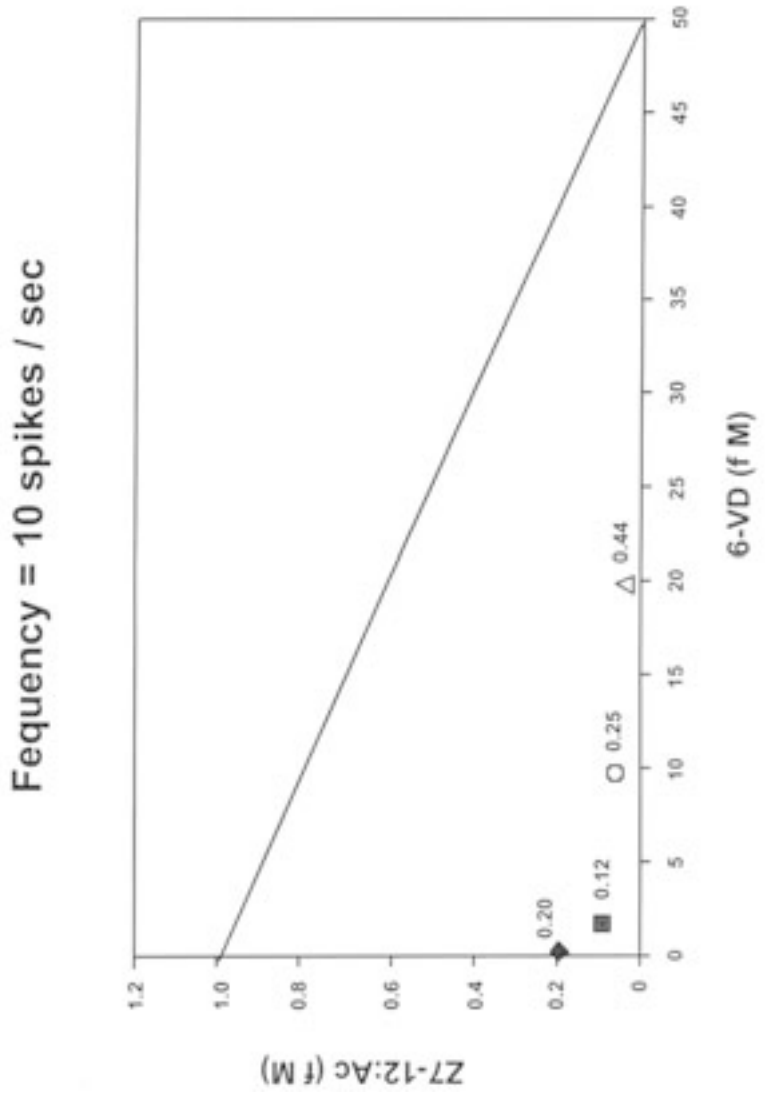


Figure 4b

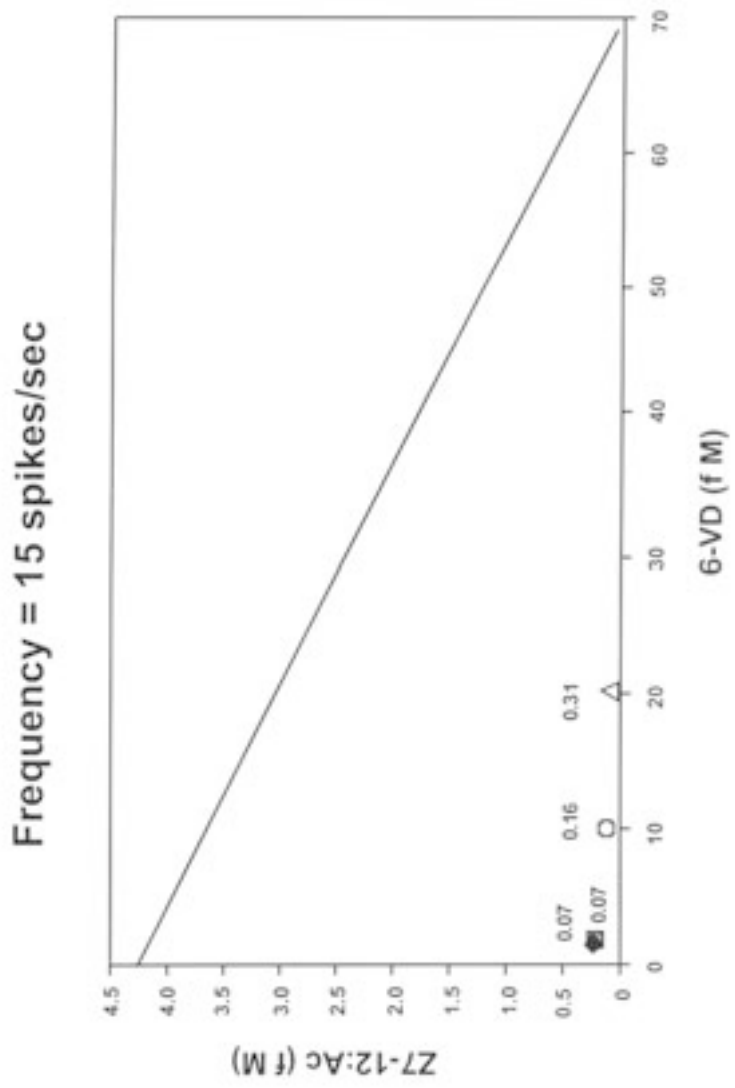


Figure 4c

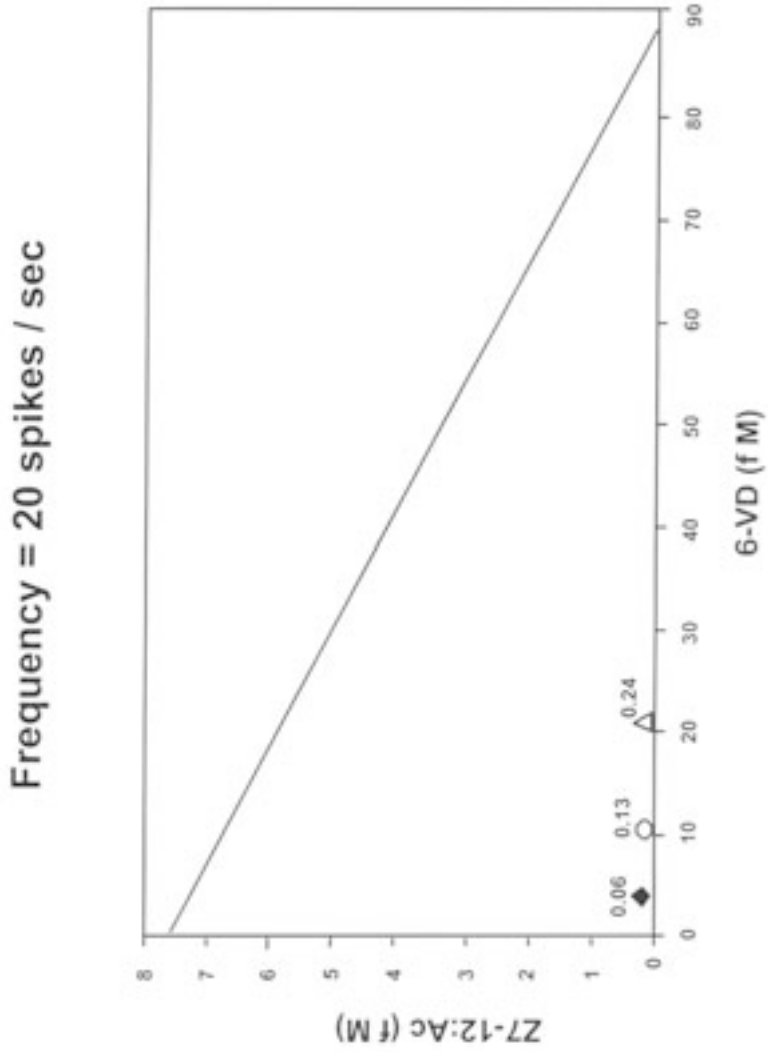


Figure 4d

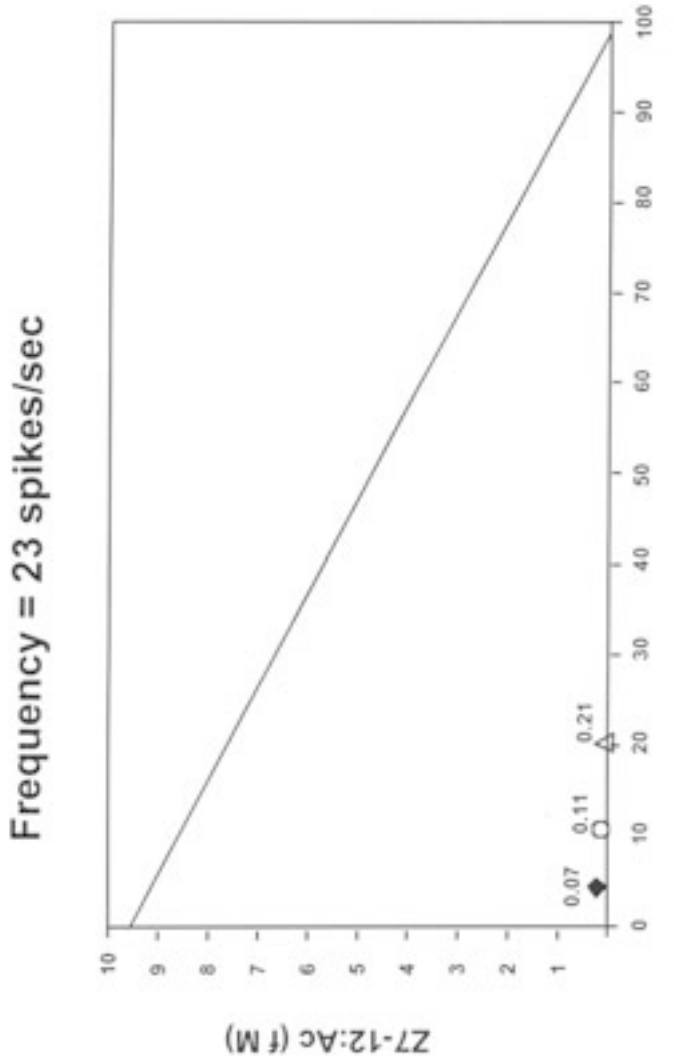


Figure 4e

various interaction studies may be important when faced with the challenge of accepting or rejecting a default non-interaction assumption for mixture risk assessments.

C. Identifying Studies That Lack Sufficient Rigor to Inform Component-Based Mixture Risk Assessments

Synergy between Steroidal Estrogens

Arnold *et al.* (1997)

Arnold and co-workers investigated responses to combinations of steroidal estrogens, the endogenous estrogenic hormones in mammals and many other vertebrates. Receptor-level interactions were evaluated using an *in vitro* transcriptional activation system in yeast consisting of the human estrogen receptor linked to a β-galactosidase reporter gene. Activation of the human estrogen receptor is measured in this system by the activity of β-galactosidase in the growth medium. Interaction of steroidal estrogens was also evaluated *in vivo* using temperature-dependent sex determination in red-eared slider turtles as a model system. Synergistic sex reversal by combinations of polychlorinated biphenyls has previously been reported in this species (Bergeron *et al.* 1994), and the results are discussed later in this paper.

A thorough discussion of the Arnold *et al.* study (Arnold *et al.* 1997a) is important because it demonstrates how applying the criteria helps to identify inadequacies in data interpretation and study design critical to evaluating interaction studies (see Table 1 for a summary of the evaluation). For transcriptional activation experiments in yeast, Arnold *et al.* (1997a) did not explicitly define a no-interaction hypothesis, but compared the arithmetic sum of enzyme activity produced by doses of the agents alone with enzyme activity elicited by the same concentrations in combination (Criterion 2). However, a simple summation function is not consistent with the Bliss independence model, and in this situation is not consistent with the Loewe additivity model because the individual DRCs are not linear in the range used for the interaction analysis. The authors do not specify any other no-interaction hypothesis. Therefore, the study failed to test an appropriate no-interaction hypothesis (fails Criterion 2).

Arnold *et al.* (1997a) used a full-factorial design to test multiple dose combinations based on the doses used to characterize the DRCs of the individual agents (Arnold *et al.* 1997a). DRCs were determined for estradiol-17β, estrone, estradiol-17α, and estriol, but it is unclear whether the largest dose of any of the steroids elicited the maximum transcription activation levels in yeast (partially satisfies Criterion 1). However, since the dose combinations are within the same range as the doses of individual agents tested, the study satisfies Criterion 3 if the no-interaction hypothesis is response addition. Synergism was reported for the following combinations at submaximal response levels: estradiol-17β plus estrone; estradiol-17α plus estrone; estradiol-17α plus estradiol-17β, but not for dose combinations in the maximal response range for the individual estrogens.

The authors discuss that at high response levels dose combinations were not synergistic due to saturation of the receptor-mediated response (Arnold *et al.*

1997a). If one assumes that maximum enzyme activity was elicited by the largest dose combinations of estradiol-17-β and estrone tested, then these combinations might be regarded as purely additive because further increases in dose could elicit no greater response. However, because Arnold *et al.* did not sufficiently characterize the maximal enzyme induction (*i.e.*, saturation) for the individual agents, there is considerable uncertainty in such conclusions. There is a second way to interpret the responses at high doses that is more consistent with the no-interaction hypotheses that was apparently understood by the authors (response-addition). If the DRC is truly linear and no maximum receptor mediated response is identified, then responses at high doses should be viewed as antagonistic.

In their yeast study, Arnold *et al.* (1997a) used a one-way analysis of variance least significant differences test, but it is unclear how this test was applied and whether results obtained with dose combinations were compared statistically with results expected under response addition. Even if the test was applied correctly (*i.e.*, Criterion 4 was satisfied), no inferences can be made about interactions because the no-interaction hypothesis was deficient (Criterion 2 was failed). The raw data from the study would have to be reanalyzed using a valid non-interaction model and appropriate statistical test to determine whether conclusions can be made on a local or global basis.

Had Arnold *et al.* (1997a) used only an *in vitro* system to assess potential interactions, without attempting to verify the interactions in an organism, the study would clearly have failed the requirements of Criterion 5. While transcriptional activation assays may be more relevant to estrogenic activity in an organism than simple equilibrium binding displacement assays, interaction studies in transcriptional activation systems alone can only suggest potential interactions that may or may not manifest changes *in vivo (*McInnes and Brodie 1988). To evaluate interactions *in vivo,* Arnold and co-workers also investigated the ability of combinations of steroidal estrogens to cause sex reversal in red-eared slider turtles (Arnold *et al.* 1997a).

In the laboratory, the temperature at which eggs are incubated during the critical period of gonadal differentiation may influence sex determination in turtles. Turtle eggs were incubated at a male-producing temperature and treated with estradiol-17α, estradiol-17β, or estriol at the beginning of the period of gonadal differentiation during which the eggs are sensitive to the effects of estradiol-17β. The sex ratio at hatch was determined for only three doses of estriol and estradiol-17β individually (Criterion 1) and for only three dose combinations of estradiol-17β given in combination with 0.01 μg of estriol (Criterion 3). The dose response relationship on sex ratio was determined using polynomial logistic regression (Arnold *et al.* 1997a).

This portion of the Arnold study satisfies Criteria 1 and 3 in that the DRCs for the individual agents were characterized sufficiently to see that they are linear in the range of the doses tested (*i.e.*, logit of sex change proportion versus dose was linear).[5] The logit fit to the combination included a quadratic term to represent the flat response beginning at about 0.2 μg. The authors employed a one-tailed t-test to show that the linear term in the DRC for estradiol-17β alone was statistically smaller than the linear term in the DRC for the combination (Arnold *et al.* 1997a). This finding was used to infer synergy for the combination. The authors did not determine the expected response under Loewe additivity or independence (fails Crite-

5 This discussion is developed from Table 3 and Figure 3 of the article (Arnold *et al.* 1997). The raw data were not available.

rion 2); therefore, their statistical test for synergy and their conclusion that the agents have a synergistic effect on sex reversal are questionable (fails Criterion 4).

While using both *in vivo* as well as *in vitro* approaches to assess interactions is preferable in this instance, it is not clear that interactions in a human estrogen receptor-reporter gene construct in yeast have any relevance to interactions in turtles. Hence, that combination of *in vivo* and *in vitro* studies would not fulfill the requirements of Criterion 5 to assess interactions at relevant levels of biological organization. Although we have not evaluated the reliability of the data, synergism between temperature and estradiol has been reported (Crews *et al.* 1994; Wibbels *et al.* 1991a, 1991b). Hormone-temperature interactions, if they occur, could confound interactive effects of steroid hormones. Because it may be necessary to understand interactions between individual sex hormones and temperature before interactions between different sex hormones or chemicals are interpretable in the turtle egg system, the *in vivo* portion of the Arnold *et al.* study is also questionable for satisfying Criterion 5. The level of uncertainty introduced would have to be carefully considered before using the Arnold *et al.* (1997a) study in lieu of default non-interaction assumptions for mixture risk assessment. The study fails two of the five criteria and only partially satisfies others (see Table 1 for a summary of the evaluation).

D. Toward Improving Reliability

Bergeron *et al.* (1994)

Kortenkamp and Altenburger (1998)

The following discussion of the Bergeron *et al.* (1994) study and Kortenkamp and Altenburger's (1998) reanalysis of the Bergeron study using isobols underscores the reasons that a clear and consistent method of evaluating interaction studies is needed. In their review, Kortenkamp and Altenburger (1998) present a clear discussion of how the recent debate over synergistic estrogenic effects has failed to take into account well-developed methods of interaction analysis developed over the last 100 years. Specifically, the authors highlight the need to characterize DRCs of the individual agents tested, the need to define the expected response assuming that chemicals do not interact, and the need to apply a valid statistical test to determine differences between observed effects and expected effects (criteria 1, 2, and 4). These authors also mention that testing dose combinations using multiple fixed mixtures is desirable (Criterion 3). Despite agreeing on the importance of certain aspects of interaction analysis, we differ with Kortenkamp and Altenburger (1998) regarding their acceptance of Bergeron and co-workers' conclusions. Accepting Bergeron's conclusions would appear to us to contradict the importance of the concepts and methods Kortenkamp and Altenburger (1998) explain so clearly.

The Bergeron *et al.* (1994) report was one of the earliest studies suggesting a synergistic estrogenic response. These authors reported a synergistic interaction between two hydroxylated polychlorinated biphenyls in producing sex reversal during development in the red-eared slider turtle. Bergeron and co-workers reached their synergy conclusion without explicit discussion of any of the issues addressed by the five criteria. Of more importance, the substance of their study fails four of the five criteria proposed here for evaluating interaction studies (see Table 1).

Bergeron and co-workers' characterization of the DRCs for the individual PCBs are inadequate (Criterion 1), and a no-interaction hypothesis is not defined or tested (Criterion 2). Their conclusion that the components synergized is based on a single fixed mixture of the chemicals administered at only two doses, an insufficient amount of data on combination responses to characterize interactions generally (Criterion 3). Statistical tests to distinguish departure from additivity or independence were not applied (Criterion 4). Had the study design and data interpretation been adequate, the level of biological organization assessed may have provided a relevant basis for the conclusions had the relevance of the fixed-temperature model been established (fails Criterion 5).

To elaborate, Bergeron *et al.* (1994) used only two doses to characterize the DRCs of the individual PCBs (fails Criterion 1). They tested only two concentrations of the mixture, and both of these were at a 1:1 ratio of the components (fails Criterion 3). For 2′,4′,6′-trichloro-4-biphenylol, 100 μg produced sex reversal in 21% of eggs spotted with the substance, while 200 μg produced sex reversal in 100% of the treated eggs. For 2′,3′,4′,5′-tetrachloro-4-biphenylol, 100 μg produced less than 10% sex reversal and 200 μg produced approximately 26% sex reversal. The two agents were said to produce a synergistic interaction based on the observation that 10 μg of each in combination produced approximately 33% sex reversal. Bergeron *et al.* (1994) made no effort to distinguish their result statistically (fails Criterion 4) from a model of non-interaction (fails Criterion 2). They did not account for measurement variability and ignored DRC estimation uncertainty.

Kortenkamp and Altenburger (1998) plotted an isobologram of the Bergeron *et al.* (1994) data by assuming the rough equivalence of the 21%, 26%, and 33% responses. Their isobologram shows that the 10 μg dose combination of each agent that produced 33% sex reversal falls below the additivity isobole for individual PCB doses of 100 μg and 200 μg (21% and 26% sex reversal), *i.e.,* the combination is synergistic by their reanalysis. Based on their isobolographic reanalysis, Kortenkamp and Altenburger (1998) make a determination opposite ours regarding the reliability of Bergeron *et al.*'s synergy conclusion. They state:

> *Although it is desirable to test combinations of agents at several mixture ratios so that isoboles can be constructed reliably, this is not a necessary prerequisite. Valid conclusions about the combination effects of mixtures can often be drawn on the basis of surprisingly few data. . . . .*
>
> *Even without complete dose response curves for the single compounds, it is obvious . . . [from their own isobolographic reanalysis] . . . that Bergeron et al. have correctly concluded that the two compounds acted synergistically in affecting sex determination in Tachemys scripta.*

We find Kortenkamp and Altenburger's approach to be a constructive application of the local assessment of additivity using limited data. However, claiming that this approach validates Bergeron's conclusions works against their convincing discussion of the importance of applying concepts and methods of interaction analysis developed over the past 100 years in the fields of pharmacology, toxicology, ecotoxicology, epidemiology, chemotherapy, and biometry (Kortenkamp and

Altenburger 1998; concepts addressed by Criteria 1 and 4). In contrast to basing their determination on a reasonable characterization of the DRCs of the agents alone, their determination requires the assumption of a monotonically increasing DRC for sex reversal by the individual PCBs. Their determination was also made in the absence of statistical tests.

Based on the published data, the behavior of the individual DRCs below 100 μg is unknown. Doses less than 100 μg were not tested for the PCBs individually. Thus, no information is provided about the slope of the individual DRCs in the range at which combination responses were assessed. Although we agree that many reasonable extrapolations of the DRCs back to 10 μg or back to 0 μg would support a finding of synergy, there also exist extrapolations of the DRCs that would not support synergy. Without additional data at lower doses, the characteristics of the DRCs (Criterion 1) and subsequent analyses of additivity are speculative.

Therefore, we hesitate to agree with Kortenkamp and Altenburger's (1998) acceptance of Bergeron *et al.* (1994) conclusions because the study lacks sufficient rigor to infer interactions. Bergeron *et al.* (1994) may have correctly identified a synergistic effect, but their conclusions cannot be accepted as valid without assuming both the shape of the individual and combination DRCs and the results of statistical analyses. It is noteworthy that a more recent publication by Bergeron *et al.* (1999) indicates that 12 of 14 binary dose combinations of steroidal estrogens were additive, only two dose combinations were synergistic, and that all three ternary dose combinations tested were additive.

IV. CONCLUDING REMARKS

Mixture risk assessments will generally be more accurate when the underlying studies are scientifically rigorous and match the goals of the risk assessment. However, studies that rely on assumptions rather than accepted principles of interaction analysis are widespread in the toxicological and pharmacological literature, as has been discussed by others (Berenbaum 1989; USEPA 1990) and illustrated by several of the studies analyzed in this paper. In contrast, reliable interaction studies are those that are interpretable without making assumptions about untested and unanalyzed parameters. Although there is debate among experts regarding which models of non-interaction, which methods of combination analysis, and which statistical tests are most appropriate, these unresolved scientific questions should not prevent adherence to accepted principles of interaction analysis. Accepted principles of analysis must be clear so that risk assessors and researchers can apply the science in a way that best supports the needs of risk assessment.

In this paper, we propose a set of five criteria to be used as a guideline for evaluating studies of chemical interactions. In Section II, we attempted to demonstrate that the criteria are grounded in accepted principles of interaction analysis that reflect the consensus of the literature developed over decades of research in pharmacology, toxicology, and biometry. In Section III, we attempted to demonstrate how the criteria advance larger scientific goals. The three examples presented in Section IIIA (Kissin *et al.* 1987; Levassuer *et al.* 1997; Taylor *et al.* 1995) addressed how the five criteria allow a consistent and objective evaluation of interaction studies that differ in goals, design, and methodology. Without such a guideline, the reliabil-

ity of such divergent studies for drawing conclusions, making predictions, and forming testable hypotheses might be less clear to a majority of risk assessors and researchers. By comparing the two studies in Section IIIB (Burger *et al.* 1990; Mayer and Doolittle *et al.* 1995), we attempted to show how the criteria clarify differences between study designs and methods of analysis. Together with the discussion of studies in Sections IIIC and D (Arnold *et al.* 1997a; Bergeron *et al.* 1994; Kortencamp and Altenburger 1998), we tried to demonstrate how the criteria help identify critical deficiencies that limit the inferences that can be drawn based on data. Through these examples, we also tried to show how the criteria help to clarify differences in the underlying assumptions and conclusions of different analysts, and how the criteria help to clarify uncertainties regarding interaction conclusions. We believe these examples also suggest how widespread application of these criteria would lead to more testable hypotheses and more consistent and reliable conclusions. The validity of the five criteria proposed in this paper depend on the extent to which they further such goals.

Although the five criteria proposed in this review are indeed prescriptive, there is wide latitude among the various methods and approaches that will satisfy them. Applying these criteria may assist risk assessors who must evaluate interaction studies for use in risk assessments of chemical mixtures by helping them to distinguish between sound, reliable studies and those for which the conclusions require an unacceptable number of assumptions. Application of the criteria would also allow refocusing the scientific discussion on how best to use interaction data in the toxicological assessment of chemical mixtures. Refocusing the discussion would encourage research aimed at improving risk and safety assessments for chemical mixtures found in products, food and the environment.

ACKNOWLEDGMENTS

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

The authors thank Terry F. Quill and Nathan J. Karch for helpful editorial suggestions and Randy Hatton for citations on epinephrine-propranolol interactions.

REFERENCES

Abdelghani AA, Tchounwou PB, Anderson AC, *et al.* 1997. Toxicity evaluation of single and chemical mixtures of Roundup, Garlon-3A, 2,4-D, and Syndets Surfactant to channel catfish (*Ictalurus punctatus*), bluegill sunfish (*Lepomis Microchirus*), and crawfish (*Procambarus* spp). Environ Toxicol Water Qual 12:237-43

Abou-Donia MB, Lipadula DM, Campbell G, *et al.* 1985. The joint neurotoxic action of inhaled methyl butyl ketone vapor and dermally applied *0*-Ethyl *O*-4-Nitrophenyl Phenylphosphonothioate in hens: Potentiating effect. Toxicol Appl Pharmacol 79:69-82

Albertini S and Gocke E 1992. Phenobarbital: does the positive result in TA1535 indicate genotoxic properties? Environ Mol Mutagen 19(2):161-6

Arnold SF, Klotz DM, Collins BM, *et al.* 1996a. Synergistic activation of estrogen receptor with combinations of environmental chemicals. [see comments] [retracted by McLachlan JA. In: Science 1997 Jul 25;277(5325):462-3]. Science 272:1489-92

Arnold SF, Klotz DM, Vonier PM, *et al.* 1996b. Synergism in estrogen-regulated gene expression. Mol Cell Endocrinol 123:19-122

Arnold SF, Bergeron JM, Tran DQ, *et al.* 1997a. Synergistic responses of steroidal estrogens *in vitro* (yeast) and *in vivo* (turtles). Biochem Biophys Res Commun 235:336-42

Arnold SF, Vonier PM, Collins BM, *et al.* 1997b. *In vitro* synergistic interaction of alligator and human estrogen receptors with combinations of environmental chemicals. Environ Health Perspect 105:615-8

Ashby J and Tennant R W. 1988. Chemical structure, *Salmonella* mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. Mutat Res 204(1):17-115

Banks D and Soliman MR. 1997. Protective effects of antioxidants against benomyl-induced lipid peroxidation and glutathione depletion in rats. Toxicology 116:177-81

Baynes RE, Halling KB, and Riviere JE. 1997. The influence of Diethyl-m-toluamide (DEET) on the percutaneous absorption of permethrin and carbaryl. Toxicol Appl Pharmacol 144:332-9

Ben-Shlomo I, Finger J, Bar-Av E, *et al.* 1993. Propofol and fentanyl act additively for induction of anaesthesia. Anaesthesia 48:111-3

Berenbaum MC. 1978. A method for testing for synergy with any number of agents. J Infect Dis 137:122-30

Berenbaum MC. 1981. Criteria for analyzing interactions between biologically active agents. Adv Cancer Res 35:269-335

Berenbaum MC. 1984. Synergy assessment with growth curves [letter]. J Infect Dis 150:304

Berenbaum MC. 1985. The expected effect of a combination of agents:the general solution. J Theor Biol 114:413-31

Berenbaum MC. 1988. Synergy and antagonism are not synonymous with therapeutic advantage and disadvantage [letter]. J Antimicrob Chemother 21:497-500

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

Bergeron J M, Crews D, and McLachlan J A. 1994. PCBs as environmental estrogens: turtle sex determination as a biomarker of environmental contamination. Environ Health Perspect 102(9):780-1

Bergeron JM, Willingham E, Osborn CTIII, *et al.* 1999. Developmental synergism of steroidal estrogens in sex determination. Environ Health Perspect 107(2):93-7

Bianchi L, Zannoli A, Pizzala R, *et al.* 1994. Genotoxicity assay of five pesticides and their mixtures in Saccharomyces cerevisiae D7. Mutat Res 321:203-11

Bianchi-Santamaria A, Gobbi M, Cembran M, *et al.* 1997. Human lymphocyte micronucleus genotoxicity test with mixtures of phytochemicals in environmental concentrations. Mutat Res 388:27-32

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

Bouakka M, Legendre P, Jouis V, *et al.* 1988. Calcium ionophore and phorbol myristate acetate synergistically inhibited proteoglycan biosynthesis in articular chondrocytes by prostaglandin independent mechanism. Biochem Biophys Res Commun 153:690-8

Boyd CA, Weiler MH, and Porter WP. 1990. Behavioral and neurochemical changes associated with chronic exposure to low-level concentration of pesticide mixtures. J Toxicol Environ Health 30:209-21

Brown MC. 1997. Multisite mucous membrane bleeding due to a possible interaction between Warfarin and 5-Fluorouracil. Pharmacotherapy 17:631-3

Burger BV, Le Roux M, Mackenroth WM, *et al.* 1990. 7-Vinyldecyl acetate, novel inhibitor of pheromonal attraction in the false codling moth, *Cryptophlebia leucotreta*. Tetrahedron Lett 31:5771-2

Calabrese EJ. Multiple Chemical Interactions. 1991. Lewis Publishers, Chelsea, MI

Carter WH Jr, Chinchilli VM, Campbell ED, *et al.* 1984. Confidence interval about the response at the stationary point of a response surface, with an application to preclinical cancer therapy. Biometrics 40:1125-30

Carter WH Jr, Stablein DM, and Wampler GL. 1979. An improved method for analyzing survival data from combination chemotherapy experiments. Cancer Res 39:3446-53

Carter WH Jr and Wampler GL. 1986. Review of the application of response surface methodology in the combination therapy of cancer. Cancer Treat Rep 70:133-40

Carter WH Jr, Gennings C, Stanswalis JG, *et al.* 1988. A statistical approach to the construction and analysis of isobolograms. J Am Coll Toxicol 7:963-73

Cassee FR, Groten JP, van Bladeren PJ, *et al.* 1998. Toxicological evaluation and risk assessment of chemical mixtures. Crit Rev Toxicol 28:73-101

Caudle RM and Williams GM. 1993. The misuse of analysis of variance to detect synergy in combination drug studies. Pain 55:313-7

Chou TC and Talalay P. 1984. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. Adv Enzyme Regul 22:27-55

Christ GJ, Goldfarb J, and Maayani S. 1990. Receptor-mediated mutual-effect amplification elicited by phenylephrine and serotonin in isolated rabbit aorta. J Pharmacol Exp Ther 252:500-6

Cohen SD, Callaghan JE, and Murphy SD. 1972. Investigation of multiple mechanisms for potentiation of malaoxon's anticholinesterase action by triorthotolyl phosphate. Proc Soc Exp Biol Med 141:906-10

Cohen SD and Murphy SD. 1971. Malathion potentiation and inhibition of hydrolysis of various carboxylic esters by triorthotolyl phosphate (TOTP) in mice. Biochem Pharmacol 20:575-87

Crebelli R, Conti L, Fuselli S, *et al.* 1991. Further studies on the comutagenic activity of cigarette smoke condensate. Mutat Res 259(1):29-36

Crews D, Bergeron JM, Bull JJ, *et al.* 1994. Temperature-dependent sex determination in reptiles: proximate mechanisms, ultimate outcomes, and practical applications. Dev Genet 15:297-312

CRARM (Commission on Risk Assessment and Risk Management) 1997. Risk Assessment and Risk Management in Regulatory Decision-Making. U.S. Congress, Washington, DC

Dent LA and Orrock MW. 1997. Warfarin-fluoxetine and diazepam-fluoxetine interaction. Pharmacotherapy 17:170-2

Dorian P, Sellers EM, Reed KL, *et al.* 1983. Amitriptyline and ethanol: pharmacokinetic and pharmacodynamic interaction. Eur J Clin Pharmacol 25:325-31

Dzubow L M. 1986. The interaction between propranolol and epinephrine as observed in patients undergoing Mohs' surgery. J Am Acad Dermatol 15(1):71-5

EDSTAC (Endocrine Disruptor Screening and Testing Advisory Committee). 1998. Final Report. EPA/743/R-93/003. U.S. Environmental Protection Agency, Washington, DC

Erickson C K, Tyler T D, Beck L K, *et al.* 1980. Calcium enhancement of alcohol and drug-induced sleeping time in mice and rats. Pharmacol Biochem Behav 12(5):651-6

Faessel H M, Slocum H K, Jackson R C, *et al.* 1998. Super *in vitro* synergy between inhibitors of dihydrofolate reductase and inhibitors of other folate-requiring enzymes: the critical role of polyglutamylation. Cancer Res 58(14):3036-50

Fed. Reg. (Federal Register) 1998a. 63 Fed Reg 42852, August

Fed. Reg. (Federal Register) 1998b. 63 Fed Reg 71542, December

Fletcher RS, Arion D, Borkow G, *et al.* 1995. Synergistic inhibition of HIV-1 reverse transcriptase DNA polymerase activity and virus replication *in vitro* by combinations of carboxanilide nonnucleoside compounds. Biochemistry 34:10106-12

Garfield JM and Bukusoglu C. 1996. Propofol and ethanol produce additive hypnotic and anesthetic effects in the mouse. Anesth Analg 83:156-61

Gaumont Y, Kisliuk RL, Parsons JC, *et al.* 1992. Quantitation of folic acid enhancement of antifolate synergism. Cancer Res 52:2228-35

Gennings C. 1995. An efficient experimental design for detecting departure from additivity in mixtures of many chemicals. Toxicology 105(2-3):189-97

Gennings C. 1996. Economical designs for detecting and characterizing departure from additivity in mixtures of many chemicals. Food Chem Toxicol 34:1053-8

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

Gessner PK. 1988. A straightforward method for the study of drug interactions: an isobolographic analysis primer. J Am Coll Toxicol 7:987-1012

Gessner P K. 1995. Isobolographic analysis of interactions: an update on applications and utility. Toxicology 105(2-3):161-79

Greco W, Unkelbach HD, Poch G, *et al.* 1992. Consensus on concepts and terminology for combined-action assessment:the Saariselka agreement. Archives of Complex Environmental Studies 4:65-9

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

Greco WR, Faessel H, and Levasseur L. 1996. The search for cytotoxic synergy between anticancer agents: a case of Dorothy and the ruby slippers? [editorial; comment]. J Natl Cancer Inst 88:699-700

Greco WR, Park HS, and Rustum YM. 1990. Application of a new approach for the quantitation of drug synergism to the combination of cis-diamminedichloroplatinum and 1-beta-D-arabinofuranosylcytosine [see comments]. Cancer Res 50:5318-27

Groten JP, Schoen ED, and Feron VJ. 1996. Use of factorial designs in combination toxicity studies. Food Chem Toxicol 34:1083-9

Hass B S, Chen J J, Chou M W, *et al.* 1987. Binary mixtures containing isomers of nitrobenzo[a]pyrene induce greater-than-additive mutational responses in *Salmonella typhimurium*. I. Analysis by the total concentration-proportional mixture model. Mutat Res 190(4):247-52

Henschler D, Bolt HM, Jonker D, *et al.* 1996. Experimental designs and risk assessment combination toxicology: Panel discussion. Food Chem Toxicol 34:1183-5

Hermann M. 1981. Synergistic effects of individual polycyclic aromatic hydrocarbons on the mutagenicity of their mixtures. Mutat Res 90(4):399-409

Hill AV. 1910. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curve. J Physiol 40:4-7

Holzmann S, Kukovetz WR, Braida C, *et al.* 1992. Pharmacological interaction experiments differentiate between glibenclamide-sensitive K+ channels and cyclic GMP as components of vasodilation by nicorandil. Eur J Pharmacol 215:1-7

Howe GE, Gillis R, and Mowbray RC. 1998. Effect of chemical synergy and larval stage on the toxicity of atrazine and alachlor to amphibian larvae. Environ Toxicol Chem 17:519-25

Hu WY, Reiffenstein RJ, and Wong L. 1987. Interaction between flurazepam and ethanol. Alcohol Drug Res 7:107-17

Jackson R C. 1993. Amphibolic drug combinations: the design of selective antimetabolite protocols based upon the kinetic properties of multienzyme systems. Cancer Res. 53(17):3998-4003

Jaworska M, Gorczyca A, Sepiol J, *et al.* 1997. Metal-metal interactions in biological systems. Part V. *Steninernema Carpocapsae* (Steinernematidae) and *Heterorhabditis Bacteriophora* (Heterorhabditidae) entomopathogenic nematodes. Water Air and Soil Pollut 93:213-23

Jonker D, Woutersen R A, and Feron V J. 1996. Toxicity of mixtures of nephrotoxicants with similar or dissimilar mode of action. Food Chem Toxicol 34(11-12):1075-82

Jonker D, Woutersen R A, van Bladeren P J, *et al.* 1993. Subacute (4-wk) oral toxicity of a combination of four nephrotoxicants in rats: comparison with the toxicity of the individual compounds. Food Chem Toxicol 31(2):125-36

Jonker D, Woutersen R A, van Bladeren P J, *et al.* 1990. 4-Week oral toxicity study of a combination of eight chemicals in rats: comparison with the toxicity of the individual compounds. Food Chem Toxicol 28(9):623-31

Kaden D A, Hites R A, and Thilly W G. 1979. Mutagenicity of soot and associated polycyclic aromatic hydrocarbons to *Salmonella typhimurium*. Cancer Res 39(10):4152-9

Kawalek J C and Andrews A W. 1981. Effect of aromatic hydrocarbons on the metabolism of 2-aminoanthracene to mutagenic products in the Ames assay. Carcinogenesis 2(12):1367-9

Kissin I, Brown PT, and Bradley EL Jr. 1990. Sedative and hypnotic midazolam-morphine interactions in rats. Anesth Analg 71(2):137-43

Kissin I, Brown PT, and Bradley EL Jr. 1992. Locomotor activity after recovery from hypnosis: midazolam-morphine versus midazolam. Anesth Analg 75(6):929-31

Kissin I, Brown PT, Bradley EL Jr, *et al.* 1989. Diazepam—morphine hypnotic synergism in rats. Anesthesiology 70(4):689-94

Kissin I, Brown PT, Robinson CA, *et al.* 1991. Acute tolerance to the hypnotic effect of morphine in rats. Anesth Analg 73(5):619-21

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

Kodell RL and JG Pounds. 1991. Assessing the toxicity of mixtures of chemicals. In: D Krewski D, and Franklin C, eds, pp 559-91, Statistics in Toxicology. Gordon and Breach, NY

Kortenkamp A and Altenburger R. 1998. Synergisms with mixtures of xenoestrogens: a reevaluation using the method of isoboles. Sci Total Environ 221(1):59-73

Kreitzer JF and Spann JW. 1973. Tests of pesticidal synergism with young pheasants and Japanese quail. Bull Environ Contam Toxicol 9:250-6

Ku HH. 1966. Notes on the use of propogation of error formulas. J Res Natl Bureau Stand 70:263-73, as cited in Kissin *et al.* (1987)

Kurtz DL, Stewart RB, Zweifel M, *et al.* 1996. Genetic differences in tolerance and sensitization to the sedative/hypnotic effects of alcohol. Pharmacol Biochem Behav 53(3):585-91

Lau CE and Wang J. 1996. Alprazolam, caffeine and their interaction: relating DRL performance to pharmacokinetics. Psychopharmacology 126:115-24

Le AT, Hasson NK, and Lum BL. 1997. Enhancement of warfarin response in a patient receiving etoposide and carboplatin chemotherapy. Ann Pharmacother 31:1006-8

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

Livingston GK and Dethlefsen LA. 1979. Effects of hyperthermia and X irradiation on Sister Chromatid Exchange (SCE) frequency in Chinese hamster ovary (CHO) cells. Radiat Res 77:512-20

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

Lu G, Jun HW, and Suh H. 1997. Percutaneous absorption and disposition studies of methotrexate in rabbits and rats. Biopharm Drug Dispos 18 :409-22

Machado SG and Robinson GA. 1994. A direct, general approach based on isobolograms for assessing the joint action of drugs in pre-clinical experiments. Stat Med 13:2289-309

Marinovich M, Ghilardi F, and Galli CL. 1996. Effect of pesticide mixtures on *in vitro* nervous cells: Comparison with single pesticides. Toxicology 108:201-6

Marinovich M, Guizzetti M, and Galli CL. 1994. Mixtures of benomyl, pirimiphos-methyl, dimethoate, diazinon and azinphos-methyl affect protein synthesis in HL-60 cells differently. Toxicology 94:173-85

Martin SH, Ottea JA, Leonard BR, *et al.* 1997. Effects of selected synergists on insecticide toxicity in tobacco budworm (*Lepidoptera*: Noctuidae) in laboratory and field studies. J Econ Entomol 90:723-31

Martinez-Irujo JJ, Villahermosa ML, Alberdi E, *et al.* 1996. A checkerboard method to evaluate interactions between drugs. Biochem Pharmacol 51:635-44

Mayer MS and Doolittle RE. 1995. Synergism of an insect sex pheromone specialist neuron: implications for component identification and receptor interactions. J Chem Ecol 21:1875-91

McCain WC, Lee R, Johnson MS, *et al.* 1997. Acute oral toxicity study of pyridostigmine bromide, permethrin, and DEET in the laboratory rat. J Toxicol Environ Health 50:113-24

McInnes G T and Brodie M J. 1988. Drug interactions that matter. A critical reappraisal. Drug*s* 36(1):83-110

McLachlan J A. 1997. Retraction: Synergistic activation of estrogen receptor with combinations of environmental chemicals. Science 277:462-3

Merlin JL. 1994. Concepts of synergism and antagonism. Anticancer Res 14:2315-9

Misslin R, Belzung C, and Vogel E. 1988. Interaction of RO 15-4513 and ethanol on the behaviour of mice: antagonistic or additive effects? Psychopharmacology 94:392-6

Monette R, Potvin L, Baines D, *et al.* 1997. Interaction between calcium ions and *Bacillus thuringiensis*-toxin activity against Sf9 cells (*Spodoptera frugiperda*, Lepidoptera). Appl Environ Microbiol 63:440-7

Naguib M and Abdulatif M. 1993. Isobolographic and dose-response analysis of the interaction between pipecuronium and vecuronium. Br J Anaesth 71:556-60

Ogawa S, Hirayama T, Nohara M, *et al.* 1985. The effect of quercetin on the mutagenicity of 2-acetylaminofluorene and benzo[alpha]pyrene in *Salmonella typhimurium* strains. Mutat Res 142(3):103-7

Ogawa S, Hirayama T, Tokuda M, *et al.* 1987. Mutagenicity-enhancing effect of quercetin on the active metabolites of 2-acetylaminofluorene with mammalian metabolic activation systems. Mutat Res 190(4):241-6

Piatti E, Marabini L, and Chiesara E. 1994. Increase of micronucleus frequency in cultured rat hepatocytes treated *in vitro* with benomyl and pirimiphos-methyl separately and in mixture. Mutat Res 324:59-64

Poch G. 1980a. Dose factor of potentiation derived from isoboles. Arzneimittelforschung 30:2195-6

Poch G. 1980b. Quantitation of increase in maximum drug response caused by potentiating agents. Arzneimittelforschung 30:1310-1

Poch G, Dittrich P, and Holzmann S. 1990a. Evaluation of combined effects in dose-response studies by statistical comparison with additive and independent interactions. J Pharmacol Methods 24:311-25

Poch G, Dittrich P, Reiffenstein RJ, *et al.* 1990b. Evaluation of experimental combined toxicity by use of dose-frequency curves: comparison with theoretical additivity as well as independence. Can J Physiol Pharmacol 68:1338-45

Poch G and Holzmann S. 1980. Quantitative estimation of overadditive and underadditive drug effects by means of theoretical, additive dose-response curves. J Pharmacol Methods 4:179-88

Poch G and Londong W. 1985. Simple approach to assess potentiated drug combinations in clinical trials: studies with pirenzepine plus H2-receptor antagonists. Int J Clin Pharmacol Ther Toxicol 23:283-7

Poch G and Londong W. 1993. Modern approach to the evaluation of combined effects of single-dose trials and clinical time-course studies, exemplified by combinations of pirenzepine and H2-receptor antagonists. Int J Clin Pharmacol Ther Toxicol 31:483-8

Poch G and Pancheva S N. 1995. Calculating slope and ED50 of additive dose-response curves, and application of these tabulated parameter values. J Pharmacol Toxicol Methods 33(3):137-45

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

Poch G, Reiffenstein RJ, and Unkelbach HD. 1990. Application of the isobologram technique for the analysis of combined effects with respect to additivity as well as independence. Can J Physiol Pharmacol 68:682-8

Porter WP, Green SM, Debbink NL, *et al.* 1993. Groundwater pesticides: Interactive effects of low concentrations of carbamates aldicarb and methomyl and the triazine metribuzin on thyroxine and somatotropin levels in white rats. J Toxicol Environ Health 40:15-34

Proudfoot JM, Puddey IB, Beilin LJ, *et al.* 1997. Unexpected dose response of copper concentration on lipoprotein oxidation in serum: discovery of a unique peroxidase-like activity of urate/albumin in presence of high copper concentrations. Free Radic Biol Med 23:699-705

Qureshi S, Laganiere S, Caille G, *et al.* 1992. Effect of an acute dose of alcohol on the pharmacokinetics of oral nifedipine in humans. Pharm Res 9:683-6

Rajamoorthi K, Singh BK, Donovan S, *et al.* 1997. Differential metal binding interactions of the imidazolinones revealed by NMR and UV stectroscopy. Pestic Sci 49:17-28

Ramakrishna N and Ramachandran BV. 1978. Malathion A and B esterases of mouse liver—III: *In vivo* effect of parathion and related PNP-containing insecticides on esterase inhibition and potentiation of malathion toxicity. Biochem Pharmacol 27:2049-54

Recker MW and Kier KL. 1997. Potential interaction between clarithromycin and warfarin. Ann Pharmacother 31:996-8

Redai I, Richards KM, England AJ, *et al.* 1995. Interaction of decamethonium with hexamethonium or vecuronium in the rat: an isobolographic analysis. Anesth Analg 81:768-72

Ribas G, Carbonell E, Creus A, *et al.* 1997. Genotoxicity of humic acid in cultured human lymphocytes and its interaction with the herbicides alachlor and maleic hydrazide. Environ Mol Mutagen 29:272-6

Riekkinen P Jr, Schmidt B, Jäkälä P, *et al.* 1997. Chronic nimodipine and acute metrifonate treatment decreases age-related cortical high voltage spindles in rats. Psychopharmacology 129:91-5

Salamone MF, Heddle JA, and Katz M. 1979. The use of the *Salmonella*/microsomal assay to determine mutagenicity in paired chemical mixtures. Can J Genet Cytol 21(1):101-7

Sanders CC, Sanders Jr WE, and Moland ES. 1993. Decimal assay for additivity of drugs permits delineation of synergy and antagonism. Antimicrob Agents Chemother 37:260-4

Scaramellini C, Bennett G, and Leff P. 1997. Analysis of agonist-agonist interactions: the crucial influence of curve shape. J Pharmacol Toxicol Methods 37:167-78

Seed J, Brown RP, Olin SS, *et al.* 1995. Chemical mixtures: current risk assessment methodologies and future direction. Reg Toxicol Pharmacol 22:76-94

Shou M, Grogan J, Mancewicz JA, *et al.* 1994. Activation of CYP3A4: Evidence for simultaneous binding of two substrates in a cytochrome P450 active site Biochemistry 33:6450-5

Siegmund W, Zschiesche M, Bohne M, *et al.* 1992. Pharmacokinetic interactions between nocloprost, acetylsalicylic acid and ethanol. Int J Clin Pharmacol Ther Toxicol 30:539-40

Siller FR, Quintanilla-Vega B, Cebrian ME, *et al.* 1997. Effects of arsenite pretreatment on the acute toxicity of parathion. Toxicology 116:59-65

Somani S M, Schaeffer D J, and Mack J O. 1981. Quantifying the toxic and mutagenic activity of complex mixtures with *Salmonella typhimurium*. J Toxicol Environ Health 7(3-4):643-53

Soto AM, Chung KL, and Sonnenschein C. 1994. The pesticides endosulfan, toxaphene, and dieldrin have estrogenic effects on human estrogen-sensitive cells. Environ Health Perspect 102:380-3

Stinecipher J and Shah J. 1997. Percutaneous permeation of N,N-diethyl-m-toluamide (DEET) from commercial mosquito repellents and the effect of solvent. J Toxicol Environ Health 52:119-35

Suhnel J. 1990. Evaluation of synergism or antagonism for the combined action of antiviral agents. Antiviral Res 13:23-39

Suhnel J. 1996. Zero interaction response surfaces for combined-action assessment. Food Chem Toxicol 34:1151-3

Sumpter JP and Jobling S. 1995. Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ Health Perspect 103(suppl)7:173-8

Takahashi H, Tanaka J, Tsuda S, *et al.* 1987. Contribution of monoaminergic nervous system in potentiation of 2-sec- butylphenyl N-methylcarbamate (BPMC) toxicity by malathion in male mice. Fundam Appl Toxicol 8:415-22

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

Tallarida RJ, Porreca F, and Cowan A. 1989. Statistical analysis of drug-drug and site-site interactions with isobolograms. Life Sci 45(11):947-61

Tallarida RJ and Raffa RB. 1996. Testing for synergism over a range of fixed ratio drug combinations: replacing the isobologram. Life Sci 58:23-8

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

Tennant RW, Margolin BH, Shelby MD, *et al.* 1987. Prediction of chemical carcinogenicity in rodents from *in vitro* genetic toxicity assays. Science 236(4804):933-41

Thompson DM and Moerschbaecher JM. 1982. Phencyclidine in combination with pentobarbital: supra-additive effects on complex operant behavior in pigeons. Pharmacol Biochem Behav 17:353-7

Thornton-Manning JR, Campbell WL, Hass BS, *et al.* 1989. Role of nitroreduction in the synergistic mutational response induced by mixtures of 1- and 3-nitrobenzo[a]pyrene in *Salmonella typhimurium*. Environ Mol Mutagen 13(3):203-10

Tverskoy M, Fleyshman G, Ezry J, *et al.* 1989. Midazolam-morphine sedative interaction in patients. Anesth Analg 68:282-5

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

Unkelbach HD and Wolf T. 1984. Drug combinations - concepts and terminology. Arzneim-Forschung. Drug Research 34:1151-3

USEPA (U.S. Environmental Protection Agency). 1986. Guidelines for the Health Risk Assessment of Chemical Mixtures. Washington, DC

USEPA (U.S. Environmental Protection Agency). 1989a. Risk Assessment Guidance for Superfund. Volume 1. Human Health Evaluation Manual (Part A). Interim final. EPA/540/1-89/002. Office of Emergency and Remedial Response, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1990. Technical Support Document on Risk Assessment of Chemical MixturesReport (dated November 1988). EPA/600/8-90/064. Office of Health and Environmental Assessment, Cincinnati, OH.

USEPA (U.S. Environmental Protection Agency). 1999. Risk Assessment Forum: Guidance for Conducting Health Risk Assessment of Chemical Mixtures. External Review Draft. Washington, DC

Vale C, Pomes A, Rodriguez-Farre E, *et al.* 1997. Allosteric interactions between gamma-aminobutyric acid, benzodiazepine and picrotoxinin binding sites in primary cultures of cerebellar granule cells. Differential effects induced by g - and d-hexachlorocyclohexane. Eur J Pharmacol 319:343-53

van Birgelen AP, Fase KM, Van der Kolk J, *et al.* 1996. Synergistic effect of 2,2',4,4',5,5'-hexachlorobiphenyl and 2,3,7,8- tetrachlorodibenzo-p-dioxin on hepatic porphyrin levels in the rat. Environ Health Perspect 104:550-7

van Birgelen AP, Van der Kolk J, Fase KM, *et al.* 1994a. Toxic potency of 3,3',4,4',5-pentachlorobiphenyl relative to and in combination with 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in a subchronic feeding study in the rat. Toxicol Appl Pharmacol 127:209-21

van Birgelen AP, Van der Kolk J, Fase KM, *et al.* 1994b. Toxic potency of 2,3,3',4,4',5-hexachlorobiphenyl relative to and in combination with 2,3,7,8-tetrachlorodibenzo-p-dioxin in a subchronic feeding study in the rat. Toxicol Appl Pharmacol 126:202-13

van Steveninck AL, Gieschke R, Schoemaker HC, *et al.* 1993. Pharmacodynamic interactions of diazepam and intravenous alcohol at pseudo steady state. Psychopharmacology 110:471-8

Verma SP, Salamone E, and Golden B. 1997. Curcumin and genistein, plant natural products, show synergistic inhibitory effects on the growth of human breast cancer MCF-7 cells induced by estrogenic pesticides. Biochem Biophys Res Commun 233:692-6

Verschoyle RD, Reiner E, Bailey E, *et al.* 1982. Dimethylphosphorothioates. Reaction with malathion and effect on malathion toxicity. Arch Toxicol 49:293-301

Vonier PM, Crain DA, McLachlan JA, *et al.* 1996. Interaction of environmental chemicals with the estrogen and progesterone receptors from the oviduct of the American alligator. Environ Health Perspect 104:1318-22

Wampler GL, Carter WHJr, Campbell ED, *et al.* 1992. Relationships between various uses of antineoplastic drug-interaction terms. Cancer Chemother Pharmacol 31:111-7

Wessinger WD and Balster RL. 1987. Interactions between phencyclidine and central nervous system depressants evaluated in mice and rats. Pharmacol Biochem Behav 27:323-32

Whelan TV.1987. Propranolol, epinephrine, and accelerated hypertension during hemodialysis. *Ann Intern Med* 106:327

Whong WZ, Lu CH, Stewart JD, *et al.* 1989. Genotoxicity and genotoxic enhancing effect of tetrandrine in *Salmonella typhimurium.* Mutat Res 222(3):237-44

Wibbels T, Bull JJ, and Crews D. 1991a. Chronology and morphology of temperature-dependent sex determination. J Exp Zool 260:371-81

Wibbels T, Bull JJ, and Crews D. 1991b. Synergism between temperature and estradiol: a common pathway in turtle sex determination? J Exp Zool 260:130-4

Yoshida D, Matsumoto T, and Okamoto H. 1979. Interaction between amino-alpha-carboline and amino-gamma-carboline on mutagenicity in *Salmonella typhimurium.* Mutat Res 68(2):175-8

Yu CM, Chan JCN, and Sanderson JE. 1997. Chinese herbs and warfarin potentiation by 'Danshen'. J Intern Med 241:337-9

NOTE FROM HERA'S CO-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. HERA's Managing Editor approved the final manuscript, as submitted for publication.

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. <douglas_crawford-brown@unc.edu>.

William R. Greco, Ph.D. Department of Biomathematics, Roswell Park Cancer Institute, Buffalo, NY. <rosgreco@acsu.buffalo.edu>.

Richard Hertzberg, Ph.D. U.S. Environmental Protection Agency, Region IV, 461 Forsyth Street, S.W.Atlanta, GA <hertzberg.richard@epa.gov> .

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 <ass1@cdc.gov>.

Comments on Borgert et al.

William R. Greco

Biomathematics/Biostatistics Resource; Department of Cancer Prevention, Epidemiology & Biostatistics; Roswell Park Cancer Institute, Elm & Carlton Streets; Buffalo, New York 14263

The overall theme of this report, the description and use of a set of five criteria for evaluating chemical interaction studies for mixture risk assessment is timely, appropriate and important. I have just a few comments on several issues that were only touched upon lightly in the report.

1. There is little discussion about what a researcher or reader of a mixture toxicity study should do with a firm conclusion of Loewe synergy, Loewe antagonism, Loewe additivity, Bliss synergy, Bliss antagonism, Bliss independence or a combination of these phenomena. The intended use of something often dictates the criteria for quality of that something. How should the final conclusion of an *in vitro* or *in vivo* chemical interaction study be translated into regulatory policy, especially when results include irregular isobols, such as seen in Figure 3 of the paper? This is a difficult and important question for which I know no clear answer.

2. Many investigators advocate the use of *a priori* knowledge of the mechanism of action of each single agent in a mixture to suggest a "no interaction" reference model for the combination: Loewe additivity for similar agents (same mechanism), and Bliss independence for dissimilar agents (different mechanisms). Although this report does not advocate this practice, neither does it condemn it. The reasons that I am strongly against this practice include: (a) we often do not know the mechanisms of single agents, and it is often these agents that are studied in mixtures; (b) our current knowledge of single agent mechanisms is often wrong; (c) most agents have more than one important action; (d) if the concept of mechanism includes all of the relevant events involving the agent: transport to and from the site(s) of action, metabolism of the agent, binding to the receptor(s), and transduction of the mechanistic signal to other cellular elements, then no two different agents can ever be strictly defined as similar; (e) it would not be clear how to analyze data for a mixture of similar and dissimilar agents; (f) conclusions are not easily comparable between agent mixtures that have been evaluated with fundamentally different data analysis methods; (g) adding an extra subjective step to an analysis of mixture data contributes to confusion.

3. In most of the paper, it is implied that an endpoint of synergy, antagonism, Loewe additivity, or a combination is a yes/no variable, with no indication of the "strength" or magnitude of the interaction. This is more in line with the hypothesis-testing point of view. Should the magnitude of the interaction, in some sense, be reported, and inferences be made that take the magnitude into account? I think so. However, this generates a major problem, since each different statistical approach to interaction analysis will have its own metric for quantifying the degree of interaction. These metrics are only occasionally interconvertible.