Draft Proposed Universe to PCCL Process

1.0 Background

This document summarizes alternatives for screening chemicals from the universe to a preliminary CCL (PCCL). (A similar approach is being developed for microbes, as described in a companion document). It considers two broad knowledge categories (health effects and occurrence) that can be used in conducting the Universe to PCCL screening. This document proposes a series of parallel gates that can be used to screen chemical contaminants (hereafter, contaminants). The nature of the available knowledge governs which gate is the appropriate one for a given contaminant. The gates are parallel because each contaminant may pass through to the PCCL by virtue of having satisfied the criteria of any one gate.

A "gate" is a path only from the universe to the PCCL. The path from the PCCL to the CCL will use a different method, and is evolving towards using a combination of an *a posteriori* (prototype) statistical algorithm and expert input. The PCCL to CCL processes are not discussed in this report.

In this document, there are two categories of knowledge that govern access to the gates: data and information. **Data** are defined as measured values that reflect adverse health effects or occurrence in drinking water. **Information** is defined as anything used to estimate or derive parameter values for health effects or occurrence, including, where appropriate, expert judgment. Some information may come from models (i.e. QSAR) or some other estimation process. Modeled or estimated values are generally linked to information that is readily available.

The proposed gate mechanism for screening contaminants from the universe to the CCL applies to both chemicals and microbes. This paper will primarily consider the knowledge base for the chemicals. The knowledge base for the microbes is addressed separately, and a distinct screening process for them is being developed.

2.0 Proposed Universe to PCCL Process

The "total universe" is of unknown size; thus, the data group has proposed constructing a CCL universe that will include all contaminants for which we are able to gather data or information. We assume that this CCL universe will change over time as more data and information become available. The Data activity group is defining the CCL universe using a principles-based approach, which is described in a separate document. The CCL universe will be the starting point for the screening to the PCCL. It is important to note that not every contaminant in the CCL universe will be characterized by having data directly reporting its known or potential adverse health effects and its known or potential occurrence. It is likely that some contaminants may be characterized by surrogate data, such as production data, or by other indirect measures.

As mentioned in the background section above, the Methods activity group is proposing a process whereby the chemicals in the CCL universe are screened for inclusion in the PCCL by

one of four or more parallel gates. The purpose of the gates is to select and possibly prioritize the contaminants that move from the CCL universe to the PCCL (prioritization possibly being based on which gate allows the contaminant to pass). The NAS/NRC approach depicted by the Venn diagram (page 82) provides a good basis for four of the proposed gates. The methods activity group proposal has added additional detail on process to the basic framework suggested by the Venn diagram. In the methods activity group approach, the words used to describe the criteria for passing through any of the gates correlate with the terms "demonstrated" and "potential" used by the NAS. Accordingly, the four primary gates for moving from the universe to the PCCL will allow passage of contaminants that meet the following knowledge criteria:

Gate I - Demonstrated health effects and demonstrated occurrence
Gate II - Potential health effects and demonstrated occurrence
Gate III - Demonstrated health effects and potential occurrence
Gate IV - Potential health effects and potential occurrence

The appropriate gate for screening a contaminant will be a function of the nature of knowledge about that contaminant. A contaminant that does not pass through its gate will remain in the CCL universe. Ones that pass through a gate will get on to the PCCL.

The methods activity group has gone beyond the words "demonstrated" and "potential" in their gate approach. "Demonstrated" will mean there are real measured **data on which the knowledge rests**. "Potential" means 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. Having established this distinction, the descriptors of the four primary gates become as follows:

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.

The methods activity group discussed the need for at one least extra gate (Gate V) that is a "nomination" process. The nomination gate will allow a contaminant to move to the PCCL because experts think it should for reasons that may not be fully captured in existing data, or where this information is unlikely to be captured by a mechanical process. Other gates can be added as well.

The challenge of the proposed gate-driven process is to "mechanically" screen the large number of contaminants in the CCL universe using a manageable rule-based approach in all or most

cases. Screening will be easiest if we quantify as much as possible the acceptable data and information elements for occurrence and adverse heath effects, plus establish clear pass/fail criteria for movement through a gate to the PCCL. Although the pass/fail criteria have yet to be established, it has been suggested that the real or estimated occurrence must be greater than the concentration associated with the critical adverse health effect in order for a contaminant to pass though a gate. The magnitude of the difference between the measures of occurrence and adverse health effects that will be applied in determining movement to the PCCL has not yet been determined, nor have the measures that will be used for occurrence and adverse health effects.

3.0 Gate Qualification Elements

3.1 Gate I.

The contaminants processed through Gate I are those that have data on both occurrence (meaning here concentration) and adverse health effects. We know they are there and that they can cause problems. The gate must ask the question: Is the contaminant there at high enough concentrations to be of concern? Contaminants identified in the universe are screened through Gate I only if they have health effects data and concentration data in ambient or finished water. The information must be measured data, not the results from modeling or estimation in some way.

Occurrence data may be any measurements in ambient or finished water. The quality of the data will vary, and the data selected for use during the gate screening process can be selected from a hierarchy of most to least relevant, as proposed by NRC. NRC recommended a hierarchical order of information to use in developing ⅓ universe to PCCL screening including observations from: (1) tap water; (2) distribution systems; (3) finished water; and (4) source water. Observations from (5) watersheds and aquifers, (6) historical candidate release data, and (7) chemical production data were recommended as criteria for demonstration of potential occurrence in drinking water.

The methods activity group proposes including occurrence data from (1) tap water; (2) distribution systems; (3) finished water; (4) source water and (5) watersheds and aquifers for Gate I. The anticipated data elements for occurrence data could include statistical measures from studies with measurements representing conditions in a particular water type (ground water, surface water, finished drinking water, source water, river, etc.), and from defined geographic regions (local, regional, national). The relevance of the data to estimates of national drinking water occurrence will vary and may need to be considered (e.g. measurements from ambient water in one state are different from temporally and spatially representative measurements from water system).

Consistent with NRC recommendations, the methods activity group suggests that toxicity data be any measurements from human epidemiological and whole animal studies. Anticipated data elements include risk-based values (derived endpoints including Reference Doses, Acceptable Daily Intakes, Drinking Water Equivalent Levels, etc.), no observable adverse effect levels

(NOAEL), lowest observable adverse effect levels (LOAEL), relative risks, population attributable risk, and perhaps, lethal doses (LD50) or other lower order endpoints.

3.2 Gate II.

The contaminants processed through Gate II are those that have occurrence (meaning here concentration) data and health effects information. We know they are there and suspect they may cause problems. The Occurrence data will be identical to that for Gate I, but the health effects information will be less quantitative. Types of information about toxicity that could be used for Gate II include:

- Presence on lists associated with a particular health effect (e.g. NTP Report on Carcinogens list);
- Structural or functional analogues: candidate is in a group that has been implicated for health effects;
- Threshold of Evaluation: a concentration that is not specific to a contaminant, but can be applied to all chemical candidates
- An estimated toxicity value from a model (e.g. structure activity relationship [SAR], or quantitative structure activity relationship [QSAR]).
- A combination of these, i.e. guilty by list or association in combination with threshold of evaluation/ modeled toxicity estimate.

3.3 Gate III.

The contaminants processed through Gate III are those that have health effects data and concentration information. We suspect they may be there, and know they can cause problems. The health effects data requirements for Gate III will be the same as for Gate I.

The types of information that could be used to indicate the potential for a contaminant to occur in water include: measurements in media other than water, information about use or release to the environment, historical release data; and chemical production data, information about degradation and transformation products, and physical property information. Some specific information to be considered includes:

Production data
Measurements in soil, air, sediment, organisms
Historical contaminant release data (NPDES, Inventory of Toxic Substances, etc.)
Degradation/transformation products
Known incidences of releases
Solubility data
Henry's law constant
Half-life
Structure activity relationships (SARs) for solubility and persistence
Other model estimation (e.g. fugacity)

For example, contaminants belonging to the following groups might be found in water and could use occurrence information in Gate III. The groups were identified by the NRC (2001) and include:

Drinking water treatment chemicals
Disinfection by-products
HPV (high production value chemicals)
Naturally occurring substances (including radionuclides)
Microbial agents/vectors Associated with water transmission (e.g. found in feces, nature, water treatment and distribution)
Pesticides
Leachates from collection, treatment and distribution system
Degradation/transformation products of already regulated contaminants
Reaction and combustion by-products

3.4 Gate IV

The contaminants processed through Gate IV are those that have information on adverse health effects and possible occurrence, but have no direct data. They may be there, and may cause problems. The information that will be considered in determining whether or not to process the contaminant though Gate IV will be the information discussed for health effects under Gate II and for occurrence under Gate III.

In order to be processed through Gate IV there should be some indication of a need to worry. (e.g., the candidate is likely to reach watersheds or sources and is soluble although we can't analyze for it, and it has suspected health effects but no direct data to demonstrate such effects. Most contaminants that pass on to the PCCL through Gate IV will likely do so with low confidence.

3.5 Gate V

The contaminants processed through Gate V are those that have no direct data or information on occurrence or effect, but someone has enough concern to nominate it for consideration. We do not know if it is there, and have no confirmation that there are adverse health effects when it is present, but have reason to be concerned.

If we have some information that is compelling for one of the gate criteria (adverse health effects or occurrence), that contaminant can be nominated through Gate V. For example, a chemical with a structure similar to dioxin but no data or information might be nominated for inclusion in the PCCL. If we have no data and no information for either adverse health effects or occurrence the contaminant does not get onto the PCCL.

The methods activity group has yet to develop the requirements for the nomination process.

4.0 Pass/Fail Criteria Suggestions

There are a number of options that can be used as criteria for determining whether a contaminant moves from the universe to the PCCL. The options range from purely qualitative to purely quantitative. These options can be described as follows

Qualitative. We have enough information for this candidate to include it on the PCCL, so it moves through the gate. The qualitative option simply requires determining that if one has the appropriate data, which then allows passage through the gate. However, this pass/fail criterion may let too many things onto the PCCL. It misses the opportunity to screen out things that do not belong on a PCCL, because it asks only whether necessary data are present, not whether they indicate concentrations likely to produce effects.

Semi-quantitative (Margin of Exposure). Compare the water occurrence level (concentration) and the toxicity at a gross level for Gate I. For example, how many orders of magnitude are there between the measurements or estimates of occurrence and the lowest levels at which adverse health effects occur (Note: this is analogous to concept of Margin of Exposure)? Establish some order of magnitude value criteria for going through the gate. One option might include binning of occurrence or effects data to identify or prioritize candidates. The semi-quantitative approach for Gate I offers both a comparative step, and allows for conservatism. However, it requires defining an acceptable margin of exposure for each type of data and also requires that all data be in the same units. This is still a fairly simple comparison.

The semi-qualitative Approach for Gate II could compare a water occurrence (concentration) level and some conversion of the toxicity information. This approach could include binning the health effects information (e.g. high, medium, low), or the toxicity information could be converted to an acceptable water concentration by some estimation process (See Section 5.0). For example, a candidate guilty by association could be assigned a threshold of evaluation level.

In Gate III the occurrence measure is based on information. The easiest approach is to simply screen for candidates that have the potential for occurrence. However, this does not provide much information about its potential for being in water long enough to get to drinking water, or to a consumer of water. Persistence and solubility measures can be used as an indicator of potential water occurrence, but often must be estimated, and the estimating models are better for some chemical classes than others. Solubility alone is a very coarse screen, and could let too many contaminants through that do not persist in the environment. A combination of persistence and solubility with information about levels released to the environment may provide a more robust estimate. However, there is no obvious way to convert that to a potential drinking water concentration. One could compare solubility to a toxicity level, but this may be a very coarse screen. A quantitative comparison of an estimated maximum concentration in water to a toxicity level may infer a greater level of confidence than warranted by the data.

Quantitative (Hazard Quotient). For Gate I, when the concentration in water is of the same or

greater magnitude as the concentrations associated with adverse health effects, move the contaminant to the PCCL. One could apply uncertainty factors to the raw data (see Section 5.0) for toxicity and water concentration data in making a comparison. The comparison could be a ratio of water concentration data to toxicity data based on:

- The reported measurements adjusted by uncertainty factors
- A comparison of adjusted occurrence and effects levels (Hazard Quotient)

For example, the questions can be asked: is the demonstrated (actual data) concentration (preferably a 95% Upper Confidence Level or could be maximum value) in water of the contaminant greater than the demonstrated (known) risk-based value[1](potency)? If yes, then it is on the PCCL with high confidence.

Alternatively, if the water concentration is less than the potency measure or threshold of concern then the contaminant is not on the PCCL. Something can be off the list if the maximum plausible concentration is less than a threshold of concern, or on the list if a plausible or likely maximum concentration is above the threshold of concern. Statistical and probability treatments could also be used, but the data required are probably not available.

The quantitative option provides an estimate of the potential importance of a candidate. However, it requires converting all data to the same units and creating a mechanism for accounting for uncertainty/lack of confidence in the exposure data (a system already exists for the toxicity data).

With either the semi-quantitative or quantitative methods some consideration is needed to account for non-detects (i.e. how to use method detection limit information since detection limits may vary). In addition, some consideration of the reliability of the available data for such an estimate would need to be factored in, particularly if based upon data with less relevance to drinking water exposure, e.g. ambient water measurements, LD_{50}s. The comparison in the semi-quantitative and quantitative approaches is similar; the difference may be in the data requirements, but principally is in the expression of results, either as a margin of exposure or as a hazard quotient.

In the quantitative approach for Gate II, uncertainty factors would be applied to the water concentration data and it would be compared to an estimated toxicity level. The toxicity information has to be convertible to a water concentration for this step. For Gate III, the question would be asked: is the potential plausible maximum estimated concentration in water of the contaminant greater than a demonstrated risk-based value? If a contaminant passes either Gate II or III, it continues onto the PCCL with medium confidence. There are likely to be no simple quantitative approaches that can be used for Gate IV.

[1] R**isk-based values** may be defined **as** data elements derived from a critical evaluation of toxicological and other data regarding a chemical's health effects. Risk-based values may be generally regarded as representing a protective exposure in a defined context.

5.0 Suggestions for Processing Data and Information

There are a number of options for processing data and information in order to examine the relationship between adverse health effects and concentrations in drinking water and to make decisions on movement to the PCCL. Tables 1 and 2 can be used to adjust various types of toxicity and exposure data/information for the purposes of making pass/fail judgements.

The data for health effects that originate from controlled or accidental, but known, exposure scenarios become the basis for risk-based values. The data that are generally available are no-observable-effects-levels or no-observable-adverse-effects levels (NOELs or NOAELS) and lowest-observable-effects-levels or lowest-observable-adverse-effect levels (LOELs or LOAELS). Uncertainty factors are applied to the highest NOAEL of the critical effect in order to estimate a dose that is likely to be without any adverse effect or alternately to the lowest LOAEL in the absence of a NOAEL (by definition this LOAEL would be for the critical effect).

By convention the uncertainty factors are generally applied in units of 1,3 or 10 to adjust for the following characteristics of the value that is the starting point for the calculation and the quality of the database. The default in situations where data are limited is 10 for each factor (up to a maximum of 4 factors). The individual adjustment factors can be categorized as follows:

Adjustment to account for variability among humans
Adjustment from animal data to humans
Adjustment for use of a LOAEL in cases where no NOAEL was observed
Adjustment for a less than lifetime study
Adjustment for limitations in the database.

An estimate that is produced by EPA using this approach is called a Reference Dose (RfD). The ATSDR's MRL, FDA's ADI, California's PHG, and WHO's TDI are comparable values. This approach is used for chemical contaminants that have a threshold for their toxicity (i.e. there are doses that will cause no effect). Table 1 shows how the individual uncertainty factors might be applied to the data from toxicity studies. Uncertainty factors are used to adjust data on potency downward to maximize the protective nature of the risk-based value.

Table 1. Uncertainty factors used in development of risk-based values*		
Type of toxicity data	Composite factor	Confidence that RfD is protective
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		

It would be possible to use a similar uncertainty factor approach in processing the concentration data. However, in this case the data would be used to adjust the concentration data upward depending on the nature of the available data. Possibilities might be as described in Table 2.

Table 2. Uncertainty factors suggested for concentration data for the Universe to PCCL process.*		
Type of data	Composite factor	Confidence in Concentration
95% UCL of finished water concentration	1	High
Mean of finished water concentration	10	Medium to High
95% UCL of source water	3	Medium to High
Mean of source water	30	Medium
95%UCL 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 are not based on an understanding of the underlying science.		

When occurrence knowledge is in the form of information rather than data, several approaches are possible as follows:

- Develop a screen for candidates from an occurrence list to determine whether it can get into water by considering persistence, solubility, mobility measurements (has ability to actually be in water). The persistence screen should be very coarse (gets a "0" if known not to be persistent enough, 1 if don't know or if know it is persistent, or blank if can't estimate).

- Screen by some consideration of solubility and persistence (half-life?) in combination with production/release data.

- Use SAR/QSAR to estimate or rank potential water concentration with physical properties such as half life, and solubility.

- Combine production/ release data with environmental fate data to estimate a water concentration and compare to health effect at a gross level (e.g. order of magnitude or bins (high, medium, low).

In the case of toxicity based purely on information, the use of "guilty by association" information

may make it possible to evaluate candidates that have not undergone traditional toxicity testing. However, depending on the information, it may be difficult to estimate a water concentration based on toxicity information, and the estimation may not be reliable. An alternative may be to use the threshold of evaluation in combination with another identifier such as group identification. A quantitative comparison of contaminant concentrations in water may infer a greater level of certainty than is warranted. A question is when we know it is there, and suspect it may cause problems, do we need to determine if there are high enough concentrations to be of concern?

Attachment A

Notes on the Pass/Fail criteria (Section 4.0)
NOT reviewed by Methods Activity Group

Three categories of criteria are proposed in this section, corresponding roughly to three levels of evidence that might be encountered for the Occurrence and Effects information. These are the Qualitative, Semi-Quantitative (Margin of Exposure) and Quantitative (Hazard Quotient) criteria. As discussed in the May 1 phone conference, however, the Semi-Quantitative and Quantitative approaches may be viewed as similar by introducing the concept of Uncertainty and Variability Factors that depend on the quality of data available. This approach is described below. The Qualitative criterion is put aside for now, as it would require direct judgments by individuals for each member of the universe, clearly impossible when sorting a large number of candidates for the PCCL. In any event, this criterion is best reserved for the Nominations Process.

To understand the unified Semi-Quantitative and Quantitative criterion proposed here, consider the following simple criterion:

A candidate goes onto the PCCL if existing data and/or correlations based on data provide a *reasonable* basis for the *possibility* that the maximum concentration in drinking water could exceed the concentration yielding adverse effects in the most sensitive subpopulation.

As a first step, imagine that the maximum concentration is C and that the concentration threshold for effects in the most sensitive subpopulation is C_T. Imagine also that both C and C_T are perfectly established quantities, with absolutely no error. If one were dealing with a carcinogen, C_T would be defined by the concentration that corresponds to a maximally acceptable lifetime excess probability of cancer. Under this condition of perfect knowledge, the candidate would not be of concern as a threat to public health so long as:

$C < C_T$ or equivalently $C/C_T < 1$

This quantity, C/C_T, is analogous to the Hazard Quotient used routinely in risk assessment, although broadened here to include carcinogens.

The question then is: *What do we do if either C or C_T or both quantities are not established with perfect confidence*? The answer lies in the terms “reasonable” and “possibility” in the simple criterion given previously.

“Possibility” in this criterion, with respect to C, means that an analysis of the variability in exposure conditions (either spatial or temporal or both) and/or the uncertainty in the methods used to establish those conditions suggests that the maximum concentration is unknown but is characterized by some sort of uncertainty distribution such as the one shown below in Figure 1. As the evidence on which the value of C is estimated degrades in quality (e.g. it is from a small

sample, taken at unrepresentative times, by unreliable methods), the "spread" or variance of this distribution increases. This is shown in the figure by the two curves. The solid curve represents the uncertainty distribution for a case where the data are relatively reliable, and the dashed curve represents a case where the data are relatively unreliable.

"Possibility" in this criterion, with respect to C_T, means that an analysis of the variability in sensitivity between individuals and species and/or the uncertainty in the methods used to establish those sensitivities suggests that the maximum sensitivity (meaning the smallest value of C_T) is unknown but is characterized by some sort of uncertainty distribution such as the one shown in Figure 2. As the evidence on which the value of C_T is estimated degrades in quality (e.g. it is from a small sample, taken at unrepresentative times, by unreliable methods), the "spread" or variance of this distribution increases. This is shown in the figure by the two curves. The solid curve represents the uncertainty distribution for a case where the data on sensitivity are relatively reliable, and the dashed curve represents a case where the data are relatively unreliable.

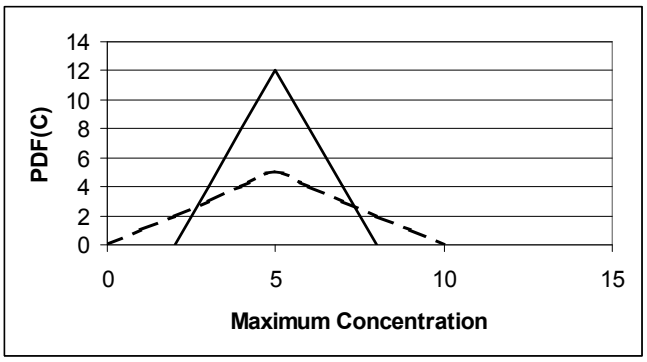

Figure 1. Two uncertainty distributions for the maximum concentration to which people might be exposed.

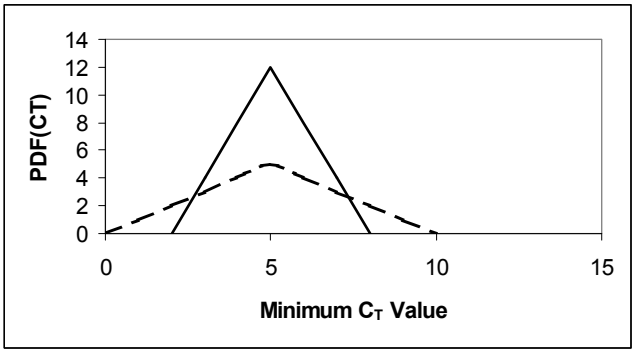

Figure 2. Two uncertainty distributions for the minimum value of C_T in an exposed population.

As a conservative, or health-protective, approach, one might decide to use a value of C_T to the

far left in the uncertainty distribution. For example, for the solid curve in Figure 2, one might use a value close to 2. This is analogous to dividing the best estimate of C_T (which is 5 in that figure) by some "factor" that reflects the "spread" of the uncertainty distribution (in this example, the factor clearly is 2.5; for the dashed curve, an equivalent point to the left of the best estimate might require division by a factor of 5 or 10, since the uncertainty is much higher.). We will call this new quantity, given by the best estimate of C_T divided by some uncertainty factor that is related to the spread (and hence the quality of the evidence), something like the "screening value of C_T". Let's show this by the symbol C_{TS}, calculated by:

$$C_{TS} = C_T / UFC_T$$

where C_T is whatever value we find based on any data or methods that might be available from the database, and UFC_T is the uncertainty factor whose magnitude increases as the quality of the data underlying the estimate of C_T declines. Note that this is completely analogous (in fact is identical) to the way in which the RfC or RfD is defined in regulatory assessments. Note also that we are NOT saying that C_{TS} is the "best estimate" of the minimum value of C_T in the population (the actual minimum value may be well above C_{TS} for all we know). We instead are saying that, for screening purposes, we are "reasonably" confident that the actual minimum value of C_T, whatever that might be, is not BELOW C_{TS}. If we want a greater degree of confidence (being more stringent in what we mean by "reasonably"), the value of UFC_T is increased. But it also is important to bear in mind that at some value of UFC_T, the confidence required is "too high"; it no longer is "reasonable".

We can do exactly the same procedure for the value of C. Now we want to ask: *Given what we have seen about C (however this was obtained from data or correlations or whatever), what is the highest plausible value C might have in the real world, taking into account variability and uncertainty summarized in Figure 1*? We again define a "screening value of C", given by the symbol C_S and calculated by:

$$C_S = C \times UFC$$

where C is whatever value we find based on any data or methods that might be available from the database, and UFC is the uncertainty factor whose magnitude increases as the quality of the data underlying the estimate of C declines. Note again that this is completely analogous to the way in which the RfC or RfD is defined in regulatory assessments. Note also that we are NOT saying that C_S is the "best estimate" of the maximum value of C to which people might be exposed (the actual maximum value may be well below C_S for all we know). We instead are saying that, for screening purposes, we are "reasonably" confident that the actual maximum value of C, whatever that might be, is not ABOVE C_S. If we want a greater degree of confidence (being more stringent in what we mean by "reasonably"), the value of UFC is increased. But it also is important to bear in mind that at some value of UFC_T, the confidence required is "too high"; it no longer is "reasonable".

So now, we define something like the "hazard quotient", which we might call the "screening

hazard quotient" and given by the symbol HQ_S:

$HQ_S = C_S / C_{TS}$

and place a candidate on the PCCL if:

$HQ_S > 1$

This only leaves us with the question: *How large should UFC_T and UFC be for specific cases of information underlying the "best estimate" judgments of C_T and C, respectively*? As the uncertainty factors increase, we increase the confidence that no candidate is being left off the PCCL because we have underestimated its potential to pose a risk. But this comes at the price of letting an increasingly large pool of candidates onto the PCCL, potentially flooding that list with too many candidates and bringing the CCL process to a halt due to shear numbers. In other words, as the uncertainty factors increase, we reduce the number of false negatives (a candidate that poses a risk but is left off the PCCL) at the expense of a larger number of false positives (a candidate that poses no risk but makes it onto the PCCL). Some balance must be found.

The key to this balance seems to lie in the concept of "reasonable possibility" mentioned previously. "Reasonable" in this instance has two meanings: the values of UF must be the result of reason, traceable to specific scientific and probabilistic arguments; and the UF values must not be SO large as to result in screening values that could not be excepted to occur (i.e. would be outside the range of uncertainty/variability shown in Figures 1 and 2). For the case of C_T, the EPA already has a procedure for assigning values of UFC_T, and the work of Dourson and others (references) has placed these on the kind of probabilistic basis discussed here. The value of UFC_T increases when:

- The measured value of C_T is obtained from a human subpopulation not expected to be the most sensitive;
- The measured value of C_T is obtained from a non-human population;
- The measured value of C_T is based on an incomplete data set (perhaps only some effects were examined, but not all; or the population follow-up was incomplete; etc).

It would be necessary to define a procedure for assigning the value of UFC_T based on which combination of these data characteristics apply. This process is not explored further in this section.

For the case of C, no such procedure has been developed for assigning values of UFC, and so new work would be needed. It would be necessary to assemble data on compounds in which the maximum concentration has been reliably measured, as well as a series of surrogate measures (e.g. average value; value during drought; value in groundwater but not surface water). One could then develop ratios (e.g. maximum value over value in groundwater only) for a variety of compounds, and produce a probability density function for these ratios. From such a PDF, one could say something like: for at least 95% of compounds where complete information is

available, the ratio of maximum over groundwater only values is no larger than 4. And so, the value of UFC to be applied to a compound in which ONLY the groundwater value is available would be something like 4. The value of UFC increases when:

- The measured value of C is obtained from a non-representative water supply;
- The measured value of C represents something other than the maximum;
- The measured value of C is based on an unreliable methodology.

It would be necessary to define a procedure for assigning the value of UFC based on which combination of these data characteristics apply. This process is not explored further in this section.

What can be done when there are NO measurements of water concentration of any kind, but where some measures of production volume, persistence, etc, are available? One possibility, mentioned at the beginning of this section, is simply to consider this candidate in a nomination process, since such a process allows for significant use of expert judgment (whereas any automated process, needed to process large numbers of candidates, cannot require such judgments).

Another possibility is to use screening models for water such as those employed routinely in the Office of Pesticides. These are very simple models that reflect maximum possible concentrations in shallow aquifers or in small ponds. Their use was reviewed on an ILSI panel several years ago (reference). They make use of some estimate of the loading into the water body (e.g. grams per day entering the body), the size of the water body, and the persistence or half-life in that water body. As in the case of the conservative procedures described above, there is no claim that the concentrations calculated are best estimates. The claim is simply that the actual concentrations are unlikely to be exceeded by the calculated values. There remain significant challenges in using such models, one of the more important being how to decide how much of a total source term or release rate or production rate of a compound to assign to the screening water body. It would not be "reasonable" (in the sense used earlier in this section) to assign it all to the one screening water body. This option should be explored for candidates with no direct water measurements. Again, the procedure could be to assemble information, on both measured maximum water concentration and the parameters needed for the screening model, over a range of compounds for which both kinds of information is available. It then would be possible to calculate the concentration predicted by the screening model and compare that to the actual measured maximum value and develop a ratio. A distribution of these ratios would be determined for a variety of compounds, and used to establish a form of uncertainty factor to be applied to the result of the screening model when only that result is available.