Methods Activity Group

Status Report for the CCL Work Group

Plenary Meeting

May 12, 2003

The Methods Activity Group has had three conference calls since the March plenary mtg.

Participants of Work Group *(not all participants on all calls) SARA CHECK*

- Laura Anderko
- Douglas Crawford-Brown
- Mike Dourson
- Alan Elzerman
- Brian Ramaley
- Colin Stine
- Craig Stow
- Lynn Thorp
- Dan Wartenberg
- Tom Carpenter and other EPA staff
- Jo Anne Shatkin and other Cadmus staff
- Steve Via, AWWA
- Jeff Rosen and Dave Drain, Perot Systems Gov't Services
- Abby Arnold and Sara Litke

Progress toward Deliverables

- At the March 27meeting the CCL Work Group reviewed the gate approach to screen from the CCL universe of contaminants to the PCCL.
- The activity group has refined the gate approach and today seeks review and comment by CCL Work Group.

Progress toward Deliverables

Gate Approach to screen from CCL Universe to PCCL

- Methods Activity Group proposes a gate approach
- Gates are “paths only from the universe to the PCCL” for both microbes and chemicals
  - Knowledge base for microbes is forthcoming

Overview of Gate Approach

Assumptions:

- CCL universe defined by data group, includes all contaminants for which we are able to gather data or information.
- CCL universe may change over time
- Not every contaminant in the CCL universe will be characterized by having data directly reporting its known or potential adverse health affects and its known or potential occurrence.
- Some contaminants may be characterized by surrogate data, such as production data, or by other indirect measures.

Overview of Gate Approach, (cont.)

- Four gates describe the criteria for passing through any of the gates (correlate with terms “demonstrated” and “potential” as used by NAS)
- However, activity group goes beyond words “demonstrated and potential”
- Definition of terms:
  - Demonstrated
    - real measured data on which knowledge rests are available. For health effects, this includes toxicology study endpoints from human or animal studies. For occurrence, data means water concentration data.
  - Potential
    - information on the contaminant or on a surrogate contaminant that is suggestive of, or generally correlates well with, a specific effect or measure of occurrence.

Overview of Gate Approach, (cont.)

- Appropriate gate for a contaminant will be a function of the nature of the knowledge about that contaminant.
- Contaminant that does not pass through a gate remains on CCL universe, those that go through a gate, are listed on the PCCL.

Overview of Gate Approach, (cont)

- Four Gates:

I. Quantitative data or measures of adverse health effects and quantitative data on concentrations in water.

II. Information that there may be adverse health effects and quantitative data on concentrations in water

Overview of Gate Approach

III. Quantitative data or measures of adverse health effects and information that suggests there may be significant presence or potential to occur in water.

IV. Information that suggests there may be adverse health effects and information that suggests there may be significant presence or potential to occur in water.

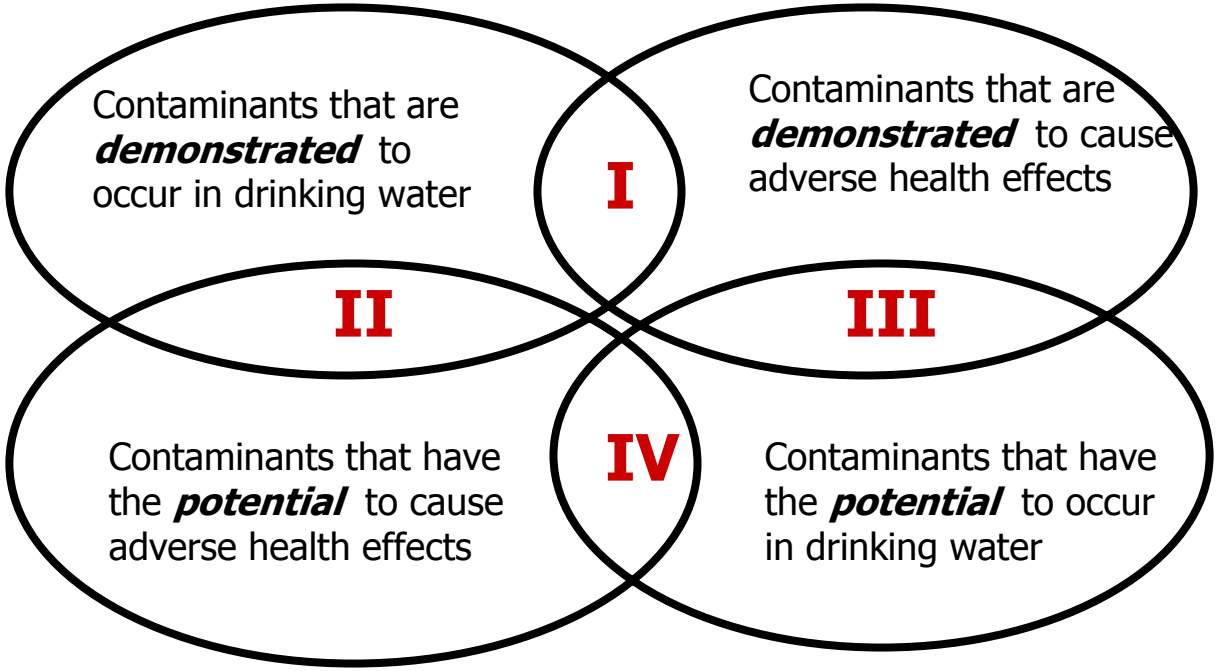
Overview of Gate Approach

V. Nomination process

A nomination gate that allows a contaminant to move to the PCCL because experts think it ought to

VI. Possible other gates?

The Universe of Potential Drinking Water Contaminants



Gate I - Quantitative data or measures of adverse health effects and quantitative data on concentrations in water

Gate II - Information that suggests that there may be adverse health effects and quantitative data on concentrations in water

Gate III - Quantitative data or measures of adverse health effects and information that suggests there may be significant presence or concentrations in water

Gate IV - Information that suggests that there may be adverse health effects and information that suggests there may be significant presence or concentrations in water.

Two Methods for Gates

- Semi Quantitative
  - Bin effects into several categories reflecting different potencies (e.g., high, medium, low)
  - Bin exposures into several categories reflecting different concentrations (e.g., NAS 1 to 10 magnitude choices)
  - Combine bins to establish a score
  - Determine cut off score for PCCL
- Quantitative
  - Estimate reasonably maximum potency by Table 1
  - Estimate reasonably maximum exposure by Table 2
  - Compare exposure to potency
  - If ratio is greater than a certain value (e.g., 1), then contaminant is on the PCCL

Table 1. Uncertainty factors used in development of potencies*

Type of toxicity data	Composite factor	Confidence in Estimate
Sensitive human NOAEL	1	High
Average human NOAEL	10	High
Experimental animal only	100	Medium to high
Less than lifetime study	1000	Medium
Lack of a NOAEL	3000	Medium to low
Insufficient studies to determine critical effect	10,000	Low
Lethal dose data only	100,000+	Not applicable
95% UCL of NOAEL based on QSAR	3000	Not applicable

* Choice of uncertainty factor, composite uncertainty factor, and confidence are as defined by the EPA, except for lethal dose and structure data. Values for these latter two categories can be found in the literature.

Table 2. Uncertainty factors suggested for concentration data for the Universe to PCCL*

Type of data	Composite factor	Confidence in estimate
99 percentile of tap water measurements	1	High
Mean of tap water measurements	10	Medium to High
99 percentile of source water…	3	Medium to High
Mean of source water…	30	Medium
99 percentile of ambient water…	10	Medium to low
Mean of ambient water…	100	Low
Detection limit…	3	Not applicable
Maximum concentration based on structure (information)		Medium

* Choice of uncertainty factor, composite uncertainty factor, and confidence are wild guesses for discussion purposes only. They should be based on a comparison of available data for these measurements.

Gate Approach

- Next steps:
  - Detailed review and discussion by the CCL Work Group and ultimately, approval of draft gates approach
  - Test the proposed approach

Progress toward Deliverables: Model for classifying from PCCL to CCL

- Deliverable: Recommended decision method and associated prototype approach(es) for classifying from the PCCL to the CCL
  - The group has reviewed a paper “Model Fit to Example Data” to understand the model classes and identify whether a preference could be determined at this point.
  - An attribute scoring approach to be used on an example data set to test various approaches.
  - Reviewed example data set and test various approaches with raw and scored data.
  - Next Steps: Identify process to select approach.

Progress toward Deliverables (cont.) Classification Models (PCCL-CCL)

- Purpose of model review was to understand how the models work and what inferences can be drawn about the models.
- The models were evaluated based on how well they replicated previous decisions. (list/no list)
- Through a cross validation analysis, a mean classification rate was calculated for each of the models, using raw and scored data.
- Multivariate Adaptive Regression Splines (MARS) had the lowest misclassification rate for both raw and scored data.

Progress toward Deliverables (cont.) Classification Models (PCCL-CCL)

- Participants noted however, that misclassification could be caused by anomalies in the data set and/or original decisions made in building the first CCL, rather than due to the models themselves.
- The group will continue to explore all four classes of models using a larger and more robust training data set.
- The Models reviewed:
  - Artificial neural networks (ANN)
  - Classification and regression trees (CART)
  - Logistic regression (a specific form of a generalized linear model (GLM) and
  - Multivariate adaptive regression splines (MARS)

Progress toward Deliverables
Classification Models (PCCL-CCL)

- Next Steps:
  - EPA is building a data set to fit the four classes of models

Attributes

- N. Kim provided overview of NRC rationale and thinking behind attributes.