

# RISK-BASED PRIORITIZATION SCHEME FOR HAZARDOUS WASTE CHEMICALS USING MODELED EXPOSURE AND TOXICITY ESTIMATES

Chandrika J. Moudgal<sup>®</sup>, ORD, NCEA, Cincinnati, OH and Raghuraman Venkatapathy, ORISE, DOE, Oak Ridge, TN

\* Contacting author; address: NCEA, USEPA, MS-117, 26 W. Martin Luther King Dr., Cincinnati, OH 45268, e-mail: Moudgal.Chandrika@epa.gov; Phone: (513) 569 7078; Fax: (513) 569 7475

EPA Science Forum

Healthy Communities and Ecosystems

## Introduction

- The Solid Waste Disposal Act (1965; 42 USC Sec. 6901-6991), as amended by the Resource Conservation and Recovery Act (RCRA) of 1976 (42 USC Sec. 6921-Subtitle C of SWDA; 40 CFR 260) and the Hazardous and Solid Waste Amendments (HSWA) of 1984 (amended SWDA/RCRA), required the U.S. Environmental Protection Agency (EPA) to institute a national program to improve the solid waste management system in the U.S.
- The HWIR-media rule established two new regulatory designations for contaminated media that would otherwise be subjected to regulation under the RCRA Subtitle C regulations. The two designations "above the bright line" and "below the bright line" distinguished media that would remain in Subtitle C and media that would be eligible for exemption from Subtitle C regulations. The rule replaced a set of generic, national management standards for waste cleanup with a more flexible regulatory framework under RCRA. This enhanced the ability of regulators to select more appropriate and cost effective remedies based on site-specific conditions at a wide variety of sites while still keeping the effects to human health and environment to a minimum (1).

## Research Objective

- The objective of this project was to develop a ranking scheme for screening and prioritizing large groups of chemicals based on modeled estimates of toxicity and exposure. This method was applied to approximately 220 HWIR chemicals, most of which lacked experimental toxicity data.
- The objective was achieved by applying a Classification and Regression Tree (CART) approach that grouped chemicals based on similarities in their exposure and toxicity potentials.

## Methods

- Exposure Estimation:** The CalTOX (2) multimedia model was selected for analyzing the chemical-specific exposure distributions for 220 HWIR chemicals. CalTOX is a spreadsheet model that can be used to estimate exposure and risk due to inhalation, ingestion, and dermal absorption of chemicals by humans.
- Non-cancer toxicity Estimation:** A commercial QSTR software, TOPKAT<sup>®</sup> (3), was used to obtain non-cancer toxicity estimates of chemicals on the HWIR list. Only chemicals that had valid toxicity estimates were selected for the final ranking.

- Risk:** The output of the model (and primary ranking metric) was referred to as a "relative ranking Hazard Quotient" (rrHQ). The rrHQ model is given by Equation (1)

$$rrHQ = \frac{U_{Exp} \cdot Q_{TOT}}{LOAEL} \quad (1)$$

here:

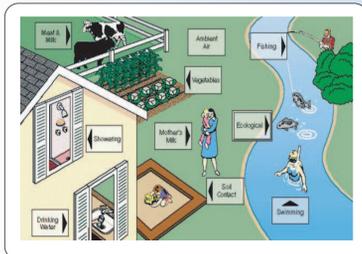
$U_{Exp}$  = unit oral exposure estimate (per unit soil concentration) in mg/kg-day

$Q_{TOT}$  = total quantity of substance in hazardous waste sites in the U.S. in Kg.

$LOAEL$  = rat chronic LOAEL estimate representing a chronic toxicity exposure level in mg/kg-day.

### Occurrence

For determining QTOT in Equation 1, the total mass (in kg) of chemical waste was taken from the 1996 NHWCS data base (1), which covered 1760 waste streams from 156 Large Quantity Generators (LQGs) and Treatment, Storage, or Disposal Facilities (TSDFs).

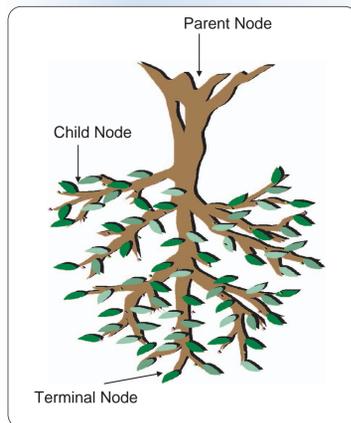
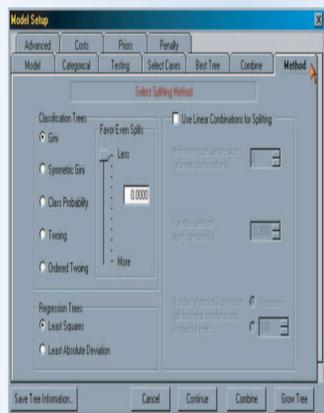


## Chemical Classification

- The CART methodology was used to rank chemicals based on their hazard potential.
- CART is technically known as binary recursive partitioning. The process is binary because parent nodes are always partitioned into two child nodes and recursive because the process can be repeated by treating each child node as a parent node.
- The primary elements of a CART analysis are a set of rules for (a) splitting each node in a tree, and (b) deciding when a tree is complete.
- The CART methodology is incorporated into a software named CART (4).

## CART Menu Options

- Model Tab- Select "Regression" under tree type; target and predictor variables under variable selection
- Method Tab- Select "least absolute deviation" under regression tree
- Testing Tab- "Choose V-fold cross validation" and set V=50.
- Best Tree Tab- "Select one Standard Error (SE) rule"
- Advanced Tab- "Minimum number of parent nodes was 1 and minimum number of child nodes set to 2"
- All other menu options were set to their default values.



## Exposure CART Variable Selection

- Depending on the correlation between the target variable (log exposure) and predictor variables (physicochemical parameters from CALTOX), the variable with a higher correlation (Pearson's coefficient, r) value between the target variable and either the variable or its log value was chosen.

The following variables were chosen as descriptor variables for the final exposure-CART analysis

- molecular weight (log(MW))
- octanol-water partition coefficient (log(Kow))
- melting point (Tm)

- vapor pressure (log(VP))
- solubility (log(S))
- Henry's Law Constant (HLC)
- diffusion coefficient in pure air (Dair)
- diffusion coefficient in pure water (Dwater)
- organic carbon-water partition coefficient (log(Koc))
- reaction half-life in air (log(ta))
- reaction half-life in ground-surface soil (log(tg))
- reaction half-life in root-zone soil (log(ts))
- reaction half-life in vadose-zone soil (log(tv))
- reaction half-life in groundwater (log(tq))
- reaction half-life in surface water (log(tw))

## Exposure CART Data Set

- For generating the exposure CART, initially, all CalTOX (2) database chemicals and 222 HWIR chemicals were considered excepting the following: inorganic chemicals, chemicals with missing parameter values for any one of the 16 chemical parameters (in CalTOX), and those that had an undefined log value for any one of the descriptors, resulting in a total of 419 data points.
- The initial data set contained some compounds that had actual/experimental data for some of the parameters while others in the data set contained parameters that were estimated using EPIWIN software.
- To eliminate errors introduced due to the incorporation of modeled data, only chemicals or data points that used actual or experimental data were considered in generating the final CART. Epiwin was used to model the missing parameters. Model estimates for half lives are based on rules that estimates the value for a number of chemicals and thus not reliable for use in CART.
- After eliminating chemicals that contained/used modeled data, there were 278 chemicals left in the CART training data set.

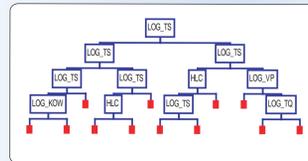
## rrHQ CART Variable Selection

- The model options for generating the rrHQ CART were similar to the one used in exposure CART except for the following: In the model tab (CART menu options), hazard quotient log(rrHQ) was chosen as the target variable while exposure (log(exp)), occurrence (log(Kg)) and toxicity (log(LOAEL)) were chosen as the dependent variables.

## rrHQ CART Data Sets

- For generating the rrHQ CART, of the roughly 500 HWIR chemicals obtained from Office of Solid Waste (OSW), chemicals with the highest occurrence and a valid LOAEL value were considered.
- Three of the HWIR chemicals (carbazole, maleic anhydride, and triethylamine) were eliminated because they had very low exposure estimates (log(exp) < 261).
- Deuterated toluene was removed because a valid LOAEL estimate could not be obtained.
- The final number of chemicals in the data set used to generate the rrHQ CART was 218.

## Results and Discussion



## Relative ranking Hazard Quotient (rrHQ) CART

- The rrHQ CART approach classifies chemicals into risk-based categories or groups + based on similarities in their occurrence, exposure and toxicity potentials.
- The objective of using this approach was to obtain maximum discrimination among the risk-based categories by generating a tree with a maximum number of non-overlapping terminal nodes at the highest possible confidence.

In generating the HQ CART, four scenarios were followed:

- point estimates were used for exposure and HQ;
- point estimates were used for HQ, but average estimates from the exposure CART were used for the exposures;
- point estimates were used for exposure and average exposures were used to calculate HQ; and
- average estimates were used for HQ and exposure.

## Exposure CART

- CART generated using 278 chemicals (data not shown).
- Used Least Absolute Deviation (LAD) regression tree.
- Optimal number of nodes chosen such that there was no overlap between the exposure bounds among the nodes.
- Exposure bounds for a particular node were calculated as the Average exposure of all chemical in a node ± standard deviation.

Table: Split criteria for the chemicals in the exposure dataset.

Node	Mean	SD	# of Chemicals	LB	UB	Split criteria
1	-88.36	0.000	1	-88.36	-88.36	$\log(T) \leq 0.270$ $\log(K_{ow}) \leq -0.525$
2	-69.43	4.74	2	-74.17	-64.69	$\log(T) \leq 0.270$ $\log(K_{ow}) > -0.525$
3	-56.51	3.57	7	-60.07	-52.94	$0.270 < \log(T) \leq 0.470$
4	-38.71	0.0968	2	-38.81	-38.61	$HLC > 21582.6$ $0.470 < \log(T) \leq 0.711$
5	-34.17	1.07	22	-35.25	-33.10	$0.0144 < HLC \leq 21582.6$ $0.470 < \log(T) \leq 0.711$
6	-31.42	1.19	12	-32.61	-30.23	$HLC \leq 0.0144$ $0.711 < \log(T) \leq 0.895$
7	-21.99	1.92	22	-23.91	-20.07	$0.711 < \log(T) \leq 0.895$
8	-17.83	0.53	3	-18.36	-17.30	$HLC \leq 0.09$ $0.895 < \log(T) \leq 0.967$
9	-15.39	1.21	26	-16.61	-14.18	$HLC > 6.085$ $0.967 < \log(T) \leq 1.42$
10	-12.15	1.58	30	-13.74	-10.57	$HLC < 6.085$ $0.967 < \log(T) \leq 1.42$
11	-6.69	1.98	150	-8.67	-4.71	$\log(T) > 1.42$

## TOPKAT<sup>®</sup> LOAEL Predictions, CalTOX Exposure Estimates and calculated Hazard Quotients:

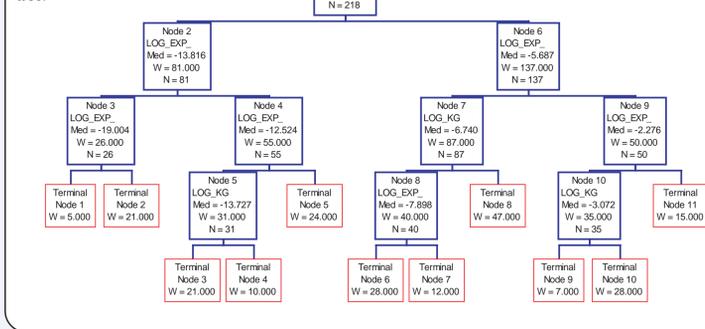
- Toxicity estimates ranged 5 orders of magnitude, while exposure estimates ranged 32 orders of magnitude. rrHQs were calculated using Equation (1). Average exposure estimates were determined using classification by the Exposure CART. Average rrHQs were calculated by substituting average exposure instead of exposure estimated by CalTOX in Equation (1).

Scenario (a): Point estimates for both HQ and exposure

- An overlap was observed in the 20 node tree that was initially generated.
- A new tree with 11 nodes was generated for the 218 chemicals in the HWIR dataset.
- This 11 node tree was pruned to obtain an eight node tree by combining five of the overlapping nodes. Nodes 3 and 4 from the 11-node tree were combined to form a single node while nodes 7, 8 and 9 were combined into a single node.
- The resultant tree had non-overlapping nodes at 50% confidence interval.

- 5 of the 8 nodes are split based on exposure values alone (nodes 1-4, 8; total number of chemicals in these nodes = 96)
- They contain approximately 45% of the chemicals in the data set.
- The remaining 55% of the chemicals are split based on both exposure and occurrence.
- None of the chemicals in the data set were split on toxicity values, thus implying that toxicity did not play an important role in ranking the health risks associated with these chemicals.
- All chemicals in nodes 1 through 6 of the 8-node tree have a hazard quotient value much lesser than one, and are hence not considered to pose a risk to human health.
- All chemicals in node 7 also have hazard quotients less than one except for hydrocyanic acid. Hence, a majority of the chemicals in node 7 are not expected to pose a risk to human health.
- Most chemicals in node 8 have hazard quotients greater than one, thus implying that they may pose a risk to human health. The chemicals in this node are split based on exposure values alone

Figure. Tree containing 11 nodes for scenario (a). The 11-node tree was pruned to generate the final 8-node tree.



## Conclusions and Recommendations

- The exposure component of the rrHQ models completely dominated the relative rankings; the toxicity and occurrence components contributed very little. The exposure dominance was due almost entirely to the extremely broad range of modeled exposure values, which in turn were strongly influenced by specific environmental half-lives. The model currently gives each of the three variables equal weight. Qualitatively, there would be somewhat greater confidence in the modeled toxicity estimates than in the modeled exposure estimates, and, on the surface, the greatest confidence in the measured occurrence data. However, the form in which the occurrence data were available, diminishes their potential utility.

- These models can be used for prioritizing the highest risk unassessed HWIR constituents for further research.

- The models could be used for identifying the most likely constituents for possible exemption (delisting) by focusing on the lowest risk groupings. In this case, additional information would serve to justify removal of chemicals from hazardous waste regulation. The ranking models, themselves, would probably not be adequate for issuing exemptions, but could be used for stimulating data submissions.

- In a broader context, the models are applicable to any scenario in which chemical concentrations are measured in soil. Thus, the ranking models could be used for preliminary assessment of uncontrolled waste sites where a large number of chemicals might be found. The models could be run under site-specific scenarios for direct relevance to a given situation. The intended use would be to provide an initial indication of risk-drivers at a site, but would not be suitable for "elimination" screening. That is, the chemicals predicted to have low risk should not be dropped from consideration, but should still be assessed with other methods.

- For even broader application, the exposure component can be modified to include other "release" modes, such as combustion or surface water. The latter would be particularly useful for application to development of the Contaminant Candidate List under the Safe Drinking Water Act (SDWA).

## References

- U.S. EPA. 1996. Environmental Fact Sheet: Hazardous Waste Identification Rule for Contaminated Media (HWIR-Media)-Proposed Rule. EPA/530/F-96/010.
- McKone, T.E. and K.G. Enoch. 2002. CalTOX<sup>™</sup>, A Multimedia Total Exposure Model Spreadsheet User's Guide, Version 4 (Beta). Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory, Berkeley, CA.
- Accelrys. 2001. TOPKAT users guide 6.1. Accelrys Inc. Burlington, MA
- CART<sup>®</sup> (Classification and Regression Tree) 128 Pro. 2003. Version 5.0, Salford Systems, San Diego, CA.

Mention of trade names or commercial products does not constitute endorsement or recommendation for use.