Comments on the OMB Proposed Risk Assessment Bulletin

Commenter: Gary Ginsberg, Ph.D., Toxicologist, Connecticut Dept of Public Health
410 Capitol Ave., Hartford CT 06134
gary.ginsberg@po.state.ct.us
860-509-7750

Date: March 27, 2006

As a risk assessor in Connecticut state government, I reviewed the Proposed Risk Assessment Bulletin with great interest. We rely heavily on the types of risk assessments mentioned on Page 9 of the proposed Bulletin, most notably the IRIS files and ATSDR Toxicological Profiles. We have a need for these risk assessments to be thorough and scientifically robust and so we have a mutual interest with the writers of this bulletin in seeing that federal risk assessments are optimized. We all strive for standardization in risk assessment so that, to the extent possible, every scenario is treated with uniform methods, and that the assessments are kept free of biased assumptions and selective judgement.

Unfortunately, the requirements put forth in the draft Bulletin are not always clear and may not be useful in achieving our common risk assessment goals. Rather that putting forth a unifying vision for expanding and improving upon current practices, it provides a vast and rather vague array of requirements (quality standards). It doesn't spell out whats wrong with current risk assessments and what problems or issues the Bulletin is trying to fix. For example, IRIS files and ATSDR Toxicological Profiles are highly respected, excellent sources of risk information that have been invaluable for many years. These have been the gold standard, providing the key inputs to site-specific risk assessment and regulation. If OMB has identified problems in the way that these documents are constructed or used, then the Bulletin should clearly state these issues and the means for their resolution. In its current form, the Bulletin represents a series of general principles, many of which are already followed to one degree or another in federal risk assessments. It is unclear how these would specifically apply to IRIS, ATSDR Tox Profiles or site-specific assessments. Therefore, more specifics with a few case examples are needed. Is

there an IRIS file, ATSDR Tox Profile, or site-specific risk assessment that OMB can point to that demonstrates what it is looking for?

Standards Related to Level of Precision and Central Estimate

Based upon goals and quality standards described in the Bulletin, it appears that OMB is overlooking the underlying nature of risk assessment. It is an inherently uncertain science because of the extrapolations typically needed in going across species, dose routes, dose levels, general population to vulnerable subgroups, exposure time frames, and from single chemical toxicology studies to mixtures exposure in people. USEPA has made considerable advances over the past 20 years in developing risk methods that are consistent and transparent, with new modeling and statistical techniques being applied as they are validated. These new techniques are helping assessments make maximal use of imperfect datasets and present full distributions of risk, thereby decreasing the level of uncertainty. What is needed is guidance on how to use these techniques in a consistent manner across a variety of risk applications. Instead, the Bulletin puts forth data quality requirements that assume some idealized risk paradigm which is capable of providing exact results and verifiable projections of risk. For example, the Bulletin asks assessors to "convey the precision associated with (risk) estimates" (Page 13), disclose "error sources affecting data quality" and to emphasize central estimates over upper bounds (Page 19). Agency risk assessments are already fairly explicit about uncertainties and how they are addressed. Further examination of precision and sources of error implies a level of exactitude not possible and would be counter-productive given the types of uncertainties typically present in risk assessments.

Central vs. Upper Bound Estimates

This point about central estimates is worth expanding. Central estimates of risk are not simply the "expected" risk as suggested on Page 19; that implies that central estimates are closer to the true risk and are thus more robust and accurate. It also implies that an upper percentile risk is a crude estimate or conservative guess that one could use to be on

the safe side, but which was not intended to be accurate. Both descriptions are false. The central estimate is still just an estimate which is subject to measurement error and population variability. The reason the Superfund program uses the 95% upper confidence limit (UCL) on the mean soil concentration is that we can't sample every inch of dirt, so we will never know the true mean. Instead we use a statistical representation that takes into account variability in the data to create a central estimate that is not precise or accurate, but rather is a reasonable upper bound on an uncertain statistic (the true mean). Similarly, upper confidence limits are used for cancer slope factors. Thus, while we typically have more data on people near the center of a distribution, these central estimates still involve considerable uncertainty. The Bulletin glosses over these points and elevates the central estimate as if it were the most or only reliable statistic.

The upper bound risk is not some high end extrapolation of the central estimate. It is intended to represent members of the population who are at greater exposure and risk due to real host and environmental factors. These people would not be protected by regulation that is only targeted at central estimates of risk. Since the data may be less robust at the tails of the distribution (fewer numbers of people and thus fewer samples), and since multiple conservative assumptions can be compounded to yield unrealistic upper bound results, careful inspection of an upper bound result is needed. However, calculation of a realistic upper bound risk is quite possible and useful for regulatory determinations. For example, lead cleanup standards are based upon the well established health goal of keeping 95% of the population of exposed children below the blood lead target of 10 ug/dl. A risk management approach which focused upon central estimates would overlook important inter-individual variability in lead exposure and toxicokinetics that causes some children to have much higher internal dose than others for a given concentration of environmental lead.

Standards that Seek Unnecessary and Excessive Information

The Bulletin is asking for levels of description and analysis that are impractical and unprecedented in risk assessment. Strict adherence to these quality standards would

greatly expand the length and complexity of risk assessments making them less, not more, transparent and useful. The following are several examples contained in the Bulletin:

Standard for Characterizing Results (page 19) wants assessments to present results for a variety of toxic effects and/or studies, including a discussion of "alternative theories, data, studies, and assessments …". This is a very open-ended request that can multiply the level of work with little benefit. Risk assessments normally screen a large body of toxicology and epidemiology information to establish the most pertinent, robust, and sensitive data sets for the evaluation of dose response, point of departure and reference level. Other data sets may be important for the weight they add to the toxicity assessment, but they generally give way to the strongest study demonstrating sensitive dose response. The logic is by setting a toxicity value on this endpoint, that you automatically protect against the chemical's other toxic effects. If the toxicology assessment is properly done, any important conflicting studies should be described and dealt with in one way or another (e.g., excluded because of weaknesses, described qualitatively, considered as an uncertainty). There is no reason to present risk results for all the endpoints and all the studies as the critical studies should rise to the top of the assessment and be the focus of the remainder of the analysis. The other studies and endpoints would be of lower utility for a variety of possible reasons and thus could only confuse the risk characterization. Presenting a risk range based upon the range of plausible exposures and vulnerabilities in the population is certainly important. Presenting a risk range based upon individual studies/endpoints in the toxicity database is unnecessary and wastes the effort of the analyst and the end-user of the assessment. In fact, it may give the risk manager the wrong impression that the analyst doesn't have a clear idea on how to calculate the risk.

Standard for Discussing Scientific Limitations (Page 20) asks the assessment to provide details for possible research that could theoretically help resolve uncertainties/limitations in the current risk assessment. This requirement is far beyond what is normally within bounds and is highly speculative – one cannot tell the time frame, costs difficulty, etc. of some possible studies that have yet to be sponsored or designed. Further, there may be numerous plausible studies. The job of the risk assessment is to identify major data gaps

and uncertainties and find ways to address them within the risk assessment if possible. Proposing new research is a possible outcome of a risk assessment, but not at the level of detail suggested by this standard. Further, this should not be a core requirement expected of all risk assessments.

Updates (Page 21) asks the assessment to disclose how fast the relevant database and assumptions are evolving and how likely they will be significantly different in the coming months to years. This is again highly speculative and outside the bounds of risk assessment. It is a basic tenet of science in general that the field is constantly moving forward with a need to revisit assessments in the future as data or methods warrant. There is no need to place a requirement on a risk assessment to guess how rapidly the field is changing and when would a new risk assessment be called for.

Additional Comments

Standards Related to Objectivity (Page 14) – it is useful to remind risk assessors to consider both positive and negative studies in considering the weight of evidence for a particular effect. However, the Bulletin is unclear how this weighting should occur. All things being equal, a positive finding is more influential than a negative finding. That's because toxicologists and epidemiologists recognize the bluntness of their instrument which can lead to false negative results, especially at low levels of exposure where toxic effects can blend into the background occurrence of disease. A clear positive finding that cannot be explained away by methodological issues, is much harder to dismiss than a negative study. The Bulletin should provide more guidance on weight of evidence and capture the principle of predominance of positive findings.

Certification (Page 21) – the certification process is not clear as to who makes this determination and how it is to be made. There are many standards placed on risk assessments in this bulletin, some more pertinent than others, and in many cases, meeting the standard might involve a range of responses, from something simple to something

highly complex. Determining compliance with most of the standards is not a simple yes or no and may be left to case-by-case value judgements. Thus, the certification process will likely be somewhat subjective and raise questions of inconsistency and bias (e.g., one risk assessment subjected to more of the Bulletin requirements than another one). Rather than a certification process, important risk assessments should be sent out to peer reviewers, as is now often done.

Standard for Characterizing Human Health Effects – Page 20 – the distinction suggested between adverse and non-adverse effects is still somewhat vague and open to interpretation. Clear examples of adverse and non-adverse effects should be provided in the Bulletin. Biomonitoring findings are obviously not a health effect in and of themselves. However, there are no examples beyond that simple one. Precursor effects such as biochemical changes on the mechanistic pathway to toxicity are important to address. They are now considered as equivalent to adverse effects for the purposes of setting RfDs. They represent perturbation of key risk pathways in a direction that could lead to increased susceptibility to other chemicals or environmental factors which act similarly, and which may actually be adverse effects in sensitive members of the population. The Bulletin should recognize the value of precursor effects in establishing reference levels.

Page 10 – Problem Formulation - The guidance calls for an iterative dialogue with the risk manager (agency decisionmaker). However, an important tenet is that risk assessment should be conducted independent of from risk management so that there is no perception that the results are influenced by the costs or feasibility of addressing the risks that are calculated. While it makes sense for there to be an open (iterative) communication in the scoping phase, such a channel would do a disservice (perhaps compromise) the risk assessment process beyond the initial phase.

Page 10 – Scientific Completeness – this is a vague term that can only leave the door open for unwarranted criticism of a risk assessment because it may have left something out. A risk assessment cannot consider everything about a chemical, its exposure,

toxicity or ability to interact with other chemicals. Rather, the assessment is purposively streamlined to highlight the most significant toxicity information, the most relevant toxicokinetics, and the most plausible and significant exposures. Significance, relevance, and plausibility are the key words, not completeness. Thus, this goal seems to be largely irrelevant, poorly defined and so should be removed.

Page 12 – Informational Needs – the example is given that users should be made aware of "…the ranges of chemical doses for which the assessment is relevant …". This gives the impression that a risk assessment will calculate risks for doses that may or may not be relevant to the public or sub-groups thereof. This is usually not an issue since an RfD or cancer potency value is independent of exposure level and the way these toxicity values are applied are always contingent on the dose in a particular receptor group and exposure scenario (e.g., the cancer risk is projected for excavation workers or school students, etc.). Therefore, there is rarely confusion over the relevance of chemical doses. The other aspect that this example may be trying to address is the gap between high dose animal studies vs. much lower environmental doses that people are experienced to. Again, this is rarely a point of confusion as the primary goal of risk assessment is to perform extrapolation across species and exposure levels to yield a risk assessment relevant to the general public exposed at environmental levels. The caveats and uncertainties associated with these projections need to be clear to the risk manager. If thresholds exist, these are identified so that the RfD indicates doses below which the toxic effects are no longer relevant. Since by definition risk assessments describe the significance of exposure at various dose levels, the standard concerning relevant ranges of chemical dose is unclear and of questionable utility.