IDENTIFYING CHEMICAL CANDIDATES FOR SUNSETTING A SCREENING AND SCORING SYSTEM

Barbara S. Glenn, MPH

and

Jeffery A. Foran, Ph.D.

Dept. of Health Care Sciences
The George Washington University
2150 Pennsylvania Ave., NW
Washington, D.C. 20037

9 September 1992

BACKGROUND

Despite over two decades of regulation, many of the problems caused by the discharge of toxic chemicals into natural ecosystems have not been adequately addressed. For example, reports suggest that over 500 chemicals in the Great Lakes basin continue to cause serious threats to health of the Great Lakes ecosystem and its human residents (IJC 1983). One of the reasons for continuation of problems associated with toxic chemicals in the Great Lakes basin and in other ecosystems is that regulatory programs have traditionally been focused on treatment at the point of chemical discharge. Few activities have evolved to address pollution prevention at the chemical or pollutant source. Further, the few regulatory programs designed to control chemical production and use have traditionally been site and chemical specific and usually not coordinated between state, national, or international jurisdictions.

Isolated chemical specific management activities focused on pollution prevention, including chemical phase-out or bans, coordinated internationally, have occurred only rarely such as for CFCs and some pesticides. Where a chemical ban has been proposed (e.g. DDT), it has been based on relatively poorly defined scientific-socio-political values which do not transcend the specific chemical ban. The lessons learned from these activities are of limited usefulness for dealing comprehensively with existing problems or anticipating future problems.

Some new initiatives for comprehensive, coordinated chemical control, based on the concept of pollution prevention, have begun through the Organization for Economic Cooperation and Development (OECD). OECD has recognized that most chemicals are

not confined within political borders and that chemical specific regulatory activities, particularly where they are based on the end-of-pipe or on disposal options, have not prevented widespread environmental pollution. The new initiatives, which include consideration of a process called **Sunsetting** (Whalstrom 1989, Foran 1990), consider exposure reductions and elimination for certain hazardous chemicals. Exposure reductions and elimination may occur through a combination of activities that include phase-out or ban of a chemical, changes in or phase-out and ban of certain manufacturing processes, and changes in or phase-out and ban of certain products.

Although proposed by a member country (Sweden), Sunsetting has not been received enthusiastically by the OECD. At the 14th joint meeting of the OECD Chemicals Group (May 1990), member countries generally opposed the concept of coordinated chemical phase-out or bans and substituted this concept with much softer proposals for international chemical assessment. Yet, support for a Sunset process that includes chemical phase-out and ban remains strong in at least two member countries - Sweden and the Netherlands. However, it is unlikely that a comprehensive program for chemical phase-out and ban will be adopted by the OECD.

A process to Sunset hazardous chemicals has been received somewhat more enthusiastically in the Great Lakes basin. Two countries, the U.S. and Canada, share a common border throughout the basin. These countries also share the wealth provided by the system as well as responsibility for its substantial degradation.

Because of the environmental degradation caused by myriad hazardous chemicals used and discharged in the Great Lakes basin, the U.S./Canadian International Joint Commission recommended in its sixth biennial report (IJC 1992) the use of a Sunsetting process to achieve the virtual elimination of hazardous chemicals. The IJC defined Sunsetting as:

> a comprehensive process to restrict, phase out and eventually ban the manufacture, generation, use, transport, storage, discharge, and disposal of a persistent toxic substance. Sunsetting may require consideration of the manufacturing process and products associated with a chemical's production and use, as well as the chemical itself, and realistic yet finite time frames to achieve the virtual elimination of the persistent toxic substance.

A comprehensive Sunsetting program provides an effective approach to managing existing and new chemicals that are considered particularly hazardous, and encouraging the development and use of safer substitutes (Foran 1990). A Sunsetting program will be most effective where potentially hazardous chemicals, processes, and products are evaluated and classified as candidates for Sunsetting via a uniform set of criteria. Ultimately, management of chemicals classified as Sunset candidates must occur through a set of coordinated regulatory activities, particularly where those chemicals are manufactured and used internationally or where they cross international borders during or after use, such as in the Great Lakes basin.

The Great Lakes Sunsetting Project

The George Washington University, Washington, D.C., in cooperation with Pollution Probe of Ontario, Canada, initiated a project in 1991 to develop the first steps in a Sunsetting process for hazardous chemicals in the Great Lakes basin. A two phase activity for development and implementation of a protocol in the Great Lakes basin is underway as part of this project.

Phase I has been devoted to development of a process (which includes specific, quantitative criteria) to identify, evaluate, and classify chemicals as candidates for Sunsetting. Phase II will be devoted to development, binational adoption, and implementation of a Sunsetting process in the Great Lakes basin. Phase II should result in consolidation of activities to determine the degree and nature of legal authority as well as the scientific basis to ban, phase out or substitute hazardous chemicals.

CRITERIA TO IDENTIFY CHEMICAL CANDIDATES FOR SUNSETTING - METHODS

The development of the Sunset scoring system involved the design of a system prototype, solicitation, and incorporation of comments, and a test of the scoring system using data for 45 substances. The design of the scoring system prototype was based on existing hazard ranking systems as well as by the goals of the Sunsetting concept. We examined several lists of environmental pollutants created by various state, provincial, and federal regulatory bodies, and the criteria used to create those lists. The examination was limited to those lists for which a document

could be obtained explaining the scientific and decision criteria for classification. The following lists were examined;

Great Lakes Water Quality Agreement Annex 1 Lists

Ontario Municipal Industrial Strategy for Abatement Effluent Monitoring Priority Pollutants List (EMPPL)

Ontario Clean Air Program (CAP) Generic Classification List

Michigan Critical Materials Registry

Society of German Chemists (GDCh) Advisory Committee on Existing Chemicals of Environmental Relevance

Netherlands Chemical Substances Act (WMS) Scoring System

U.S. EPA Office of Toxic Substances Chemical Scoring System for Hazard and Exposure Identification.

In addition to these lists, criteria and other mechanisms used by the U.S. Environmental Protection Agency to make decisions on banning or phasing out chemical substances or specific chemical uses under the Toxic Substances Control Act (TSCA) Section 6 were evaluated.

The first draft of the Sunset scoring system and its rationale was sent to a group of individuals who were knowledgeable about the Sunsetting concept as well as our project goals and to individuals experienced in the development of hazard ranking systems. We received many helpful comments and made several modifications to the scoring systems as a result of critical review.

The performance of the Sunset scoring system was evaluated using a list of 45 substances. We compiled this list from lists devloped by agencies in the Great lakes region and

other sources including the U.S./Canadian International Joint Commission Annex I list, the Michigan Critical Materials Register, the Ontario Effluent Monitoring Priority Pollutant List, and the Toxic Release Inventory substances used, stored or released in the Great Lakes basin. A 45-chemical subset of these approximately 800 substances was selected for screening and scoring. The subset included chemicals selected randomly from the list of 800 chemicals as well as the IJC eleven critical pollutants, and the seventeen EPA "33/50" chemicals.

Data sources, including references texts and electronic databases were identified to gather data on the toxicity and exposure potential of each of the 45 substances. The electronic databases included AQUIRE, ENVIROFATE, HAZARDOUS SUBSTANCES DATABASE, RTECS, and the INTEGRATED RISK INFORMATION SYSTEM. Summary documents, including the ATSDR Toxicological Profiles, EPA Environmental Health Assessments, and the EPA Ambient Water Quality Criteria Documents were used to supplement information gathered from electronic databases. All data were entered into an electronic database (Paradox) created for this study for analysis and scoring.

The Screening Process

The screening system presented here has a very specific objective - the identification of chemicals for which a ban or phase-out should be considered. Therefore, while parameters for toxicity and exposure similar to those used by other classification schemes are utilized in this system, the dose

ranges which separate compounds into different categories of concern are weighted more heavily toward the most potent substances.

The hazard of individual chemicals, and their potential to be classified as Sunset candidates, is determined by evaluation of the toxicity of, and the potential for exposure to the chemical. Potential exposure is assessed in categories that address a chemical's propensity to accumulate in the tissues of biota (bioaccumulation), the persistence of the chemical in the environment (as expressed by the chemical's half life or persistence based on fate and transport models), and the amount of chemical that is produced and/or released to the environment. Toxicity is assessed in categories that address adverse impacts (death, impairment of growth and reproduction, etc.) to aquatic biota upon acute (much less than the lifetime of an organism) and chronic (most of the lifetime of an organism) exposure. It includes toxicity to terrestrial and avian, non-mammalian species, and to mammalian species upon both acute and chronic exposure. The category also includes reproductive and developmental toxicity and cancer in mammalian species (including humans).

The screening and scoring system does not evaluate chemicals for genotoxic potential. Evidence of genotoxicity indicates that a substance may be a carcinogen or developmental toxicant, but genotoxicity data are generally given less weight than direct evidence of cancer or developmental toxicity in humans or laboratory animals. Individual **in vitro** tests or even

selected test batteries are also not considered to be sufficiently predictive of the **in vivo** response. Furthermore, we believe that genotoxic effects even observed **in vivo** are not adequate of themselves to list a chemical as a Sunset candidate. However, we suggest that where information on genotoxic potential is available, it should lead to other risk management activities or data collection to further characterize potential adverse effects.

Each toxicity and exposure component includes a set of triggers to allow determination of whether a chemical, relative to other chemicals, can be classified as high concern, moderate concern, or low concern for that component. These triggers, and the rationale for their selection, are presented in the next section.

Once chemicals have been screened and scored using the criteria described above, they will be included on the Sunset candidate list if they meet the following scoring conditions:

Chemicals, exclusive of pesticides, scoring **HIGH** in any toxicity category **and** **HIGH** in release and production,

or;

Chemicals scoring **HIGH** in any acute or chronic toxicity category **and** **HIGH** in persistence **or** bioaccumulation (including pesticides).

These conditions are based on two exposure and toxicity scenarios: The first addresses hazard from chemicals that are non-persistent, highly toxic upon acute or chronic exposure, and are released or could potentially be released in large quantities. These chemicals pose risks to human and nonhuman organisms through large, short-term intentional or accidental releases. An example of such a release is the methyl-isocyanate disaster in Bhopal, India. Hazard addressed under the first condition also results from long term (chronic), low level exposures. Such exposures might occur from continuous releases from a frequently used household consumer product or from a continuous emission to the ambient environment.

We have proposed to exclude pesticides from classification in this first scenario. Pesticides are intentionally released into the environment to elicit short-term (acute) toxicity to target organisms. Such releases, in theory, are assessed by the Federal Insecticide, Fungicide, Rodenticide Act and their effects controlled to levels deemed "acceptable" by the Act. The most recent generation of pesticides have been developed to be highly toxic upon acute exposure but relatively non-persistent and non-bioaccumulative. Many are also released into the environment in large quantities. Because of these characteristics, many if not most pesticides would be classified as Sunset candidates under our first scoring condition. Thus, their consideration through the chemical screening and scoring process was postponed pending discussion of how pesticides should be treated in the proposed process.

The second scenario recognizes hazard from chemicals, including pesticides, that are highly toxic via either acute or chronic exposure and are, at the same time, highly persistent or bioaccumulative. These chemicals do not have to be released in large quantities to impose their toxic effects. Rather, the hazard related to exposure to these chemicals results from their capacity to persist in the environment or accumulate in tissues of nonhuman organisms (bioaccumulation), along with their capacity to elicit effects at low concentrations or doses.

CLASSIFICATION AND SCORING SCHEME FOR BIOACCUMULATION

Score	Definition
HIGH	BAF > 5,000
MODERATE	BAF = 1,000 - 5,000
LOW	BAF < 1,000

DEFINITIONS

BAF - Bioaccumulation factor is defined as a measurement (concentration in tissue/concentration in water) of the ability of a chemical to accumulate in edible fish tissues, typically measured in the field where uptake is from exposure through diet, sediments, and other sources, as well as through the gills and integument.

BCF - Bioconcentration factor is defined as a measurement (concentration in tissue/concentration in water) of the ability of a chemical to concentrate in edible fish tissues as determined in a laboratory study where the uptake is through the gills or other external membranes.

BACKGROUND

Chemicals concentrate from surface water to tissues of aquatic biota via a process called **bioaccumulation**. Two major routes of uptake exist that result in accumulation of toxic chemicals in tissues - uptake through the skin and across gills, and uptake though the food chain. Uptake directly from the water through the skin and gills is called **bioconcentration** and uptake through the food chain is called **biomagnification**. Both play important roles in the accumulation of toxic chemicals in the tissues of aquatic biota.

Chemical-specific bioconcentration factors (BCF) address the magnitude of the uptake of a chemical by aquatic biota from surrounding surface waters. Traditionally, the BCF has represented the fraction of the uptake that occurred through bioconcentration; thus, the term bioconcentration factor. The BCF can be measured either directly or it can be estimated by specific properties of the chemical. Chemical-specific bioaccumulation factors (BAF) address the uptake of a chemical from surrounding surface waters and via the food chain. The BAF is usually measured from field observation although an estimation procedure (presented later in this section) has been developed.

Direct measurement of the BCF requires data that are derived from experiments where non-contaminated biota such as fish are placed

in a controlled aquatic setting with a known concentration of a chemical contaminant. The concentration in fish tissue is measured after a period of time sufficient to allow the chemical to reach equilibrium between the concentration in water and in fish tissue (uptake = depuration). The BCF is calculated directly as:

$$\text{BCF (kg/l)} = \frac{\text{concentration in tissue (mg/kg)}}{\text{concentration in water (mg/L)}}$$

Species and chemical-specific bioconcentration information derived from laboratory or field studies is available for only a few chemicals. Thus, quantitative structure activity relationships are often utilized to predict bioconcentration. The concentration of a chemical in aquatic biota occurs primarily in the fatty tissues and is a function of the lipid-solubility of the chemical. The higher the lipid solubility of a chemical the greater the propensity of the chemical to accumulate in fatty tissues. Generally, lipid-soluble chemicals are also soluble in solvents such as octanol and are relatively insoluble in water. Thus, the bioconcentration factor, or the propensity of a chemical to concentrate in tissues of aquatic biota, may be predicted by the octanol-water partition coefficient (K_{ow}). The higher the K_{ow} the greater the propensity of the chemical to concentrate in tissues of aquatic biota. The statistical relationship between BCF and the K_{ow} can be expressed as:

$$\text{Log}_{10}\ \text{BCF} = 0.79\ \text{Log}_{10}\ K_{ow} - 0.40 - \log_{10}$$

(Veith and Kosian 1983, U.S. EPA 1991a). Other similar equations are also available to describe the relationship between log BCF and log K_{ow} (also called log P). We will rely on the Veith-Kosian equation to predict BCF from K_{ow} for screening purposes. Where another regression equation is more appropriate, that is, where it addresses a compound or class of compounds specifically or more accurately, that regression equation will be used to predict BCF.

The relationship between BCF and K_{ow} is quite effective in predicting BCF for chemicals with K_{ow} between 3 and 6. However, the relationship is less useful for chemicals with K_{ow} above six since the size and structure of these chemicals may interfere with their transport across biological membranes. Further, chemicals with K_{ow} greater than six may partition or adsorb to particles which may inhibit transport across gill membranes or through the integument (LaKind and Rifkin 1990). In these cases, use of the equation above to predict BCF for chemicals with K_{ow} greater than 6 may overestimate actual BCF. However, other factors may result in substantial underestimation of BCF as described below.

Biomagnification may play a considerable role in determining the concentration of a chemical in tissues of aquatic biota. For example, Thomann and Connolly (1984) suggested that more than 99% of the observed concentration of PCB (Log K_{OW} = 6.4 to 6.8) in Lake Michigan lake trout resulted from exposure through the food chain. Use of the octanol-water partition coefficient to predict lake trout tissue concentration underestimated observed concentrations by a factor of four. In this case, consideration of only bioconcentration underestimates total accumulation of a chemical in tissues of aquatic biota.

The bioaccumulation factor (BAF) reflects the uptake and concentration of a chemical in the edible tissues of fish from both the food chain and from surrounding water through the gill and integument. Ideally, the BAF is calculated directly from laboratory or field studies where exposure is to a known concentration of a chemical and where exposure has occurred for a period long enough to ensure equilibrium concentrations in water and fish tissue. Direct measurement of bioaccumulation to calculate a bioaccumulation factor (BAF) is likely the most accurate method to assess the potential for a chemical to accumulate in the tissues of aquatic biota. However, direct measurement requires data that address the ambient water concentration as well as the concentration of the toxicant in prey species and the accumulation of the toxicant from both sources. Because of the difficulty in measuring bioaccumulation directly, the BAF for a chemical is rarely measured from laboratory or field studies. However, one such study (Oliver and Niimi 1988) did examine the bioaccumulation of PCBs and other organic contaminants in salmonids from Lake Ontario. The study resulted in a statistical relationship between BAF and K_{OW} for PCBs and other chlorinated hydrocarbons expressed as:

$$\text{Log BAF} = 1.07 \log K_{OW} - 0.21.$$

This relationship, as it is field derived, includes consideration of both bioconcentration and biomagnification. Thus, where information on the bioaccumulation potential of a compound is available from direct measurement such as for PCBs in salmonids, the information will be utilized in the Sunset screen to score a chemical in this category.

Where information from direct measurements of bioaccumulation are not available, it may be desirable to predict bioaccumulation from QSAR. New techniques allow the BAF to be calculated from the bioconcentration factor (BCF) or from the log of the octanol/water partition coefficient. To calculate a BAF from BCF or log P data, the U.S. EPA (1991a,b) has recommended the use of a food chain multiplier (FCM) to account for bioaccumulation of chemicals in tissues of aquatic biota. For those compounds with

log K_{OW} of greater than approximately 3.5 and less than 6.0, the U.S. EPA (based on the work of Thomann, 1989) suggests that food chain uptake, and thus the BAF, may be predicted through adjustment of the BCF or log K_{OW} with an appropriate food chain multiplier (FCM).

Where a BCF is either determined from laboratory studies or calculated from regression, it is multiplied by the food chain multiplier to obtain a BAF. Food chain multipliers are based on the K_{OW} of a compound and on the trophic level of the animal from which a BAF is calculated. For purposes of the screening portion of this study, we have chosen the most conservative food chain multiplier derived from species at the highest trophic level (usually a piscivorous fish). A BAF calculated from the BCF or the log K_{OW} are determined from BCF such that:

BAF = BCF X FCM

where BCF = Bioconcentration Factor calculated directly or from Log P in the regression equation, and

FCM = Food Chain Multiplier for trophic level 4 derived from the following table (U.S. EPA 1991b).

LOG P	FCM
3.5 - 4.0	1.0
4.1 - 4.4	1.1
4.5	1.2
4.6	1.3
4.7	1.4
4.8	1.6
4.9	2.0
5.0	2.6
5.1	3.2
5.2	4.3
5.3	5.8
5.4	8.0
5.5	11
5.6	16
5.7	23
5.8	33
5.9	47
6.0	67
6.1	75
6.2	84
6.3	92
6.4	98
6.5	100
>6.5	100

Development of a BCF without a food chain multiplier may not provide adequate protection for biota that concentrate toxic chemicals and ultimately for humans or wildlife which consume contaminated biota. Therefore, we rely on the procedure proposed in EPA's Technical Support Document (EPA 1991b) to calculate the bioaccumulation factor (BAF) and we use this factor to score potential Sunset candidate chemicals.

Finally, the relationships between K_{ow}, bioconcentration, and bioaccumulation have been established for compounds and classes of compounds that are not readily metabolized by aquatic organisms. Caution is advised in use of the regression equations to predict BCF or BAF for chemical classes that were not used in the development of the equations or for chemicals that are highly reactive or readily metabolized. Thus, the prediction of BCF and BAF from K_{ow} will be made in the screening phase of this project from the regression equations presented above for all compounds. Further analysis of the bioaccumulation potential of Sunset candidate compounds, where it is predicted from the regression equations, will then be conducted during the second phase evaluation to determine whether use of regression to predict BCF was appropriate.

DATA QUALITY

Studies used for scoring in the bioaccumulation category must be properly conducted, producing statistically and biologically significant results that are adequately reported. Generally, the guidelines provided in EPA's Assessment and Control of Bioconcentratable Contaminants in Surface Waters (1991) and in Rand and Petrocelli (1985) should be followed to calculate bioconcentration factors.

RATIONALE

The triggers for BAF scoring are relatively arbitrary. It is recognized that chemicals that have been detected in Great Lakes biota and that pose hazards to human and non-human organisms generally have BAFs greater than 1000. To this end, a number of scoring procedures have classified the BCF of highest concern to be at or above 1,000. For example, the following represent the highest BCF/BAF triggers, that is the BCF/BAF levels of greatest concern for numerous chemical scoring systems:

Great Lakes Initiative	BAF > 1000		(unpublished)
TSCA Testing Requirements	Log P > 5	(BAF > 9,000)*	(Walker 1990)
Michigan CMR	log P > 5	(BAF > 9,000)*	(Michigan DNR)
Ont. MOE	log P = 4.0	(BAF > 1,000)*	(MOE 1992)
Great Lakes WQA	log P = 4.2	(BAF > 1,000)*	(IJC)
BUA	log P > 3	(BAF > 100)*	(cite)
WMS	log P > 4	(BAF > 1,000)*	(cite)

* BAF calculated from regression equation above (1% lipid) and use of a food chain multiplier of 1 (log P = 3), 1.1 (log P = 4 or 4.2), or 2.6 (log P = 5), from U.S. EPA (1991b).

As the scoring system for Sunset chemicals will be used to determine, in part, chemical candidates for ban or phase-out, a system has been developed whereby chemicals with a BAF of greater than 5,000 would receive the highest score. We have chosen these triggers with consideration of limiting the number of chemicals that are ultimately classified as Sunset candidates (discussed further in the introduction to this section). Further, a high score under this section alone would not be enough to classify a chemical as a Sunset candidate; rather, as described elsewhere, a chemical's toxicity must be considered along with bioaccumulation to classify it as a Sunset candidate.

REFERENCES

LaKind, J. and E. Rifkin. 1990. Current method for setting dioxin limits in water requires reexamination. Environ. Sci. Technol. 24:963-965.

Oliver, B.G. and A.J. Niimi. 1988. Trophodynamic analysis of PCB congeners and other chlorinated hydrocarbons in the Lake Ontario ecosystem. Environ. Sci. Technol. 22:388-397.

Ontario Ministry of the Environment. 1992. Candidate Substances List for Bans or Phase-outs. Hazardous Contaminants Branch and Water Resources Branch, MOE. Rpt. #PIBS 1921E.

Rand, G.M. and S.R. Petrocelli. 1985. Fundamentals of Aquatic Toxicology. Hemisphere Publishing Corp. Washington, D.C. Pages 495-525.

Thomann, R.V. 1989. Bioaccumulation model of organic chemical distribution in aquatic food chains. Environ. Sci. Technol. 23:699-707.

Thomann, R.V. and J.P. Connolly. 1984. Model of PCB in the Lake Michigan lake trout food chain. Environ. Sci. Technol. 18:65-71.

U.S. EPA. 1991a. Assessment and control of bioconcentratable contaminants in surface waters. Draft, U.S. EPA, Office of Water, Washington, D.C.

U.S. EPA. 1991b. Technical Support Document for Water Quality-based Toxics Control. EPA/505/2-90-001.

Veith, G.D. and P. Kosian. 1983. Estimating bioconcentration potential from octanol/water partition coefficients. chapter 15 in : PCBs in the Great Lakes. Mackay, D., R. Patterson, S. Eisenreich, and M. Simmons (eds.). Ann Arbor Science.

Walker, J.D. Chemical fate, bioconcentration, and environmental effects testing: Proposed testing and decision criteria. Toxicity Assessment: An International Journal. 5:103-134.

CLASSIFICATION AND SCORING SCHEME FOR PERSISTENCE

Score	Definition
HIGH	Persistence in critical medium > 56 days
MODERATE	Persistence in critical medium = 7 - 56 days
LOW	Persistence in critical medium < 7 days

RATIONALE

Persistence of a compound in the environment, for example in surface waters, determines in part the potential for the compound to migrate throughout the environment, to establish elevated concentrations in some environmental compartments, and to accumulate in non-human and human biota. Unfortunately, the measurement of persistence is difficult since it can be influenced by many chemical, physical, and biological factors such as temperature, presence or absence of dissolved oxygen, light, humic materials, pH, and degrading microorganisms (Mackay 1991).

Persistence in the critical medium as it is used in this scoring system draws first on the half-life ($t_{1/2}$) of a compound in soil, sediments, air, or in water. Where only half-life data are available, they will be utilized to score a chemical in this category. However, knowledge of the fate and transport of specific chemicals within and between various environmental compartments is useful in determining the persistence of a compound within a compartment. We will utilize the Level III fugacity model of Mackay (1991, 1992) to assess and predict the persistence of chemicals within an environmental compartment (water, air, sediments, etc.).

The choice of the Mackay fugacity model is derived from its simplicity of use, its minimal data requirements, and its ability to predict the period of time that a chemical will persist in an environmental compartment. The model requires data on the molecular weight, water solubility, vapor pressure, and octanol-water partition coefficient (K_{OW}) of a chemical. For purposes of this screen, these data are incorporated into the computer program for a Level III fugacity model and persistence calculated for each environmental compartment. A chemical is then scored using the triggers for persistence as indicated above.

Specific definitions for time scales associated with persistence have not been developed under existing U.S. statutes, regulations, or policy and guidance documents (EPA, 1991). However, the time scale for persistence is described under the Great Lakes Water Quality Agreement (Annex 12) as "any toxic

substance with a half-life in water of greater than eight weeks." We have relied on this description as it is likely a period adequate to result in uptake of bioaccumulative chemicals in fish and other organisms to levels that pose hazards to humans and non-human biota that consume those contaminated organisms. Further, it is an adequate amount of time for a toxicant to elicit an adverse effect based on acute exposures and, for many organisms, an adverse effect based on chronic exposures.

REFERENCES

U.S. EPA. Revised Great Lakes Initiative antidegradation materials. Memo from R.F. Vaughn, Water Standards and Planning Branch, 10 January 1991.

Mackay, D. 1991. Multimedia Environmental Models. Lewis Publishers, Chelsea, MI. 257 pp.

Mackay, D., S. Paterson, and W.Y. Shiu. 1992. Generic models for evaluating the regional fate of chemicals. Unpublished MS.

CLASSIFICATION AND SCORING SCHEME FOR RELEASE AND PRODUCTION VOLUME

OPEN SYSTEMS	SCORE		
	HIGH	MODERATE	LOW
Amt. Released to Envrnmt. (annual)	>500,000 kg	100,000 - 500,000 kg	<100,000 kg
or			
Production Volume (annual)	>1,000,000 kg	100,000 - 1,000,000 kg	<100,000 kg

CLOSED SYSTEMS	SCORE		
	HIGH	MODERATE	LOW
Production or Use Volume (annual)	>1,000,000 kg	100,000 - 1,000,000 kg	<100,000 kg

Chemicals manufactured or used in closed systems can be scored HIGH in this category only if there is significant potential for release (accidental or intentional) via use patterns or via transport.

RATIONALE

Release of chemicals to the environment is the first step in exposure to human and non-human organisms outside of the workplace. Release of chemicals through production or manufacturing processes, use, transport, accidents, disposal, or in consumer products may result in contamination of human and non-human biota and of various environmental compartments. The potential for contamination should be related to the amount of a chemical that is released, the number of release points, and the chemical and physical characteristics of the material. In this scoring category we evaluate the potential for release to the environment.

We are concerned with two types of exposure scenarios to classify chemicals as Sunset candidates. One scenario involves compounds that are highly toxic and persistent or bioaccumulative. These chemicals pose hazards in even very small quantities; thus,

release and production information will not influence their classification in the Sunset screening process. The other scenario involves compounds that possess toxicologic characteristics that would result, when exposure occurs, in substantial hazard to humans or the environment but that are not persistent. Exposure to this class of chemicals is likely a function of the amount of chemical produced and released.

We consider two types of systems from which chemicals may enter the environment - from open systems or those from which chemicals are intentionally released either directly or via products, and from closed systems or those from which chemicals are not intentionally released. Discharges from open systems may occur via a waste stream to air, land, ground water, or to surface waters or via release from a consumer product. Release of chemicals from open systems can, in many cases be quantified through the Toxics Release Inventory or similar databases. However, information on the quantities of some chemicals that are released to the environment may not be readily available, for example chemicals that are not reported on the TRI, chemicals for which use information is considered confidential business information, or chemicals that are released or emitted from consumer products. Where information on the release of chemicals used in open systems is not readily available, we propose to substitute information on the amount of chemical produced for release information to score chemicals in this category. Chemical production may not be directly related to chemical release, therefore, the threshold amounts used for scoring production volume differ from threshold amounts used for scoring release volume.

The U.S. EPA has proposed some general criteria to determine substantial production, environmental release, and exposure (56 FR 32294, 15 July 1991). These criteria would trigger testing requirements for chemicals under Section 4(a)(1)(B) of TSCA. EPA has proposed that substantial production refers to chemicals produced in quantities of 1 million pounds (454,000 kg) and that substantial release refers to chemicals released to the environment in quantities greater than 1 million pounds per year. EPA states that, according to the TRI, 37% of the listed chemicals have releases over 1 million pounds. However, EPA also states that only 11% of all chemicals in commerce (including many not reported on TRI) are produced in quantities that exceed one million pounds and that the percentage of those released in quantities greater than one million pounds will be much smaller.

These production and release quantities are generic thresholds for most chemicals. However, EPA states that additional factors should also be considered in determining whether substantial production and release have occurred. Additional factors include bioaccumulation and persistence which EPA suggests are characteristics of particularly great concern for chemicals released to the environment. The occurrence in human adipose

tissue is also an important additional factor. For persistent, bioaccumulative chemicals, or chemicals that occur in human adipose tissue, EPA suggests that release to the environment is of greater concern than the release of non-persistent, non-bioaccumulative substances. Therefore, release of these substances to the environment may be considered substantial even if they do not meet the 1 million pound threshold.

We have chosen scoring thresholds for chemical release that are similar to the substantial release criterion proposed by EPA to trigger testing under Section 4 of TSCA. We have chosen thresholds for chemical production that are above the EPA threshold for substantial production. We have chosen these high thresholds as they likely represent the release and production volume of a relatively few chemicals. Combined with high toxicity, these chemicals may pose a great hazard to humans and the environment and should be classified as Sunset candidates.

Thresholds for chemicals used in closed systems have been chosen arbitrarily but again to favor selection of a relatively few chemicals as Sunset candidates. Chemicals with moderate or low production and potential for release through accidents receive the lower scores in this category. Chemicals with high production (> 1,000,000 kg) combined with potential for accidental release receive a high score. Potential for release for closed-system chemicals may result from accidents at the point of manufacture or use, or accidental release during transport or storage. For example, a chemical used in large volumes only in closed systems, but transported great distances to many different locations or stored at those locations may be classified as having potential for release and receive a high score in this category while the same chemical used at only one location (no transport or storage) would receive a lower score.

REFERENCES

EPA, 1991. 56 Fed Regist. 32294.

CLASSIFICATION AND SCORING SCHEME FOR ACUTE TOXICITY TO AQUATIC ORGANISMS

Score	Definition
HIGH	LC50 or EC50 < 1.0 ug/L (0.001 mg/L)
MODERATE	LC50 or EC50 = 1.0 - 10.0 ug/L
LOW	LC50 or EC50 > 10.0 ug/L

BACKGROUND

This scoring system evaluates the toxicity to aquatic organisms of acute exposures to chemicals. It is based on the common endpoints - the LC50 or EC50 - for acute exposures. The LC50 or EC50 (concentration that is lethal to, or effectively immobilizes 50% of a test population within a specified time period) is used as a measure of acute toxicity to aquatic organisms and is usually expressed over a time period ranging from 48 to 96 hours depending on the species tested.

DATA QUALITY

Studies used for scoring aquatic toxicity must be properly conducted, producing statistically and biologically significant results that are adequately reported. Generally, the guidelines provided in Rand and Petrocelli (1985) and by ASTM and the U.S. EPA (see complete citations in Rand and Petrocelli) should be followed to determine toxicity of chemicals to aquatic organisms.

RATIONALE

Determining triggers for acute toxicity to aquatic organisms is largely a subjective process. We have chosen the trigger for the highest score as it represents the lower range of toxicity for a large number of chemicals that have been tested.

No single aquatic organism is proposed as the species for which data are to be used to develop LC50 data for scoring purposes. Rather, scores in this category should be based on LC50 data for the most sensitive species tested. For many chemicals, only one or a few species have been tested. We believe that a battery of tests that covers a range of species is the most desirable to determine the score for specific chemicals. However, a full database may not be available for many chemicals and may, in many cases not be necessary to gain insight into the toxicity of a chemical to sensitive aquatic biota. Minimum database

requirements and triggers for scoring are discussed below.

Meyer and Ellersieck (1986) examined the acute toxicity of 410 chemicals to 66 species of freshwater animals conducted in nearly 5,000 tests. They concluded that, of the 66 species tested, the four most commonly tested forms were daphnids, rainbow trout, bluegills and fathead minnows. More important, the LC50 of chemicals for which toxicity data existed for each of these species was within one order of magnitude of the most sensitive species 75% of the time for daphnids, 35% of the time for rainbow trout, 28% of the time for fathead minnows, and 38% of the time for bluegill sunfish. The LC50 for these species was within two orders of magnitude of the most sensitive species 90%, 65%, 58% and 72% of the time for daphnids, rainbow trout, fathead minnows, and bluegill sunfish. Testing of daphnids in combination with one of the three fish species increased the frequency to 85% within one order of magnitude of the most sensitive species and 98 to 100% within two orders of magnitude of the most sensitive species.

Ideally, a database for scoring under this heading should include acute toxicity data from one species of either Daphnia, rainbow trout, fathead minnow, or bluegill sunfish. However, scoring can be based on toxicity data derived from other, less sensitive species with an accompanying increase in the likelihood of underestimating the toxicity of chemicals. Less sensitive species (e.g. carp or catfish) can trigger a high score in this category. Where the highest score is based on a less sensitive species, the concern for a false negative is ameliorated. However, where a moderate or low score is triggered in this category by toxicity data generated by a less sensitive species, concern for a false negative is heightened. In this case, scoring in this category should be tentative and reconsideration of the category should occur when a larger database that includes either Daphnia, rainbow trout, fathead minnow, or bluegill sunfish is available.

Our scoring triggers are based on toxicity distribution data collected by Meyer and Ellersieck (1986). Cutoffs that are based on distribution data are not necessarily any more scientifically justifiable than triggers chosen arbitrarily. However, our goal in this project is to limit the number of chemicals that are included on the Sunset list (as discussed in the Section introduction). Cutoffs based on distribution data allow some determination of the number of chemicals that may be included in the "high" classification; thus, the number of chemicals that may be included on the Sunset candidate list prior to the chemical screening process can be limited by use of triggers that are sufficiently stringent to exclude the majority of toxic chemicals.

Meyer and Ellersieck suggest that toxicity data for the 410 chemicals examined are generally bimodal. The toxicities of

insecticides are primarily in the <100 ug/L category while the toxicities of herbicides, fungicides and industrial chemicals are in the > 1 mg/L category. The lower portion of LC50 or EC50 frequency distribution for daphnids, rainbow trout, fathead minnows and bluegill sunfish ranges from 0.1 to 1.0 ug/L. Therefore, we have chosen LC- or EC50s of less than 1.0 ug/L as representing the most severe acute toxicity level for scoring in this category.

The most stringent trigger used in this category (1.0 ug/L) is substantially more stringent than triggers used in the other scoring procedures. This stringency derives from the analysis of frequency distributions for acute toxicities of several hundred chemicals, as discussed above, and from our desire to limit the number of chemicals that are classified as Sunset candidates (as discussed previously).

Cutoffs are listed below as the LC50s that result in granting a chemical the highest score in that class (signifies the level of greatest concern or highest toxicity).

GLWQA	< 1 mg/L
Ont. MOE	< 0.1 mg/L
Michigan CMR	< 1 mg/L
BUA	< 1 mg/L
WMS	< 1 mg/L
TSCA	< 1 mg/L

REFERENCES

Meyer, F.L. and M.R. Ellersieck. 1986. Manual of acute toxicity: Interpretation and data base for 410 chemicals and 66 species of freshwater animals. U.S. Dept. of the Interior, Resource Publication #160. 506pp.

Rand, G.M. and S.R. Petrocelli. 1985. Fundamentals of Aquatic Toxicology. Hemisphere Publishing Corp. Washington, D.C. Pages 495-525.

CLASSIFICATION AND SCORING SCHEME FOR CHRONIC TOXICITY TO AQUATIC ORGANISMS

Score	Definition
HIGH	NOAEC < 0.1 ug/L (0.0001 mg/L)
MODERATE	NOAEC = 0.1 - 1.0 ug/L
LOW	NOAEC > 1.0 ug/L

BACKGROUND

This scoring system evaluates the toxicity to aquatic organisms of chronic exposures to chemicals. It is based on a common endpoint - the No-Observable-Adverse-Effect-Concentration (NOAEC) for chronic exposures.

The No-Observable-Adverse-Effect-Concentration (NOAEC) is the desired measure of chronic toxicity and is expressed as a concentration to which a population of organisms is exposed for the majority of their lifetimes at, and for all concentrations below which no statistically measurable adverse effect is observed based on comparison of effect levels elicited in control organisms.

DATA QUALITY

Studies used for scoring aquatic toxicity must be properly conducted, producing statistically and biologically significant results that are adequately reported. Generally, the guidelines provided in Rand and Petrocelli (1985) and by ASTM and the U.S. EPA (see complete citations in Rand and Petrocelli) should be followed to determine toxicity of chemicals to aquatic organisms.

RATIONALE

Determining triggers for and chronic toxicity to aquatic organisms, like acute toxicity triggers, is largely a subjective process. We have chosen again the trigger for the highest score as it represents the lower range of toxicity for a large number of chemicals that have been tested.

No single aquatic organism is proposed as the species for which data are to be used to develop NOAEC data for scoring purposes. Rather, scores in this category should be based on NOAEC data for the most sensitive species tested. For many chemicals, only one or a few species have been tested. We believe that a battery of tests that covers a range of species is the most desirable to

determine the score for specific chemicals. However, a full database may not be available for many chemicals and may, in many cases not be necessary to gain insight into the toxicity of a chemical to sensitive aquatic biota. Minimum database requirements and triggers for scoring are discussed below.

Meyer and Ellersieck (1986) did not examine chronic toxicity in their study of 410 chemicals; thus a frequency distribution of response to chronic toxicant exposures is not available. Walker (1990) suggests that a chronic toxicity level, measured by the MATC, of less than 0.1 mg/L (100 ug/L) for aquatic biota used for TSCA testing and decision criteria is based on studies suggesting that the MATC is likely to be at least one order of magnitude lower than a chemical's acute EC- or LC50.

The MATC (maximum acceptable toxicant concentration = geometric mean of LOAEC and NOAEC) is a concentration used in some regulatory programs which recognize that complete protection of aquatic biota is not required. Suter et al. (1987) examined the relationship between the MATC and no effect levels in 176 tests on 93 chemicals with 18 species of freshwater fish. They determined that "MATCs are concentrations that cause substantial effects." For example, they found that the MATC corresponded to mean reductions in parental survival of 20%, mean reductions in fecundity of 42%, mean reductions in hatching of 12%, a mean reduction in larval survival of 19%, and a 20% reduction in weight of fish species examined in the study. Therefore, we reject the use of the MATC as representative of a no-effect level. Rather, we have chosen to assess chronic toxicity via information on chemical-specific NOAECs.

Since frequency response data are not available for chronic toxicity similar to those in Meyer and Ellersieck (1986), we cannot choose NOAEC triggers that represent the lowest range of chronic responses in aquatic organisms. In many cases (for example, see the U.S. EPA Water Quality Criteria documents), the NOAEC will be at least 1/10 of the concentration associated with acute toxicity (EC- or LC50). Therefore, we have arbitrarily set the triggers for the NOAEC at 1/10 of the triggers used in acute toxicity.

The most stringent trigger used in this category (0.1 ug/L) is substantially more stringent than triggers used in most other scoring procedures. Similar to the rationale for triggers in the acute toxicity portion of this scoring category, stringency derives from the analysis of frequency distributions for acute toxicities of several hundred chemicals conducted by Meyer and Ellersieck as well as application of an order of magnitude acute-chronic toxicity adjustment. Further, our desire to limit the number of chemicals that are classified as Sunset candidates (as discussed previously) has influenced our choice of triggers for this category.

Cutoffs are listed below as the NOAEC that results in granting a chemical the highest score in that class (signifies the level of greatest concern or highest toxicity).

GLWQA	< 400 ug/L
Ont. MOE	< 0.2 ug/L
Michigan CMR	< 100 ug/L
BUA	< 100 ug/L
WMS	< 1.0 ug/L
TSCA	< 100 ug/L

REFERENCES

Meyer, F.L. and M.R. Ellersieck. 1986. Manual of acute toxicity: Interpretation and data base for 410 chemicals and 66 species of freshwater animals. U.S. Dept. of the Interior, Resource Publication #160. 506pp.

Rand, G.M. and S.R. Petrocelli. 1985. Fundamentals of Aquatic Toxicology. Hemisphere Publishing Corp. Washington, D.C. Pages 495-525.

Suter, G.W., A.E. Rosen, E. Linder, and D.F. Parkhurst. 1987. Endpoints for responses of fish to chronic toxic exposures. Environ. Toxicol. Chem. 6:793-810.

Walker, J.D. 1990. Chemical fate, bioconcentration, and environmental effects testing: Proposed testing and decision criteria. Toxicity Assessment: An International Journal. 5:103-134.

CLASSIFICATION AND SCORING SCHEME FOR ACUTE TOXICITY TO TERRESTRIAL AND AVIAN, NON-MAMMALIAN SPECIES

Score	Definition
HIGH	LD50 < 1.0 mg/kg
MODERATE	LD50 = 1.0 - 10.0 mg/kg
LOW	LD50 > 10.0 mg/kg

BACKGROUND

This scoring system evaluates the toxicity to terrestrial and avian, non-mammalian species of acute exposures to chemicals. It is based on a common endpoint - the LD50 for - acute exposures.

The LD50 (dose that is lethal to, or effectively immobilizes 50% of a test population within a specified time period) is used as a measure of acute toxicity to terrestrial and avian wildlife and is expressed over a time period ranging from one to several days depending on the species tested and the nature of the test.

DATA QUALITY

Studies used for scoring toxicity to terrestrial and avian non-mammalian species must be properly conducted, producing statistically and biologically significant results that are adequately reported. Generally, the guidelines provided by the U.S. EPA (Subdivision E, 1982) should be followed to determine toxicity of chemicals to terrestrial and avian species.

RATIONALE

Determining triggers for acute toxicity to terrestrial and avian organisms, like triggers for aquatic organisms, is largely a subjective process. We have chosen the highest score to represent the lower range of toxicities for a large number of chemicals that have been tested. We have also chosen the criteria that trigger different scores to be logarithmic (base 10) after Konemann and Visser (1988); that is, a change between each score value reflects a change of one order of magnitude in the effect level.

We do not propose minimum database requirements to determine the score in this category. Rather, toxicity data derived from any non-mammalian terrestrial or avian species, assuming data have been collected from a properly conducted study, may be used to

determine the score in this category. The species most commonly tested to determine the acute toxicity of chemicals to wildlife include the mallard duck, bobwhite quail and other quail species, and the ring-neck pheasant.

Acute Toxicity Cutoffs

Hudson, et al. (1984) examined the acute toxicity of over 200 compounds (191 pesticides and 15 other environmental pollutants including TCDD and lead) to several terrestrial and avian species. Approximately three percent of the 206 compounds tested had LD50s for one or more species lower than 1 mg/kg. (A frequency distribution of toxic responses was not presented for these data.) Schafer et al. (1983) examined the oral toxicity of 998 chemicals (mostly pesticides) to wild and domestic birds. Less than one percent of the pesticides had LD50s for redwing blackbirds, starlings, or quail that were lower than 1 mg/kg. Finally, Konemann and Visser (1988) suggested that the majority of subacute effect levels for rats is expected to lie between 1 and 1,000 mg/kg. Therefore, we have chosen 1 mg/kg as the most stringent trigger for this acute toxicity category.

The most stringent triggers used in this category for acute (1.0 mg/kg) toxicity are more stringent than triggers used in the scoring procedures of the Great Lakes Water Quality Agreement, Michigan's CMR, the BUA, the WMS, and in TSCA. The most stringent trigger for acute toxicity is similar to the trigger used by the Ontario MOE. Cutoffs listed below for acute toxicity are the LD50 that result in granting a chemical the highest score in that class (signifies the level of greatest concern or highest toxicity).

	ACUTE TOXICITY
GLWQA	< 50 mg/kg
Ont. MOE	< 1.0 mg/kg/day (subchronic)
Michigan CMR	< 5 mg/kg
BUA	< 25 mg/kg
WMS	< 25 mg/kg
TSCA	< 50 mg/kg

REFERENCES

Hudson, R.H., R.K. Tucker, and M.A. Haegele. 1984. Handbook of toxicity of pesticides to wildlife. U.S. Dept. of Interior, Resource Publication #153. 90pp.

Konemann, H. and R. Visser. 1988. Selection of chemicals with high hazard potential: Part 1: WMS-Scoring System. Chemosphere 17:1905-1919

Schafer, W.E., W.A. Bowles, and J. Hurlbut. 1983. Arch. Environ. Contam. Toxicol. 12:355-382.

U.S. EPA. 1982. Pesticides Assessment Guidelines. Subdivision E. EPA-540/9-82-024.

CLASSIFICATION AND SCORING SCHEME FOR CHRONIC TOXICITY TO TERRESTRIAL AND AVIAN, NON-MAMMALIAN SPECIES

Score	Definition
HIGH	Concentrations of a compound in tissues of naturally occurring terrestrial or avian species that have been determined to cause death or result in impairment of reproduction, growth, behavior, or other features through properly conducted field studies. or NOAEL < 0.1 mg/kg
MODERATE	NOAEL = 0.1 - 1.0 mg/kg
LOW	NOAEL > 1.0 mg/kg

BACKGROUND

This scoring system evaluates the toxicity to terrestrial and avian, non-mammalian species chronic exposures to chemicals. It is based on a common endpoint - the No-Observed-Adverse-Effect-Level (NOAEL) (or No-Observed-Adverse-Effect-Concentration, NOAEC) - for chronic exposures. The NOAEL is used in some testing protocols to address acute as well as chronic exposures to toxicants. However, in this screening protocol we do not utilize the NOAEL as a measure of acute toxicity.

The No-Observed-Adverse-Effect-Level is the desired measure of chronic toxicity and is expressed as a concentration to which a population of organisms is exposed for the majority of their lifetimes at, and for all concentrations below which no statistically measurable adverse effect is observed based on comparison of effect levels elicited in control organisms.

DATA QUALITY

Studies used for scoring toxicity to terrestrial and avian non-mammalian species must be properly conducted, producing statistically and biologically significant results that are adequately reported. Generally, the guidelines provided by the U.S. EPA (Subdivision E, 1982) should be followed to determine toxicity of chemicals to terrestrial and avian species.

RATIONALE

Determining triggers for chronic toxicity to terrestrial and avian organisms, like triggers for acute toxicity, is a subjective process. We have used a technique to choose triggers in this category based on the technique utilized for acute toxicity.

We do not propose minimum database requirements to determine the score in this category. Rather, toxicity data derived from any non-mammalian terrestrial or avian species, assuming data have been collected from a properly conducted study, may be used to determine the score in this category. The species most commonly tested to determine the acute toxicity of chemicals to wildlife include the mallard duck, bobwhite quail and other quail species, and the ring-neck pheasant.

Chronic Toxicity Cutoffs

Although Hudson et al. (1984) examined the acute toxicity of over 200 compounds and Schafer et al. (1983) examined over 990 compounds, chronic toxicity data were not presented in either analysis. Therefore, we have chosen triggers for chronic toxicity at levels 1/10 of triggers for acute toxicity.

The most stringent trigger used in this category for chronic (0.1 mg/kg) toxicity is more stringent than triggers used in the scoring procedures of the Ontario Ministry of the Environment, Michigan CMR, the WMS, and in TSCA. Cutoffs listed below for chronic toxicity are the NOAEL that results in granting a chemical the highest score in that class (signifies the level of greatest concern or highest toxicity).

	CHRONIC TOXICITY
GLWQA	-
Ont. MOE	< 0.5 mg/kg/day
Michigan CMR	< 5.0 mg/kg
BUA	-
WMS	< 0.5 mg/kg
TSCA	< 5.0 mg/kg

REFERENCES

Hudson, R.H., R.K. Tucker, and M.A. Haegele. 1984. Handbook of toxicity of pesticides to wildlife. U.S. Dept. of Interior, Resource Publication #153. 90pp.

Konemann, H. and R. Visser. 1988. Selection of chemicals with high hazard potential: Part 1: WMS-Scoring System. Chemosphere 17:1905-1919

Schafer, W.E., W.A. Bowles, and J. Hurlbut. 1983. Arch. Environ. Contam. Toxicol. 12:355-382.

U.S. EPA. 1982. Pesticides Assessment Guidelines. Subdivision E. EPA-540/9-82-024.

CLASSIFICATION AND SCORING SCHEME FOR ACUTE LETHAL MAMMALIAN TOXICITY

Score	Definition
HIGH	LD50 < 1 mg/kg
MODERATE	LD50 1 - 10 mg/kg
LOW	LD50 > 10 mg/kg

PARAMETER SELECTION

Scoring chemicals on the basis of acute lethality (LD50) provides information concerning their relative hazards presented in the event of an accidental or large-scale release to the environment.

DATA QUALITY

The studies used for scoring acute lethality in mammals must be properly conducted, producing statistically and biologically significant results that are adequately reported.

The National Research Council in 1984 estimated that 59% of pesticides and inert ingredients had been evaluated for acute toxicity (NRC/NAS, 1984). The chemical category most thoroughly evaluated for acute toxicity was drugs and excipients in drug formulations (75%). Acute toxicity data were predicted to be available for 42% and 39% of food additives and cosmetic ingredients. Data were predicted to be available for 20% of chemicals in commerce produced in quantities over 1 million pounds per year.

Data on acute toxicity represent the largest body of toxicity information available. While LD50 data may have limited use in predicting chronic effects in exposed humans, they do indicate relative toxicity and local effects in the event of an accidental exposure or intentional release.

MINIMUM DATA SET

No minimum number of species is required for this parameter. If data are available for more than one species, the lowest LD50 will be selected.

RATIONALE

The rationale for selecting the dose ranges for this parameter are the same as those described for the LD50 in the scoring

scheme for toxicity to non mammalian species. The dose ranges increase on a logarithmic scale as proposed by Konemann and Visser (1988).

One approach to the selection of dose levels for classification purposes has been to identify the highest LD50 of significance or concern which would place a chemical into the lowest concern category. The rest of the dose ranges automatically follow according to a predetermined relationship, such as decreasing by an order of magnitude. The CERCLA ranking process was designed in this manner. In contrast, the Sunset scoring system focuses on the other end of the dose scale, the category of highest concern.

The toxicity rating system for pesticide labelling in the United States is a conventional hazard ranking system and includes four categories based on the LD50 (Code of Federal Regulations, 1985). The highest toxicity category contains pesticides with an LD50 of 50 mg/kg or less. The second highest category includes pesticides with LD50s ranging from 50 to 500 mg/kg. The next two lower toxicity categories are dependent on LD50 values that increase in a similar manner, by an order of magnitude between categories. The scoring system described here has an additional category involving compounds with an LD50 one order of magnitude lower than the highest pesticide toxicity category. This category was included to distinguish extremely toxic substances.

The most stringent cutoff used in this parameter for acute toxicity, 1 mg/kg, is the same as the "extremely toxic" dose level presented by Loomis (1978). It is lower than those used in the scoring procedures of the Great Lakes Water Quality Agreement, Michigan's CMR, the BUA, the WMS, and in TSCA. It is twice as high as that used by the Ontario EMPPL.

	LD50
GLWQA	< 50 mg/kg
EMPPL	< 50 mg/kg
Ontario Ban List	< 0.5 mg/kg
Michigan CMR	< 5 mg/kg
BUA	< 25 mg/kg
WMS	< 25 mg/kg
TSCA	< 50 mg/kg
CERCLA	< 0.1 mg/kg

REFERENCES

Code of Federal Regulations. 1985. Title 40, Part 162.10. Office of the Federal Register, Washington, D.C.

Konemann, H. and R. Visser. 1988. Selection of chemicals with high hazard potential: Part 1: WMS-Scoring System. Chemosphere 17:1905-1919.

Loomis, T.A. 1978. Essentials of Toxicology, 3rd Edition. Lea and Farber, Philadelphia. p. 18.

National Research Council/National Academy of Sciences. 1984. Toxicity Testing: Strategies to Determine Needs and Priorities. National Academy Press. Washington, D.C. 382 pp.

CLASSIFICATION AND SCORING SCHEME FOR SYSTEMIC TOXICITY TO MAMMALS

Score (A)	Severity
3	High - Lethal or life threatening effects; irreversible histopathology and/or permanent organ dysfunction.
2	Moderate - Changes in structure or function or biochemical changes related to adverse effects or altered function; reversible or irreversible.
1	Low - Mild, transient effects.

Score (B)	Effective Dose (mg/kg/day)
4	< 1.0
3	> 1 - 10
2	> 10 - 100
1	> 100

Final Score	(A x B)*
HIGH	10 - 12
MODERATE	5 - 9
LOW	1 - 4

*Add 1 point if based on > 2 species; maximum score remains at 12.

PARAMETER SELECTION

The systemic toxicity parameter is intended to evaluate the potential of a chemical to cause a wide variety of adverse effects in humans. All health effects excluding carcinogenicity, reproductive and developmental toxicity, and mutagenicity are included in this parameter. This parameter is included because

the ability of environmental pollutants to exert long-term detrimental effects on target organs, especially the immune and nervous systems is being met with growing public concern (OTA, 1991; OTA, 1990).

DATA QUALITY

The studies used for scoring systemic toxicity must be properly conducted, producing statistically and biologically significant results that are adequately reported. Studies reporting effects resulting from acute nonlethal, subchronic and chronic exposures are included.

The concept that effects in animals, when properly qualified, are applicable to humans is a fundamental principle of toxicology (Doull, 1987). Moreover, it is a principle critical to the goal of public health, that of disease prevention. Therefore, while epidemiological evidence is most appropriate for predicting risk to humans exposed to an environmental pollutant, data generated from animal studies are equally appropriate when information on effects in humans is not available.

The following table presents the quantity of health effects data obtained from subchronic and chronic laboratory tests for various chemical classes estimated by the National Academy of Sciences in 1984 (NAS, 1984). As the table shows, more data are available for subchronic exposures compared to chronic exposures. In addition, more data have been collected on drugs and excipients in drug formulations and pesticides and inert ingredients of pesticide formulations. Relatively few data have been collected on chemicals in commerce.

	Percent With Laboratory Health Effects Data	
Category	**Subchronic**	**Chronic**
Pesticides and inert ingredients of pesticides formulations	51	23
Cosmetic ingredients	29	16
Drugs and excipients in drug formulations	62	39
Food additives	34	13

Category	Percent With Laboratory Health Effects Data: Subchronic	Chronic
Chemicals in commerce		
> 1 million lb/yr	10	4
< 1 million lb/yr	8	3
unknown or inaccessible	7	3

MINIMUM DATA SET

The scoring system for systemic toxicity is designed to give greater weight to effects observed in multiple species. However, very severe effects observed in one species at a low dose (less than 1 mg/kg/day) also are given the highest score. Studies evaluating life-time toxic exposures are preferred, however studies using shorter exposure regimens also are used to score this parameter. The effective dose for systemic effects based on studies of duration less than 90 days are divided by a factor of ten.

RATIONALE

The scoring system for systemic toxicity draws heavily on the Michigan CMR, TSCA, and CERCLA scoring systems for this parameter. These scoring systems evaluate the severity of effect in combination with the dose level required to achieve the adverse effect. The use of information on both severity and potency provides a more detailed picture of the toxic potential of a chemical.

Systemic toxicity involves adverse effects on a wide variety of organ systems including the liver, kidney, respiratory system, cardiovascular system, gastrointestinal system, endocrine system, nervous system, and immune system. The types of physiological effects and the degree of severity caused by toxic exposures also are variable. An evaluation of severity must include the integration of information on organ system disability and organismal disability (DeRosa et al., 1989). Organismal disability involves a continuum of conditions ranging from health to disturbed function to disease and then death. Organ system impairment can be judged to be adaptive, compensatory, or adverse. The link between organ system impairment and organismal disability is not always clear. The complexity of the analysis required for systemic toxicity requires that the categorization of severity of effect must be general and depend to a large degree on professional judgment.

The MCMR and CERCLA support documents present lists of types of effects ordered by increasing severity. Classification of substances by severity in this scoring system draws upon these schemes and the following rationale. Highly severe effects are life-threatening, resulting in death or permanent dysfunction, may significantly decrease quality of life, or decrease the capacity of an organ system. Generally, an irreversible change in structure or function is considered to be highly severe. Even if an effect on function is not observed, cell death in an organ that does not replicate, such as the nervous system, represents reduced reserve capacity. An organism with reduced reserve capacity in the brain or other parts of the nervous system is more likely to experience disability at an earlier stage in its lifecycle as nerve or brain cells continue to mature and die as a result of the normal aging process. In addition, a severe effect that is not permanent will also be classified as highly severe. An example of a reversible severe effect is a serious allergic reaction or hypersensitivity that eventually goes away. Another example is a serious nervous system disorder that eventually goes away suggesting that adaptation has occurred.

Moderately severe effects are altered structure or physiological changes correlated to an adverse effect. These effects may be reversible or irreversible. They might include biochemical alterations indicating altered organ function, moderate, localized tissue damage (e.g. fatty liver), increased or decreased organ weights, depressed immunity that is reversible, etc.

Effects of low severity are transient, mild effects not necessarily correlated with an adverse effect. Such effects might include hepatic enzyme induction, slight respiratory irritation, etc.

The categories for low, moderate, and high severity are roughly equivalent to the categorization of dose levels as No Adverse Effect Levels, Lowest Adverse Effect Levels, and Frank Effect Levels by the Environmental Protection Agency (De Rosa et al., 1985).

The cutoffs and dose ranges used to score the relative potency of systemic toxicants reflect the dose levels used in laboratory animal tests as well as exposure considerations. Most environmental exposures are expected to be less than 1 mg/kg/day. The highest dose level of a chemical causing severe effects leading to its classification into the highest concern category used for the CERCLA ranking process is 0.0004 mg/kg/day for the rat and 0.0007 mg/kg/day for the mouse. Thus, the CERCLA ranking process has the lowest dose requirements for classification into the highest concern category. This low dose may have the effect of decreasing the overall weight given to chronic toxicity for determining reportable quantities relative to other endpoints (i.e., cancer and aquatic toxicity). Of those chemicals for

which reportable quantities were assigned based on the chronic toxicity score, none had reportable quantities of one pound, the category of highest concern. This may be an unfair assessment because not all of the CERCLA hazarous substances were scored for chronic toxicity. Even so, a higher dose level has been selected associated with severe toxic effects for the Sunset scoring system.

It is difficult to compare the relative stringency of this scoring system compared to the other programs that do not integrate severity of effect and potency. The GLWQA, EMPPL, and WMS programs rely only on potency to classify compounds for chronic toxicity. The Ontario list of Sunset candidates and the WMS scoring system apply a no observed adverse effect level (NOAEL) of 0.1 mg/kg/day or no observed effect level (NOEL) of 0.5 mg/kg/day to classify substances into the highest concern category. It should be noted that while the dose levels are lower than that used by our Sunset scoring system, these are no effect levels while the Sunset dose level is equivalent to a lowest adverse effect level (LOAEL). Often, when a NOAEL is not known for a chemical, the LOAEL is divided by a factor of ten to estimate the NOAEL. Scoring for severity as well as potency is a more stringent approach to chemical classification and should assist in the identification of the most hazardous substances.

Of the systems that combine the two criteria, MCMR and TSCA, score compounds the highest that cause severe adverse effects in animals at a dose level of 5 mg/kg/day or less. The dose level used by the CERCLA ranking process is an animal dose of 0.0004 mg/kg/day for the rat or 0.0007 mg/kg/day for the mouse. The use of 1 mg/kg/day by this scoring system to trigger the highest score for potency is slightly more stringent than the MCMR and TSCA scoring systems, but less stringent than the CERCLA ranking process by four orders of magnitude.

REFERENCES

De Rosa, C.T., J.F. Stara, and P.R. Durkin. 1985. Ranking Chemicals Based on Chronic Toxicity Data. Toxicol. Ind. Health 1(4):177-191.

De Rosa, C.T., M.L. Dourson, and R. Osborne. 1989. Risk Assessment Initiatives for Noncancer Endpoints: Implications for Risk Characterization of Chemical Mixtures. Toxicol. Ind. Health 5(5):805-824.

Doull, J. 1987. Introduction: Historical Perspectives and General Concepts. Toxic Substances and Human Risk. R.G. Tardiff and J.V. Rodricks, eds. Plenum Press. New York, N.Y. 445 pp.

National Research Council/National Academy of Sciences. 1984. Toxicity Testing: Strategies to Determine Needs and Priorities. National Academy Press. Washington, D.C. 382 pp.

U.S. Congress, Office of Technology Assessment. 1990. Neurotoxicity: Identifying and Controlling Poisons of the Nervous System. OTA-BA-436. U.S. Government Printing Office, Washington, D.C. 361 pp.

U.S. Congress, Office of Technology Assessment. 1991. Identifying and Controlling Immunotoxic Substances - Background Paper. OTA-BP-BA-75. U.S. Government Printing Office, Washington, D.C. 93 pp.

CLASSIFICATION AND SCORING SCHEME FOR CARCINOGENICITY

WOE Group	Definition
A	Sufficient evidence of cancer in humans, or
B	Evidence of cancer in 2 or more laboratory animal species or strains, or replicate studies, and
C	Evidence of cancer in one laboratory animal species, or benign tumors only, or
D	Inadequate data suggest carcinogenic potential, or no data.

Potency Group	$1/ED_{10}$ *
1	> 100
2	1 - 100
3	< 1

* $1/ED_{10}$ is the potency factor used to assign reportable quantities for carcinogens by the CERCLA Section 102 program. ED_{10} is the dose (mg/kg/day) calculated to result in 10% tumor incidence.

		Potency Group 1	2	3
WOE Group	A	H	H	M
	B	H	M	L
	C	M	L	L
	D	No Hazard Ranking		

PARAMETER SELECTION

This scoring system evaluates relative carcinogenic hazard for humans in a separate category. The prevention of cancer is a critically important public health goal in the United States.

Most regulations promulgated by the U.S. EPA under TSCA Section 6 and FIFRA have had as a primary goal cancer prevention.

A substance is defined as a carcinogen in this scoring system if it causes a statistically significant dose-related increase in malignant tumors in a specific tissue; or malignant and benign tumors combined in the same tissue at the same organ site in an animal bioassay.

Sufficient evidence of cancer in humans is defined as data indicating a causal association between exposure to an agent and cancer in humans. Limited evidence of cancer in humans is defined as data indicating that a causal interpretation is credible, but alternative explanations, such as chance, bias, or confounding could not be adequately excluded. These are the same definitions as those used by the U.S. EPA (1986) and are compatible with those used by the International Agency for Research on Cancer (IARC, 1987).

DATA QUALITY

The studies used for scoring carcinogenic effects must be properly conducted, producing statistically and biologically significant results that are adequately reported. Guidelines were developed for the purpose of evaluating published studies reporting carcinogenicity in humans and laboratory animals (U.S. EPA, 1986; OSTP, 1985).

The National Research Council, in a study of a sub sample of 100 substances, found that carcinogenicity tests in rodents had been conducted on 20 - 29 percent of pesticides and inert ingredients and 40 - 49 percent of cosmetic ingredients (NAS/NRC, 1984). These data were available for 10 - 19 percent of drugs and excipients, food additives, and chemicals in commerce. Data on genetic toxicity were available for a higher proportion of these substances.

IARC (1987) determined that of the 44 agents for which there is sufficient or limited evidence of carcinogenicity to humans, the 37 substances that had been tested adequately in animals produced cancer in at least one species. Therefore, IARC concluded that, "Although this association cannot establish that all agents that cause cancer in experimental animals also cause cancer in humans, nevertheless in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents for which there is sufficient evidence of carcinogenicity in experimental animals as if they presented a carcinogenic risk to humans." Given that an important goal of the Sunset protocol is to prevent disease in humans, data derived from tests using laboratory animals are given equivalent consideration to epidemiological evidence.

RATIONALE

The classification scheme for carcinogenicity is a priority setting system using criteria based on strength of evidence and potency considerations. Chemicals placed in the HIGH category are considered to pose the greatest threat to humans. All three of the scoring categories, high, moderate, and low, reflect some level of concern and, we feel, merit some type of risk management activity. No effort is made to identify chemicals which present little or no carcinogenic hazard to humans.

Chemicals that cause cancer in two or more laboratory animal species are given a higher score in this screen than chemicals with carcinogenicity data in one species. This is based on the assumption that evidence in multiple animal species increases the weight-of-evidence for cancer in humans (OSTP, 1985, EPA, 1986). Several weight-of-evidence classifications currently in use define sufficient evidence of cancer in animals or give greater weight to positive evidence in multiple species, strains, or experiments (EPA, 1986; IARC, 1987; Calif. DHHS; IJC, 1989; MOE, 1990; GDCh-Advisory Committee, 1989). The EPA system also defines sufficient animal evidence as being data which indicate the induction of cancer to an unusual degree in a single experiment with regard to high incidence, unusual site or type of tumor, or early age at onset. Other classification systems, including the Michigan CMR and the Netherlands WMS, do not give greater weight to evidence in multiple studies.

It has been suggested that tumor incidence in certain organs in animal oncogenicity studies may have less relevance to human cancer hazard assessment. Liver cancer in certain species of mice that have a high background incidence of liver cancer has been singled out as having less significance than other sites or species when predicting carcinogenicity hazard in humans. In fact, the GDCh-Advisory Committee scoring system gives less weight to substances found to cause only liver tumors in mice. Further, an increase in the incidence of liver tumors in mice, but not in other species, is not regarded as sufficient evidence of carcinogencity in the Netherlands (Vermeire and van der Heijden, 1990).

The U.S. EPA cancer risk assessment guidelines (1986) state that an increased incidence of only mouse liver tumors will be regarded as sufficient evidence of carcinogenicity if all other conditions for a classification of "sufficient" evidence in animal studies are met. Factors that would lower the evidence classification to "limited" include lack of replication, increased incidence at high dose only, or at the end of an experiment, no dose-related increase in malignancies, predominantly benign responses, no dose-related shortening of the time to the appearance of tumors, negative or inconclusive results from a spectrum of short term tests for mutagenic

activity, or the occurrence of excess tumors only in a single sex.

The Environmental Protection Agency is reviewing its use of mouse liver tumor data in cancer risk assessments (Beal, 1990). The Agency used mouse liver tumor data in approximately one-third of the agency's cancer risk assessments as of October, 1988 (39/109). Eight out of 34 (24%) compounds tested in both mice and rats caused only liver tumors in mice. The agents responsible for inducing liver cancer in mice are primarily chlorinated compounds, although nitrogen-containing compounds have also produced liver tumors. More than half of substances classified as B2 carcinogens (56%) were classified with the use of mouse liver tumor data. Four of the 52 (8%) B2 chemicals were classified solely on the basis of liver tumors in mice. All of the compounds, aldrin, bis(2-chloroethyl) ether, heptachlor, and hexachlorocyclohexane, are chlorinated substances.

It is evident that a consensus does not exist regarding the relevancy of tumors that arise in organs with a high background tumor incidence to human cancer risk. On the other hand, U.S. regulatory agencies have consistently implemented a science policy that incorporates this tumor type in weight-of-evidence schemes for carcinogenic potential. This scoring system also uses evidence of this nature to indicate carcinogenic potential in humans, but such scores are flagged to indicate their dependence on these data. Generally the scoring system will consider relevant tumors at any site unless the EPA has determined otherwise, as was the case for kidney tumors in rats (reference).

Our scoring system gives a lower score to compounds causing only benign tumors in an animal species. The U.S. EPA risk assessment guidelines for cancer and the OSTP (1985) recommend that benign tumors of the same histogenic origin should be combined with malignant tumors if they would be expected to progress to malignancy. The U.S. EPA Risk Assessment Forum concluded that foci and nodules in the rat liver fall into this category. The California Department of Health Services follows the EPA procedure reasoning that in addition to the potential to progress to malignancy, the induction of benign tumors in experimental animals reflects the biological activity of the carcinogen (CDHS, 1985). The same chemical may cause malignant tumors in other species. However, the EPA weight-of-evidence scheme gives less weight to substances causing only benign tumors in a laboratory species.

The classification of substances with cancer incidence in an animal experiment only at the highest dose level given, the Maximum Tolerated Dose (MTD), are flagged in the Sunset screening system. The possibility that metabolic pathways were saturated, or some other alteration of normal physiology occurred leading to the induction of tumors is higher if effects were observed only

at the MTD (OSTP, 1985). However, the data should be evaluated to assess metabolic differences between the animal species tested and human metabolic capacity to determine if the same response may be possible in humans. It was suggested that methylene chloride causes cancer in mouse liver and lungs and not in rats because of species differences in the importance of the pathways used to metabolize the substance (ECETOC, 1987).

The MOE scoring system gives a lower score for a compound that is carcinogenic in animal bioassays at levels shown to saturate enzymes involved in its metabolism. Professional judgment is required to determine if enzyme saturation has occurred.

This scoring system does not differentiate between carcinogens based on the possible mechanism of action. Two existing scoring systems, those used by the Ontario Ministry of the Environment and the Michigan Department of Natural Resources, use mechanistic considerations to score compounds for carcinogenic potential. The Ontario MOE system gives a higher score to animal carcinogens with evidence of direct interaction with genetic material. Such evidence might include electrophilic activity, direct alkylating activity, the production of DNA adducts, and the induction of cell transformation. The Michigan CMR also provides for differential scoring for substances considered to be initiators or promoters. In addition, risk assessment in the Netherlands differentiates between genotoxic carcinogens and carcinogens without genotoxic properties (Vermeire and van der Heijden, 1990).

On the other hand, federal and international policy discourages the categorization of carcinogens based on mechanistic assumptions. While the U.S. EPA weight-of-evidence scheme allows the use of data on genotoxicity as supporting evidence for carcinogenic potential when the data regarding a substance are otherwise limited, it does not treat compounds differently based on mechanistic considerations. Similarly, IARC (1987) has concluded that the state of the science does not allow the differentiation between initiators and promoters in classification schemes.

Perera (1991), in a review of the state of the science for cancer risk assessment, concluded that prioritizing carcinogens based on simplistic determinations of mechanism or stage of action is not adequate. A large body of evidence argues against the assignment of the existence of a threshold for action based on mechanism or the stage of involvement. This includes indications that inter-individual variability makes the existence of a population threshold for promoters not likely, evidence of indirect genomic changes induced by tumor promoters, the variety of mechanisms influencing tumor generation exhibited by single chemicals, and the effects of multiple factors, including chemicals and viruses, acting in combination.

Finally, this scoring system ranks chemicals with carcinogenicity data on the basis of potency. We have adopted a ranking methodology developed by the EPA Carcinogen Assessment Group for the CERCLA Section 102 program to establish reportable quantities for hazardous wastes. This program classifies carcinogens on the basis of weight of evidence and potency. Potency is evaluated by calculating the ED_{10} using the multistage model of dose response. The potency is defined as the reciprocal of the estimated dose associated with a lifetime increased cancer risk of 10 percent (ED_{10}).

This measure of potency was chosen by CAG in preference to the q_1^*, the usual potency factor used in EPA risk assessments, because it is relatively insensitive to the choice of the dose-response model, it does not require extrapolation beyond the observed data, and it is a statistically stable estimate. The q_1^* requires the use of upper bounds to ensure stability while the $1/ED_{10}$ does not. The relationship between the $1/ED_{10}$ and q_1^* was analyzed empirically and the two potency measures were found to be closely correlated. The $1/ED_{10}$ was found to be about 6 times the q_1^* on average.

Generally, studies are selected to evaluate dose-response in accordance with EPA's cancer risk assessment guidelines. The $1/ED_{10}$ is calculated using data from the most sensitive species tested and the substance is categorized into one of three potency groups. The ranges that trigger classification into potency groups 1, 2, or 3 were selected because they were predicted to place about 25% of scored substances into the lowest and highest group and 50% into the middle category (personal communication, Jim Cogliano, April 22, 1992). Two situations are handled in the following manner. If the dose-response data are not suitable to calculate the potency measure, the substance is automatically assigned to Potency Group 2. If all animals in a selected study developed tumors, the substance is assigned to Potency Group 1.

The methodology has been used to rank 194 substances for the CERCLA program. The same ranking methodology is being used to rank toxic air pollutants under the Clean Air Act Amendments of 1990. The potency factor, $1/ED_{10}$, calculated by the Agency for these programs is used by the Sunset screening system. If the $1/ED_{10}$ has not been calculated for a potential carcinogen scored by this system, the q_1^* will be used to calculate the $1/ED_{10}$ if it is available.

Weight of evidence and potency are combined in the CERCLA Section 102 program using a matrix to assign a substance to a high, medium, or low category and its associated reportable quantity. The Sunset screening system combines the potency measure with the same types of evidence as that used in the CERCLA program. Substances are assigned to categories labelled high, moderate, and low. Substances with a weight of evidence based on epidemiologic data are not assigned to the low category. Potency

groups 1 and 2 are categorized as high and potency group 3 is categorized as moderate. Substances with evidence of carcinogenic activity from one study or species are not assigned to the high category.

Most existing classification schemes do not include considerations of potency. Two scoring systems were described in the literature which combine weight-of-evidence and potency. The Safety and Health Index System (SHIS) developed by PPG Industries, Inc. reduces the SHIS Rating from 4 to 3 if a route specific dose level is exceeded (Henry and Schaper, 1990). These dose levels are as follows.

1. Inhalation: 1 mg/m3 (or equivalent ppm) in 6 to 7 hour daily exposures throughout lifetime.

2. Intratracheal dose: 1 mg of particulates or liquid per 100 ml or less of animal minute respiratory volume.

3. Dermal application: 2 mg/kg body weight twice weekly or total dose equal to 1.5 mg.

4. Feeding study: daily intake at 1 mg/kg body weight, total dose 50 mg (rat) or 3.5 mg (mouse).

Szejnwald-Brown et al. (1986) described a methodology for classifying carcinogens based on weight-of-evidence and potency. Compounds were classified according to both a weight-of-evidence scheme adapted from the IARC scheme and a potency scale using values derived from National Toxicology Program or Carcinogen Assessment Group data expressed as unit risks. The unit risks convey the lifetime excess cancer risk for experimental species per unit dose for each chemical. The scale used to classify compounds by potency is shown below.

Unit Risk Range	Risk Level
10-3 < Unit Risk	Very High
10-4 < Unit Risk < 10-3	High
10-6 < Unit Risk < 10-4	Moderate
Unit Risk < 10-6	Low

The Szejnwald-Brown scheme does not calculate unit risk in cases where there is conclusive evidence of human cancer. The unit risk is calculated for all animal bioassays showing statistically significant, positive results. The linearized multistage model is used to calculate the animal potency value and the upper 95% and lower 5% confidence limits. The potency value is then used

to calculate the unit risk, defined as the excess probability of cancer for the animal tested after lifetime exposure to 0.3 ug/kg/day of body weight of the substance per day. The unit dose was selected because the same level had been used by the Carcinogen Assessment Group (CAG). The weight-of-evidence and unit risk are combined in a two-dimensional matrix and are scored into groups A through E. Substances in a lower weight-of-evidence category are required to have a higher unit risk value to be given a specific score.

While the methodology has some attractive features, it was not selected for this screening system for two major reasons. First, The use of the q_1^* has some disadvantages that can be avoided by the use of the $1/ED_{10}$. These are the statistical instability of the q_1^* requiring that the upper bound be used, and the need to extrapolate beyond the observed data range. The use of the upper bound and extrapolation to very low doses make sense and are necessary when the goal of assessment is to ascertain the risk posed to a human population in connection with a particular exposure scenario. However, when the goal is to rank substances according to their hazard relative to one another, as is the goal of this screening system, it is not necessary to accomodate the limitations of the q_1^*.

The second reason for not adopting the unit risk is that the origin of the selected dose level is uncertain. While the dose level was attributed to the Carcinogen Assessment Group, the purpose for its use was not known. The Sunset Protocol should be applied to substances with a variety of properties causing exposure in different media at a wide range of potential concentrations. Therefore, the relevance of 0.3 ug/kg/day is uncertain. Finally, the relationship of the unit dose and its associated unit risk with the risk ranges causing classification in one of the three hazard categories is unclear. The ranking system adopted from the CERCLA Section 102 program is more straightforward for ranking purposes and does not suffer from these limitations.

Other ways of prioritizing chemicals based on potency which do not use low dose extrapolation models have been proposed. Clayson et al. (1983) proposed that the strength of a carcinogen can be measured by the dose rate which leads to the development of specific tumors in 50% of treated animals. Peto et al. (1984) proposed a standard numerical index for potency, the TD50 (tumerigenic dose 50). The TD50 was defined as a dose rate in mg/kg body weight/day which will halve the probability of remaining tumor free to the end of the standard life span of the species. Gold et al. (1984) calculated the TD50 for a large number of chemicals generating the Carcinogenic Potency Database. The TD50 was corrected for intercurrent mortality and background tumor rate. The TD50 was not correlated with target site, the induction of tumors at multiple sites in the same sex-species of test animal, tumor which may have been lethal, or tumor which metastasized to the lung (Gold et al., 1986).

Ames et al. (1987) used the TD50 index and human exposure data to propose the HERP (Human exposure dose/rodent potency dose). Human exposure in mg/kg/day lifetime dose was estimated from published data. Another index, PERP, the Permitted Exposure/Rodent Potency index, also used the TD50 to evaluate occupational exposure to carcinogens (Gold et al., 1987; Hooper and Gold, 1986). The TD50 was compared to a Maximum Occupational Dose Rate (MOD). The MOD was defined as the average lifetime daily dose a worker can legally receive by inhaling airborne concentrations of a chemical at the OSHA PEL for 5 days/week over a 40-year work life. A problem with the use of indices which combine exposure and potency data is that the ranking is dependent on the quality and quantity of information on exposure. The use of published concentration data or the OSHA PEL as surrogate measures of population exposure may be applicable to only a small subset of the exposed population.

The proposed advantage to the use of the TD50 was that it would often be included in the experimental dose range and had a useful analogy to the LD50, a measure of potency for acute lethality. The use of the TD50 also avoided the need to use a model to extrapolate to low doses expected in the general population. On the other hand, criticisms of the use of the LD50 also apply to the TD50. While it does use data based on chronic exposure, the TD50 does not reveal the shape of the dose-response curve at lower doses. Two compounds may have the same TD50, but the dose-response curve for one may have a steeper slope than the other and is considered more dangerous. The relative carcinogenic potency in the low dose region, the relevant area for most environmental exposure situations, may be different. The ED_{10} also suffers from this limitation, but it is a better measure because it integrates the dose-response data in the range of the observed data range through use of the multistage model and focuses on the carcinogenic response at the lowest dose within that range.

REFERENCES

Beal, D.D. 1990. Use of Mouse Liver Tumor Data in Risk Assessments Performed by the U.S. Environmental Protection Agency. Mouse LIver Carcinogenesis Mechanisms and Species Comparisons. Alan R. Liss, Inc. pp. 5-18.

Brown, S. H., 1986. A Methodology For Assessing Carcinogenic Hazards of Chemicals. Toxicol. Ind. Health 2:205-218.

California Department of Health Services. 1985. Guidelines for Chemical Carcinogen Risk Assessments and Their Scientific Rationale. State of California. Department of Health Services.

Cothern, C. R. 1989. Some Scientific Judgments in the Assessment of the Risk of Environmental Contaminants. Toxicol. Ind. Health 5:479-491.

ECETOC. 1987. The Assessment of Carcinogenic Hazard For Human Beings Exposed To Methylene Chloride. ECETOC Technical Report No. 26. European Chemical Industry Ecology and Toxicology Centre, 250 Ave. Louise (Bte63), 1050-Brussels, Belgium. 62 pp.

Gold, S. L. 1986. Association between Carcinogenic Potency and Tumor Pathology in Rodent Carcinogenesis Bioassays. Fund. Appl. Toxicol. 6:677-690.

Henry, B.J. and K.L. Schaper. 1990. PPG'S Safety and Health Index System: A 10-Year Update of an In-Plant Hazardous Materials Identification System and Its Relationship to Finished Product Labeling, Industrial Hygiene, and Medical Programs. Am. Ind. Hyg. Assoc. J. 51(9):475-484.

Hooper, K. and L.S. Gold. 1986. The Exposure-Potency Index (EPI): Ranking Carcinogenic Hazards of Volatile Industrial Chemicals. In:Cancer Prevention: Strategies in the Workplace, C.E. Becker and M.J. Coye, Eds.; Washington, D.C., Hemisphere Publishing Corporation. pp. 11-21.

International Agency for Research on Cancer. 1987. IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Volume 1 to 42, Supplement 7. pp. 322-326.

Office of Science and Technology Policy (OSTP). 1985. Chemical Carcinogens; a Review of the Science and Its Associated Principles. FR March 14, 1985. pp. 10371-10442.

Perera, F.P. 1991. Perspectives on the Risk Assessment for Nongenotoxic Carcinogens and Tumor Promoters. Environ. Health Perspect. 94:231-235.

U.S. Environmental Protection Agency. 1986. Proliferative Hepatocellular Lesions of the Rat; a Review and Future Use in Risk Assessment. Feb. 1986.

U.S. Environmental Protection Agency. 1986. Guidelines for Carcinogen Risk Assessment. FR 51:33992-34003.

Vermeire, T. and K.V.D. Van Der Heijden, 1990. Health Assessment of Hazardous Air Pollutants the Netherlands. Toxicol. Ind. Health, 6:234-243.

CLASSIFICATION AND SCORING SCHEME FOR REPRODUCTIVE AND DEVELOPMENTAL TOXICITY

Score (A)	Definition
8	Sufficient evidence of reproductive or developmental toxicity in humans, or Very severe effects in two or more laboratory animal species, or Very severe effects in one laboratory animal species with compelling, but limited, epidemiological evidence.
6	Moderately severe effects in two or more laboratory animal species, or Moderately severe effects in one laboratory animal species with compelling, but limited, epidemiological evidence.
4	Severe or moderately severe effects in one laboratory animal species.
2	Positive short-term in-vitro or in-vivo tests for teratogenicity.

Score (B)	Lowest Effective Dose (mg/kg/day)
4	< 1
3	> 1 - 10
2	> 10 - 100
1	> 100

Final Score	A x B
HIGH	32
MODERATE	12 - 24
LOW	1 - 8

PARAMETER SELECTION

The societal importance placed on reproductive health requires that any information on reproductive and developmental toxicity

be given special attention. For this reason, reproductive and developmental toxicity is scored as a separate parameter.

DATA QUALITY

The studies used for scoring reproductive and developmental effects must be properly conducted, producing statistically and biologically significant results that are adequately reported. Guidelines have been developed for the purpose of evaluating published studies reporting reproductive and developmental toxicity in humans and nonhuman mammals (IRLG, 1981; IRLG, 1986; U.S. EPA, 1986; U.S. EPA, 1988a; U.S. EPA, 1988b; U.S. EPA, 1989). These guidelines were selected to serve as the basis for evaluating the evidence used to score effects on the reproductive system and offspring.

The National Research Council estimated in 1984 that between 4 and 7 percent of chemicals in commerce had some test data available on reproductive or developmental effects (NRC/NAS, 1984). Reproductive or developmental toxicity data were reported in the literature for 45% of drugs and excipients in drug formulations. Data existed for 20% of food additives, 22% of cosmetic ingredients, and 34% of pesticides and inert ingredients.

While basic species differences do occur, reproductive or developmental effects observed in animals are believed to predict toxic potential in human beings (IRLG, 1986; U.S. EPA, 1986). Comparisons of human and animal data show that for a limited number of agents known to be teratogenic or causing adverse reproductive effects in humans, there is almost always concordance of effect between humans and at least one non-human species (Schardein et al., 1983; Nisbet and Karch, 1983). A 1980 U.S. FDA literature review identified 38 compounds for which birth defects were reported in humans associated with their intake during pregnancy (Schardein, 1983). One hundred sixty-five compounds for which teratologic effects were not reported in humans were also identified. Eighty percent of the known or suspected human teratogens tested positive in multiple nonhuman species. Eighty-five percent and eighty percent of the compounds tested positive in mice and rats, respectively. The rabbit tested positive for 60% of the human teratogens, while the hamster and monkey showed a teratogenic response to 45% and 30% of the compounds. Unfortunately, no single species can be expected to be predictive for humans for every compound tested.

Similar types of adverse effects in humans and animals have been reported. A comparison of 10 teratogenic substances found that eight of the materials produced qualitatively similar effects in humans and laboratory animals (Nisbet and Karch, 1983). Moreover, human sensitivity to these agents compared to animals was usually greater, ranging between 2 and 50 times.

Participants at a workshop evaluating human and animal developmental neurotoxicants concluded that there was close agreement for effects between humans, other primates and rodents for seven neurotoxic materials and classes reviewed (Stanton and Spear, 1990). Effects were similar in all relevant functional categories, which included sensory, motivation/arousal, cognitive, motor, and social.

MINIMUM DATA SET

Reproductive or developmental effects must be reported in at least two mammalian species before a compound can be given the highest score possible. This requirement is based on the rationale that if a material causes adverse effects in multiple animal species, it more likely will cause adverse effects in humans. Further, the most sensitive endpoint and the lowest effective dose more likely will be identified.

The requirement of evidence in two animal species is more restrictive than that recommended by the U.S. EPA in evaluating the weight-of-evidence in risk assessments for reproductive effects (U.S. EPA, 1988b; U.S. EPA, 1989). A more stringent criterion is justified by the goal of this scoring system, to prioritize compounds as potential candidates for Sunsetting. A system developed by Brown et al. (1986) requires at least two positive animal tests to place a substance into the "substantial evidence" category for developmental toxicity. At least one positive animal test and some positive (although not conclusive) human evidence is required to place a compound into an equivalent category for reproductive toxicity. The following table presents the minimum data requirements used by other scoring systems that result in the highest score possible.

Scoring System	Minimum Data Requirement
GLWQA	One mammalian assay (developmental only)
EMPPL	Not specified (developmental only)
MCMR	Two mammalian species, strains or replicate tests
TSCA	Two mammalian species (developmental only)

The U.S. EPA requires the testing of two species for developmental effects under TSCA and FIFRA (U.S. EPA, 1982; U.S. EPA, 1985; U.S. EPA, 1987). Under TSCA, the rat, mouse, rabbit,

or hamster are acceptable, while under FIFRA, the rat and rabbit are preferred. One multi generational reproduction study is required by both TSCA and FIFRA. The rat is the preferred species under TSCA, while either the rat or mouse are allowable under FIFRA.

Several types of tests are conducted to study reproductive and developmental effects in laboratory animals including single and multi generational reproduction tests, studies of reproductive endpoints with exposures 90 days or less, tests of developmental effects using exposure periods over the entire period of organogenesis or shorter, more specific exposures. If properly conducted, any test that results in adverse effects can be used to score a compound. However, this has implications on the scoring of potency because not all of these tests are able to identify the most sensitive endpoint. Therefore, the lowest adverse effect dose will merely reflect the type of test from which the data were obtained.

RATIONALE

This scoring system evaluates reproductive and developmental effects together in one effect parameter. This is because both types of effect bear on the ability to produce children and the survivability and well-being of children. The endpoints are difficult to separate. Laboratory animal studies often measure both types of effect in the same experiment. Moreover, an equivalent level of societal concern is placed on reproductive capability and congenital problems in children.

This scoring system places a higher priority on very severe reproductive and developmental effects compared to moderately severe reproductive and developmental effects. Severe reproductive effects include infertility or reduced fertility, permanent effects on spermatogenesis, or extensive cellular damage in reproductive organs. Moderately severe effects involve biochemical changes related to reproductive function, reversible effects on spermatogenesis, organ weight changes and moderate histochemical alterations in reproductive tissues. Very severe developmental effects result in death, decreased longevity, or restricted functional capability in offspring. Moderately severe developmental effects are the result of delayed growth and development, are reversible with age, and should not result in permanent harm.

This system is modeled after a classification system developed by Brown et al. (1986) that places greater weight on teratogenic events and severe embryo/fetal effects. According to this system teratogenic effects include major and minor malformations, and behavioral and functional abnormalities. Examples of teratogenic effects include;

encephaly
spina bifida
cleft palate
acaudia (short tail)
omphalocele (congenital hernia of the navel)
missing organ
malformed organ
displaced organ
abnormal organ weight
functional alterations - altered biochemistry, physiology, etc.
aortic arch
imperforate anus
micrognathia (abnormal smallness of lower jaw)
agnathia (lower jaw absent)
oligodactyly (abnormal number of fingers or toes)
syndactyly (fusion of two or more toes or fingers)
hydroencephaly
anophthalmia (absence of eyes)
mental retardation
abnormal motor ability, sociability, learning ability.

Embryo/fetal toxic effects are classified by Brown et al. into severe and minor according to the seriousness of the effect and whether the effect is reversible. Severe embryo/fetal effects include lethality, resorptions, individual skeletal variants (missing or poorly ossified sternebrae, vertebral centers, skull), abnormal umbilical cord length, transumbilical distance, post implantation loss, and minor malformations or variations that are common in the species tested. Minor embryo/fetal toxic effects include decreased crown-rump length, reduced birth weight or weight gain, retarded physical development, and increased total skeletal variants (no individually increased incidence that is statistically significant).

Of the classification systems evaluated for this project, one differentiates between the relative severity of developmental effects caused by an environmental agent. The scoring system developed by the Ontario Ministry of the Environment and used by the EMPPL program gives a higher score to pollutants causing teratogenic effects in experimental animals compared to developmental abnormalities. While the EMPPL process for chemical selection does not take advantage of this capability, the Ontario list of proposed Sunset substances was compiled using the most stringent trigger.

Our scoring system does not differentiate between substances that cause maternal toxicity at dose levels resulting in adverse developmental effects and those that do not. There is no professional consensus regarding the treatment of substances which cause adverse developmental effects at maternally toxic dose levels. The U.S. EPA Guideline for the Health Assessment of Suspect Developmental Toxicants (51 FR 34028-34040, 54 FR 9386, March 6, 1989) states that "when adverse developmental effects

are produced only at minimal maternally toxic doses, they are still considered to represent developmental toxicity and should not be discounted as being secondary to maternal toxicity." It is very difficult to distinguish between effects that occur as a result of maternal toxicity and those that occur independently at the same dose. Moreover, the effects on the dam and the offspring may not be equivalent in severity and permanence. On the other hand, the EPA guidance also states that effects produced in offspring at doses that do not result in maternal toxicity are of greatest concern. In addition, participants at a recent workshop on developmental neurotoxicity produced opposite recommendations regarding the treatment of maternal toxicity in interpreting animal tests of developmental neurotoxicity (Levine and Butcher, 1990; Tyl and Sette, 1990).

A classification system for developmental toxicants developed by Brown et al. (1986) gives greater weight to compounds that produce adverse effects below the maternally toxic dose. Other scoring systems also differentiate between substances that produce effects in the offspring alone or in both the dam and offspring at a particular dose level. Of the systems evaluated for this project, the MOE and TSCA scoring systems make this distinction.

The cutoffs and dose ranges for scoring the relative potency of reproductive and developmental toxicants reflect the dose levels used in laboratory animal tests and consideration of anticipated environmental exposure. Brown et al. (1986) evaluated 110 chemicals for male and female reproductive toxicity or developmental effects. No data were found for 62 (56%) of the compounds. Of the 48 chemicals with data, eight compounds (17%) had a LOAEL of < 1 mg/kg/day. LOAELS for 8 (17%), 16 (33%), and 11 (23%) of the substances were in dose categories > 1 - 10, > 10 - 100, and > 100 mg/kg/day, respectively. One chemical (1%) had a LOAEL of 1000 mg/kg/day or greater. Schardein published two evaluations of the potency of drugs in animals compared to humans. The relative distribution of LOAELs for teratogenic effects reported in mammals for 35 (Schardein, 1976) and 37 (Schardein, 1983) drugs is presented below.

Dose Range (mg/kg/day)	Schardein, 1976		Schardein, 1983	
	#	(%)	#	(%)
< 1.0	7	(20)	1	(3)
1 - 10	5	(14)	6	(16)
10 - 100	11	(31)	16	(46)
> 100	12	(34)	14	(38)

The lowest dose category used by this scoring system, 1.0 mg/kg/day, is one order of magnitude higher than that used by Ontario to develop its list of proposed Sunset substances. It is two orders of magnitude less than the lowest dose level used by the Michigan CMR to score industrial chemicals (500 mg/kg/day).

REFERENCES

Brown, H. S. 1986. A Methodology For Assessing Developmental And Reproductive Hazards Of Chemicals. Toxicol. Ind. Health, 2:183-203..

Inter Agency Regulatory Liaison Group (IRLG). 1986. Interagency Regulatory Liaison Group Workshop On Reproductive Toxicity Risk Assessment. Environ. Health Perspect. 66:193-221.

Interagency Regulatory Liaison Group (IRLG). 1981. Guidelines For Documentation Of Epidemiologic Studies. Am. J. Epid. 114:609-618.

Levine, T. E. and R. E. Butcher. 1990. Workshop On The Qualitative And Quantitative Comparability Of Human And Animal Developmental Neurotoxicity, Work Group IV Report: Triggers for Developmental Neurotoxicity Testing. Neurotox. Teratol. 12:281-284.

National Research Council/National Academy Science. 1984. Toxicity Testing: Strategies to Determine Needs And Priorities. National Academy Press, Washington, D.C. 332pp.

Nisbet, I. C. T. and N. J. Karch. 1983. Chemical Hazards To Human Reproduction. Noyes Data Corporation. Park Ridge, New Jersey. 245pp.

Schardein, J. L. 1976. Drugs As Teratogens. CRC Press, Inc. Cleveland, Ohio.

Schardein, J. L. 1983. Teratogenic Risk Assessment: Past, Present, and Future. In: Issues and Reviews in Teratology 1:181-214. H. Kalter, ed. Plenum Publishing Corporation.

Stanton, J. E. and L. P. Spear. 1990. Workshop On The Qualitative And Quantitative Comparability Of Human And Animal Developmental Neurotoxicity. Work Group I Report: comparability of Measures of Developmental Neurotoxicology in Humans and Laboratory Animals. Neurotox. Teratol. 12:261-267.

Tyle, R. W. and W. L. Sette. 1990. Workshop On The Qualitative And Quantitative Comparability Of Human And Animal Developmental Neurotoxicity. Work Group III Report: Weight of Evidence and Quantitative Evaluation of Developmental Neurotoxicology Data. Neurotox. Teratol. 12:275-280.

U.S. Environmental Protection Agency. 1986. Guidelines For The Health Assessment Of Suspect Developmental Toxicants. FR 51:34028-34040.

U.S. Environmental Protection Agency. 1988a. Proposed Guidelines For Assessing Female Reproductive Risk and Request For Comments. FR 53:24834 - 24847.

U.S. Environmental Protection Agency. 1988b. Proposed Guidelines For Assessing Male Reproductive Risk and Request For Comments. FR 53:24850 - 24869.

U.S. Environmental Protection Agency. 1989. Proposed Amendments To The Guidelines For The Health Assessment Of Suspect Developmental Toxicants. FR 54:9386.

CLASSIFICATION AND SCORING FOR ECOLOGICAL DISRUPTION

A few chemical compounds have adversely affected ecosystems or their components to the extent that they could be classified as large scale ecological disrupters. Examples of such chemicals and their effects include the chloro-fluoro carbons (CFC) and stratospheric ozone depletion, sulfur/nitrous oxides and acid precipitation, and greenhouse gasses (e.g. carbon dioxide) and global climate change. In some of these cases, ecological disruption has resulted from effects elicited initially on abiotic components of ecosystems (e.g. stratospheric ozone, atmospheric CO_2) rather than directly on their flora and fauna. Criteria that address toxicant impacts on biota will not capture these adverse effects. Further, the criteria that have been developed to assess chemical impacts on biota have focused on individual organisms, not on higher levels of biological organization such as populations, species, or communities. Therefore, we have developed this category to attempt to capture compounds and classes of compounds that may cause widespread ecological disruption through effects elicited on abiotic components of ecosystems or on higher levels of biological organization.

Actions to prevent ecological disruption have not occurred, for most chemicals, based on **prediction** of adverse effects; rather, actions have occurred only after chemicals have elicited their impacts on natural ecosystems and ecosystem components. For example, the destruction of stratospheric ozone by chlorine was hypothesized in 1973 by University of Michigan scientists. In 1974, University of California scientists discovered that CFCs persisted in the lower atmosphere and gradually migrated to the stratosphere. They concluded that CFCs in the stratosphere are broken down by radiation and release large quantities of chlorine. Together, the results of the University of Michigan and University of California research suggested a profound adverse effect of CFCs on an extremely fragile and immensely important part of the global ecosystem - the stratospheric ozone layer.

The U.S. began responding to the threat of CFC mediated ozone depletion in 1977 through the Clean Air Act and CFCs were banned as propellants for nonessential aerosol sprays in 1978, four years after the troubling hypotheses generated by university researchers but prior to evidence of actual stratospheric ozone depletion. Other CFC uses in the U.S. were unaffected by the ban of nonessential aerosol sprays. Although the European Community (EC) followed suit with a CFC aerosol cutback in 1980, six years after the initial hypotheses of ecological disruption, the EC essentially increased output of CFCs by over 60 percent during this period (Benedick 1991) and global output had increased by the mid 1980s. Global cooperation on CFC reductions from all sources began to occur only after substantial evidence of ozone depletion had occurred over the Antarctic in 1985. Yet, even

with extensive scientific evidence for CFC-mediated ozone depletion derived from the Antarctic as well as from regions over North America, agreements to reduce CFC production under the Montreal Protocol were not reached until 1987, with U.S. ratification occurring in 1988, fourteen years after initial predictions of adverse effects on stratospheric ozone.

Effects assessments for ecological disrupters that are conducted, either intentionally or unintentionally, in natural ecosystems provide information with a relatively high level of certainty. Such is the case for CFC-mediated ozone depletion and global controls on CFC production and use. Yet, even with a high level of certainty, cause-effect linkages are still hotly debated such as in the CFC case and in the cases of global climate change associated with elevated levels of CO_2 and other greenhouse gases and of the effects of acid precipitation. Unfortunately, the uncertainty associated with predictions based on theory or on results from surrogate systems can be substantial, and the predictions of ecological disruption derived from such studies is likely to be considered untenable.

Minimizing uncertainty in predictions or assessments of ecological disruption is clearly desirable. But where uncertainty is minimized by examining effects in natural ecosystems (in effect conducting "real-world" experiments), uncertainty reduction may come at a substantial cost. Natural or real world experiments require adverse effects to be elicited and the costs of such adverse effects may include damage to ecosystems or important components of ecosystems, and to human health.

A mechanism to assess a broad array of chemicals to predict their potential for ecological disruption has not been developed although a process to assess chemical effects at higher levels of biological organization has received some attention (Brown and Reinert 1992, Suter 1990a, Suter 1990b, Hunsaker et al. 1990, Schaeffer et al. 1988). Predictions of ecological disruption have been made **prior to the actual disruption** only in a few specific cases. And even in these cases, action to ameliorate the disruption was not taken until a cause-effect linkage had been made between the disruption and the chemical disrupter in natural systems.

There are several reasons for lack of a mechanism to identify ecological disrupters or to predict ecological disruption. First, it may be difficult to predict damage at the ecosystem or global level. This difficulty stems in part from a lack of understanding of the structure and functions of natural ecosystems and how structure and functions of such systems are affected by hazardous compounds. Further, the significance or magnitude of disruption may be difficult or impossible to predict even when an effect is predicted.

A lack of understanding of global fate and transport characteristics of hazardous compounds that may interact with components of global ecosystems such as stratospheric ozone also may inhibit the development of a predictive mechanism for ecological disrupters. Characteristics of a chemical compound such as reactivity, volatility, or persistence may be necessary but not sufficient to determine the potential for a compound to interact with critical components of ecosystems.

Even if a suite of such characteristics were adequately predictive of potential reactivity, they may not be predictive of the potential of a compound to cause widespread ecological disruption. Detailed information of the chemical's use and release patterns is also critical to predict the potential for the chemical to cause ecological disruption. Such information is usually not available at the time of creation of the compound, the time when prediction of ecological disruption potential is most useful from a regulatory standpoint. Nor is it likely that individuals involved in the creation, production, or use of a chemical compound will have insight into all future uses and releases of the compound into the environment. A new or existing compound may have only limited uses and releases early in its life cycle; thus, its potential for ecological disruption, despite its chemical and physical characteristics, may be minimized. Only later in its life cycle might uses be found that result in the release of the compound in quantities that pose threats to the ecosystems or important components.

The objective of a mechanism to characterize a chemical compound as an ecological disrupter is to predict disruption before it occurs. To make such a prediction, sufficient information must be obtained on the chemical's properties so that the potential for reactivity with the critical habitat or system (e.g. stratospheric ozone, unbuffered lakes, etc.) can be assessed. Further, information on existing as well as future use and release patterns must be available to determine the quantities of a compound that could potentially be released into the environment.

Development of a comprehensive process to predict the potential for any chemical compound to cause widespread ecological disruption seems unlikely in the near term. Therefore, we do not include such a predictive process to assess chemical compounds quantitatively as part of Sunset screening. However, evidence of ecological disruption will be used to classify chemicals as Sunset candidates, where that evidence has been documented in studies conducted by authoritative scientific bodies such as the National Academy of Sciences or in published reports in peer reviewed journals. Unfortunately, evidence for ecological disruption will be derived most often from studies of effects in natural ecosystems. However, information that predicts potential ecological disruption on a chemical specific basis, such as the early 1970s studies and hypotheses generated at the Universities

of Michigan and California on CFCs, will also support inclusion of a chemical on the Sunset candidate list. Evidence for ecological disruption derived from laboratory studies or theoretical evaluations should be supported by information on known or predicted use and release patterns of the compound. Further assessment of any compound, where it is classified as a Sunset candidate based on potential for ecological disruption, will occur in the second tier of our classification project.

REFERENCES

Benedick, R.E. 1991. Ozone Diplomacy: New Directions in Safeguarding the Planet. Harvard Univ. Press, Cambridge, Mass.

Brown, S.S. and K.H. Reinert. 1992. A conceptual framework for ecological risk assessment. Environ. Toxicol. Chem. 11:143-144.

Hunsaker, C.T., R.L. Graham, G.W. Suter, R.v. O'Neill, L.W. Barnthouse, and R.H. Gardner. 1990. Assessing ecological risk on a regional scale. Environ. Managemt. 14:325-332.

Schaeffer, D.J., E.E. Herricks, and H.W. Kerster. 1988. Ecosystem Health: I. Measuring Ecosystem Health. Environ. Managemt. 12:445-455.

Suter, G.W. 1990a. Endpoints for regional ecological risk assessments. Environ. Managemt. 14:9-23.

Suter, G.W. 1990b. Environmental risk assessment/environmental hazard assessment: Similarities and differences. Aquatic Toxicology and Risk Assessment: Thirteenth Volume. ASTM STP, 1096. W.G. Landis and W.H. van der Schalie, Eds., Amer. Soc. Testing Mater., Philadelphia, PA. pp 5-15.

DISCUSSION

The result of application of the screening process described above is a list of chemical Sunset candidates, substances that upon initial analysis have very high toxicity and potential for long term or wide-spread exposure in the environment. This list will be presented and discussed at the workshop.

A next step in a comprehensive Sunsetting activity should be focused on verification of this initial conclusion using more detailed data sources on toxicity and exposure. This step is necessary when toxicity or exposure data are derived from chemical databases, particularly where information from these databases may not be peer reviewed or clearly meet all requirements for data adequacy.

Following verification of toxicity and exposure information and classification of chemicals as Sunset candidates, further analysis of chemical candidates for Sunsetting may be appropriate prior to final decisions to ban or phase-out the chemical or associated processes and products. Such analysis should consider specific uses of the chemical and potential to ban or phase-out those specific uses. The analysis should also consider whether incremental phase-down or total chemical ban or phase-out is appropriate. For example, chemicals classified as Sunset candidates because they score high in the release and production category (as well as high in one or more toxicity categories) may be good candidates for use or release reductions, recycling, or other processes that reduce the amount of chemical

produced and used as well as the potential for the chemical to be released to the environment. Other considerations that should occur after screening and scoring may include evaluation of economic impacts of ban or phase-out options as associated with incremental phase-down or complete phase-out, as well as use, process, or product-specific bans or phase-out, analysis of substitute availability as it relates to items mentioned above, and perhaps others. Particularly important in this process is examination of potential chemical substitutes with the same criteria used to screen existing chemicals. Such screening should result in avoidance of substitution of a hazardous chemical (a Sunset candidate) with a chemical that is even more hazardous (and would ultimately be classified as a Sunset candidate itself).

A screening and scoring process based on quantitative criteria that results in identification of chemical candidates for Sunsetting or other management activities will be successful only if it is politically, economically, and socially feasible. A determination of political feasibility is problematic as it is not amenable to direct, quantitative analysis. Rather, political feasibility means development of a process that is acceptable to most of the many stake holders that will ultimately be impacted by chemical management activities and that is relatively easy to implement and manage. To this end, chemicals that are identified as Sunset candidates must be those that pose the greatest hazard to human and environmental health. Further, the number of hazardous chemicals identified by the screening and scoring

process, and identified ultimately for ban or phase-out, should be limited.

The scoring system presented here has been focused on chemicals of anthropogenic origin and where exposure to vulnerable populations or ecosystems occurs via the ambient environment. The system is not designed for, nor will it address many chemicals where occupational exposures are predominant. Consideration of the adverse human health effects of occupational exposures is critical to ensure complete understanding of the hazard of a chemical. Such consideration may also be appropriate as part of a decision process to control or ban a chemical.

A process to select chemical candidates for management activities, including Sunsetting, that is based on quantitative criteria has several important advantages compared with other selection processes. First, the process is predictable; that is, criteria are clearly defined and can be easily anticipated. Criteria are also modifiable, particularly where conditions such as management goals or activities to be applied to selected chemicals change. Criteria can also be adjusted for local, regional, or other geographic conditions. Perhaps as important as identifying candidates for management activities from a pool of existing chemicals, however, is the use of criteria to allow assessment of new chemicals to determine whether and how a new chemical may be managed. For example, an industry may choose to modify a new chemical before its extensive production or use where that chemical would be classified as a Sunset candidate via existing screening criteria. In this sense then, criteria are

highly important for planning purposes in industry as well as for governmental agencies, and for grass roots and other advocacy organizations.

REFERENCES

Foran, J.A. 1991. The sunset chemicals proposal: Managing hazardous chemicals through the OECD. Internat. Environ. Affairs, 2(4):303-308.

International Joint Commission (IJC). 1983. Water Quality Board Report to the Commission.

International Joint Commission, (IJC). 1992. Sixth Biennial Report on Great Lakes Water Quality. IJC, Washington, D.C., Ottawa and Windsor, Ontario, Canada. 59pp.

Mackay, D. 1991. Multimedia Environmental Models. Lewis Publishers, Chelsea, MI. 257 pp.

Wahlstrom, B. 1989. Sunset for dangerous chemicals. Nature 341:276.