Human and Ecological Risk Assessment: Vol. 7, No. 2, pp. 443-457 (2001)

LETTER TO EDITOR

Multiple Lines of Evidence: Hall and Giddings (2000)

To the Editor[1]—We read with interest Hall and Giddings' (2000) article on 'The Need for Multiple Lines of Evidence for Predicting Site-Specific Ecological Effects'. While we vigorously support weight of evidence and multiple lines of evidence approaches, there are errors in fact and in perspective with their article, which we point out in this reply. These errors have potentially serious adverse implications for ecosystem protection.

The stated intent of the Hall and Giddings review is to establish the need for multiple lines of evidence for confirming impacts on aquatic ecosystem biota prior to initiating corrective actions. However, individual lines of evidence should not be minimized. For instance, surface water toxicity testing and toxicity identification evaluation (TIE) results were the basis for identifying chlorpyrifos and/or diazinon as toxicants impairing many California waterways. Hall and Giddings focus on California situations and examples with chlorpyrifos and diazinon. The impression given in the Hall and Giddings article is that such individual lines of evidence are not useful unless combined with other evidence. This is not the case as we discuss below.

Because the phrase 'single species (SS) toxicity tests' is used throughout our discussion and in Hall and Giddings it requires clarification. Many regulatory agencies throughout the U.S. use a suite of tests with several different aquatic species. These are deemed SS tests since exposure chambers contain only a group of one species at a time. Multiple species tests (*e.g.*, micro- and meso-cosms) consist of a few to many species exposed simultaneously.

USE OF REFERENCE PUBLICATIONS

Contrary to what Hall and Giddings imply (p. 688), the literature is replete with studies and reviews that demonstrate that ambient water and sediment toxicity testing results can indeed predict impacts on aquatic biota (*e.g.*, Waller *et al.* 1996; DeWitt *et al.* 1998; de Vlaming and Norberg-King 1999). Hall and Giddings reference Waller *et al.* (1996), but disregard one of the major conclusions: "... it is unmistakable and clear that WET procedures, when used properly and for the intended purpose, are reliable predictors of environmental impact, provided that the duration and/or magnitude of exposure is sufficient to affect resident biota...". Hall and Giddings contend (p. 689) that Versteeg *et al.* (1999) found that SS toxicity

[1] The views in this letter are those of the authors and do not necessarily reflect the policies and positions of the U.S. Environmental Protection Agency or the State Water Resources Control Board.

1080-7039/01/$.50

test results overestimate adverse effects in mesocosms. On the contrary, Versteeg *et al.* concluded, "This analysis suggests that laboratory generated single-species chronic studies can be used to establish concentrations protective of model ecosystem and likely whole ecosystems effects."

Emans *et al.* (1993), according to Hall and Giddings, reported that SS toxicity testing results underestimate NOECs derived in multiple species tests. Actually, Emans *et al.* wrote, "For most comparisons the multiple species NOEC to single species NOEC ratio varied between 0.2 to 5.0." The correlation coefficient between the two variables was 0.935 ($P<0.001$), with the authors concluding that the two NOECs were similar. Hall and Giddings interpret Slooff *et al.* (1986) as concluding that SS toxicity test results are not reliable predictors of impacts on aquatic biota, having a tendency to overprotect. However, Slooff *et al.* wrote, "This information suggests that testing at the ecosystem level does not lead to results that are dramatically different from those obtained with single-species tests" and "Further, there is no solid evidence that predictions of ecosystem effect levels from acute tests are unreliable and so there is no reason to propose expensive and complex tests as additional or alternative research tools for routine hazard assessment."

On pages 679 and 688 Hall and Giddings state, "Microcosm studies have *consistently* demonstrated" that SS toxicity testing results produce lower effect concentrations than mesocosm adverse effect concentrations. The three references (Slooff *et al.* 1986; Emans *et al.* 1993; Versteeg *et al.* 1999) from the above paragraph do not support this contention. De Vlaming and Norberg-King (1999) compared the relationship between effect concentrations in SS toxicity tests and mesocosm experiments; their findings also contradict the Hall and Giddings conclusion. The view that SS toxicity test results underestimate impacts on aquatic biota (*i.e.*, SS effect concentrations higher than mesocosm effect concentrations) was advanced by Crossland *et al.* (1992), as well as Sarakinos and Rasmussen (1998). La Point and Waller (2000), in addition to others (Farris *et al.* 1991; Cook *et al.* 1999), note that test results with *Ceriodaphnia dubia* and *Pimephales promelas* may be under-protective of most freshwater species. Conrad *et al.* (1999) documented that toxicity of the insecticide permethrin in mesocosms could be effectively predicted by laboratory SS toxicity test results. Comparing NOECs determined in SS toxicity tests with those in mesocosms, Girling *et al.* (2000) concluded that they were equivalent or that the SS values were higher. In addition, an inherent assumption of Hall and Giddings is that microcosm results provide a 'valid' prediction of impacts on aquatic biota. This assumption is far from universally accepted (*e.g.*, Crane 1997) and is discussed below (see Inaccurate Definition of False Positive).

Hall and Giddings write, "No single water column or sediment test species was consistently the most sensitive." We concur with this conclusion and agree that every line of information available contributes accuracy to assessments. Varying sensitivity to multiple toxic chemicals is the basis for having a battery of test species and endpoints for toxicity testing. However, the Hall and Giddings outlook appears to be that even abundant and consistent toxicity testing data from a single species are inadequate for management/corrective action.

OMISSIONS

Hall and Giddings overlooked many review articles and publications that examined the relationship between toxicity testing and bioassessment results. The extensive review by de Vlaming and Norberg-King (1999) and other pertinent reviews (*e.g.*, Stewart 1996; Chapman 2000; de Vlaming *et al.* 2000) that identify effective uses of ambient water toxicity testing data were omitted. De Vlaming and Norberg-King examined 77 independent studies in which SS tests were used to assess ambient water or effluent toxicity and in which aquatic biota surveys were gathered for the purpose of exploring the correspondence between the toxicity data and impacts on aquatic biota. According to these authors, a preponderance of evidence demonstrated that SS toxicity testing results are, in a majority of cases, reliable qualitative predictors of impacts on aquatic biota (74% of the studies). The de Vlaming and Norberg-King review also revealed that, when SS toxicity test results were not effective predictors, they most frequently underestimated impacts (21% of the studies) on aquatic biota rather than overestimated them. Many relevant studies and reviews examined by de Vlaming and Norberg-King were excluded from the Hall and Giddings article (more than 75% of the more than 150 references).

Many of the examples presented in the Hall and Giddings article are with the two OP insecticides, chlorpyrifos and diazinon. Several studies (Kersting and van Wijngaarden 1982; Crossland 1984; Stephensen and Kane 1984; Eaton *et al.* 1985; Clark *et al.* 1987; Siefert *et al.* 1989; Brock *et al.* 1992; Leeuwangh *et al.* 1994; van der Hoeven and Gerritsen 1997) on OP insecticides (especially chlorpyrifos) that discuss the effectiveness of SS toxicity testing results in predicting mesocosm adverse effect concentrations or aquatic ecosystem impacts were reviewed by de Vlaming and Norberg-King (1999). All of these studies were omitted by Hall and Giddings.

Hall and Giddings advise that SS toxicity test results (which mostly measure lethality) "cannot be used in isolation to predict ecological effects." This is in contrast to other researchers (*e.g.*, Sibley *et al.* 2000) who propose that tissue enzyme activity in aquatic invertebrates may be a useful predictor of ecosystem population-level responses to OP insecticides. Furthermore, other evidence is available documenting that diazinon concentrations below the acute lethality LC50 for *Ceriodaphnia dubia* have significant effects on the nervous system and behavior of salmonid fishes (Moore and Waring 1996; Scholz *et al.* 2000; Scholz and Collier 2000).

Many of the supporting examples in the Hall and Giddings article come from Hall and colleagues' work in the Chesapeake Bay area. Some of the more critical references overlooked are those of Hartwell and colleagues performed in watersheds of the Chesapeake Bay. Extensive studies (*e.g.*, Hartwell 1997, 1999; Hartwell *et al.* 1997; Hartwell *et al.* 1998; Hartwell *et al.* 2000) conducted by this group demonstrate that fish and benthic community impairment can be predicted with ambient water and sediment, SS toxicity testing results, respectively.

Hall and Giddings appear to embrace bioassessments as the perfect monitoring and assessment tool. While we energetically advocate the use of bioassessments, they are not without limitations. While emphasizing limitations of SS toxicity testing results, Hall and Giddings failed to give equal consideration to limitations of bioassessment results (or for that matter results from any other monitoring and

assessment tools). Those using or intending to use bioassessments should also be familiar with their limitations and uncertainties. Further, since bioassessments seem to be considered 'validation' tools for toxicity-testing results (see de Vlaming and Norberg-King 1999 for discussion of this issue) their strengths and limitations *must* be evaluated. Strengths and limitations/uncertainties associated with bioassessments have been discussed in several reviews (*e.g.*, Barbour *et al.* 1996; Clements and Kiffney 1996; de Vlaming and Norberg-King 1999; LaPoint and Waller 2000).

Notable limitations in field assessments include low resolution/sensitivity (limited ability to discern actual impacts on biota) and inability to discriminate among different causes of perturbations. "Most importantly, inability to establish a direct cause-and-effect relationship between contaminants and selected endpoints greatly limits instream biomonitoring", Clements and Kiffney (1996). Ambient water toxicity testing, in combination with TIEs, is very effective when it comes to identification of cause and geographic source of contamination (*e.g.*, de Vlaming *et al.* 2000). Variance in natural systems also is recognized as a confounding factor hindering reliable bioassessments. Multiple controversies exist regarding performance and interpretation of bioassessments. There is certainly no concurrence on what constitutes an ecologically significant effect in bioassessments.

A major strength of SS toxicity tests is their use as predictive water quality management tools. This predictive paradigm has been established by water quality regulations in the United States. The 'reactive' nature of bioassessment results cannot achieve this forward-looking protection and can only play a role at, or above, the threshold of instream impairment. That is, bioassessments are retrospective rather than prospective. Thus, if bioassessments are required in all situations, regulatory agencies would be transformed from protecting ecosystems to attempting to *restore* damaged systems.

From our perspective, Hall and Giddings underplay variability in bioassessment results (*e.g.*, Barbour *et al.* 1996; LaPoint and Waller 2000), and with this variability comes less ability to distinguish effects. Based on data from Hall's Chesapeake Bay Program, these authors indicate that there is low temporal variability for indices of biological integrity (IBI) from *large numbers* of undisturbed sites. With large numbers of undisturbed sites (in many systems, such sites do not exist), with thoroughly performed field surveys, and with data expressed as IBIs, bioassessments will be expensive. For more equivalence in comparisons of toxicity testing and bioassessment results variability, Hall and Giddings should compare toxicity testing control performance with bioassessment data from undisturbed sites (see following section for further discussion of toxicity test and bioassessment variability).

Finally, costs for thorough and reliable bioassessments through time exceed the costs for most other monitoring and assessment procedures. Whether regulatory agencies, with their limited monitoring budgets, should always be required to shoulder the costs of bioassessments is debatable. That entities, responsible for chemical contamination of aquatic ecosystems will willingly finance *independent* bioassessments is uncertain. While we are not arguing against bioassessments, corrective actions cannot be suspended in the absence of these assessments.

ASSESSMENT OF TOXICITY TEST VARIABILITY

The Hall and Giddings discussion on SS toxicity test variability is incomplete because they (1) inaccurately portray references on toxicity test variability and (2) focus mostly on references published prior to 1990, omitting recent pertinent reviews. Debate has occurred regarding assertions that inherent variability in toxicity test data is too high to be reliable in regulatory programs. However, several publications propose that current WET test methods are sound and that precision is within the range of other monitoring procedures required in regulatory programs (Ausley 1996; Burton *et al.* 1996; DeGraeve *et al.* 1998; USEPA 2000).

Ausley (1996) compared coefficients of variation (CVs) of chemical analyses and aquatic toxicity tests conducted by North Carolina NPDES permittees. Ausley reported that CVs for chemical analyses ranged from 11.8 to 291.7%. CVs for SS toxicity test parameters ranged from 14.8 to 67.6%. He concluded, " *precision of toxicity analyses is within the range of that being reported for commonly analyzed and regulated chemical parameters.*" Ausley highlighted the difficulty in comparing precision estimates for chemical and toxicity test analyses noting that, while chemical precision is often determined well above analytical detection, toxicity testing precision is often based on the minimum detection level.

Denton and Norberg-King (1996) cited several studies that revealed precision of toxicity testing methods compares favorably with chemical analytical procedures. These authors proposed that improvements in test result consistency could be achieved by limiting within-test variability and through controls of upper and lower statistical power (*e.g.*, limits on test minimum significant difference—MSD). Practices to control within-test variability include controlling within-test sensitivity and following well-defined test methods.

DeGraeve *et al.* (1998) presented results from a waste treatment facilities and state regulatory program survey regarding toxicity testing issues. This project team concurred with the conclusions of the Society of Environmental Toxicology and Chemistry (SETAC) Pellston workgroups (Burton *et al.* 1996; Chapman *et al.* 1996) that (1) between-laboratory CVs for toxicity test methods were low, (2) the tests can be routinely completed successfully by well-trained, competent laboratories, but there is a wide-spread need for training of those involved in toxicity testing who lack such expertise and training, and (3) strengthened quality assurance (QA)/quality control (QC) practices could improve performance of analyses.

Warren-Hicks *et al.* (1999) compared within- and between-laboratory results of reference toxicant test variation as measures of reproducibility and comparability, respectively. The authors concluded that some laboratories could consistently reproduce test results, while others could not and inferred that test precision is a factor of *laboratory experience* and not inherent methodological weakness. To address this laboratory expertise and experience issue, they recommended that additional test acceptability criteria, such as upper and lower bounds of MSD, be established and incorporated in the NPDES process. The latter recommendation corroborates recommendations discussed by Denton and Norberg-King (1996); DeGraeve *et al.* (1998); and USEPA (2000).

The U.S. Environmental Protection Agency (USEPA) (2000) recently published a document in which test variability is evaluated and that discusses approaches to

address test method variability. Data analyzed were from 75 laboratories for 23 SS test methods conducted 1988 to 1999. This study documented that for the lethality LC50 in acute toxicity tests, 75% of laboratories had a CV of 19 to 29%, depending on method. For the LC50 survival endpoint in chronic toxicity tests, 75% of laboratories had a CV of 12 to 35%, depending on method. This study documented that for EC25s for growth and reproduction endpoints in chronic toxicity tests, 75% of laboratories had a CV of 14 to 45%, depending on method. The USEPA (2000) concluded that "comparisons of WET method precision for analytes commonly limited in the NPDES permits clearly demonstrate that the variability of the promulgated WET methods is within the range of variability experienced in other analyses." The USEPA (2000) also recommended establishment of test-specific variability limits.

According to Yoder and Rankin (1995), in bioassessments data 'variability is compressed through the use of multimetric mechanisms such as the index of biological integrity (IBI).' IBIs are an aggregate of metrics that may result in loss of information, mask effects, and reduce sensitivity of bioassessments. Because IBIs are aggregates, comparing CVs from these indices to those from SS toxicity test results seems inappropriate (see p. 702 in Hall and Giddings). Nonetheless, IBI CVs at sites designated as impaired can range from 10 to 55% (Yoder and Rankin 1995). Hall and Giddings refer to the low variability in IBIs at reference/unimpaired sites (p. 703). Bioassessment results are less variable at unimpaired than at impaired sites. Thus, if Hall and Giddings intend to make comparisons between IBI and toxicity test (especially LC50) CVs, the comparison should, at the very least, be bioassessment CVs at impaired sites with toxicity test results.

Hall and Giddings (p. 682) suggest that 'Uncertainty in the test endpoint reduces the reliability' of SS toxicity test results. To the contrary, they (p. 684) summarize one of Moore *et al.*'s (2000) findings as 'there were no false positives when the 7-day survival data were used as an endpoint.' We find these points contradictory and, also interesting in that most of the Clean Water Act (CWA) 303(d) impaired waterway listings in California, due to chlorpyrifos and/or diazinon, were based on *C. dubia* lethality tests. Furthermore, most of the ambient water samples from these waterways produced 100% mortality within 24 to 48 hours and contained several toxic units of one or both of the OP insecticides.

Hall and Giddings conclude (p. 684) that reviews of SS toxicity testing results disclose that 'variations in percent survival were greatest at concentrations of intermediate toxicity which is where NPDES toxicity limits are often set; and comparisons of coefficients of variation show that survival is more variable than LC50 values.' It is a well-known toxicological fact that responses are more variable at intermediate levels of toxicity. However, the USEPA (1991) recommends that toxicity limits and compliance evaluation be based on an effect concentration (*e.g.*, NOEC, IC25, LC50). Fulk (1996) stated that the CVs of multiple LC50s or IC25s provide appropriate information regarding the repeatability and reproducibility of WET tests. This point is critical because permits for acute and chronic tests are evaluated on statistical endpoints (*e.g.*, TUc=/00/NOEC, TUa = 100/LC_{50}). The significance of Hall and Gidding's second point to their theme that toxicity testing variability is unacceptably high is unclear. In the discussion above, we observe that CVs for toxicity testing results are equivalent to other monitoring procedures.

When the variability of toxicity testing analyses is viewed in the context of regulatory programs, these methods provide data that are as precise as those from other monitoring procedures. As with other monitoring methods, lack of expertise/experience in performing the analyses and failure to adhere to prescribed QA, or good laboratory practices, will reduce precision. Refinements of test acceptability criteria can improve test power to detect effects (or the lack thereof) and increase the statistical confidence in results (USEPA 2000).

Hall and Giddings fail to make a convincing case that variability in toxicity testing data is unacceptably high or greater than in other types of monitoring results. There is no more uncertainty in toxicity testing results than in other monitoring procedures. We do advocate that toxicity testing facilities (including those in the Chesapeake Bay Program, see below) provide information on control performance, control charts, reference toxicant data, and MSDs.

DEFINITION OF FALSE POSITIVE

Excessive numbers of 'false positives' is one of the major factors used by Hall and Giddings to indicate limitations of SS toxicity testing data. Considerable caution must be exercised when applying the concept of false positive/negative when comparing toxicity test results with bioassessment indices. In the Hall and Giddings use, any *single* ambient water or effluent sample that shows significant toxicity, when bioassessment data suggest a healthy biota (bioassessment measurements are assumed to be 'correct'), is designated as a false positive. This designation is inappropriate (see discussion in de Vlaming and Norberg-King 1999) and would not be construed in this fashion by a regulatory program. This issue is significant because many of Hall and Giddings' conclusions rest on this single toxicity test designation.

Most water samples tested represent a point in time whereas bioassessment indices represent biota-integrated responses over time. Obviously, toxicological responses are a function of magnitude, as well as duration and frequency of exposure. Therefore, if we are to apply the concepts of false positive/negative when comparing toxicity and bioassessment data, multiple water sample test results must be compared to the biota indices. In the California ambient water monitoring programs (and likely elsewhere), a single toxic ambient water sample is not interpreted as proof, or a prediction, of impacts on aquatic biota. Moreover, in the California program ecological relevance of ambient water toxicity testing results is judged on the basis of magnitude, duration, frequency, and geographic extent.

Even in the case where a single water sample indicates significant toxicity and the biota survey suggests no impacts, caution must be exercised in labeling the toxicity test a 'false positive'. For example, this could be a river sample from the 'front' of a pulse (duration) of toxicity that will have impacts on the aquatic biota—the toxicity was 'real', but the critical exposure duration for impacts had not yet been achieved. On the other hand, the toxicity in the sample could have been 'real', but the pulse duration was insufficient to impact the aquatic biota. As indicated above, a concern regarding bioassessments is that they lack sensitivity to discern all adverse impacts. In Hall and Giddings' designations, bioassessments never yield 'false negatives.' Likewise, biota at a given site may no longer have the sensitivity to 'realize' a toxic insult expressed by sensitive SS surrogates.

We draw attention to the fact that Hall and Giddings target particularly 'false positives' with *Ceriodaphnia dubia* (Table 4). Ambient water toxicity testing throughout California (with associated TIEs) using *Ceriodaphnia* has identified chlorpyrifos and/or diazinon as significant contaminants (*e.g.*, de Vlaming *et al.* 2000). This chlorpyrifos- and diazinon-caused toxicity led to many watersheds throughout the state being placed on the CWA 303d list of impaired waterways. Hall and Giddings exclude subsets (results from impaired sites) of data, from several review articles/studies, that do not support their hypothesis (p. 686 and Table 4) and, by doing so, reach misleading conclusions.

Hall and Giddings claim (p. 686) that Barbour *et al.* (1996) reported a 31% false positive rate when comparing toxicity testing and bioassessment data. Based on a single sample (which we argue is insufficient to make a conclusion or connection) Barbour *et al.* actually reported 11% false positives, so Hall and Giddings censored data (results from 65% of the sites studied were excluded). Barbour *et al.* did not identify single sample toxicity test results that did not agree with instream biota measurements as 'false positives'. Barbour *et al.* did not conclude that toxicity testing results are invalid predictors of impacts on aquatic biota or that the results of toxicity testing, on their own, are invalid regulatory tools.

Diamond and Daley (2000), according to Hall and Giddings, reported that WET tests indicated toxicity at 38% of unimpaired sites. To achieve this percentage, Hall and Giddings censored data from more than 50% of the sites studied. The Diamond and Daley designations for toxicity testing results (that fail to 'match' with bioassessment data) do not take into consideration the points we make in paragraphs two and three of this section. Furthermore, it is important to understand that agreement between toxicity testing and bioassessment results in the Diamond and Daley study was based on effluent tests, not ambient water tests. Hall and Giddings do not distinguish between the results of the two. Three significant observations made by Diamond and Daley were ignored by Hall and Giddings. That is, Diamond and Daley reported that (1) agreement between toxicity testing and field survey results *increased* with the inclusion of more than one toxicity test result. Diamond and Daley also acknowledged, as have several others (Waller *et al.* 1996; de Vlaming and Norberg-King 1999; LaPoint and Waller 2000), that (2) ambient water toxicity testing results show greater agreement with bioassessment data than do effluent data. The Diamond and Daley analysis revealed that (3) when toxicity testing and bioassessment results did not agree, the frequency of 'false negatives' and 'false positives' was equivalent. Another pertinent example is the recent work of Maltby *et al.* (2000) demonstrating that SS test results on effluents effectively predict impacts on aquatic biota.

We find Hall and Giddings's proposal that *Ceriodaphnia* tests yield a 'false positive' rate of 57% unfounded. In the California ambient water toxicity testing program essentially all toxic samples are re-tested, treated with piperonyl butoxide (which inhibits the toxicity of metabolically activated OP insecticides such as chlorpyrifos and diazinon), analyzed for the two OP insecticides with enzyme linked immunosorbant assays or by GC/MS, and put into the TIE process. Using these procedures there has been less that 5% false positives (as defined in USEPA 2000). Further, many of these ambient water samples demonstrated many (3-70) toxic units (TUs) of one or both of the OP insecticides (their toxicity is additive).

In the Hall and Giddings article considerable emphasis is placed on the Chesapeake Bay Program work centered in Hall's laboratory. As stated above, we contend that it is highly inappropriate to compare only a single toxicity test result with a given species against bioassessment indices. In Hall *et al.*'s (1998) work on Chesapeake Bay watersheds during 1996, water column samples from each site were taken only once (October 1-9) and sediment collected only once from each site (September 26-27). So there was only one toxicity test for each site with each of the SS tests for the entire year. For both the fish and benthic bioassessments (for calculating IBIs) collections were gathered three times during the summer (July, August, and September) and data were pooled. Essentially the same sampling and testing procedures were used in the Chesapeake Bay Program study during 1997 (Hall *et al.* 2000). Inferences should not be made regarding the relationship between water column or sediment toxicity with IBIs based on one sample per year, per site and per species. Further, sampling for toxicity tests and bioassessments was not concurrent. Although *Eurytemora affinis* (a copepod) and larval fish toxicity tests were conducted, no plankton (*e.g.*, zooplankton) bioassessments were conducted for comparisons.

In Hall *et al.* (1998), water samples from five separate sites resulted in mortality of 89, 65, 64, 50, and 48% and a 94, 84, 82, 81, and 77% reduction of reproduction in *Eurytemora.* Hall *et al.* concluded that none of these samples/sites were toxic. Two factors could contribute to the inability to detect such responses—(1) high/unacceptable variability and/or (2) inappropriate statistical analyses. In any case, if the toxicity testing conducted in the Chesapeake Bay Program continues to show this low level of sensitivity, it is not surprising that relationships between toxicity test results and bioassessment indices will be uncommon. We propose that toxicity test sensitivity (*e.g.*, minimum significant difference—MSD) goals be set for all SS species protocols in the Chesapeake Bay Program. Without adequate sensitivity, the tests will provide an abundance of false negatives (as defined in US EPA 2000).

Table 6 of the Hall and Giddings article reveals a 65 and 75% agreement between ambient toxicity results and fish community data for water column and sediment toxicity results, respectively. In Table 7, water column toxicity tests agreed with the fish index *or* sediment toxicity tests agreed with the benthos index at 13 of 16 (81%) sites. With the one time sampling/testing and the low sensitivity of some of the toxicity tests in Hall's laboratory, these are striking relationships. Given the sampling and toxicity testing limitations, we suggest that Hall's Chesapeake Bay Program data are insufficient for inferences regarding the predictiveness of SS toxicity testing results.

Hall and Giddings (p. 705, also p. 686) state, "A positive result from a single species test suggests the need for additional investigations such (*sic*) repeated testing with the same species." As stated in the first paragraph of this section, we concur that no single toxicity test should be used to determine compliance or to predict impacts to aquatic biota. However, throughout their article, Hall and Giddings use a single toxicity test result for comparisons with bioassessment data. This greatly weakens their proposal that SS testing is not effective means for predicting impacts to aquatic biota.

IMPLICATIONS TO AQUATIC ECOSYSTEM PROTECTION

Hall and Giddings state (p. 705) that results from toxicity test results "*cannot* be used in isolation to predict ecological effects." They advocate that multiple lines of

evidence are essential. These recommendations have major implications for environmental protection as well as current regulatory laws and activities. A continuous theme in the Hall and Giddings article is that adverse impacts on species and populations are acceptable so long as aquatic ecosystem health appears intact. This theme is inconsistent with U.S. federal and state regulatory codes. The federal CWA states that '....it is the national policy that the discharge of toxic pollutants in toxic amounts be prohibited.' To achieve these national goals the USEPA individual states have established enforceable water quality standards, for both individual chemicals and toxicity. Many states and regions *require* that compliance with toxicity water quality standards, as well as with NPDES effluent limitations, be assessed with USEPA methods (these are SS toxicity testing procedures). Significant omissions from the Hall and Giddings article include (1) the relationships of their recommendations to current regulatory codes, especially the CWA, and (2) their visions for correcting these codes.

When Hall and Giddings conclude that toxicity test results alone are inadequate, the implication is that water quality standards and effluent limitations for toxicity should not be enforced without other lines of evidence. Following their line of reasoning, if water quality standards and effluent limitations for toxicity are not enforceable, why should they exist? Adopting the Hall and Giddings recommendation would necessitate major changes in regulatory statutes. That is, water quality standards and effluent limits for toxicity, as well as for individual chemicals, would have to be significantly modified.

Hall and Giddings referenced Karr and Chu (1999), yet neglected to underscore the key premise of their book (pp. 4-9), which is the accelerated and pervasive degradation of aquatic biota in the United States. Contrary to the Hall and Giddings hypothesis, Karr and Chu contend (p. 25) that toxicological studies underestimate contaminant effects in the field. Aquatic species extinctions and precipitous losses of biodiversity are well documented; chemical pollution contributes to these phenomena (Christian 1995). In fact, Ricciardi and Rasmussen (1999) presented evidence that (1) freshwater biota in the U.S. are disappearing five times faster than terrestrial species and three times faster than coastal marine mammals, (2) extinction rates of freshwater animals are accelerating, and (3) North American freshwater biodiversity is being depleted at the same rate as that of tropical rain forests. Richter *et al.* (1997) concluded that the three leading threats to aquatic species are, in order (1) agricultural non-point pollution, (2) alien species, and (3) altered hydrologic regimes. Wilcove *et al.* (2000) propose that the three leading causes of the decline of aquatic biota are, in order (1) habitat degradation/loss, (2) pollution, especially from agricultural origin, and (3) alien species.

Requiring multiple lines of data in all situations of probable aquatic ecosystem degradation would boost the income of some who are involved in environmental monitoring and assessment, as well as provide a reprieve to polluting entities, but it also would result in delayed environmental protection efforts and further aquatic ecosystem deterioration. *Prevention of aquatic ecosystem degradation is (1) preferable, (2) the charge of regulatory agencies, and (3) almost certainly less expensive than attempts at restoration (proactive rather than retroactive actions).* While we enthusiastically support weight of evidence and multiple lines of evidence approaches, we are certain that there have been and will be many cases where one line of monitoring data is

sufficient to initiate corrective actions. It is critical for regulatory agencies to continue utilizing early warning monitoring signals of impacts to aquatic biota and acting upon solid single lines of evidence. As an example, if the magnitude, duration or frequency, and geographic extent of toxicity (including when these data are provided by SS testing) meet necessary criteria, we assert that corrective actions are not only warranted, but essential.

If overprotection of aquatic ecosystems exists or if protective efforts are initiated prematurely based on insufficient data, why are aquatic biota so imperiled? If aquatic ecosystems are resilient, robust, and healthy as Hall and Giddings suggest (p. 688), why are so many California streams and rivers on the CWA 303d of impaired waterways? Because prospective prevention of degradation is not what it should be, many aquatic ecosystem/water quality *restoration* projects are underway in California. As de Vlaming (2000) observed, "The concept of uncertainty has been used permissively to postpone environmental protection actions. We should not continue along this pathway given continuing species extinctions and biodiversity losses." Corrective action should not have to be reactive (damage exists) such that restoration, rather than protection and prevention, is the operational mode.

On pages 689-690, Hall and Giddings identify aquatic ecosystems as 'not fragile' and conclude that aquatic populations are 'resilient' to toxic inputs. These conclusions are inconsistent with information presented above. The implication is that toxic concentrations of chemicals are permissible in aquatic ecosystems because of their 'robustness' and 'resilience'. This line of reasoning verges on proposing that a goal of environmental science is to define and allow an upper limit for toxic chemicals in aquatic ecosystems. That is, aquatic ecosystems become disposal sites and treatment facilities for toxic chemicals. In relation to this line of reasoning de Vlaming (2000) wrote, "Rather than center on attempting to (1) predict ecosystem toxic chemical assimilation capacity or (2) predict the losses that biotic communities can sustain and possibly rebound, we should focus on identification and development of practices that minimize or eliminate entry of toxic chemicals entry into aquatic ecosystems." Further, the arguments are anthropocentric—while advocating for *prevention* in the areas of human health and crop protection, *restoration* is deemed acceptable for aquatic biota.

Hall and Giddings (pp 696-702) emphasize that physical habitat is critical for the health of aquatic biota. We concur with this recognized reality. However, this understanding does not alter the fact that contaminants and other factors impact aquatic biota and must be addressed. Environmental regulatory structure in some states, including California, is such that different agencies have jurisdiction over various components of aquatic ecosystems. Therefore, remediation, when there are multiple stressors, may not be completely synchronized. That remediation efforts have not been initiated for one aquatic ecosystem aggravation is not justification for inaction on other contributors. We understand that there are multiple stressors to aquatic life. Contaminants are not the only stressor, but in many waterways, they are a major or a significant contributor. Furthermore, 'disproving' contaminant effects on aquatic biota with a bioassessment is difficult. An 'optimal' design to dissociate potential contaminant effects from bioassessment-predicted insults would situate sites above and below contaminant input locations. Additionally, the bioassessment sites should be relatively equivalent with regards to habitat and other non-contaminant, potential impacting factors.

Santillo *et al.* (1998) argue that too often a decision to control a substance is deferred until the results of further research become available. In regard to this argument, the Precautionary Principle is being increasingly adopted in Europe. The Precautionary Principle states (according to Eduljee 2000), "Preventative action must be taken when there is reason to believe that harm is likely to be caused, even when there is no conclusive evidence to link cause with effect: if the likely consequences of inaction are high, one should initiate action even if there is scientific uncertainly."

CONCLUSIONS

The article by Hall and Giddings inappropriately downgrades the use of SS toxicity testing results as predictors of impacts on aquatic ecosystem biota, and as a basis for corrective actions. Adopting their recommendations would thwart and delay aquatic ecosystem protection efforts. Chapman (2000) provided a more balanced review of the strengths and limitations of using SS toxicity testing data.

While multiple lines of evidence, advocated by Hall and Giddings, are desirable or necessary for some corrective actions, they must not be required in all situations. We believe that any solid single line of evidence, including toxicity testing/TIE data, indicating significant insults to aquatic populations requires corrective action. Failure to act in such cases would constitute environmental irresponsibility.

Victor de Vlaming,[1] Debra Denton,[2] and Mark Crane[3]

[1]State Water Resources Control Board Division of Water Quality, 1001 I Street, Sacramento, CA 95814; Tel (off): (916)341-5548; Tel (fax): (916)341-5463; e-mail: de*vv@dwq.swrcb.ca.gov; [2]USEPA, Region 9, 75 Hawthorne Street, San Francisco, CA 94105; e-mail: denton.debra@epa.gov; [3]School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0BW; United Kingdom; e-mail: m.crane@rhnc.ac.uk

REFERENCES

Ausley LW. 1996. Effluent toxicity testing variability. In: Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts, pp 157-71. SETAC Press, Pensacola, FL

Barbour MT, Diamond JM, and Yoder CO. 1996. Biological assessment strategies: applications and limitations. In: Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Predictions of Receiving System Impacts, pp 245-270. SETAC Press, Pensacola, FL

Brock TCM, Crum SJH, van Wijgaarden R, *et al.* 1992. Fate and effects of the insecticide Dursban 4E in indoor *Elodea*-dominated and macrophyte-free freshwater model ecosystems: I. Fate and primary effects of the active ingredient, Chlorpyrifos. Arch Environ Contam Toxicol 23:69-84

Burton GA, Raymon A, Ausley LA, *et al.* 1996. Effluent toxicity test variability. In: Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impact, pp 131-156. SETAC Press, Pensacola, FL

Chapman GA, Anderson BS, Bailer AJ, *et al.* 1996. Methods and appropriate endpoints: discussion synopsis. In: Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts, pp 51-82. SETAC Press, Pensacola, FL

Chapman PM 2000. Whole effluent toxicity testing—usefulness, level of protection, and risk assessment. Environ Toxicol Chem 19:3-13

Christian J. 1995. Ecosystem Protection Through the Endangered Species Act. Proceedings: 4th National Conference on Water Quality Criteria and Standards for the 21st Century, pp 4.25-4.32. Arlington, VA

Clark JR, Borthwick PW, Goodman LR, *et al.* 1987. Comparison of laboratory toxicity test results with responses of estuarine animals exposed to fenthion in the field. Environ Toxicol Chem 6:151-60

Clements WH and Kiffney PM. 1996. Validation of whole effluent toxicity tests: integrated studies using field assessments. In: Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Predictions of Receiving System Impacts, pp 229-244. SETAC Press, Pensacola, FL

Conrad A, Fleming R, and Crane M. 1999. Laboratory and field responses of *Chironomus riparius* (Meigen) to a pyrethroid insecticide. Water Res 33:1603-10

Cook RB, Suter GW II, and Swain ER. 1999. Ecological risk assessment in a large river-reservoir. 1. Introduction and background. Environ Toxicol Chem 18:581-8

Crane M.1997. Research needs for predictive multispecies tests in aquatic toxicology. Hydrobiologia 346:149-55

Crossland NO.1984. Fate and biological effects of methyl parathion in outdoor ponds and laboratory aquaria. Ecotox Environ Safety 8:482-95

Crossland NO, Mitchell GC, and Dorn PB. 1992. Use of outdoor artificial streams to determine threshold toxicity concentrations for a petrochemical effluent. Environ Toxicol Chem 11:49-59

DeGraeve GM, Smith G, Clement W, *et al.* 1998. WET Testing Program: Evaluation of practices and implementation. Water Environment Research Foundation. Project 94-HHE-1. Alexandria, VA

Denton DL and Norberg-King TJ. 1996. Whole effluent toxicity statistics: A regulatory perspective. In: Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impact, pp83-102. SETAC Press, Pensacola, FL

DeWitt T, Long E, Swartz, R, *et al.* 1998. Sediment toxicity can indeed predict environmental effects. SETAC News 18(3):30-1

de Vlaming V. 2000. Probabilistic risk assessments—a magic bullet for all situations, they are not. SETAC Globe 1(3):27-8

de Vlaming V and Norberg-King TJ. 1999. A Review of Single species Toxicity Tests: Are the Tests Reliable Predictors of Aquatic Ecosystem Responses? EPA 600/R-97/11. Office of Research and Development, Duluth. July

de Vlaming V, Connor V, DiGiorgio C, *et al.* 2000. Application of whole effluent toxicity test procedures to ambient water quality assessment. Environ Toxicol Chem 19:42-62

Diamond J and Daley C. 2000. What is the relationship between whole effluent toxicity and instream biological condition? Environ Toxicol Chem 19:158-68

Eaton J, Arthur J, Hermanutz R, *et al.* 1985. Biological effects of continuous and intermittent dosing of outdoor experimental streams with chlorpyrifos. In: Bahner RC and Hansen DJ (eds), Aquatic Toxicology and Hazard Assessment: Eighth Symposium. ASTM STP 891, pp 85-118. American Society for Testing and Materials, Philadelphia, PA

Edjuljee GH. 2000. Trends in risk assessment and risk management. Sci Total Environ 249:13-23

Emans HJB, van de Plassche EJ, Canton JH, *et al.* 1993. Validation of some extrapolation methods used for effect assessment. Environ Toxicol Chem 12:2139-54

Farris JL, Cherry DS, and Neves FJ. 1991. Validation of Copper Concentrations in Laboratory Testing for Site-Specific Copper Criteria in the Clinch River Plant. Final Report. American Electric Power Company, Columbus, OH

Fulk FA. 1996. Whole effluent toxicity testing variability: A statistical perspective. In: . Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts, pp 172-9. SETAC Press, Pensacola, FL

Girling AE, Pascoe D, Janssen CR, *et al.* 2000. Development of methods for evaluating toxicity to freshwater ecosystems. Ecotoxicol Environ Safety 45:148-76

Hall LWJr and Giddings JM. 2000. The need for multiple lines of evidence for predicting site-specific ecological effects. Human Ecolog Risk Assess 6:679-710

Hall LWJr, Anderson RD, Alden RW, *et al.* 1998. Ambient Toxicity Testing in Chesapeake Bay—Year 6. Final Report Chesapeake Bay Program Office, Annapolis, MD

Hall LWJr, Anderson RD, Messing A, *et al.* 2000. Ambient Toxicity Testing in Chesapeake Bay—Year 7. Final Report. Chesapeake Bay Program Office, Annapolis, MD

Hartwell SI. 1997. Demonstration of a toxicological risk ranking method to correlate measures of ambient toxicity and fish community diversity. Environ Toxicol Chem 16:361-71

Hartwell SI. 1998. Empirical assessment of an ambient toxicity risk ranking model's ability to differentiate clean and contaminated sites. Environ Toxicol Chem 18:1298-303

Hartwell SI, Alden RW, Wright DA, *et al.* 2000. Correlation of measures of ambient toxicity and fish community diversity in a Chesapeake Bay tributary, Maryland, USA: A biological, chemical, and geochemical assessment. Environ Toxicol Chem 19:1753-63

Hartwell SI, Dawson CE, Durell EQ, *et al.* 1999. Integrated measures of ambient toxicity and fish community diversity in Chesapeake Bay tributaries. Ecotoxicology 7:19-35

Hartwell SI, Dawson CE, Durell EQ, *et al.* 1997. Correlation of measures of ambient toxicity and fish community diversity in Chesapeake Bay, USA, tributaries—Urbanizing watersheds. Environ Toxicol Chem 16:2556-67

Karr JR and Chu EW. 1999. Restoring Life in Running Waters: Better Biological Monitoring, pp 4-25. Inland Press, Covelo, CA.

Kersting K and van Wijngaarden R. 1982. Effects of chlorpyrifos on a microecosystem. Environ. Toxicol Chem 11:365-72

LaPoint TW and Waller WT. 2000. Field assessments in conjunction with whole effluent toxicity testing. Environ Toxicol Chem 19:14-24

Maltby L, Clayton SA, Yu H, *et al.* 2000. Using single species toxicity tests, community-level responses, and toxicity identification evaluations to investigate effluent impacts. Environ Toxicol Chem 19:151-7

Moore A and Waring CP. 1996. Sublethal effects of the pesticide diazinon on olfactory function in mature male Atlantic salmon parr. J Fish Biol 48:758-775

Moore TF, Canton SP, and Grimes M. 2000. Investigating the incidence of Type I errors for chronic whole effluent toxicity testing using *ceriodaphnia dubia.* Environ Toxicol Chem 19:118-22

Ricciardi A and Rasmussen JB. 1999. Extinction rates of North American freshwater fauna. Conservation Biol 13:1081-93

Richter BD, Braun DP, Mendelson MA, *et al.* 1997. Threats to imperiled freshwater fauna. Conservation Biol 11:1081-93

Santillo D, Stringer RL, Johnston PA, *et al.* 1998. The Precautionary Principle: Protecting against failures of scientific method and risk assessment. Mar Pollut Bull 36:939-50

Sarakinos HC and Rasmussen JB. 1998. Use of bioassay-based whole effluent tests to predict benthic community response to a complex industrial effluent. J Aquatic Ecosyst Stress 6:141-57

Scholz NL, Truelove NK, French BL, *et al.* 2000. Diazinon disrupts antipredator and homing behaviors in Chinook salmon (*Oncorhynchus tshawyscha*). Can J Fish Aquat Sci 57:1911-8

Scholz NL and Collier TK. 2000. Endangered salmon and probabilistic risk assessments for organophosphate pesticides. SETAC Globe 1(6):32-3

Sibley PK, Chappel MJ, George TK, *et al.* 2000. Integrating effects of stressors across levels of biological organization: Examples using organophosphorus insecticide mixtures in field-level exposures. J Aquatic Ecosyst Stress Recov 7:117-30

Siefert RE, Lozano SJ, Brazner JC, *et al.* 1989. Littoral enclosures for aquatic field testing of pesticides: Effects of chlorpyrifos on a natural system. Entomolog Soc Amer Misc Publ 75:57-73

Slooff W, van Oers JAM, and de Zwart D. 1986. Margins of uncertainty in ecotoxicological hazard assessment. Environ Toxicol Chem 5:841-52

Stephenson RR and Kane DF. 1984. Persistence and effects of chemicals in small enclosures in ponds. Arch Environ Toxicol 13:313-26

Stewart AJ. 1996. Ambient bioassays for assessing water quality conditions in receiving streams. Ecotoxicology 5:377-93

USEPA (U.S. Environmental Protection Agency).1991. Technical Support Document for Water Quality-based Toxics Control. EPA/505/2-90-001. Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2000. Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System. EPA 833-R-00-003. Office of Wastewater Management, Washington, DC. June

van der Hoeven N and Gerritsen AAM. 1997. Effects of chlorpyrifos on Individuals and populations of *Daphnia pulex* in the laboratory and field. Environ Toxicol Chem 16:2438-47

Versteeg DJ, Belanger SE, and Carr GJ. 1999. Understanding single-species and model ecosystem sensitivity: Data-based comparison. Environ Toxicol Chem 18:1329-45

Waller WT, Ammann LP, Birge WJ, *et al.* 1996. In: Grothe DR, Dickson KL, and Reed-Judkins DK (eds), Whole Effluent Toxicity Testing: An Evaluation of Methods and Predictions of Receiving System Impacts, pp 271-286. SETAC Press, Pensacola, FL

Warren-Hicks W, Parkhurst BR, Moore D, *et al.* 1999. Whole Effluent Toxicity Testing Methods: Accounting for Variance. Water Environment Research Foundation. Project 95-PQL-1. ISBN 1-893664-01-5

Wilcove DS, Rothstein D, Dubow J, *et al.* 2000. Leading threats to biodiversity. In: Stein BA, Kutner LS, and Adams JA (eds), Precious Heritage: The Status of Biodiversity in the United States, pp 239-254. Oxford Univ. Press, Oxford and New York

Yoder CO and Rankin ET. 1995. Biological Criteria Program Development and Implementation in Ohio. In: Davis WS and Simon TP (eds), Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making, pp 109-44. CRC Press, Lewis Publishers, Boca Raton, FL