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

May 12-13, 2003
Washington, DC

Meeting Summary

-Final-

Table of Contents

Attachments

The fifth meeting of the NDWAC CCL Classification Process Work Group was held on May 12-13, 2003. The meeting objectives were to

- report and provide feedback on activity group activities to date;
- review and agree on
  - gate approach for screening from the CCL universe to the PCCL,
  - data guidelines;
- identify questions and issues the work group needs to address;
- agree on tasks to be conducted to prepare for subsequent meetings;
- review work plan in light of activity group progress and evaluation of tasks;
- determine activity group tasks and conference call schedule between May 12-13 and July 16-17, 2003.

Welcome and Introductions

Facilitator Abby Arnold, RESOLVE, welcomed everyone to the meeting and asked the work group members and other meeting participants to introduce themselves (see attachment A). Following introductions, the work group reviewed and approved the meeting agenda (see attachment B).

Work group member Ed Thomas, National Rural Water Association (NRWA), shared with the work group a document outlining NRWA's policies, comments, and priorities related to the CCL (see attachment C).

General Approach for Moving Forward

Through the course of its discussions at this meeting, the work group decided on a general approach to tasks to prepare for the July 16-17 meeting. The group decided to 1) begin developing initial drafts of some sections of the recommendation report for the topics on which there has been sufficient discussion to define options and 2) continue discussions by conference call for other topics. Specific next steps for various tasks and issues are outlined in the relevant sections below.

Draft Table of Contents for the Work Group's Report to the NDWAC

Ms. Arnold presented a draft table of contents prepared by the facilitation team for the work group's review. With some revisions, the group decided to move forward with the table of contents as a working draft (see attachment D).

Members discussed the audience for the report. A work group member who also serves on the NDWAC commented that the report should be sufficiently scientifically oriented to allow the NDWAC to pass it with confidence, but it should also be understandable to the public. Another member suggested that the report could include a technical appendix.

Several members commented on incorporating microbial issues into the report. A member noted that the group should clearly communicate upfront that the report addresses both chemicals and microbes. Another member commented that to the extent genomics can be used in the near term, the group's recommendations related to genomics should be incorporated throughout the report

rather than reserved to a separate section at the end. He added that a discussion of genomics technology that is not yet feasible for use in the CCL process and recommendations for further research could appropriately be a separate section in the report.

A member commented that the report should include a discussion of limitations and caveats where appropriate. For example, it should explain that the CCL universe was not defined as "everything" but was built with practical considerations in mind, which has implications for the process. Another member added that the report should clearly explain what it does and does not mean for a contaminant to be included in the CCL universe. A third member generalized that the report should include a discussion of the implications and trade-offs of the various decisions the group makes in its recommendations.

A member noted that since developing its guiding principles the group has had little discussion focused explicitly on transparency and risk communication. Several members offered to begin drafting the report section on transparency and risk communication to prompt the group's discussion. The work group will consider whether to bring in additional expertise to assist in this area.

A member commented that any recommendations the group makes regarding research should relate directly to the CCL screening and classification process. He noted that although research may be needed on myriad contaminants, recommending such research is beyond the scope of the work group.

Updates and Proposals from Activity Groups

Data Activity Group

Building the CCL Universe

A member of the Data Activity Group presented an update on the group's work since the March 27-28 plenary meeting and an overview of the draft document "Next Steps – Continuing Discussion to Describe Guidelines for Building the CCL Universe" (see attachments E and F). He reminded participants of the key points from "Building the CCL Universe," which the work group reviewed at the March 27-28 meeting. Based on the principles presented in that document, the activity group is exploring a four-stage process for building the CCL universe:

1. compile the most relevant data sources on contaminants with known or potential occurrence in drinking water or known or potential health effects
2. compile less directly relevant data sources
3. use additional data sources to fill data gaps associated with stages 1 and 2
4. use surrogates to fill important information gaps

In stage 1, some of the data sources selected would be used for lists of contaminants, and others would also provide data elements. Examples of possible data sources include the National Contaminant Occurrence Database, High Production Volume Chemical lists, and the World Health Organization's Drinking Water Quality Guidelines.

In stage 2, some of the data sources selected would be used for lists of contaminants, and others would also provide data elements. Examples of possible data sources include the National

Sediment Inventory, Occupational Safety and Hazard Administration (OSHA) work place hazard information, and data sources on ecological endpoints. EPA would filter relevant information from these data sources. For example, ergonomic hazard information from the OSHA data sources would not be included.

In stage 3, data sources that are not directly relevant to the occurrence and health effects principles would be used to provide data elements and fill data gaps, but they would not be used to add contaminants to the CCL universe. For example, the High Production Volume Chemical Lists provide a list of contaminants in stage 1; in stage 3, the chemical abstract services database might be used to provide solubility data for those contaminants. In stage 4, appropriate surrogates would be used to model or estimate information to fill remaining data gaps. For example, quantitative structure-activity relationship (QSAR) modeling might be used to estimate solubility.

EPA staff presented a diagram to illustrate the proposed process (see attachment G).

Discussion

Some work group members commented that it may not be necessary for stages 3 and 4 to be separate, as both stages involve filling data gaps. A member observed that both data from various data sources and information estimated from surrogates can have uncertainty, depending on the specific data element. A member responded that the distinction between stage 3 and stage 4 is not specifically about uncertainty. Rather, the activity group proposes separating the task of filling data gaps into two stages in order to indicate that EPA would first try to fill gaps with directly measured data and then try to fill remaining gaps by making inferences from surrogates. Another member added that separating the step of using surrogates to estimate information helps to make it transparent that an additional tool is being used in the process.

A member questioned whether "known contaminants" as defined in the "Next Steps – Continuing Discussion" document would encompass contaminants that would be included in the CCL universe based on surrogate information. He suggested that a separate definition of "known contaminants" is not necessary when "contaminants," "emerging contaminants," and "new contaminants" are defined. Another member suggested changing "known contaminants" to "known and potential contaminants."

Some members questioned the distinction between stage 1 data sources and stage 2 data sources and noted that "relevant" is a value-laden term. A Data Activity Group member explained that the group had not defined relevant beyond "relevant to occurrence in water or health effects." A member observed that the distinctions between data sources in the various stages may be the same as the distinction between "data" and "information" proposed by the Methods Group in the approach for screening from the CCL universe to the PCCL. Several members commented that the work group should use consistent terminology throughout the process. A member suggested using the term " primary data" for data that reflect a direct measurement of occurrence in water or a direct analytical conclusion regarding a health effect and "secondary data" for data that can be used to estimate or derive health effects or occurrence parameters. Another member questioned whether such a distinction is necessary at the universe stage. He asked whether a hierarchy of data was being proposed. A member responded that using the terms "primary" and

"secondary" (or "data" and "information") is not intended to imply that one kind of data is inherently better than the other.

A member noted that an important question remaining for the work group to address is how to assess the adequacy of the CCL universe. Another member suggested that the work group may not need to reach a decision on what constitutes an adequate CCL universe. Instead the work group could decide on a process to recommend – including something similar to the proposed four-stage process along with the surveillance and nomination processes discussed at the March 27-28 meeting – and rely on the process to result in an adequate CCL universe.

A member suggested that the work group consider using surrogates to build the whole universe rather than just to fill data gaps in a final stage. He suggested that surrogate measures for chemicals and microbes might be used as an early screen for the "universe of everything" to identify contaminants on which to focus. If a relatively simple but reliable screen such as this could be developed, it would allow the process to begin with the more broadly defined universe than the proposed CCL universe. Also, it might allow a more efficient use of resources by reducing the number of contaminants *before* extensive data gathering efforts were undertaken. A member commented that this approach may be just as likely to miss an important contaminant as the approach proposed by the Data Group. Another member questioned how EPA could determine whether something had been missed without having data on the contaminants.

A member questioned how the data sources would be searched to create the CCL universe data set, to create the PCCL data set, and to create the CCL data set. Noting that there are no "health effects" databases or "occurrence" databases per se, he questioned how the process to create the data set for each stage would differ from the other stages. He acknowledged that the questions being asked at each stage differ, but observed that the data set to answer the questions may be the same. He commented that if "compile data sources" in the proposed four-stage process for building the CCL universe means developing a computer program to pull all the data from the various sources together, the result of the process will be the data set needed for the PCCL. Another member responded that because the questions asked to screen from the CCL universe to the PCCL focus on health effects and occurrence, EPA would need to upload fewer data into the CCL universe data set. Then with fewer contaminants at the PCCL stage, EPA would upload more data to apply to the wider range of considerations (the attributes) to classify from the PCCL to the CCL. Another member added that for the fewer contaminants at the PCCL stage, more rigorous data quality considerations could be applied. The original member responded that while the distinctions sound reasonable conceptually, the distinction on a practical implementation level remains unclear. An EPA staff member observed that any database EPA creates is subject to high degree of scrutiny, which would make it very difficult to construct a detailed CCL universe for tens of thousands of contaminants. He commented that if the work group recommends a process that requires such a universe, it is possible EPA would not be able to implement the process completely as recommended. He suggested that it would be helpful for the work group to articulate the principles and objectives of what should be achieved with the CCL universe and with the PCCL so that if EPA needs to alter the approach it can be guided by those principles and objectives.

A member of the facilitation team explained that EPA has begun to develop an example CCL universe data set based on 23 data sources. The Methods Activity Group will use the data set to further explore possible types of models for classifying from the PCCL to the CCL. Developing the example data set, however, also may provide insights to answer some of the work group's questions about the proposed approach for building the CCL universe.

A member noted that few databases exist for microbes and suggested that the work group should define what information is needed for the CCL process, identify possible sources, and suggest a method for assessing the information.

Next Steps

The work group decided that Rick Becker and Wendy Heiger-Bernays would begin to draft four sections of the report to the NDWAC: 5.1. Fundamental Principles for Building the CCL Universe; 5.2. Guidelines for Selecting Data Sources to Build the CCL Universe; 5.5. Process for New and Emerging Contaminants; and 5.7. Process for Updating the CCL Universe. To write the draft sections, Dr. Becker and Dr. Heiger-Bernays will use the "Next Steps – Continuing Discussion" document (with revisions based on the work group's discussion), "Building the CCL Universe," and other work group documents.

The work group decided to continue exploring data sources and data to build a CCL universe, but also keep open options, such as using surrogate measures more extensively, especially for microbes. Discussion of issues will continue on "CCL Universe" and "Microbes/VFAR" conference calls as outlined in the call schedule below.

Methods Activity Group

<u>Draft Proposed Universe to PCCL Screening Process – The Gate Approach</u>

A member of the Methods Activity Group presented an update on the group's work since the March 27-28 plenary meeting and an overview of the document "Draft Proposed Universe to PCCL Process" (see attachments H and I). Noting that the work group reviewed the draft universe to PCCL screening process at the March 27-28 meeting, he explained that the activity group had refined the approach and was again seeking work group comments. The document "Draft Proposed Universe to PCCL Process" focuses on the knowledge base for chemical contaminants; a proposal for incorporating microbial contaminants is forthcoming.

The Methods Group proposes a set of parallel "gates" to screen from the CCL universe to the PCCL. The activity group is developing criteria for each gate, and a contaminant would pass from the CCL universe to the PCCL by satisfying the criteria for any one gate. The criteria correlate to the NRC terms "demonstrated" and "potential." For the gate approach, the group has defined "data" as measured values that reflect adverse health effects or occurrence in drinking water and "information" as anything used to estimate or derive parameter values for health effects or occurrence. Using these definitions, "demonstrated" will mean there are data on which the knowledge of a contaminant rests, and "potential" will mean there is information on the contaminant or a surrogate contaminant that is suggestive of, or generally correlates well with, a specific effect or measure of occurrence. The proposed descriptors of the four primary gates are then as follows:

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
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

These four gates correspond to the areas of intersection of the four ovals in the NRC Venn diagram (*Classifying Drinking Water Contaminants for Regulatory Consideration*, NRC 2001, page 82). The appropriate gate for a contaminant will be a function of the nature of the knowledge about that contaminant. A contaminant that does not pass through its appropriate gate remains in the CCL universe; a contaminant that does pass through is listed on the PCCL.

The group proposes a fifth gate that would be a nomination process to allow a contaminant to move to the PCCL because experts believe it ought to. The group also is leaving open the option of defining additional gates.

The activity group has begun to define options that could be used as criteria for determining whether a contaminant moves from the universe to the PCCL. The options can be described as qualitative, semi-quantitative, and quantitative. Acknowledging that the group has done the least thinking about the qualitative option, the member outlined the group's ideas thus far on the options:

- Qualitative: Determine whether enough data or information exists for a candidate; if so, it passes through the gate onto the PCCL. For this option, the question asked is whether the necessary data are present, not whether they indicate concentrations likely to produce health effects.
- *Semi-quantitative:* Bin health effects into several categories reflecting different potencies (e.g., high, medium, low). Bin exposures into several categories reflecting different concentrations (e.g., NRC's 1 to 10 magnitude categories). Combine effects and exposure bins to establish a score. Determine a cut-off score for inclusion on the PCCL.
- *Quantitative:* Estimate maximum potency using toxicity data that are adjusted by uncertainty factors to normalize among different types of data. Estimate maximum exposure using concentration data that are adjusted by uncertainty factors to normalize among different types of data. Compare exposure to potency: if the ratio is greater than a certain value (e.g., 1), then the contaminant is included on the PCCL.

The member noted that one drawback of the semi-quantitative option is that it would not, for example, distinguish between water and dioxin. If a scale of 1 to 10 were used for both health effects and exposure, water would score 1 in effects and 10 in exposure, and dioxin would score 10 in effects and 1 in exposure. Combining effects and exposure, each of the two substances would receive the same overall score of 10. The member commented that a disadvantage of the quantitative option is that it is more labor intensive, but it would distinguish between water and dioxin.

The member presented two tables as an example of how composite uncertainty factors could be used to standardize among different types of toxicity and concentration data. He stressed that the composite factors for concentration data are "wild guesses," included in the table only for discussion purposes.

Discussion

Members discussed which would be the appropriate gate for a contaminant that had both data and information. A member suggested initially that the data would take precedence. Noting, however, that for some contaminants, the information may be more persuasive than the data, the group decided that contaminants with both data and information could be tried through two gates. A member also noted that some contaminants in the CCL universe would not be considered for any of the four primary gates, since making it to a gate requires having information or data for health effects *and* information or data for occurrence. Another member noted that contaminants without both kinds of data/information could be nominated through gate V.

A member commented that it may not be possible to obtain concentration measures for microbial contaminants. Microbe occurrence is usually measured categorically as "present" or "not present." The member also observed that unlike most chemicals, which tend to have fairly constant concentrations where present, concentrations of microbes can vary be orders of magnitude. He commented that because of a range of complicating factors with microbes (e.g., variations and changes in concentrations, presence of sensitive humans), if a pathogen is detected, generally it is considered to be of concern. He added that due to problems with culturing and with detecting small amounts of microbes, occurrence data do not exist for most microbes.

Several members voiced support for not pursuing the quantitative option. Some commented that for screening from the CCL universe to the PCCL, a quantitative method is not necessary, and it would require too much extra effort for too little extra value. Another member asked whether anything would be lost in using the semi-quantitative option rather than the quantitative option. Referring to table 1 of types of toxicity data and uncertainty factors, he noted that for most contaminants, the data available will be based on QSARs. He questioned how the quantitative option, based mostly on QSAR information, would be better than the semi-quantitative option. A member responded that if quantitative data are available, the process should not use less than the quantitative data; the data should not be converted to scores because information would be lost in the conversion, which may result in missing some contaminants that should be on the PCCL. Another member observed that the problem faced with a semi-quantitative approach is a dynamic range. He suggested that the group could use the quantitative approach to determine the dynamic range, and then use that range in the semi-quantitative approach.

A member suggested using the quantitative option when possible, and the semi-quantitative option otherwise. Another member suggested that the quantitative option could be used for gate 1, where both effects data and occurrence data are available. A member commented that using a

combination of the quantitative option and the semi-quantitative option makes the process more cumbersome and resource intensive.

A member observed that selecting an option is largely a matter of deciding what the group is willing to accept in terms of false positives (i.e., including contaminants on the PCCL that do not need to be included) and false negatives (i.e., missing contaminants that should be included on the PCCL). Another member observed that at the universe to PCCL screening stage, false negatives are the primary concern. He suggested that the work group test the semi-quantitative option to determine if it results in more false negatives than the quantitative option.

A member suggested another semi-quantitative option using three scores each – high, medium, low – for both health effects and occurrence. Contaminants scoring high in occurrence and high in effects would be included on the PCCL; contaminants scoring low in occurrence and low in effects would not be included; and so on as indicated by "yes" and "no" in table 1 below. The member commented that whether to include contaminants that fall in the boxes marked "?" would need to be determined, possibly with consideration to the desired size of the PCCL. He noted that an advantage to this semi-quantitative option is that the scoring bands for health effects are wider: the difference between an animal NOAEL and a human NOAEL is less of a concern because both would be captured in the band. Another member commented that the key questions with this approach would be where the scoring bands are bound and how different combinations of scores are prioritized (e.g., high potency and low occurrence versus low potency and high occurrence). Another member suggested that the approach would need to incorporate data quality considerations.

Table 1: Example Semi-quantitative Option

		Potency (µg or organisms/ liter)		
		High	Medium	Low
Concentration (µg or organisms/ liter)	High	yes	yes	?
	Medium	yes	?	?
	Low	?	?	no

A member questioned how a computer could be "taught" to perform the semi-quantitative option. He explained that a computer can search for high concentrations in terms of high micrograms per liter. However, what the work group really means by "high concentration" is "high enough to make someone sick"; the definition of "high" must be tied to a health effect, which will be different for different contaminants. Another member commented that quantitative data are not available for most contaminants, so if the work group cannot determine a way to use a semi-quantitative approach will not work, screening from the CCL universe to the PCCL cannot be automated. A member responded that a qualitative or categorical approach is still an option. He added that a qualitative or categorical option may be necessary also because for many contaminants, data will not be available to allow even a semi-quantitative option to be used.

Next Steps

The work group decided to continue to develop the gate approach for screening from the CCL universe to the PCCL. Further discussion is needed on the details (pass/fail criteria; qualitative/categorical, semi-quantitative, and quantitative options). As the group desires and as

time allows, these issues will be discussed on "CCL Universe" calls. Issues also will be discussed at the July 16-17 meeting. The VFAR/Microbe Group will continue to explore additional approaches for screening microbial contaminants.

Mike Dourson and Doug Crawford-Brown will revise the "Draft Proposed Universe to PCCL" document, incorporating points discussed at this meeting, to form the beginning of chapter 6 of the draft report to the NDWAC.

Classifying from the PCCL to the CCL

A member of the Methods Activity Group presented an update on the group's work on classification models since the March 27-28 meeting (see attachment H). Preliminary runs were performed with four classes of models (artificial neural networks (ANN), classification and regression trees (CART), logistic regression, and multivariate adaptive regression splines (MARS)) using example data to help the group understand how the different classes of models work and what inferences can be drawn about them. The models were run with both raw and scored data, and the analysis results were summarized in the paper "Model Fit to Example Data."

The models were evaluated based on how well they replicated previous decisions to list of not list contaminants. Through a cross validation analysis a mean classification rate was calculated for each of the models. MARS had the lowest misclassification rate for both raw and scored data, and ANN had the highest misclassification rate. The member noted that ANNs perform well with very large data sets, so one reason why it faired worst in this analysis is that a small data set was used. He also explained that all of the models reviewed are pattern recognition models. They do not "learn" or seek a fundamental truth; they simply discern patterns and mimic what has been done with the training data set.

In reviewing the analysis results, activity group participants noted that misclassification by the models could be caused by anomalies in the data set rather than inherent shortcomings in the models themselves. The activity group will continue to explore all four classes of models (and possibly others) using a larger and more robust training data set.

Noting some of the problems identified with the example training data set used in this analysis, a member noted that an important question for the work group is what should be used for the real training data set when EPA implements the classification process.

Next Steps

The work group will continue to explore various model types for classifying contaminants from the PCCL to the CCL. Between now and the July 16-17 meeting, the group's discussion on PCCL to CCL classification will focus on the attributes (see attributes section below). EPA and the technical consultants will continue developing the example training data set to test the model types.

VFAR/Microbe Activity Group

A member of the VFAR/Microbe Activity Group presented an update on the group's work (see attachment J). He reported that several pilot projects are underway and preliminary data should

be available soon. EPA and the technical consultants are continuing the nucleic acid and protein sequence searches of plasmids and pathogenicity islands. Based on the searches thus far, activity group members believe "islands" of genes look promising as a VFAR candidate.

The member commented that genomics may be helpful in determining some of the attributes for microbes, such as prevalence and potency. He reported that the activity group aims to provide both recommendations on how to incorporate genomics into the CCL process in the near term and recommendations on a research strategy to pursue VFAR techniques for the longer term.

Next Steps

The group will prepare a discussion paper on the potential of using genomics and the potential of VFAR techniques. The group also will continue its discussions of defining the universe for microbial contaminants, defining the attributes and data elements for microbial contaminants, and incorporating the use of genomics into the attributes.

Attributes

Tom Carpenter, EPA, presented a summary of the April 25th conference call on attributes (see attachment K). He noted that work group members from all three activity groups participated. On the call, Nancy Kim reviewed the approach she used to score the attributes for chemicals for the NRC process. She summarized the discussion of NRC committee members, her data evaluation, and difficulties encountered with scoring. Mr. Carpenter noted that Dr. Kim did the scoring for the NRC committee on her own free time, while the work group has technical resources it can use to further explore some of the attribute issues.

Observations made by Dr. Kim and other conference call participants included the following:

Potency – "How much of a contaminant causes illness?"

- The scoring approach worked fairly well. Data were available for most contaminants.
- Scoring based on no observable adverse effect levels (NOAELs) or lowest observable adverse effect levels (LOAELs) avoided dependence on uncertainty factors.
- Nutrients may need different scoring approach from xenobiotics.

Severity – "How bad is the health effect?"

- Severity was scored based on the most sensitive health effect (i.e., the same effect as used to score potency).
- Most chemicals scored high the first time, so the approach was reevaluated.
- Scores may need to be adjusted where severe effects occur above the LOAEL.

Prevalence – "How commonly does a contaminant occur in water?"

- The committee intended prevalence to include temporal and spatial aspects, but temporal and spatial data are sparse. The scoring approach used population exposed and number of detects to derive prevalence with water data, or used production data.
- The approach used a hierarchy of data types.

- The preference would be to use percent of detects rather than number of detects, but data on number of samples are not available to determine percents.
- Detection limits decrease over time and could affect detection frequency; this needs to be addressed.

Magnitude – "What is the expected concentration relative to the level causing a health effect?"

- The approach used median detections among detects only.
- The work group may want to consider using other statistics.
- The issue of redundancy was raised over using the potency attribute in two scores. A suggestion was made to link magnitude to severity.
- The work group should consider what the added value of magnitude is.

Persistence/Mobility – "What is the likelihood that a contaminant will be found in the aquatic environment?"

- Persistence/mobility was scored based on amplification, solubility, and stability (average x 10/3).
- Naturally occurring chemicals may need a different scale from xenobiotics.

Questions for further work group discussion include the following:

- Does the work group agree that the five attributes identified by the NRC are the correct attributes? If not, why not, and is there an alternative?
- If the group agrees that these are the correct attributes, does the group agree with the NRC definition of each? If not, what is an alternative definition?
- What data elements should be used? Which data elements should be considered "data" and which should be considered "information"? What is the hierarchy of data elements?
- Should raw or scored data be used?

Discussion

A member asked whether the concept of a safe dose applies to microbial contaminants. Another member commented that the potency issue and determining the level of risk associated with a particular dose are not difficult for microbes. He suggested, though, that the question the work group should seek to answer is "how should we approach potency for microbial contaminants?" rather than "how can we develop a measure of potency for microbes similar to the measure for chemicals?"

A member commented that the work group should propose an approach that presents the data and allows the policy makers to make decisions based on the data. Another member responded that using uncertainty factors is not a matter of policy, it is a part of science and the nature and source of data. A member commented that an explanation should be provided of how the data were developed (e.g., what uncertainty factors were applied, what scoring approach was used).

A member observed that an important scientific consideration is how many significant figures the data support. He commented that therefore, scientifically, it may be better to score the

attributes rather than to use raw data. Another member added that scoring provides a consistent form of input for future use of the model if the kinds of available data change.

The work group agreed to pursue a scored approach for the attributes. Through the discussion, the group continued to identify issues and questions to be addressed to develop a scoring approach, building on the issues and alternatives outlined by EPA in two tables: "NRC Attributes, Potential Data Elements, Issues, and Alternatives" and "NRC Attributes, Potential Data Elements, Issues, and Alternatives for Microbes" (see attachments L and M).

A member suggested that for microbes, occurrence of a waterborne disease outbreak should be sufficient information to place the responsible microbe on the CCL, without needing first to determine a potency score. Another member suggested that rather than specifying a waterborne disease outbreak as a reason to bypass the process, the group should make sure that the process it recommends captures microbial contaminants that cause waterborne disease outbreaks (as well as chemicals that cause waterborne incidents). A member noted that information on an outbreak would be epidemiological data rather than toxicological data. She commented that the group will have to consider whether and how to incorporate epidemiological data into the attribute scoring approach.

The group discussed some of the practical aspects of scoring the attributes. A member asked whether the scoring process could be automated. Some members responded that it could, though one member noted that for data poor contaminants, the only option may be to have experts do the scoring. A member commented that the ability to automate the process is an important consideration, since one of the group's fundamental questions is how to effectively develop a CCL from the thousands of contaminants on the PCCL. Another member noted that the alternative to having people score the attributes is having people search for data, which may be more resource intensive. A member suggested that the work group could test the proposed model to determine its sensitivity to small variations in the attributes. If it were found to be very sensitive, the group could modify the approach to make it more robust.

A member again raised the question of whether the PCCL is a necessary step if a CCL universe database is built and the classification procedure is automated. A member responded that the CCL universe likely would not include as much data for each contaminant as would be included at the PCCL stage. Another member noted that the concern about false negatives and false positives changes from one step to another: false negatives are the primary concern in building the universe and screening to the PCCL, while the concern about false positives increases at the CCL stage.

Members generally agreed that the five attributes defined by the NRC are the right attributes with which to proceed, though further discussion is needed on some details and on a scoring approach. A member commented that EPA will not have full knowledge for all – or perhaps any – contaminants, so the process should include all five attributes and use partial knowledge. Another member noted the concerns raised in previous discussions that the definition of potency is based on the concept of threshold dose, which may not apply to microbes and all chemicals. She commented that if the concept does not apply to microbes, the group needs to address this issue for microbes, but overall the group should not discard the established threshold-dose

paradigm. Other members noted that further discussion is needed regarding magnitude and whether it is redundant.

Next Steps

The work group decided to continue the discussion of attributes through conference calls (see schedule below) focused on the following questions:

- Are the NRC attributes the right attributes? What are the issues for these attributes?
- Which data elements apply to which attributes?
- Do we use scores of raw data? How do we go about scoring? What is the hierarchy of data elements?

The VFAR/Microbe Activity Group will continue to discuss attributes issues for microbes.

Public Comment

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

Next Steps

Building the "CCL Universe"

- By May 21, the facilitation team will combine "Building the CCL Universe," "Next Steps – Continuing Discussion to Describe Guidelines for Building the CCL Universe," and portions of "Approach to the Development of an Example CCL Universe (Data Set)" into one document and distribute it to the work group.
- Rick Becker and Wendy Heiger-Bernays will use the combined document to prepare the discussion drafts of the appropriate recommendation report sections. (The drafts will then follow the general schedule outlined below.)
- Discussion of issues will continue on "CCL Universe" and "Microbes/VFAR" conference calls as outlined below.

Screening from the CCL Universe to the PCCL

- Mike Dourson and Doug Crawford-Brown will revise the "Draft Proposed Universe to PCCL" document to form the beginning of section 6 of the draft recommendation report (see schedule below).
- As the group desires and as time allows, universe to PCCL screening issues will be discussed on "CCL Universe" calls. Issues also will be discussed at the July 16-17 meeting.
- Discussion of screening issues for microbes also will continue on the "Microbes/VFAR" conference calls as outlined below.

Classifying from the PCCL to the CCL

- Members will contact Sara Litke (slitke@resolv.org) if they have any additional model types they would like the group to consider.
- Between now and the July 16-17 meeting, the group's discussion on PCCL to CCL classification will focus on the attributes (see conference calls below).

- EPA and the technical consultants will continue developing the example training data set to test the model types.

Discussion Draft Sections of Recommendation Report

As noted above, the work group decided to begin discussion drafts of some sections of the recommendation report. The drafts will be circulated among members for review and comment and will be discussed at future meetings. In this early stage of drafting, the documents will reflect various options and perspectives. They will frame a) the NRC recommendation, b) the work group discussion thus far, c) possible options, and d) agreement reached by work group, if any.

Schedule:

- By May 21, RESOLVE will circulate general standards for document naming and formatting, etc.
- By June 6, writers will submit drafts to Sara (slitke@resolv.org).
- By June 9, RESOLVE will distribute (post on website) drafts to full work group, along with protocols for reviewing.
- By June 19, members will submit comments to Sara. Facilitation team will work with writers to incorporate comments.
- By July 8, RESOLVE will distribute (post on website) revised drafts to full work group, reflecting the range of comments received.
- Work group will discuss drafts at July 16-17 meeting.

Transparency/Risk Communication

Primary participants: Daniel Wartenberg, Ed Thomas, Ken Merry, Benson Kirkman, Laura Anderko

Task schedule:

- By May 21, Daniel will draft outline and distribute to Ed, Ken, Benson, and Laura.
- By May 21, Sara will send EDSTAC report chapter on transparency/risk communication to the group for reference/ideas.
- By May 28, participants will comment on outline and identify portions they would like to take lead in drafting.
- By June 4, participants will send drafts to Daniel.
- By June 6, Daniel will compile into single document and send to Sara to distribute to full work group. (Draft then follows general draft document schedule outlined above.)

Draft Glossary

- By May 30, current draft of glossary will be posted on the website (click on "Document Index" and go to "G").

Conference Calls

CCL Universe (and CCL Universe to PCCL Screening, as necessary)

Primary participants: Wendy Heiger-Bernays, Rick Becker, Graciela Ramirez-Toro, Doug Crawford-Brown

Tasks:

By May 21, the facilitation team will propose a work plan for participants to review and modify on the May 27 conference call.

Call/Meeting schedule:
- Tuesday, May 27, 2003, 1:00-2:00 Eastern
- Friday, June 20, 2003, 11:00-1:00 Eastern
- Wednesday, July 2, 2003, 2:00-4:00 Eastern [or July 1, 11:00-1:00 if Graciela N/A July 2]
- In-person meeting: July 15, time TBD

Attributes

Primary participants: Brian Ramaley, Mike Dourson, Jeff Griffiths, Gary Lynch, Craig Stow, Alan Elzerman, Nancy Kim, Colin Stine

Conference Call Topics:
- *Call 1:* Summary of where the group is based on discussions thus far. Are these the right attributes for chemicals? (Issues for microbes will first be discussed by VFAR/Microbe Group.) What are the issues for these attributes?
- *Call 2:* Which data elements apply to which attributes?
- *Call 3:* Do we use scores of raw data? How do we go about scoring? What is the hierarchy of data elements?

Call/Meeting schedule:
RESOLVE will identify times for three calls based on participants' schedules.

Microbes/VFAR

Tasks:
- White paper on VFAR potential and potential of using genomics
    - EPA begins outline, with input from Jeff
    - Circulate outline among activity group participants for comment
    - Members and EPA begin drafting portions of paper
- Universe definition
- Attributes/data elements issues
- How to incorporate genomics in attributes

Call/Meeting schedule:
- Friday, May 23, 2003, 1:00-3:00 Eastern
- Monday, June 16, 2003, 1:00-3:00 Eastern

Future Meetings

The work group chose dates for meetings through 2003 as listed below. It is expected that all meetings will be held at the RESOLVE offices.
- July 16-17, 2003
- September 17-18, 2003
- November 13-14, 2003

Attachment A

Work Group Members in Attendance

Dr. Laura Anderko
Dr. Rick Becker
Dr. Douglas Crawford-Brown
Dr. Michael Dourson
Dr. Alan Elzerman
Dr. Jeff Griffiths
Dr. Wendy Heiger-Bernays
Mr. Buck Henderson
Dr. Nancy Kim
Mr. Ephraim King
Ms. Carol Kocheisen (alternate for Dr. Benson Kirkman)
Mr. Gary Lynch
Mr. Ken Merry
Mr. Brian Ramaley
Dr. Graciela Ramirez-Toro
Dr. O. Colin Stine
Dr. Craig Stow
Mr. Ed Thomas
Ms. Lynn Thorp
Dr. Daniel Wartenberg

Attachment G
Proposed Process for Building the CCL Universe