National Drinking Water Advisory Council (NDWAC) Contaminant Candidate List (CCL) Classification Process Work Group

July 16-17, 2003
Washington, DC

Meeting Summary

-Final-

Table of Contents

Attachments

The sixth meeting of the NDWAC CCL Classification Process Work Group was held on July 16-17, 2003. The meeting objectives were to

- Review example CCL Universe data set
- Review results of AWWA Workshop on data
- Review draft documents
  - Building the CCL Universe
  - Screening from CCL Universe to the PCCL
  - Classifying from the PCCL to the CCL
  - Transparency/Risk Communication
  - Applications of Genomics to the CCL Process
  - Draft Glossary
- Review recommendations from discussion of attributes
- Decide on next steps for completing the report to the NDWAC
  - Identify questions and issues the work group needs to address
  - Agree on tasks to be conducted between July and September 2003

Welcome and Introductions

Facilitator Abby Arnold, RESOLVE, welcomed everyone to the meeting (see attachment A for list of work group members in attendance). Following introductions, the work group reviewed and approved the meeting agenda (see attachment B).

Review of Tasks Conducted Since Last Meeting

Tom Carpenter, EPA, presented an overview of the work conducted since the May work group meeting (see attachment C).

Current Proposed Definition of the CCL Universe for Microbes

A member of the Microbe Activity Group presented the group's proposal for defining the CCL universe for microbes (see attachment D). The challenge with microbes, he explained, is that no databases exist. He commented that though there are known to be a very large number of microbes in the world, those of concern for the CCL process are only those dangerous to humans and found in water. The task of identifying these microbes has been simplified somewhat by the publication of a list of 1415 species of infectious organisms known to be pathogenic to humans identified by L.H. Taylor, et al. ("Risk Factors for Human Disease Emergence," *Philosophical Transactions: Biological Sciences*, 2001).

The Microbe Group proposed to consolidate the CCL universe and the PCCL into a single list of microbes. The process would begin with the list of 1,415 recognized human pathogens, and from that list, those microbes that are known to be associated with source water, recreational water, or treated drinking water would be selected for inclusion in the PCCL.

The proposed steps for constructing the microbial PCCL are as follows:

1. Edit the list of known human pathogens (Taylor et al 2001) to add missing pathogens and to eliminate those whose biological properties are incompatible with water transmission by ingestion, inhalation or dermal contact.

2. Compile a list of microbes that occur in water, and that are known or have the potential to cause adverse health effects in humans, and reconcile this list with the edited list of known human pathogens.
3. Monitor CDC/WHO surveillance data and information for new and emerging pathogens associated with water.
4. Establish a mechanism for adding microbes to the CCL universe based upon genomic or proteomic information suggestive of emerging pathogenicity and biological properties that may facilitate waterborne transmission.
5. Establish a nomination process for expeditious addition of pathogens to the CCL universe based upon new data or information provided by stakeholders or experts.

The group member noted that step 4 begins to move toward a process we are not currently ready to use that would rely more and genomics and proteomics.

Work group members expressed general support for the proposed process and decide to incorporate the appropriate sections of "Defining the Microbial Universe" into chapters 5 and 6 of the draft report to the NDWAC.

Example CCL Universe Data Set of Chemicals

Jo Anne Shatkin, Cadmus Group, reported on the progress of preparing and analyzing the example CCL Universe data set of chemicals (see attachment E). The purpose of the data set is to support the work group process, to test the approaches being considered by the work group, and to gain insights about available data and identify data elements, issues, and gaps. Twenty-three data sources were chosen for inclusion in the data set based on three criteria: high-quality data, easily accessible electronically, and contained information relative to building the CCL Universe. Dr. Shatkin explained how the data from the twenty-three sources were compiled and linked and how the data elements were classified as occurrence data, occurrence information, health effects data, health effects information, or other, using the definitions of "data" and "information" proposed in the draft approach for screening from the CCL Universe to the PCCL.

Dr. Shatkin reported that of the 10,360 contaminants in the data set, 774 (7.5%) had health effects data or information *and* occurrence data or information. These 774 contaminants, therefore, could be tested at one of the proposed gates for screening from the CCL Universe to the PCCL. Of the 10,360 contaminants, 6,245 have no data or information on either health effects or occurrence.

Using the 62 contaminants with health effect data and occurrence data, Dr. Shatkin illustrated how various criteria comparing occurrence levels to health effect levels might be used to determine which contaminants pass through gate I onto the PCCL.

In closing, Dr. Shatkin shared some insights gained from developing and analyzing the example data set. Her comments included the following:

- Data availability is better for health effects than for water occurrence.
- Additional gate I contaminants could be identified using other natural water surveys (e.g., USGS National Water-Quality Assessment Program (NAWQA)).

- Additional data sources may provide additional data elements that could be helpful for attribute scoring and classification, though these sources may require manual manipulation and judgment.

The group discussed what lessons and conclusions could be drawn from the example data set and analysis. Members acknowledged that because the data set is only an example it should not be used to draw extensive conclusions. Rather, it can help to identify remaining decisions and judgments that need to be made.

Comments from individual members included the following:

- The proposed methods for screening and classification should not be judged based on whether or not they capture the contaminants on the first CCL. The work group needs to decide how to evaluate proposed methods.
- Several of the contaminants that were on the first CCL but are not captured in the example data set are pesticide degradates. The group should discuss possible ways to capture types of contaminants that are missed.

Report on the American Water Works Association (AWWA) Workshop

Jeff Rosen, Perot Systems Environmental Services, reported on the workshop held by AWWA to explore the data sources identified through the work group process (see attachment F). The workshop objectives included the following:

- Review the information resources to identify relevant data sources (expand the characterization of the data sources);
- Identify data elements and each information source that can be used to score the attributes;
- Evaluate the logistics of using these information sources to score the attributes;
- Evaluate the metadata available for these information sources;
- Identify specific fields and each data source that contain the required data elements.

At the workshop about twenty experts from a range of fields reviewed the data sources to evaluate their potential usefulness for moving from the PCCL to the CCL, specifically for scoring the attributes. Workshop participants added 12 additional sources to the list already identified in the CCL Work Group process.

Mr. Rosen listed about 20 data sources judged by workshop participants to be useful and easily accessible. He also shared some of the observations, conclusions and recommendations made by participants, including the following:

- Most of information that will be most useful to the CCL process is in text format.
- Characteristics of available sources need to be explored in greater depth using the best available research and extraction technology (not wed interfaces).
- Automation of data location, extraction, manipulation, and attribute scoring faces significant challenges. Expert participation and judgment will be needed.
- Significant automation is an important goal for future CCLs. It can best be met through ongoing expert data development.

- Better understanding of the planned attribute scoring is needed in order to evaluate data elements availability. The logistics of combining data elements will depend on the planned attribute scoring.
- Review of available information should focus on the logistics of extraction, normalization, and attribute scoring.
- Review of the available information sources should include identification of significant data gaps.

Report on the July 15 Working Session

Several work group members participated in a working session on July 15 to review in more detail the example CCL Universe data set and analysis and the products from the AWWA workshop. These work group members reported on the July 15 session (see attachment G). They observed that the findings from the example CCL Universe data set and the AWWA workshop are consistent relative to data availability and extraction. They suggested that the group should begin discussing data quality. They also identified several questions for the work group to address. As a next step they proposed a workshop to conduct an example attribute scoring exercise. Prior to the workshop, the technical/facilitation team would extract a sample data set. At the workshop the team would report on lessons learned and challenges related to data extraction. Work group members would then participate in the scoring exercise, evaluate the challenges and issues, and synthesize lessons learned to begin to develop recommendations for the NDWAC.

The work group decided to go ahead with the workshop as part of the September 17-18 plenary meeting. The group identified tasks to prepare for the workshop and to continue to explore data sources and develop the example CCL Universe data set (see attachment H).

Classifying from the PCCL to the CCL: Attribute Scoring

Chemical Contaminants

Dr. Shatkin reported on the three conference calls on the attributes for chemical contaminants (see attachment I). The conference call participants decided to recommend that the work group proceed with the five National Research Council (NRC) attributes and the NRC definitions of those attributes. Overall lessons and conclusions that emerged from the calls included the following:

- Consistency in scoring (i.e., getting a good range of scores among the data) is more important than assuring that the scales are normally or lognormally distributed for classification models.
- Scales should be integers; any higher resolution (e.g., using two or more significant figures) implies that we know more than we do.
- Models require using the same attributes for all contaminants, even if there are data gaps.

Dr. Shatkin summarized lessons and conclusions on potency, severity, and prevalence and observations and remaining questions related to magnitude and persistence/mobility.

Following the presentation, a member expressed concern about considering the treatability of a health effect in determining severity scores, given that some people do not have access to health

care. Another member commented that in some cases the NRC approach considered whether an effect was reversible or nonreversible. Members agreed that severity should be included among the attributes, but further discussion of the scoring scheme is needed.

Microbial Contaminants

Nelson Moyer, Cadmus Group, presented an overview of the scoring schemes being discussed by the Microbe Activity Group (see attachment J). He noted that of the five attributes, the most information is available for severity. The activity group is considering a weighted scoring protocol that takes population susceptibility categories into account. For potency the group is considering a three tiered scoring approach using primary, secondary, and tertiary data elements. The scheme for persistence-mobility is still in the early stages of discussion, and the group has not yet focused on prevalence or magnitude.

A member asked whether the weighted scoring scheme being considered for severity would result in a lower score for a microbe that causes an effect only in a sensitive subpopulation. Dr. Moyer responded that the weighting scheme is a way to normalize information in order to include sensitive subpopulations.

A member noted that "documented outbreak" was listed as a primary data element for potency scoring and asked whether it would serve better as a data element for occurrence. Another member commented that little information on potency is available for microbes, so the group is considering possible ways to estimate potency. Dr. Moyer explained that the rationale is that if a microbe has survived in water systems to cause an outbreak, it is probably fairly potent.

Members of the Microbe Activity Group stressed that schemes are still in draft stage and under discussion. The group will continue developing and refining the schemes to prepare for the September plenary meeting.

Next Steps

As noted above, the work group decided to have a workshop on September 17 to further explore possible attribute scoring schemes and related data issues (see attachment H). In preparation for the workshop, the Microbe Activity Group will continue developing and refining the schemes to prepare for the September plenary meeting. Several other members will work with the technical/facilitation team to develop the scoring schemes for chemicals and identify the critical issues to address at the workshop. Where possible, alternative approaches for addressing the issues will be presented at the workshop for work group members to explore.

Classifying from the PCCL to the CCL: Evaluating Classification Approaches

The group briefly discussed next steps for evaluating classification approaches. In addition to the attribute scoring schemes, a data set is needed to train, or calibrate, the models. The group decided to begin to review the list of contaminants for the training data set at the September meeting. A member commented that the training set should include about 200 contaminants, some that would be on a CCL and some that would not. He noted that if the training set includes microbes, the work group can test how well the models work with both chemicals and microbes. He explained that a given model will be evaluated on whether it is equally successful for microbes and chemicals; subtle differences in the thought processes for scoring microbes and chemicals will not matter if the success rate is equal. The next steps for evaluating classification approaches are outlined in attachment H.

Chapter 3: Transparency and Risk Communication

The work group reviewed the draft outline for chapter 3 of the report to the NDWAC and discussed some general issues of transparency and risk communication (see attachment K). A member commented that the approach does not have to be simple, it has to be transparent. He explained that transparency requires that caveats are clear, the language of the report is clear, and the concepts of the recommended approach are understandable even to those without advanced degrees in the technologies being recommended. Another member added that the group should not underestimate the ability of interested members of the public to learn and understand technical issues and processes.

The work group discussed the audience for the recommendation report. Members agreed that the primary audience is the NDWAC, and EPA through the NDWAC. The broader audience for the report includes a range of interested parties, such as the constituencies of work group members, elected officials, policy decision makers, owners of small systems, and members of the public who might attend information sessions about the CCL and the CCL process.

Members offered several suggestions for preparing the recommendation report:

- The NRC recommended that the public should be involved in helping to define the problem and articulate the questions. The report should include a discussion of the work group process and how the recommendations were developed (i.e., how members were selected, the extent to which other members of the public were involved through public comment at meetings or member outreach).
- The group needs to keep the audience in mind as it drafts the report.
- The report should include a full discussion of uncertainties and explain that the work group did its best to develop the best possible process, building on the NRC recommendation to move beyond contaminants that are easy to identify and classify.
- Each chapter should include a discussion of uncertainties, issues, and limitations.
- The report needs to explain the task and the process without creating undue concern among the public.
- The report needs to explain the context and ultimate goals of the CCL process.
- All technical terms should be defined in the glossary.
- Use diagrams, examples, and stories to help illustrate points.
- The report should be clear up front and throughout as to what the work group's intent was.

A member commented that the group also needs to consider transparency from the technical perspective. He suggested that in selecting among possible classification models, the group may want to give consideration not only to which model gives the optimal results, but also to how transparent and easy to explain each model is.

The work group agreed that matters of transparency and public involvement related to the work group process will be discussed in the introductory chapters of the report, and chapter 3 will provide recommendations of what EPA could do to be transparent if the recommended CCL process is adopted. Several members volunteered to revise the outline for chapter 3 and draft sections of text for the work group to review at the September meeting.

Chapter 6: Screening from the CCL Universe to the Preliminary CCL (PCCL)

The group reviewed the chapter 6 draft (see attachment L). A member explained that he drafted the possible foreword to the chapter to provide context and to explain what it means for a contaminant to be on the PCCL and what it means for a contaminant to be on the CCL. Members commented that the draft foreword presented several important concepts clearly and understandably. The group decided to include the piece, revised as necessary, in the introduction or other earlier section of the report. A member suggested a specific revision: define "validated data" or choose another term for validated, since validation has a specific meaning in statistics. Another member commented that definition B on page 2 should say ". . . may result in contaminants being placed on the CCL" rather than ". . . will result in the contaminant being placed on the CCL."

The group discussed possible criteria for screening chemical contaminants from the CCL Universe to the PCCL. The chapter 6 draft outlines three options: qualitative, semi-quantitative, and quantitative. Table 1 summarizes which options could be applied to which gate, as proposed in the chapter 6 draft, based on whether "data" or "information" is available for the contaminants at the gate.

Table 1: Criteria options for gates I-IV.

Gate	What we have: Data (D) or Information (I)	Possible Criteria
	Occurrence	Health Effects
I	D	DQualitative, Semi-quantitative, Quantitative
II	I	DQualitative, Semi-quantitative
III	D	IQualitative, Semi-quantitative
IV	I	IQualitative, Semi-quantitative

Members noted that for the majority of contaminants, data or information, as currently defined, will not be available for both health effects and occurrence. They acknowledged that the challenge is both to improve and expand the data available and to determine a method that can be used if and when data are not available. A member suggested that the group should consider the less limiting option of screening contaminants that have data or information on either health

effects *or* occurrence. Another member noted that using surrogate information to estimate health effects and occurrence could greatly expand the number of contaminants that can be considered at the gates. She suggested that the group should explore ways to estimate effects and occurrence based on surrogate information before deciding whether to require data or information on health effects or occurrence or both. A member suggested that quantitative structure activity relationships (QSARs) could be used to estimate lethal dose levels and solubility for chemicals without other data or information, and a comparison of lethal dose level and solubility could then be used as a screen.

The group discussed the option of using qualitative criteria for all of the gates (i.e., if a contaminant simply has data or information on health effects and occurrence it passes on to the PCCL). Some members suggested that at the PCCL stage, more rigorous criteria may not be necessary. Another member pointed out to that while qualitative criteria would provide a relatively simple screen, they may result in an unmanageable number of contaminants on the PCCL.

A member asked where consideration of sensitive populations would come into the proposed screening approach and how epidemiological measures such as odds ratios might be used. She suggested that for health effects, measures of severity could be used rather than or in addition to the proposed measures of potency.

A member summarized that there are two options for increasing the number of contaminants to be screened: 1) require contaminants to have data or information on health effects or occurrence rather than both health effects and occurrence, or 2) expand the definition of usable information.

A member questioned how much the screening approach will rely on expert judgment in the end. He suggested that if the recommended process is essentially one of expert judgment, the work group should not mask it with overly analytical calibrations.

Next Steps

The group decided that additional analysis is needed to inform its decisions on the various issues raised (see attachment H). The analysis will evaluate 1) the feasibility of using QSARs to estimate the lethal dose levels and solubility, and 2) the feasibility of using solubility and lethal dose levels to rank or bin contaminants for the PCCL.

The group also discussed the terminology used in the chapter 6 draft and decided to revise the language to clarify terms such as data, information, potential, and known and make them more consistent with standard English usage.

Chapter 8: Applications of Genomics to the CCL Classification Process

Dr. Moyer summarized that chapter 8 recommends several applications of genomics for the CCL process and discusses the limitations, possible timeframe, and preparatory steps for incorporating those applications (see attachment M). The work group members agreed that they were comfortable with including the chapter in the report to the NDWAC.

Public Comment

No members of the public expressed an interest in making comments to the work group at this meeting.

Next Steps

See attachment H for a detailed list of next steps.

Future Meetings

The remaining work group meetings in 2003 are listed below. The group decided to reserve dates for two additional meetings in early 2004 in case they are necessary. RESOLVE will collect members' schedules and identify possible dates. It is expected that all meetings will be held at the RESOLVE offices.

- September 17-18, 2003
- November 13-14, 2003