Accounting for Certainty/Confidence in Attribute Scoring

Report for the NDWAC CCL CP Work Group

September 17, 2003

Overview

- CCL Classification Process recommended by NRC requires data for attributes:
  - contaminant occurrence
  - adverse health effects

- Many CCL contaminants are relatively unknown, emerging, or new -
  - Data availability and quality will vary
  - Different types of data/data elements (surrogates) will have to be used to represent the attributes for different contaminants

Overview

- Attribute scoring process is an approach to
  - "normalize" the different types of data elements
  - assign an attribute score for each data element from its own calibrated scale for that attribute
- Scoring approach needs to address
  - differences in data quality (some implied in data element hierarchy)
  - that some data and scores will have higher level of certainty/confidence than others
- NRC-NDWAC -- the scoring approach should
  - avoid complex rule-making
  - be based on the data

Overview

- NDWAC CCL CP Workgroup posed – should some indication or measure of the level of certainty/confidence be captured in the process?
  - Various issues and options have been discussed to account for varying levels of certainty/confidence
- This discussion is not dealing with quantitative or statistical measures of uncertainty or variance

Rather, it focuses on

- NDWAC's concern to express expert judgement of certainty/confidence because of the nature or quality of the data used for scoring

Perspectives

- Certainty/Confidence – a Paradox
  - Group discussion has suggested you lower the Attribute score because the data/data element was of lower quality
  - In Risk Assessment often err on the side of caution. Under some circumstances might raise the score
    - Example, if a contaminant scores quite high (i.e., it may be of significant concern because of its high *potential* for occurrence and health effects), but there is also low c/c in the data, an expert judgment might be to place it on the list because of the uncertainty that it might be a bad actor, instead of "lowering" its score and not listing it.
  - Statistical analysis shows that either approach are appropriate for different data.
  - Do you lower or raise the score because of lower certainty/confidence?
  - Or does that depend on the results for all attributes, and/or whether the score is high or low?

Perspectives

- BIAS
  - As a component of evaluating certainty/confidence, it has been suggested that bias might be specifically evaluated as well.
  - In addition to, or as part of, a certainty/confidence score, for example, can/should a bias indicator (a directional score) be used?
- Biased Studies
  - Related bias issue is how to handle biased, targeted study data
    - For example: the results of a local, targeted water quality monitoring study may only have "worst-case" results from a very small area
  - Scoring protocols cannot be designed to handle *every unique* case
  - Unique, biased studies/data will likely have to be dealt with on the parallel track of expert review, such as part of the evaluation of data sources

Perspectives

- Prototype Classification System
  - CCL is a judgement process
  - Not a rigorous numerical process
  - A classification or sorting process
  - Simplicity vs. complexity? Transparency
  - Misleading appearance of precision?
- Whatever approach considered must be accommodated in calibration and training of model
- Certainty/Confidence concerns inherent in the process
- Some components of certainty/confidence are linked to the data source quality

Options

- 1) Include certainty/confidence factors in scoring data
- 2) Assign 5 separate certainty/confidence attribute scores to the data
- 3) Assign 1 combined certainty/confidence attribute score to the data
- 4) Assign separate certainty/confidence “flags” to the data for each attribute
- 5) Ignore certainty/confidence at this stage of the process (Attribute Scoring)

Options

- 1) Include certainty/confidence factors in scoring data
  - A weighting or adjustment factor could be included in computation of the score for the attributes.
  - This approach is analogous to setting weighting factors in a rule-based system
    - Would require further expert opinion to establish weights
    - Difficult to preset rules for every situation (e.g. whether to lower or raise attribute score by factor)
    - Incorporating adjustments in score may obscure transparency
    - Approach may still slant outcome of processing by identifying contaminants with less certain scores over others, without clearly identifying the adjustment in the end result

Options

- 2) Assign 5 separate certainty/confidence attribute scores to the data
  - The certainty/confidence score could be treated as a separate measure for each contaminant for each attribute and be processed in the algorithm
    - In essence, this creates a companion certainty/confidence “attribute”
    - Doubles the number of attributes and their actual use/affect, as half of the variables, in a prototype model is not clear.
    - Each uncertainty score may increase the size of the required training set by a factor of 3 (more or less). At some point, training becomes infeasible.

Options

- 3) Assign 1 combined certainty/confidence attribute score to the data
  - The certainty/confidence values for each attribute could be summed, or averaged, into one composite certainty/confidence value for each contaminant
    - Only adds one “attribute”
    - Use/affect in model still not clear
    - Would not differentiate confidence of individual attributes, which becomes important since attributes will not likely be weighed equally in the process.

Options

- 4) Assign separate certainty/confidence “flags” to the data for each attribute
  - The certainty/confidence scores could simply be stored and carried in the system (as “flags”) and evaluated at the end by EPA/experts when the resultant classification has been completed.
    - as noted by the Methods Activity Group, output from any prototype model will require some level of expert review in the final analysis
    - certainty/confidence scores could provide some additional information for review of the outcome of the classification processing

Options

- 5) Ignore certainty/confidence
  - Certainty/confidence is inherent in the process
    - much of the data that will be used with upcoming CCL contaminants will lack certainty; accept that fact?
  - Not a regulatory determination; a classification process to aid a decision whether or not to list the contaminant
  - Records will be kept on the information used to develop the scores that could always be evaluated at the end of the process
    - As discussed, some (expert) review of the model output would be needed at the end of the process.
    - the final review of the top contenders could further evaluate the data used, and the c/c of the data, as part of the final decision
    - There would not be a need to deal with uncertainties of contaminants that are graded far below the top contenders.

Workgroup Findings

- 1) Include certainty/confidence factors in scoring data
  - Incorporating adjustments in the score obscures transparency
- 2) Assign 5 separate certainty/confidence attribute scores to the data
  - Doubles the number of attributes and increases the size of the training set – too complex
- 3) Assign 1 combined certainty/confidence attribute score to the data
  - Does not differentiate confidence of individual attributes
  - May obscure transparency
- 5) Ignore certainty/confidence
  - Certainty/Confidence not accounted for in CCL process but could be reviewed before listing or in regulatory determination process
  - Not favored

Workgroup Findings

- 4) Assign separate certainty/confidence “flags” to the data for each attribute
  - Certainty/Confidence not accounted for in scoring and algorithm but instead, “flagged” for review by experts
- Further assess bias concerns
  - Biased/unique studies/data dealt with at beginning of process; parallel track of expert review, evaluation of data sources