



Department of Pesticide Regulation



Paul E. Helliker
Director

MEMORANDUM

Gray Davis
Governor
Winston H. Hickox
Secretary, California
Environmental
Protection Agency

TO: Randy Segawa, Senior Environmental Research Scientist **HSM-02031**
Environmental Monitoring Branch

VIA: Joseph Frank, Senior Toxicologist
Worker Health and Safety Branch

FROM: Sally Powell, Senior Environmental Research Scientist [original signed by S. Powell]
Worker Health and Safety Branch
445-4248

DATE: September 11, 2002

SUBJECT: COMMENTS ON PRIORITIZATION SCHEME FOR CANDIDATE TOXIC
AIR CONTAMINANTS

These comments address the prioritization of pesticides for risk evaluation as Toxic Air Contaminants in general, and DPR's draft response to the Scientific Review Panel's (SRP) recommendations for the prioritization scheme (Attachment).

Purpose of prioritization scheme

Two distinct, though not mutually exclusive purposes have been articulated. The first is to rank order pesticides on inhalation risk. A scheme meeting this purpose ranks the pesticides relative to each other, but does not indicate absolute risk. In other words, the ranking does not indicate whether the risk associated with any pesticide is unacceptable or acceptable. The second and more demanding purpose is to predict absolute risk, in other words, to identify the pesticides that pose unacceptable risks.

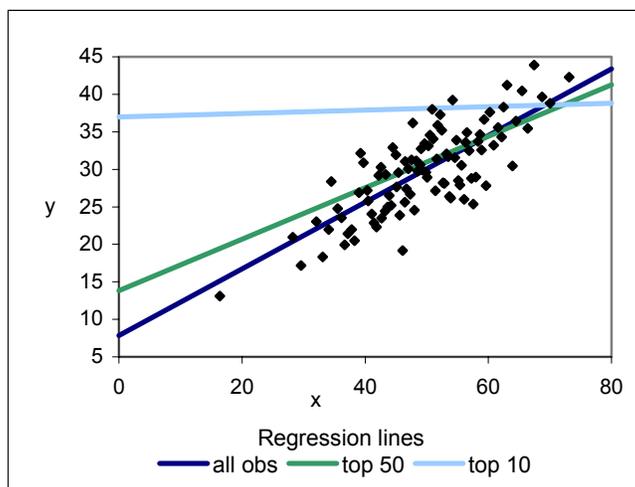
A rating scale can be developed for the first purpose without any external validation criteria, simply by building it from variables known to be correlated with risk. If criterion measurements of risk are available for a representative set of pesticides, the scale can also be validated empirically. In order even to construct a rating scale for the second purpose, however, criterion measurements of risk acceptability are required. Moreover, to validate this scale empirically, criterion measurements on an independent set of pesticides are required. Since risk, in the present case, is a function of pesticide toxicity and ambient air concentrations, criterion measurements of risk must reflect both. MOEs and known cancer incidence are examples of such criteria.

Impossible to validate using currently available data

There are currently no appropriate criterion data for either constructing or validating rating scales. The few pesticides for which ambient air concentrations have been measured were selected precisely because they were judged to present high potential risks. The selections were based on the same variables that are now proposed to form a rating scale. Preselection



introduces the statistical artifact known as range restriction. Range restriction is most easily explained with an illustration. The figure below is a scatter plot of 100 (X, Y) observations.



In the complete dataset, X and Y have $R^2 = 0.57$. If the top 50 percent of the data are selected on variable X ($X > 50$), R^2 drops to 0.21 and the slope of the regression line becomes less steep. If only the top 10 percent of the data on variable X ($X > 63$) are used, R^2 drops to zero and the slope of the regression line becomes almost flat. This means that it would be virtually impossible for a rating scale to perform well in a highly preselected dataset. Worse, one might be misled to add variables to the scale that predicted well among the preselected pesticides, but gave incorrect results when applied to the full range of pesticides.

For the reasons just explained, it is not desirable to attempt to construct a risk rating scale with empirical predictive validity. Nonetheless, a scale can still be constructed to rank order pesticides. If criterion data on pesticides representing the full range of potential risk become available in the future, they can then be used to refine and validate the scale.

Constructing rating scales

Three factors go into the risk ranking: use, volatility and toxicity. The role of use and volatility is to predict air concentrations (exposure). The role of toxicity is to predict risk associated with the air concentrations.

Within each of the four toxicity end-points, use and toxicity interact multiplicatively. For example, if *either* acute toxicity *or* the high-single-application use is low, then there is less acute risk than if both are high. Across toxicity types, the risks interact additively. That is, being carcinogenic does not make a chemical's acute toxicity more serious. Volatility also interacts multiplicatively; if a chemical is not at all volatile, then no matter how high the use or the

toxicity, the risk will not be high. Unlike use, volatility is not specific to toxicity type. This leads to the following schematic risk index:

$$Risk\ Score = Volatility \times \sum_{i=1}^4 (Toxicity_i \times Use_i)$$

where the summation is over the four toxicity end-points.

A rating scale must be constructed for each variable to give it the desired weight in the risk score. The SRP suggested that the four toxicity end-points be weighted equally. To achieve this, the four toxicity-rating scales must have the same number of points and the four use scales must have the same number of points (use and toxicity do not have to be equal). Volatility, if it is not to get as much weight as all the other variables combined, needs to have a proportionately smaller number of points.¹

Piecewise validation of the ranking scale

Although data are not available to empirically validate the risk ranking, it is possible, as you have indicated, to validate some of the individual elements that go into the risk score. As you mentioned, for example, the validity of the LC₅₀ as a surrogate for the acute inhalation NOEL can be investigated. The role of use and volatility in the ranking is to predict air concentration (exposure). The four toxicity end-points are associated with different exposure durations or averaging periods. In principle, therefore, different use information is relevant to each end-point. For acute toxicity, the relevant use information might be the maximum single-day total within any section (or township). For subchronic toxicity, it might be the maximum 1-, 2- or 3-month total in any county. If surrogate use periods and use regions are to be used, they can be validated against the target period/use. To avoid the effects of range restriction, empirical validation should only be done if data are available for pesticides representing the full range of potential risks.

cc: Jay Schreider

¹This isn't the whole story. Variables with larger standard deviations carry greater weight in a score. A variable with a higher mean does not get greater weight, but another variable that is multiplied by it does. So, to give equal weight to the four types of toxicity in the score formula above, all four toxicity rating scales and all four use scales would have to have the same mean and standard deviation. The standard deviation of a rating scale is completely determined by the number of points on the scale (more points \Rightarrow greater standard deviation) *if* the rated items are distributed equally across the scale categories. If the items are mostly clumped in one or two consecutive categories, the standard deviation will be smaller, while if items fall mostly in the two ends of the scale, the standard deviation will be larger than if they are distributed evenly. Thus, in order to weight the variables properly, it is necessary to know how the population of pesticides is distributed on the rating scales. This would require data on a set of pesticides representative of the population. Lacking this, weighting cannot be controlled completely.

ATTACHMENT

DRAFT 5/16/02

To: Tobi Jones
From: Randy Segawa

Subject: RESPONSE TO THE SCIENTIFIC REVIEW PANEL RECOMMENDATIONS FOR THE TOXIC AIR CONTAMINANT PRIORITIZATION

The Department of Pesticide Regulation (DPR) prioritizes pesticides that are candidate toxic air contaminants based on several criteria from three main categories: toxicity, volatility, and use. At its April 26, 2002 meeting, the Scientific Review Panel (SRP) discussed DPR's prioritization and provided several recommendations. The following are the SRP recommendations, DPR's staff responses, and a plan to test various prioritization schemes.

SRP Recommendations and DPR Staff Responses

Recommendation 1: Create two separate prioritizations. Create the first prioritization using the total toxicity score. Create the second prioritization by multiplying the total toxicity score by the volatility score and use score.

Response to Recommendation 1: DPR will create these prioritization schemes and compare to other schemes. See plan for testing prioritization schemes below.

Recommendation 2: Use acute No Observed Effect Levels or LC5 values instead of LC50 values to determine the acute toxicity scores.

Response to Recommendation 2: DPR used LC50 values because these are available for virtually every pesticide. The studies used to determine the LC50 are not designed to produce a NOEL and almost never do. Acute inhalation NOELs are fairly rare and are generally present only for fumigants. Likewise, the acute inhalation lethality studies are designed to lead to a LC50. In virtually all of these studies, there will not be enough animals, data, or observations to derive an LC5.

Recommendation 3: Increase the weighting for cancer and reproductive toxicity to the same as acute toxicity and subchronic/chronic toxicity.

Response to Recommendation 3: DPR will increase the weightings and compare to prioritization other schemes. See proposal for testing prioritization schemes below.

Recommendation 4: Alter the volatility criterion to include the octanol/water partition coefficient, in addition to vapor pressure.

Response to Recommendation 4: Woodrow, et al. (1997, 2001) correlated pesticide flux with various physicochemical factors. They found that flux from pesticides applied to plants is

correlated with vapor pressure. Flux from pesticides applied to water is correlated with vapor pressure x water solubility. Flux from pesticides applied to soil is correlated with vapor pressure x water solubility x soil adsorption. DPR proposes to use these factors for the volatility criterion.

Recommendation 5: Assign prioritization scores to hazardous air pollutants for comparison.

Response to Recommendation 5: DPR will include this information.

Plan for Testing Prioritization Schemes

The April 26 discussion revealed several uncertainties in prioritizing pesticides for the toxic air contaminant program, such as the relative weighting of the different criteria and if some criteria should be additive or multiplicative. DPR's prioritization and the SRP's recommendations are largely based subjective judgment regarding the significance and predictive ability of the different criteria. The following plan outlines a process to objectively compare various prioritization schemes and select the most appropriate one.

The SRP discussion revealed a more focused goal of the prioritization: to rank pesticides according to their anticipated margin of exposure or cancer incidence. If this is the goal, it should be possible to select a group of pesticides for which the margins of exposure and cancer incidence are known and to correlate these values with the scores of several alternative prioritization schemes. The prioritization scheme with the highest correlation to margins of exposure and cancer incidence should be used to determine the prioritization for all candidate toxic air contaminants. DPR proposes to use pesticides that have draft or final toxic air contaminant risk assessments as the test group, including:

- 1,3-dichloropropene
- Azinphos-methyl
- Chlorpyrifos
- Ethyl parathion
- Methyl bromide
- Methyl isothiocyanate
- Methyl parathion
- Molinate
- Tribufos (DEF)

DPR proposes to test the following individual criteria:

- Acute Toxicity (Jan 02 proposed score 1 - 4)
 - LC50
- Subchronic and Chronic Toxicity (Jan 02 proposed score 1 - 4)
 - No Observable Effect Level
- Carcinogenicity (Jan 02 proposed score 0 or 2)
 - Prop 65 listing
- Reproductive Toxicity (Jan 02 proposed score 0 or 2)
 - Prop 65 listing
- Use Amount (Jan 02 proposed score 1 - 8)
 - Statewide annual total
- Volatility (Jan 02 proposed score 1 - 8)
 - Vapor Pressure, adjusted for water or soil applications

To estimate the predictive ability of the individual criteria, DPR will conduct the following data analyses for the nine test pesticides. To estimate the predictive ability of LC50 data, DPR will correlate the LC50 with the acute No Observable Effect Level. To estimate the predictive ability of statewide annual use for subchronic exposure, DPR will correlate the statewide annual use with the 3-month county use for the year and county of ARB monitoring. To estimate the predictive ability of statewide annual use for acute exposure, DPR will correlate statewide annual use with the 95th percentile individual application amount for the year of ARB monitoring. Since the subchronic/chronic NOEL is used to determine the margin of exposure, no additional analyses are necessary for this criterion. Only two of the nine test pesticides are carcinogens or reproductive toxins, insufficient data to evaluate these criteria. Woodrow, et al. demonstrated the predictive ability of vapor pressure for flux, with water solubility and soil adsorption adjustments; no additional analyses are necessary for this criterion.

DPR proposes to test the following overall scoring alternatives:

- Previous Alternatives
 - 1996 score
 - Jan 2002 score
- Toxicity
 - Combined (acute + subchronic/chronic + cancer + reproductive)
 - Exclude reproductive toxicity (acute + subchronic/chronic + cancer)
 - Highest individual toxicity score
- Toxicity and Exposure Relation
 - Toxicity only (volatility and use excluded)
 - Additive (toxicity + volatility + use)
 - Multiplicative (toxicity x volatility x use)
- Toxicity:Volatility:Use Weighting
 - 1:1:1
 - 2:1:1
 - 1:2:1
 - 1:1:2

Other scoring schemes are possible and may be tested, but DPR proposes these schemes for the following reasons. The previous scoring alternatives (1996 score and Jan 2002 score) will be tested for comparison to other schemes. The combined toxicity score was recommended by the SRP. Excluding reproductive toxicity may result in a better correlation because it is captured by the other toxicity criteria. Including a separate score for reproductive toxicity gives it greater weight in comparison to other criteria. The margin of exposure is determined by a single toxic effect. Therefore, using the highest individual toxicity score may result in a better correlation than a combined toxicity score. Using toxicity only (excluding volatility and use) for the overall score was suggested by the SRP. The additive score (toxicity + volatility + use) will be tested for comparison. The multiplicative score (toxicity x volatility x use) was suggested by the SRP. Since the margin of exposure is a multiplicative value, this may result in a better correlation. The different weightings will determine if one factor has greater influence on the margin of exposure than others. The total number of schemes that will be tested is:

$$2 \text{ previous alternatives} + (3 \text{ tox options} \times 3 \text{ relations} \times 4 \text{ weights}) = 38$$

The scoring ranges will be adjusted to ensure the proper weightings are tested. For example, when testing the 1:1:1 additive scheme all criteria should have a maximum score of 4. However, when testing the 1:1:1 multiplicative scheme the toxicity score should have, for example, a maximum score of 9 and volatility and use should have maximum scores of 3.