Comments on the

OFFICE OF MANAGEMENT AND BUDGET *PROPOSED RISK ASSESSMENT BULLETIN (2006)*

©2006 Paolo F. Ricci, Ph.D., MPA, LLM[1]

We comment on the *Proposed Risk Assessment Bulletin*, as developed by the OMB (2006), and thus respond to their notice for public comments. Although their separating risk assessment from and management is somewhat artificial, because regulatory risk assessment is mandated by legislation and thus inherently imbedded in risk management, we think that the separation is useful. We do not comment on permitting or quasi-judicial agency actions because these are excluded by the OMB (OMB, 2006). Our concerns and the recommendations to alleviate those concerns are given in the text, as discussions of the OMB's proposals for standards.

Regulatory risk assessment requires a prior statutory command before it can be undertaken. Moreover, after it is completed, it must be able to sustain peer and judicial reviews. The latter is often deferential to an agency's scientific findings on different grounds than a scientist may think and thus further adds to the need for the proposed Bulletin. This, with other considerations developed in the proposed Bulletin, supports a risk guidance that is second to none, particularly when such guidance is part of the regulatory process. Overall, we suggest that any guidance – such as this Bulletin gives -- to the way in which risk assessment is conducted at the federal level in the United States must be of *sufficient* specificity to:

- Reflect the state-of-knowledge about the *representation* of risks under different levels of information available to the risk assessor (e.g., be able to use appropriate measure of uncertainty, such as upper and lower probabilities, fuzzy numbers, and so on) and *causal modeling* (e,g., include thresholds, *U*- or *J*-shaped dose-response models as well as the defaults assumptions of the LNT when the scientific information is either at or above the scientific evidence for the linear, no threshold default model, simultaneities (when these exist), and so on),

- Provide the necessary probabilistic and statistical guidance relevant to accounting for various forms of bias in estimation,

- Provide the necessary statistical guidance relevant to choice of models, including Bayesian (e.g., traditional, robust, and Bayesian model averaging), and selected criteria to judge the performance of those models,

- Account for the extent to which uncertainty and variability must be described and analyzed, including stopping rules and value of information criteria,

[1] These comments can only be attributed to its Author: Dr Paolo F. Ricci, 685 Hilldale Ave, Berkeley, CA, 94708; apricci@earthlink.net; I thank Drs A. L. Cox Jr. (Cox and Associates) and F. Dayritt (Ateneo de Manila University) for discussions and comments. I have used few references to keep the length of this paper consistent with the OMB's *Bulletin*.

- Account for model specification and the appropriate conditionalization in probabilistic networks and tree models, and

- Account for the limitations to consensus judgments through formal and thus replicable analysis.

The reason for this listing is that regulatory risk assessment should answer the following question:

Can risk assessment-based judgments and methods correctly – i.e., formally, soundly, and replicably -- predict the probability of adverse outcomes to inform risk managers of the soundness of the options developed by risk assessors?

This question raises two corollary questions. These are:

1) **Can the combination of choice of data and methods used in regulatory risk assessments be shown to be the best possible, given the state of information and knowledge, with what level of probative value, and under what circumstances?**

2) **Because risk assessment results combine qualitative choices of data and techniques to assess this data within the scientific method and with scientific judgments: Are those regulatory results equitable on those who are at risk but may not benefit form the activity to be regulated**?

Clearly, the OMB's final *Bulletin* cannot be an encyclopedia of terms-of-art, techniques, and examples. Nonetheless, to be useful and to meet its objectives relative to the quality of the information, transparency, and so on, the guidance should be sufficient for sound risk assessments that are credible, transparent, and use the best possible science. The importance of the task should be met by a commensurate set of guidelines based on clearly enunciated principles, measures, and methods.

INTRODUCTION

The word *risk* is generally equated with a probability of an adverse event or outcome, over a specific period of time and space; the term *hazard* describes the context generating the risk and thus it is a physical event, not a probability. To clarify these two distinctions, we let the risk be the LHS and the hazard (in terms of its risk factors), and let the RHS of the canonical risk model accounts for the factors that are associated with the hazard and thus result in the risk[2]:

Risk(probability of an adverse outcome) = g(risk factors, adjustments, time, random or other forms of error).

This model, like most models attempting to portray some amount of empirical causation, requires theoretical (e.g., biological) justification before being mathematically specified

[2] We use model instead of function or relation; the context defines which is which. As we discuss, other measures of risk may be used, depending or the information context (e.g., fuzzy measures, intervals, Dempster-Shafer possibilities). However, we limit our discussion to representations of uncertainty based or probability measures and their calculus. The term canonical means, in this paper, conforms to its lexicographical definition of standard method.

and then statistically estimated (either the parameters of that model or the model itself). Moreover, causal models obtain from sampling results, and thus require inference from sample results to the population or populations of concern to the policy-maker. Thus, the usual situation that risk assessors face is one in which working hypotheses are accepted or rejected from experimental evidence and prior information that are combined via Bayesian updating rules (or other forms of probabilistic conditioning such as Jeffrey's) (Jeffrey, 1983; Howson and Urbach, 1993). Probabilities can be degrees of belief (e.g., de Finetti, 1975; Ramsey, 1990; vonPlato, 1994; Hampel, 2001), likelihoods are conditional probabilities from experimental or observational studies, as well as frequencies (long-run), among other definitions.

A more complete example of the causal chain running from a hazard to the risk – encompassing the ambit of the proposed *Bulletin* -- is depicted below (in which a hazard can generate risk (as a probability of response) from exposure to pressure, heat, radiation, chemical interactions, and so on):

Hazard(s) as source(s) of risk → agent(s) release → fate, transport, storage, and release, possibly as other agent(s) from hydrolysis, microbial action and so on → exposure to agent(s) → dose (possibly as by-product(s)) → response(s), such as fetal malformations, cancer, and so on → individual risk(s)possibly represented by as percentiles of the output distribution(s)from exposure to $PM_{2.5}$ → aggregate risk(s) (e.g., expected annual incidence of an exposure-specific population of males older than 65).

Although this chain is shown as deterministic, it is in fact probabilistic. Indeed, the chain consists linked random variables (e.g., exposure to agent *XYZ* varies over time and is sampled locally) resulting in an output distribution (e.g., the distribution of the risks of a adverse outcome such as death due to leukemia) (Pearl, 2000; Jensen, 1996). This is part of the reason why, generally, when we speak of risk, we mean distributions rather than just a probabilistic statement.[3] When a value is known with certainty, it is represented by a distribution having its probability mass equal to one (a *degenerate* distribution) (Granger-Morgan and Henrion, 1991; Ricci, 1996).

This proposed OMB *Bulletin* is concerned with both technological risks and with environmental risks. A logical (and probabilistic) way to link these can use fault-trees and event-trees with exposure-response models, which are not shown in the Figure below, but can be surmised at the end of the event–tree as damage functions (and value or utility functions, although we do not show these for brevity), as can be required by an actual risk analysis as can be found in the technological risk analyses conducted by US federal agencies such as DOE, NRC, NASA and others:

[3] A response surface may also be a very useful description of risky situations.

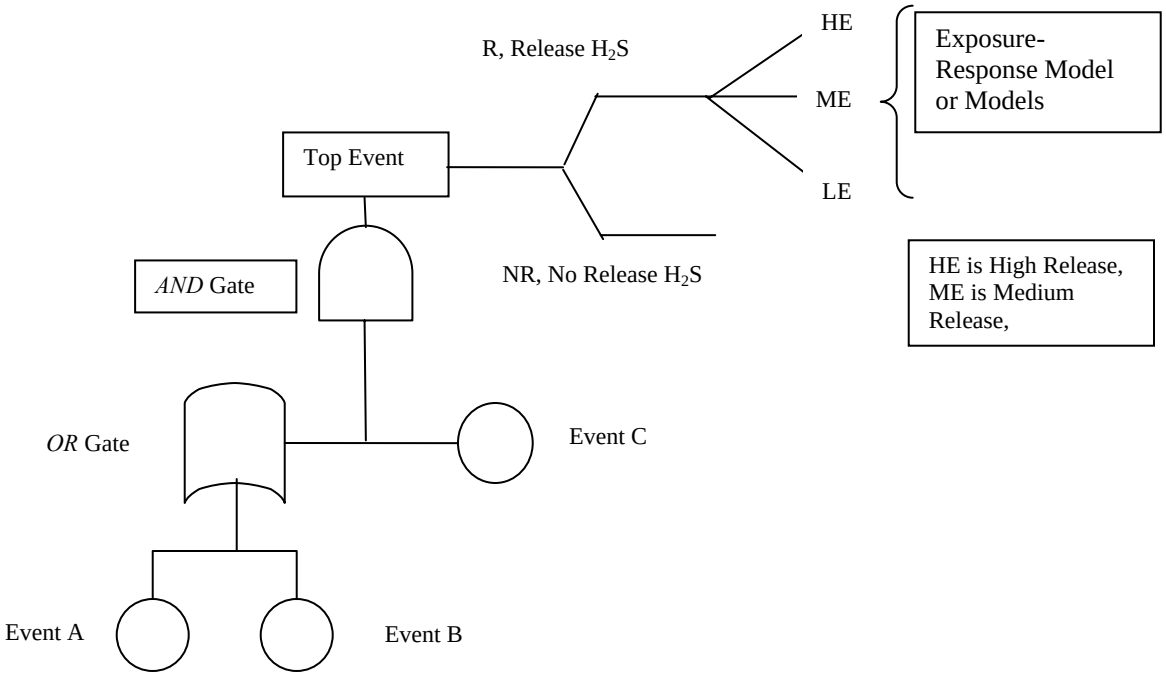


Figure 1. Combined Fault-Tree and Event-Tree

The example below (Ricci, 2006) extends the logical aspect of the fault- and event-tree to include probabilities. Because the construction and analysis of these trees is well established and used by many federal regulatory agencies such as NASA, NRC, DOE, and so on, we do not go beyond these simple examples and diagrams.

Example. Suppose that the probabilities of the *independent* events *A*, *B* and *C* are rare, say $pr = 0.01$. Use the rare event approximation: $pr(D) \approx pr(A \text{ AND } B)+pr(A \text{ AND } C) = pr(A)pr(B)+pr(A)pr(C)$. That is: $pr(D) \approx (0.01)(0.01)+(0.01)(0.01) = 0.0002$. Suppose that gate *B* is a *NOT* gate. What is the probability at that gate? Define the probability of success as pr_s and the probability of failure is $(1-pr_s)$. For an *AND* gate leading to a success and several sub-events the probability is $\Pi_i(pr_i)$; letting *B* be the complement of *A*, the probability of the *NOT* gate is $[1-\Pi_i(pr_i)]$.

Technological risk for one or more technologies are generally depicted through the complement of the cumulative distribution (Apostolakis, 1974; Andrews and Moss, 1993; Ricci, Sagan, and Whipple, 1985):

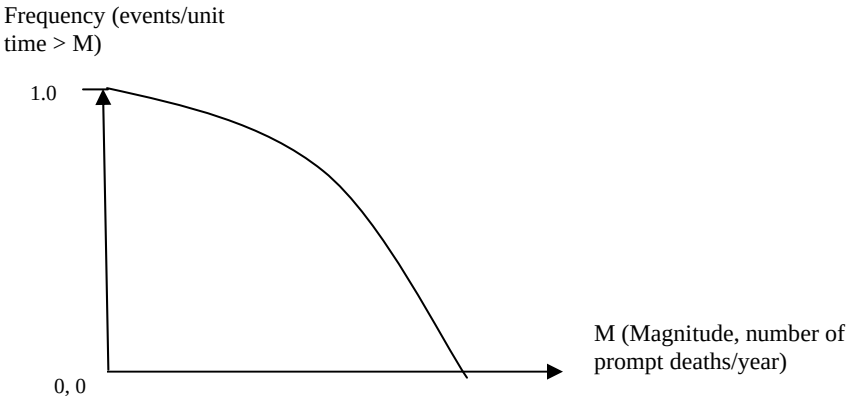


Figure 2, Complement of the Cumulative Distribution of the Random Variable M

Using continuous data, this curve is based on the density function of the random variable *X, f(x)*, which might measure the number of prompt deaths (namely, their magnitude, $M \geq 0$) from a specific and well identified event, e.g., single explosion of a certain yield. In practice, additional analyses are required including have data to which the complement of the cumulative distribution can be fit (Ricci, 2006). Specifically, the data to fit the complement of the distribution are often available near the high probabilities of events of small magnitude, but not often near low probability events of high magnitude. The parameters of this curve, given data on *M* and the frequency, *fr(M)*, can be estimated using the maximum likelihood estimator, MLE. To summarize these concepts consider Systems *A* and *B* as alternatives to reach a specific objective. The random variable *X* measures the magnitude of the adverse outcomes, as counts of prompt deaths. As depicted below, Figure 3, the MLE curve of System *A* is uniformly riskier than System *B* because the risk curve for *A* dominates the risk curve for *B* everywhere. However, the *95%* confidence intervals do not show that dominance because the confidence bounds about each curve overlap. The variability about the risk curve *B* (the straight line crossing it) overlaps with the uncertainty about the risk curve *A* (straight line crossing it). Developing these curves is a statistical estimation problem, in which the complement of the cumulative distribution appropriate to the events is fit to the data; their use in risk management deals with the dominance of one curve over another, for given choices of technological options (Apostolakis, 1974; Ricci, Sagan, and Whipple, 1985; Bedford and Cooke, 2001).

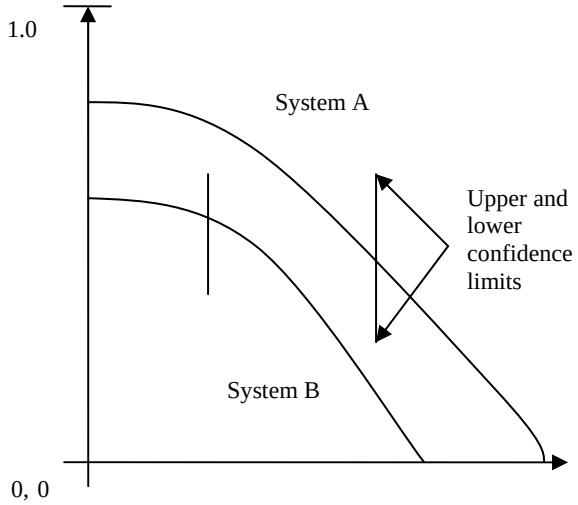


Figure 3, Comparison of Complements of Cumulative Distributions (the curves are not constrained to start at 1.0)

We will discuss how to include concepts of dominance to resolve choices where the uncertainty bounds overlap, as shown in Figure 3.

GENERAL COMMENTS

In its draft proposal for a bulletin on risk assessment, the OMB is concerned with human and ecological risks from routine and non-routine events leading to adverse effects, the cost of control and compliance with federal regulations; however, it does not consider financial risk. The OMB (2006) states that:

... "risk assessment" means a scientific and/or technical document that assembles and synthesizes scientific information to determine whether a potential hazard exists and/or the extent of possible risk to human health, safety, or the environment. For the purposes of this Bulletin, this definition applies to documents that could be used for risk assessment purposes, such as an exposure or hazard assessment that might not constitute a complete risk assessment as defined by the National Research Council. (Citations omitted)

This is a procedural understanding of risk assessment guided by administrative reasoning. Risk assessment goes well beyond the assembly of information; it is a scientific process requiring the selection and use of qualitative and quantitative information and knowledge, abstraction and prediction for the express purpose of describing rational outcomes while accounting for their uncertainty. This observation seems to be corroborated by the fact that (OMB, 2006):

. . . risk assessment disseminated by a Federal agency is subject to OMB's Information Quality Guidelines and the agency's Information Quality Guidelines. These guidelines require risk assessments to meet the three key attributes of utility, objectivity, and integrity.

It is well known that most US federal and state regulatory analyses are based on forms of risk assessment predicated on the concept of *acceptable* or *tolerable* risk (for instance, measured by a percentile on the cumulative probability of response, over a specific time

period) (Ricci and Molton, 1981). Recently, regulatory risk assessments have had to include a new science-policy dimension: an emphasis on the overall quality of the information provided by a regulatory risk assessment. As the OMB (2006) states, the quality of risky information is a concern that is part of the federal legislative and executive commands (accomplished through legislation and executive orders (E.O.s)):

> [T]he "Information Quality Act," Congress directed OMB to issue guidelines to "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility and integrity of information" disseminated by Federal agencies. Pub. L. No. 106-554, § 515(a). … Section 624 of the Treasury and General Government Appropriations Act of 2001, often called the "Regulatory Right-to-Know Act," (Public Law 106-554, 31 U.S.C. § 1105 note) directs OMB to "issue guidelines to agencies to standardize ... measures of costs and benefits" of Federal rules. Executive Order 12866, 58 Fed. Reg. 51,735 (Oct. 4, 1993), establishes that OIRA is "the repository of expertise concerning regulatory issues, including methodologies and procedures that affect more than one agency," and it directs OMB to provide guidance to the agencies on regulatory planning. E.O. 12866, § 2(b). The Order requires that "[e]ach agency shall base its decisions on the best reasonably obtainable scientific, technical, economic, or other information." E.O. 12866, § 1(b)(7). The Order also directs that "[i]n setting regulatory priorities, each agency shall consider, to the extent reasonable, the degree and nature of risks posed by various substances or activities within its jurisdiction." E.O. 12866, § 1(b)(4). Finally, OMB has additional authorities to oversee the agencies in the administration of their programs.

The OMB makes a regulatory distinction between the forms of risk assessment that require federal government concern. One of these is the *influential risk assessment*, which is defined as (OMB, 2006):

> ... a risk assessment the agency reasonably can determine will have or does have a clear and substantial impact on important public policies or private sector decisions. The term "influential" should be interpreted consistently with OMB's government-wide Information Quality Guidelines and the Information Quality Guidelines of the relevant agency. A risk assessment can have a significant economic impact even if it is not part of a rulemaking. … Alternatively, the federal government's assessment of risk can directly or indirectly influence the regulatory actions of state and local agencies or international bodies.

The OMB's distinction reflects i) the level of detail in a risk assessment (regulatory analyses, such as those included in formal rulemaking, can be different from those used for informal procedures or even under guidelines), and ii) the magnitude of the effect associated with regulatory choice designed to reduce, mitigate, or otherwise decrease the probability of injury or death.

Although the distinction is appropriate for managing risks without incurring in undue foreseeable social costs, the phrase and the paragraph that explains it raise several issues because at least four of the terms included are vague (i.e., *reasonably, clear, substantial,* and *significant*). For example, what is *reasonable* (often used as a legal term of art) today may not be so *reasonable* tomorrow nor for whom an event *should* be reasonable. Second, the OMB's proposal merges legal and scientific terms of art (e.g., *clear* and *substantial* are common terms in administrative law and have specific meaning that is not equivalent to their colloquial or scientific uses) in a way that may be counterproductive: the cart is placed before the horse. In other words, the definition of *reasonable*, *clear*, and *substantial* must be formalized to include quantitative guidance (even though formal analysis is applied on case-by-case basis). In particular, we think that legal terms-of-art should not be used in scientific explanations because they tend to introduce unneeded

vagueness and thus obfuscate the scientific discourse, rather than facilitate it. Although there has to be a convergence between terms-of-art, that convergence should, in our opinion, occur in the risk management phase. There, already established interpretations of probability numbers as lexicographical terms can help the convergence.

SPECIFIC COMMENTS ON THE STANDARDS IN THE OMB's (2006) PROPOSED BULLETIN

The scope of the OMB's *Bulletin* is to provide agency-wide standards for risk assessment, under federal law. As will become apparent from our comments, regulatory risk assessment and management cannot truly be separated: each reinforces and benefits the other regardless of whether the legislation that demands it is risk-based or allows for risk-cost-benefit balancing. Regulatory laws manage hazards – the danger imposed by them is measured by risk -- and their success can be measured by the reduction in risk associated with a regulatory standard.

OMB's *Standards Relating to Scope of Risk Assessment*

We concur with the OMB's view of risk assessment that:

> ... for an assessment to be complete, the assessment must address all of the factors within the intended scope of the assessment. For example, a risk assessment informing a general regulatory decision as to whether exposure to a chemical should be reduced would not be constrained to a one-disease process (e.g., cancer) when valid and relevant information about other disease processes (e.g., neurological effects or reproductive effects) are of importance to decision making.

This statement is relevant to the completeness of an assessment, discussed later. We however note that the OMB does not explicitly accounts for events such as fires, peak pressure and other forms of physical exposure, although these can be concomitant with routine and non-routine events leading to the release of chemicals and other agents. In this context, fault- and event-trees can be used as depictions of the initial reasoning, and then followed by more explicit analysis. Moreover, there is an implied concept that all exposure are noxious, which is a presumption based on conjectured events while the factual basis may be otherwise. Thus, it seems that the OMB should account for all possibilities that are scientifically credible and based on facts; this can include instances of hormetic behavior modeled by *J*- and *U*-shaped exposure (or dose) response models as well as traditional dose-response models used in regulatory risk assessments (Calabrese and Baldwin, 2001).

Risk assessments are seldom conducted without prior managerial or regulatory objectives and numerical targets: they are designed and conducted to inform the decision process (via the methods of risk management) about potential adverse consequences or benefits. The results from a *risk assessment* can be scientifically sound per se, that is, regardless of the regulatory framework within which they are produced. However, given that the risk assessments under consideration by the OMB are part of regulatory law, the effect of prior legal or regulatory considerations on the conduct of an assessment requires explicit discussion of their effect on the completeness of the results obtained from that

assessment. A statute can mandate a theoretically incomplete risk assessment, for example, by commanding zero risk. Although legally possible, the assessment leading to the conclusion of zero risk (depending on how the statute is framed) can be incomplete because it does not include an assessment of alternatives that are based on risk-risk trade-off analysis, or risk-cost-benefits analysis. The point is that limiting a risk assessment to just considering risks is incomplete because – regardless of the statutory command -- costs will be incurred in meeting the zero-risk standard. Moreover, reducing the hazard of concern to the legislators my cause larger risks that can go either unnoticed because they are diffuse or have some other indirect and intangible effects that are unaccounted for by the single-minded focus on the elimination of the direct health hazard. A relevant question is: Should a regulatory agency consider alternatives to a specific command even though the statute either proscribes those considerations or is silent on that point? A paradox seems to be: not conducting the appropriate analysis is technically incorrect (on decision-theoretic grounds), while extending the boundaries of the analysis envisioned under the statute beyond the risk component can be technically correct but "illegal."

Second, requiring *valid and relevant information about other disease processes*, imposes perhaps unneeded constraints on causal analysis, which is a component of the legal basis of those assessments. An unanswered question is: When is a statistically significant result (e.g., with a *p*-val = *0.06*) valid and relevant for regulatory risk assessment?[4] Moreover, the need for *completeness* (a term discussed shortly by adding more specificity) can lead to regulatory stasis. The reason is that the process of assessing the state-of-knowledge, determining areas of remarkable incompleteness (so that it can be asserted that the information remaining is *valid and relevant*), developing and implementing appropriate study or studies to compensate for the weakness of earlier studies or theories, waiting for results, and then proceeding with assessing the options for managing hazards, and thus reduce risks, takes time and money. Viewed prospectively, within that the stasis period, the hazards of concern can increase the burden of disease; but, the opposite can also be true. This is a classical problem that can be addressed by decision-theoretic methods.

It also seems that the distinction between *influential* and its complement, *non-influential*, risk assessments is problematic because, whether the assessment is influential on not depends, on the state-of-information and knowledge, at the time the risk assessment takes place, and on the decision rules used to assess the penalty (e.g., via loss functions) associated with each choice, before the choice leaves the analysis stage and is implemented. We also note that a collection of non-influential risk assessments can add up to become as influential as influential assessments. This eventuality should be explicitly accounted for in the OMB's *Bulletin*.

If a risk assessment under the OMB's guidance in the *Bulletin* (OMB, 2006) is prospective -- e.g., it is both *ampliative* and *predictive* of future health or environmental burdens -- it must allow updating as new information becomes available via formal and thus replicable rules. Moreover, the OMB should provide an indication of the appropriate stopping rules that would guide federal agencies to determine when the threshold

[4] We do not discuss the controversies about *p*-values for the sake of brevity; yet, we suggest that the OMB might consider addressing this issue and thus provide some initial guidance.

between *influential* and *non-influential* occurs, if that distinction were to be retained. For example, criteria based on the expected value of sample information, with specific loss functions that, depending on the context, may be asymmetric, are well established in decision theory (Grimmet and Stirzaker, 1982). Importantly, given the heterogeneity of the information used in health and environmental risk assessments, the state-of-the-information at the time the predictive risk assessment takes place can be assessed by statistical and logical methods (e.g., data and text mining; evidence-based reasoning). These facts should be reflected in any standard of performance involving heterogeneous information and complicated fact patterns leading to an overall causal network. Forensic (ex post) applications of risk assessment could also benefit from these concepts, with the difference that prediction should substituted by ex post reconstruction of a causal network, given a known outcome.

These considerations lead us back to the concept of *completeness* again (OMB, 2006). This concept is inherent to data analysis and causal model building. In the proposed *Bulletin*, *completeness* is addressed as follows (OMB, 2006):

> Although scientific completeness may entail analysis of different health effects and multiple target populations, the search for completeness will vary depending upon the nature of the assessment. In a fault-tree analysis of nuclear power accidents, an aspect of completeness may be whether pathways to accidents based on errors in human behavior have been addressed as well as pathways to accidents based on defects in engineering design or physical processes.

We suggest that the OMB's statement that:

> …the search for completeness will vary depending upon the nature of the assessment.

Should be clarified and expanded. For example, if an assessment is partially completed, perhaps because of paucity of information or because an experiment is not yet reported, then its unbiasedness must be assessed and reported. It is insufficient to fill-in the blanks with assumptions, perhaps grounded on conservatism or on some version of the precautionary principle, because doing so may obfuscate transparency (Ricci, Rice, and Ziagos, 2003). The effect of partial completeness of the information, as a function of missing or otherwise unavailable information, and the impact of these on the risk estimates, must be assessed and reported. In this context, for instance, completeness can be defined as that amount of scientific guidance that is sufficient for the purpose of regulatory risk assessment.

The OMB adds that incompleteness can be a result of the fact the:

> Affected entities can include populations, subpopulations, individuals, natural resources, animals, plants or other entities. If a risk assessment is to address only specific subpopulations, the scope should be very clear about this limitation. An analytic product may be incomplete when it addresses only risks to adults when there is information suggesting that children are more exposed and/or more susceptible to adverse effects than are adults.

This distinction seems to mix *heterogeneity* and *causal modeling* with *completeness* (which we take to be a general aspect of any risk assessment). In other words, the fact

that an assumption is incorrect (e.g., assume homogeneity, when heterogeneity governs) does not rise to the level of completeness of an analysis, but it does question the accuracy of the risk predictions (Day, 1985). Importantly, the OMB's statement implicitly considers the equitable distribution of risk (heterogeneity being a special case). If the OMB wished, it could consider *completeness* as it relates to the canonical risk model we began with. If so, completeness could be parsed as:

1) *Modeling completeness*: consistency in the best available science described by mathematical and statistical models,
2) *Technological completeness* and *boundary completeness*: consistency with scenarios that may be unusual a priori but that, a posteriori predictable from historical or other extrapolations that were not used ex ante,
3) *Data completeness*: consistency with the several available studies that provide positive and negative results in animal or epidemiological results,
4) *Analysis completeness:* consistency with best probabilistic and statistical practices, including estimation,
5) *Default, bias* and *representation completeness*: substitutions of defaults by probabilistic representations, analysis of biases and causes of biases, use of the appropriate measure or structure to represent uncertainty in its possible forms (e.g., randomness and epistemological uncertainty),
5) *Definitional completeness*: use *probable*, *likely*, *possible* and so on according to a specific schedule of formal definitions, wherever the literature indicates a standard usage in the context of risk analysis, or de novo.

Thus, *completeness* in regulatory risk assessments also relates to the budgetary means and constraints to achieve it that; in turn, relates to the value of that information. Because the tools exist to estimate the value of sample information, perfect information, and so on. More guidance could be very useful here as well. These VoIs are consistent with, and rely on, the probabilistic analyses discussed throughout this work.

OMB's *Standards Related to Characterization of Risk*

The OMB is concerned with the accuracy and precision of risk results. It states that:

> Every risk assessment should provide a characterization of risk, qualitatively and, whenever possible, quantitatively. When a quantitative characterization of risk is provided, a range of plausible risk estimates should be provided. Expressing multiple estimates of risk (and the limitations associated with these estimates) is necessary in order to convey the precision associated with these estimates. (Citations omitted)

To be sure, risk analysis, which uses qualitative and quantitative information, involves different scales of measurement and statistical methods of analysis. It follows that risk assessors must explain and account for violations of the assumptions inherent to using each method (e.g., for statistical methods: non-normality, dependence, small samples, heterogeneity, incomplete specification of causal models, and so on), completeness of the data (e.g., account for missing data, near singularities, etc.), and the propagation and fusion of the uncertainties (e.g., through Monte Carlo simulations, Bayesian networks, or other method) in the network of random variables that are comprised in each risk assessment (Pearl, 2000; Jensen, 1996). In this context, the assumptions of stochastic independence must be assessed very thoroughly, relative to possibly irrelevant components, because the multiplication of probabilities over the inflated components of the causal chain will decreases the overall estimate of risk unduly. Moreover, it is

particularly important to make sure that the proper (theoretically correct) condizionalization) is used in any risk network or chain. Some guidance is needed to be able to deal with those multiple estimates, particularly if there is a combinatorial explosion of estimates. Moreover, if these estimates do not dominate cardinally, the OMB should provided criteria (e.g., first order stochastic dominance) for resolving issues that arise in this context.

The OMB (2006) summarizes Congressional guidance on these points as follows:

> In the 1996 amendments to the Safe Drinking Water Act (SDWA), Congress adopted a basic quality standard for the dissemination of public information about risks of adverse health effects. Under 42 U.S.C. 300g – 1(b)(3)(B), the agency is directed ''to ensure that the presentation of information [risk] effects is comprehensive, informative, and understandable.'' The agency is further directed ''in a document made available to the public in support of a regulation [to] specify, to the extent practicable— (i) each population addressed by any estimate [of applicable risk effects]; (ii) the expected risk or central estimate of risk for the specific populations [affected]; (iii) each appropriate upper-bound or lower-bound estimate of risk; (iv) each significant uncertainty identified in the process of the assessment of [risk] effects and the studies that would assist in resolving the uncertainty; and (v) peer-reviewed studies known to the [agency] that support, are directly relevant to, or fail to support any estimate of [risk] effects and the methodology used to reconcile inconsistencies in the scientific data.'' These SDWA quality standards should be met, where feasible, in all risk assessments which address adverse health effects.

In §42 U.S.C. 300g – 1(b)(3)(B) the Congress has a clear policy view of the role of risk assessment (which explicitly consists of both as analysis and evaluative judgments) and of the implication of unequal and inequitable distribution of risks. Although in some situations the distribution of risks and benefits may not be as clear as under the SDWA, the OMB should provide administrative and scientific guidance beyond what the Congress promulgated.

We find that the OMB's *Bulletin* can provide such guidance. For example, it does not define the upper or lower bounds of the risk estimates (recollect that a risk is a probability) and thus leaves the choice ambiguous, although it cold have provided specificity using probabilistic numbers (e.g., using well established levels of significance, such as $\alpha = 0.05$ or $\alpha = 0.01$, depending on the importance of the errors associated with the statements to which these probabilities apply). We also suggest that the agency provide additional guidance to meet the Congress':

> v) peer-reviewed studies known to the [agency] that support, are directly relevant to, or fail to support any estimate of [risk] effects and the methodology used to reconcile inconsistencies in the scientific data.

An example of quantitative guidance would be to use meta-analytic tests and diagrams, such as funnel plots (Ricci, Cox, and MacDonald, 2006). These methods provide information that can be used in assessing certain biases in the "peer-reviewed studies known to the [agency]" by accounting for the "file-drawer" problem, and other issues.

OMB's *Standards Related to Objectivity*

The OMB (2006) states that risk assessment results:

... must be scientifically objective, neither minimizing nor exaggerating the nature and magnitude of the risks. On a substantive level, objectivity ensures accurate, reliable and unbiased information. When determining whether a potential hazard exists, weight should be given to both positive and negative studies, in light of each study's technical quality. The original and supporting data for the risk assessment must be generated, and the analytical results developed, using sound statistical and research methods.

We have already discussed aspects of completeness operationally to achieve it. The cited statement requires additional explanation because it states what competent risk assessors should know and face routinely. The issues that require guidance include, for instance, how to access negative results that may not be published or be available, due to the bias towards positive results in the peer-reviewed literature, the unwillingness to share original data, and so on. On the other hand, statistical methods have increased in number and complexity, thus requiring constant skill development. Theoretical -- mechanistic -- understanding of the causal basis of disease are increasing at an increasing rate, demanding familiarity with molecular biology, epidemiology, cell biology and other disciplines that may not easily be available to develop the sort of causal understanding of exposure and response required even in *influential* risk assessments. All of these instances of difficulties suggest that the OMB should develop the minimum requirements of soundness for research and analysis.

To achieve *objectivity*, the OMB (2006) proposes that:

[b]eyond the basic objectivity standards, risk assessments subject to this Bulletin should use the best available data and should be based on the weight of the available scientific evidence. The requirement for using the best available scientific evidence was applied by Congress to risk information used and disseminated pursuant to the SDWA Amendments of 1996 (42 U.S.C. 300g-1(b)(3)(A)&(B)). Under 42 U.S.C. 300g–1(b)(3)(A), an agency is directed ''to the degree that an agency action is based on science,'' to use ''(i) the best available, peer-reviewed science and supporting studies conducted in accordance with sound and objective scientific practices; and (ii) data collected by accepted methods or best available methods (if the reliability of the method and the nature of the decision justifies use of the data).'' Agencies have adopted or adapted this SDWA standard in their Information Quality Guidelines for risk assessments which analyze risks to human health, safety, and the environment. We are similarly requiring this as a general standard for all risk assessments subject to this Bulletin. (Citations omitted)

The OMB's intentions are commendable. Because a risky choice presumes a comparison (a limiting case is a choice against the *do nothing* action), what weights should be used in determining the best data or results? What is the criterion and what stopping rule should be used in making the choice? But, without clear instructions, the discussion is inchoate. What is *best available science*? How is the available scientific evidence weighted and by whom? What is an accepted method or best available method? Who says with credibility that it is acceptable? How is credibility established? How does an agency justify a choice over another -- keeping in mind the earlier admonition of objectivity? Clearly, *best* is an absolute but the OMB provides no discernible way formally and replicably to determine it.

It is well established that consensus and peer reviews can be erroneous and biased. Consensus-based methods have been demonstrated to lead to incorrect results either because of framing issues (Kanheman and Tversky 1979; Granger-Morgan and Henrion, 1991) or for theoretical reasons (Ricci, Cox and MacDonald, 2004), or both. Without

such guidance, how can the public be able make informed and accurate judgments? The OMB believes that the (lay and scientific) public can assess as follows:

In addition … information must be presented in proper context. The agency also must identify the sources of the underlying information (consistent with confidentiality protections) and the supporting data and models, so that the public can judge for itself whether there may be some reason to question objectivity. Data should be accurately documented, and error sources affecting data quality should be identified and disclosed to users. … A risk assessment report should also have a high degree of transparency with respect to data, assumptions, and methods that have been considered.

Objectivity is sine qua non for risk-based regulatory decisions because of the pervasive use of defaults, conjectures, arbitrary safety factors, and the difficult union of practical judgments. The result of some regulatory risk assessments can be seen as essentially a form of expedience in the center of the law and science. This can cause confusion between an obvious aspect of objectivity – such as that invoked by the OMB -- and the perhaps subtle *lack* of objectivity, such as that can be achieved by not reporting negative results or using defaults that go unchallenged. More specifically, if the LNT is used (by the US EPA, for example), its choice and theoretical justification would appear to be objective to most well informed, but non-specialists because it uses experimental and observational data for fitting that model, and has a biological (molecular) basis. Yet, it is – at low dose of interest to regulatory analysis– still a pervasive conjecture. Moreover, if an agency does not recognize other theoretically and empirically persuasive models, such as the *J*- and *U*-shaped models that are not based on conjectures, it is unlikely that lay individuals have the time and expertise (let alone the funds) to reassess a risk assessment and attack an agency's choice of defaults, even though those defaults are part of regulatory guidelines and thus are not legally enforceable standards!

As the US Supreme Court stated in *Chevron USA* v. *Natural Res. Def. Council*, (467 US 837 (1984)), in a case under the CAA:

We have long recognized that considerable weight should be accorded to an executive department's construction of a statutory scheme it is entrusted to administer and the principle of deference to administrative interpretations

"has been consistently followed by this Court whenever decision as to the meaning or reach of a statute has involved reconciling conflicting policies, and a full understanding of the force of the statutory policy in the given situation has depended upon more than ordinary knowledge respecting the matters subjected to agency regulations. (citations omitted)

". . . If this choice represents a reasonable accommodation of conflicting policies that were committed to the agency's care by the statute, we should not disturb it unless it appears from the statute or its legislative history that the accommodation is not one that Congress would have sanctioned." (Citations omitted)

In this case, the Court also concluded that:

Judges are not experts in the field, and are not part of either political branch of the Government. Courts must, in some cases, reconcile competing political interests, but not on the basis of the judges' personal policy preferences. In contrast, an agency to which Congress has delegated policymaking responsibilities may, within the limits of that delegation, properly rely upon the incumbent administration's views of wise policy to inform its judgments. While agencies are not directly accountable to the people, the Chief Executive is, and it is entirely appropriate for this political branch of the Government to make such policy

choices - resolving the competing interests which Congress itself either inadvertently did not resolve, or intentionally left to be resolved by the agency charged with the administration of the statute in light of everyday realities. When a challenge to an agency construction of a statutory provision, fairly conceptualized, really centers on the wisdom of the agency's policy, rather than whether it is a reasonable choice within a gap left open by Congress, the challenge must fail. In such a case, federal judges - who have no constituency - have a duty to respect legitimate policy choices made by those who do. The responsibilities for assessing the wisdom of such policy choices and resolving the struggle between competing views of the public interest are not judicial ones: "Our Constitution vests such responsibilities in the political branches." (Citations omitted)

The Chevron case also prohibits a lower court from substituting its judgment for that of the federal agencies, even if the lower court's judgment is more reasonable than the agency's (Richards, Smith, and Kritzer, 2006). This (*Chevron*) deference by the US Supreme Courts still holds (see, e.g., *US* v. *Haggar Apparel, Inc*., *cert*. to US S. Ct., No. 97-2044, 1999, resulting in the judgment by court below being vacated and remanded). We think that deference reinforces the requirement that the OMB be more explicit in its guidance. The issue here is that, from the beginning to the end, the process of risk assessment is strictly a component of the regulatory process (Ricci and Gray, 1998; Ricci and Gray, 1999). As such, it is a legal process that uses science but can go beyond science in way that is often neither obvious nor intelligible without specific guidance.

OMB's *Standards Related to Critical Assumptions*

The OMB addresses the use of assumptions (often some essential scientific conjectures as when a dose-response model is chosen by default) in risk assessment. The standard is (OMB, 2006):

Risk assessments should explain the basis of each critical assumption and those assumptions which affect the key findings of the risk assessment. If the assumption is supported by, or conflicts with, empirical data, that information should be discussed. This should include discussion of the range of scientific opinions regarding the likelihood of plausible alternate assumptions and the direction and magnitude of any resulting changes that might arise in the assessment due to changes in key assumptions. Whenever possible, a quantitative evaluation of reasonable alternative assumptions should be provided. If an assessment combines multiple assumptions, the basis and rationale for combining the assumptions should be clearly explained.

It is not clear what the *reasonable alternative assumptions* might be and how this standard is applied outside judicial proceedings, whence the term has specific meaning (the term *reasonable* is part of objective legal standards). This, in conjunction with other OMB's standards stated in their proposed *Bulletin*, can result in the following logical conundrum. *Completeness*, *best available science* now conflict with the *reasonable alternative assumptions*. The latter can be used to deny or affect the former. Yet, at the same time, they can be useful in expanding the envelope of regulatory default models (for these, see, e.g., Crump (1984) and the US EPA predominant use (in cancer risk assessment) of that approximate solution to the multistage model as the linearized multistage model, which is linear, non threshold) and include *U*- and *J*-shaped dose-response models.

It is not clear how *the range of scientific opinions regarding the likelihood of plausible*

alternate assumptions is developed when the causal models are incomplete. Is the *likelihood* judgmental; what is it conditioned on and by whom? The methods to meet the OMB's *critical assumptions* standard range from elicitations to scenario building (and can include sensitivity analysis or Monte Carlo simulations) but are not limited to these methods.

This standard lacks specificity and leaves to possibly unverifiable or irreproducible judgment the choice of the assumptions for ranking. This problem is exacerbated when, as is normally the case in risk assessment, *critical assumptions* encompass multiple issues and are multidisciplinary. We suggest that the OMB consider more specific language that reflects the variety of possible assumptions, as exemplified below:

- *Theoretical I* -- e.g., when is a biomarker an appropriate predictor of an adverse health effect?
- *Theoretical II* -- e.g., when is a conjectural model (such as interpolation to zero from the responses at the observed data) supplanted by the more recent factual knowledge (e.g., hormesis, as opposed to threshold or the LNT hypothesis), which is amply based on experimental data and biological theory?
- *Probabilistic I* -- e.g., what probabilistic conditioning is used in the analysis?
- *Probabilistic II* -- e.g., what is the rationale for the choice of a specific distribution or distributions in an assessment? Equally important, how are situations where the theoretical distributions are not known, assessed?
- *Statistical* -- e.g., how is heterogeneity analyzed?
- *Mechanistic I* -- e.g., how is human error in coding a critical sub-routine accounted for?
- *Mechanistic II* -- e.g., how is the magnitude of a rare event (in technological risk assessments) excluded from analysis, given that it is an element in the set of potential events?
- *Policy I* -- e.g., why has new evidence not become part of the regulatory risk assessment defaults?
- *Policy II* -- e.g., why rely on seemingly arbitrary factors of safety and their multiplication, when distributions (and the percentiles of the distributions) can be used instead of arbitrary factors of safety that are combined multiplicatively, unlike engineering factors of safety that are calculated from known properties of materials and then related to the load-bearing characteristic of the material?[5]
- *Policy III* -- eg; why does the scientific assessment become constrained by the legal process, given the various standard contained in this proposed *Bulletin*?

Note on Consensus Rankings

Many practical risky choices are based on qualitative assessments, consensus, and the rating of attributes and attribute levels. As Ricci (1985, 2006) and Cox (2002) discuss, mathematical analysis can help identify the limitations of what any risk rating or risk ranking system – qualitative or quantitative – can achieve. For example, suppose that a rating system is to be used to compare two different options, *A* and *B*, to determine which should be ranked higher, e.g., in competing for scarce risk-management resources or in a priority order for regulatory concern or intervention. If the overall rating of risk is to be based on component ratings of several risk components or factors, as in all of the above examples, then how should the overall risk rating of alternatives *A* and *B* depend on the component ratings?

Some properties might include the following (Ricci, Cox, and MacDonald, 2004).

[5] A rationale may be statistical, but it is not articulated in this proposal, nor is it justified.

Axiomatic Properties for Aggregating Component Scores into Final Risk Scores

1. Which of alternatives A and B is rated higher in the overall risk rating should depend only on their component ratings. Thus, the components used to rate risk should be sufficient to do the job: together, they determine whether A is assigned a higher, equal, or lower rating than B.
2. Which of A and B is rated higher on overall risk depends on each of their component ratings. Specifically, if A and B are identical in all respects except that A rates higher or worse than B on one factor (e.g., exposure), then B should not be rated higher than A in the overall risk rating. This property should hold for all the risk components: none of them is irrelevant.
3. If A rates higher (or lower) than B on every component rating, then B should be rated no higher (or lower) than A in the overall risk rating.
4. Risk ratings of A and B should be based only on their own data, i.e., whether A is rated higher or lower than B should not depend on what other alternatives (other than A and B) are also being rated, if any.
5. If one or more component ratings are zero (e.g., for exposure potential or for human health impact potential of exposure), then the overall risk rating should be zero (or "Negligible" in systems with that category).
6. If the rating for a component is uncertain (e.g., if it has a 0.2 probability of being "L", 0.5 probability of being "M", and 0.3 probability of being "H"), then the single "equivalent" rating assigned to that component (i.e., H, M, or L after considering its uncertainty) should not depend on the ratings assigned to the other components.

Such logical relations among the component ratings and the overall risk rating put strong constraints on rating systems. For example, if quantitative ratings are used, then conditions such as *5* and *6* imply that the aggregation formula used to combine component ratings into an overall risk rating must be multiplicative, i.e., the overall risk rating is proportional to a product of its component ratings. Such multiplicative aggregation of quantitative ratings satisfies properties *1* through to *4*. On the other hand, if only qualitative rankings are used for the components, then it turns out that there is no qualitative ranking system that can assign coherent overall risk rankings (meaning complete, transitive rank-orderings with ties allowed) based on arbitrary component rank-orderings in such a way that principles *1* through to *4* are satisfied. Similar limitations hold for aggregating fuzzy (Klir and Folger, 1988) ratings of linguistic labels or scales (e.g., *H*, *M*, *L*, and *N*), depending on how they are formalized. In other words, qualitative component ratings alone, without further mathematical operation (e.g., multiplication) may not contain enough information to be coherently aggregated into an overall qualitative risk rating that is related to them in normatively desired ways.

Another concern is that a risk rating system with only a few possible outcome categories may not produce enough information to assist making a decision if it is not inclusive enough to support effective decision-making. For example, a *3*4* matrix assigning a label of *H*, *M*, *L*, or *N* to each of three components (Hazard, Exposure, and Impact) can provide only a small amount of information (technically, at most six bits of information, equal to the information content of six tosses of a fair coin) to guide decision-makers.

Of the much larger quantities of potentially useful and relevant information collected and entered into such a rating scheme (several hundred bits at a conservative estimate), almost all is lost in aggregation during the rating process. The small fraction that remains (6 bits in this case, or even less if the probabilities of the *12* cells are not all equal) may be insufficient for effective decision-making, which typically requires at least enough information to discriminate among alternatives that have very differently preferred

outcomes. The minimum amount of complexity and information required for a classification system (including a risk rating system) to make few errors can be rigorously analyzed via techniques from information theory and statistical learning theory. A key insight from such formal analysis is that a classification system that lacks enough complexity to discriminate well among essentially different situations may lead to poor decisions, i.e., ratings with high error rates and high expected losses from decision errors.

Rule-based risky decision procedures that do not explicitly identify or optimize the quantitative human health impacts of recommended interventions and classifications run the risk of triggering pre-specified actions (e.g., interventions to withdraw or restrict exposure) that unintentionally do more harm than good, creating unintended adverse consequences for human health. Rational risk management requires comparing the probable consequences of alternative risk management actions and then choosing the available action with the most desirable probability distribution of consequences. Substituting "importance" in human medicine, or other non-consequential criteria, for actual human health consequences as a guide to risk management decision-making, may lead to recommended actions that create far more harm to human health than they prevent. Methodologically, no small number of qualitative labels for risk and its components can suffice to make effective risk management decisions. This claim may be sharpened and formalized in at least the following ways.

- Axiomatic Approach. A minimal requirement for effective decision-making might be that actions with HHI(act) > HHI(status quo), i.e., those that do more harm than good to human health, should not be recommended. (A stronger requirement would be that no act should be selected for implementation if an alternative act with preferred consequences is available.) To decide whether HHI(act) – HHI(status quo) > 0, it is necessary to assess both the human health risks and benefits of the act well enough to decide whether their difference exceeds HHI(status quo). The ability systematically to compare such differences implies that risks and benefits can be represented numerically (on a "difference scale" that is unique up to choice of unit), under well-known conditions [the "axioms of difference measurement" in representational measurement theory (Luce and Suppes, 2001)] for coherent qualitative ranking of differences. Under these conditions, ability to compare differences implies that risks and benefits can be represented quantitatively.

- Error Probability Approach. Suppose that we want to identify acts for which BENEFIT – RISK > 0 with high probability. For simplicity, suppose that RISK and BENEFIT are modeled as independently uniformly distributed random variables (with bounded ranges) for the set of alternative risk management acts being considered. Each act corresponds to a pair of (RISK, BENEFIT) attribute values drawn uniformly from the entire rectangle of possible values. If we use N qualitative labels to classify the RISK and BENEFIT of each act, corresponding to partitioning their continuous ranges of possible values into N contiguous intervals (such as H, M, L), then the error rate in classifying acts with BENEFIT – RISK > 0 will be 0.5/N (since a rectangular grid of N x N cells will separate cases with BENEFIT – RISK > 0 perfectly except along the diagonal of N cells with the same rating level on each attribute, where the error rate is 50%.) To identify "good" actions with an error rate of no more than 5% would require at least N = 10 levels. (On the other hand, with N levels, one has a (1 – 1/N) probability that an act will be correctly classified with certainty.) Similar analyses can be extended to more factors and more complicated decision boundaries, e.g., to quantify the error probability as a function of N when the goal is to determine whether the product of exposure, illness-per-exposure, and consequence-per-illness, factors exceeds a certain level. In general, using too few qualitative labels for the factors leads to excessive error rates.

- Information Theoretic Approach. Even without formal axioms or quantitative analysis of decision error probabilities, it is perhaps obvious that determining whether the net human health benefits of an act achieves at least a certain target threshold (e.g., the net benefit of the status quo or the best act identified so far) requires at least enough information to determine whether any of the following fractions is zero.

Consider several risk factors stated as fractions in a multiplicative risk model. If *any* of those fractions is zero, the corresponding risks or benefits calculated using it is also zero. But, if the fractions are all logically independent, i.e., whether one is zero is not determined or constrained by whether others are, then, there is no way to represent the answers to which ones are non-zero by *any* set of three-level (e.g., *H*, *M*, *L*) or four-level (e.g., *H*, *M*, *L*, *N*) ratings of three components, as there are only $3^3 = 27$ possible configurations of qualitative rating (or $4^3 = 64$ configurations when a four-level rating is used) compared to $2^8 = 256$ configurations of yes-no answers to whether each of the above *8* factors is non-zero). Similarly, there is no way to use a *3*-attribute, *3*-level rating system to show whether each of the last *6* factors is large enough so that their product can exceed a specified level. Even without further refinements, it is clear that *any* such *qualitative* rating system, regardless of the exact design and interpretation of its component attributes and rating scores, will in general be too limited to represent the information needed to decide whether there are positive net benefits from a proposed action, let alone to decide which action is best or whether a given action is worth undertaking. More generally, any qualitative rating system can be interpreted as a classifier (producing qualitative labels as output) and evaluated by comparing the average bits of information required to make a correct risk management decision with high probability (using information theory bounds for classifiers) to the number of bits of information actually provided by the rating system. The calculations illustrate the more general point that insufficiently informative ratings cannot support decision-making with low error rates.

OMB's *Standards Related to Regulatory Analysis*

The OMB (2006) states that:

For major rules involving annual economic effects of $1 billion or more, a formal quantitative analysis of the relevant uncertainties about benefits and costs is required. In this Bulletin, we highlight important aspects of risk assessments useful for regulatory analysis:

1) The scope of the risk assessment should include evaluation of alternative options, clearly establishing the baseline risk analysis and the risk reduction alternatives that will be evaluated. When relevant, knowledge of the hazard and anticipated countermeasures should be understood in order to accurately capture the baseline risk.

2) The risk assessment should include a comparison of the baseline risk against the risk associated with the alternative mitigation measures being considered, and describe, to the extent feasible, any significant countervailing risks caused by alternative mitigation measures.

3) The risk assessment should include information on the timing of exposure and the onset of the adverse effect(s) as well as the timing of control measures and the reduction or cessation of adverse effects.

4) When estimates of individual risk are developed, estimates of population risk should also be developed. Estimates of population risk are necessary to compare the overall costs and benefits of regulatory alternatives.

5) When a quantitative characterization of risk is made available, this should include a range of plausible risk estimates, including central estimates. A "central estimate" of risk is the mean or average of the distribution; or a number which contains multiple estimates of risk based on different assumptions, weighted by their relative plausibility; or any estimate judged to be most representative of the distribution. The central estimate should neither understate nor overstate the risk, but rather, should provide the risk manager and the public with the expected risk. (Citations omitted)

More guidance seems necessary here. The OMB reviews established material for risk assessment. For example, consider the average rather then the median as a measure of central tendency; this is a minor point. The OMB then points to the bias of an estimator (e.g., the mean) with no further guidance. The matter of biased and unbiased estimation is far from trivial. Consider, for example, their point 4, above, which we repeat:

When estimates of individual risk are developed, estimates of population risk should also be developed. Estimates of population risk are necessary to compare the overall costs and benefits of regulatory alternatives.

This is an excursion into the use of damage functions in risk management. Is the OMB differentiating between sample results and inference to a population? In other words, having estimated the sample mean, does anyone truly need to be reminded that inference is what most people would be interested in, such as inferring the magnitude of the (unknown) population mean? The OMB introduces a new term, *plausibility*. It is not defined and the reader is left to her own devices to figure out if the OMB meant *probability* or if the OMB actually meant to refer to Dempster–Shafer possibility theory[6]. It then asks for a method, to be used with that measure, to assess if an assumption is plausible. Whatever happened to probability theory as a plausible means to deal with the OMB's concern?

OMB's Special Standards for Influential Risk Assessments

The OMB adds that *all influential risk assessments should meet certain additional standards*. These are reproduced below (citations omitted).

1. Standard for Reproducibility

Influential risk assessments should be capable of being substantially reproduced. … [T]his means that independent reanalysis of the original or supporting data using the same methods would generate similar analytical results, subject to an acceptable degree of precision. Public access to original data is necessary to satisfy this standard, though such access should respect confidentiality and other compelling considerations. It is not necessary that the results of the risk assessment be reproduced. Rather, someone with the appropriate expertise should be able to substantially reproduce the results of the risk assessment,

[6] Heuristically, a Dempster-Shafer measure is a set of real values located at a single probability mass point: therefore there is uncertainty about the location of those values, unlike a probability mass value that is single-valued. *Plausibility* is a function that provides an upper bound, while a *belief* function provides a lower bound using these measures. Our question is not rhetorical: the OMB provides guidance and thus should use *probability* and possibility *advisedly*.

given the underlying data and a transparent description of the assumptions and methodology.

We think that this standard should apply to most regulatory risk assessments – and not just influential risk assessments unless their aggregate is below the threshold that makes that aggregate less that influential -- for at least *4* reasons. These are: 1) public funds are used, 2) those at risk (or their agents) must be sufficiently informed, 3) reproducibility is required for scientific studies and not just regulatory risk assessment, and 4) the addition (but not their multiplication) of a few insignificant risks can result in a significant one.

…

3. Standard for Presentation of Numerical Estimates

When there is uncertainty in estimates of risk, presentation of single estimates of risk is misleading and provides a false sense of precision. Presenting the range of plausible risk estimates, along with a central estimate, conveys a more objective characterization of the magnitude of the risks. Influential risk assessments should characterize uncertainty by highlighting central estimates as well high-end and low-end estimates of risk. The practice of highlighting only high-end or only low-end estimates of risk is discouraged.

In a sample the range is the difference between the maximum and the minimum values in that sample. If the sample is too small or is the result of an experimental or observational study that was incorrectly designed, the usefulness of the mean and the range become questionable unless the statistical analyses associated with those numbers are provided independently to assess their validity and appropriateness. The mere mechanical characterization that seems to be advocated here is insufficient, unless more guidance is given that deals with outliers, missing data, robustness of estimators, and so on so that the spirit of the standard, namely avoiding overestimation, can be met.

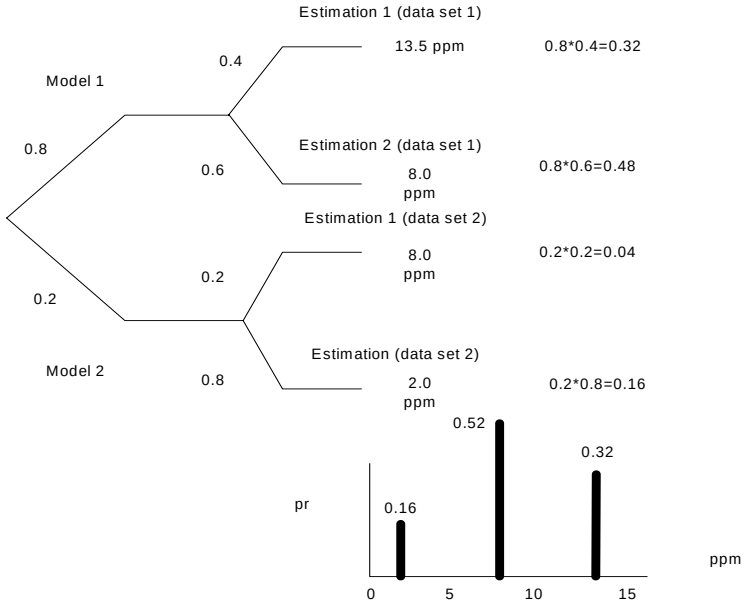
This Bulletin uses the terms "central" and "expected" estimate synonymously. When the model used by assessors is well established, the central or expected estimate may be computed using standard statistical tools. When model uncertainty is substantial, the central or expected estimate may be a weighted average of results from alternative models. Formal probability assessments supplied by qualified experts can help assessors obtain central or expected estimates of risk in the face of model uncertainty.

The statement *[w]hen model uncertainty is substantial; the central or expected estimate may be a weighted average of results from alternative models* raises an important issue. Yet, the answer to the question implicit in this statement requires additional guidance. For example, Bayesian model averaging (Madigan and Raftery, 1994; Cox, 2002; Ricci, 2006).) has been deemed to be a new method that lacks regulatory (but not scientific) credibility. Second, is the OMB suggesting that agencies should weigh Model *1* (M_1) and Model *2* (M_2) and then the data set used to estimate the parameters of M_1 and M_2 ? If so, How and using what sort of weights? One theoretical way for such weighting consists of calculating the *lottery* below in which the probabilities (*pr*) are subjective beliefs (and thus weights) of the realism of these *two* models :

$pr_1 * M_1 + (1 - pr_1) * M_2$,

But this presumes that all models are known to the experts and that those experts can (are willing to) provide their probabilistic judgments (by some encoding procedure, perhaps). We provide a simple example that extends the lottery above:

Example. A risk assessor has used two data sets and two models to estimate the tolerable exposures, measured in parts per million to a toxic agent found in soil. He has used $1*10^{-5}$ as the tolerable risk level, but is unable to determine which of these results is most credible. The risk manager has hired an independent consultant who, based on her experience with the substance, its biological effects and knowledge of exposure-response, has developed the decision tree that follows.



Distribution of exposures at an individual lifetime tolerable risk level 1/100,000, resulting from fitting two different exposure-response models to two different data sets. ppm is parts per million.

On the basis of these results, he recommends to the risk managers that the most plausible exposure level is *8.00* ppm. Is this sort of representation what the OMB has in mind? And, if not, how far into the technical aspects of these analyses should an agency engage to satisfy public policy, judicial review, and the public?

In this example, the probabilistic weights *are* based on scientific judgment that is replicable, open to scrutiny (by questioning the expert who provided those weights), replication by others after the results are achieved, and by changing the number of objects included in the tree as a function of the state-of-knowledge. The unanswered questions include: how were the models and data sets selected, which ones were not selected and why, who selected the weights, and do the weights provide the full characterization required by the regulatory analysis? The OMB's should provide specific guidance so that for transparency and so on are met. These sorts of analyses can be used to portray the full state of the information and knowledge, summarized in probability density or mass functions (of which the lottery is a canonical example), and thus represents the entirety of the uncertainty about a random variable. Moreover, the agencies can use methods such Monte Carlo or other simulation techniques to encompass other relevant forms of uncertainty in a coherent whole.

4. Standard for Characterizing Uncertainty

The OMB (2006) states that:

Influential risk assessments should characterize uncertainty with a sensitivity analysis and, where feasible, through use of a numeric distribution (e.g., likelihood distribution of risk for a given individual,

exposure/event scenario, population, or subpopulation). Where appropriate, this should include sufficient description so that the lower and upper percentiles and the median, mean, mode, and shape of the uncertainty distribution are apparent. When one or more assumptions are used in a risk assessment, the assessor may evaluate how plausible changes in the assumptions influence the results of the assessment. An assumption may be used for a variety of reasons (e.g., to address a data gap or to justify the selection of a specific model or statistical procedure). Professional judgment is required to determine what range of assumptions is plausible enough to justify inclusion in the sensitivity analysis. Sensitivity analysis is particularly useful in pinpointing which assumptions are appropriate candidates for additional data collection to narrow the degree of uncertainty in the results. Sensitivity analysis is generally considered a minimum, necessary component of a quality risk assessment report.

It appears that the OMB does not clearly address the fact that the representation of uncertainty and variability is intrinsic to the term risk, although there are differences between decision-making under uncertainty and decision-making under risk, and ambiguity (Anand, 2002).

There is no doubt that the OMB's ideas are useful and important; we think that, nonetheless, more guidance must be given relative to the scientific treatment of the uncertainty (as variability and uncertainty), as well as to its characterizations under the situation encountered in practical instances. Specifically, because the OMB's guidelines are applicable to several agencies of the US federal government, uniformity of treatment -- as desiderata -- should be enunciated more clearly. The important problems with applied statistical analysis are not in the calculation of the mean or median or other percentiles but, rather, in the development of causal models and estimation of the parameters of those models from possibly messy data that may require robust, non-parametric methods or other methods.

We think that, although the OMB correctly discusses sensitivity analysis as a *necessary* method, by the same token it does not confront the problems that can arise in uncertainty analysis and the biases that result from model and variable specification. The OMB (2006) adds that:

A model is a mathematical representation -- usually a simplified one -- of reality. Where a risk can be plausibly characterized by alternative models, the difference between the results of the alternative models is model uncertainty. For example, when cancer risks observed at high doses of chemical exposure are extrapolated to low doses (i.e., doses below the range of empirical detection of cancer risk), a dose-response model must be employed to compute low- dose risks. Biological knowledge may be inadequate to predict the shape of the dose-response curve for cancer in the low-dose region. While it is common for risk assessors to use a model where cancer risk is proportional to dose (even at low doses), there are cases where it has been demonstrated, through huge epidemiological studies or detailed biologic data from the laboratory, that a non-linear dose-response shape is appropriate. When risk assessors face model uncertainty, they need to document and disclose the nature and degree of model uncertainty. This can be done by performing multiple assessments with different models and reporting the extent of the differences in results. A weighted average of results from alternative models based on expert weightings may also be informative.

Here, the devil is not only in the details but also in their aggregate. The issue regarding the choice of model issues requires clearer guidance and considerable advanced analyses that generally are not done by practical risk assessors. The OMB uses *nature and degree of model uncertainty*. Precisely, what is the *nature* of uncertainty? And, what is the

degree of uncertainty (for example, if that degree is actually a probability number, meaning that a model is more probable than another, with know properties, and thus not a degree)? Or is the OMB referring to the likelihood ratio? If so, what about the inconsistency with Bayesian methods for model choice? In short, more guidance is needed. Specifically, legal acceptability in administrative or other civil law is very different from scientific acceptability. It follows that it is incumbent on the OMB to provide specific scientific guidance and to introduce the appropriate state-of-the-art methods in regulatory analysis, as it recognizes. The advantage of our suggestion is that it maintains consistency of methods in regulatory application throughout federal agencies. Is the *nature* component of the uncertainty presumably related to the correct specification of the model and to the choice of uncertainty measure? Thus, for example, a fuzzy measure of can be used, as can a probability measure (including a hybrid measure), depending on the *nature* of the uncertainty that the assessor faces (using our understanding of the OMB's terms). The *degree* of uncertainty (unless the OMB refers to the magnitude of the measure) might even relate to goodness of fit (as might be appropriate for some statistical models) or to the AIC (the Akaike Information Criterion), or to Bayesian Information Criteria, as criteria to chose between models (Linhart and Zucchini, 1986).

The point being made is that the *nature* of any measure to represent real world uncertainty (which is what the OMB seems to mean) dictates the calculus for analysis: this where guidance is much needed to meet the legal acceptability. Dealing with uncertainty requires understanding the scale of measurement. Take a probability: it is measured between zero and one, included. However, if we wish to compare individual risks and say something to the effect that *A*'s risk is twice that of *B*'s, we can run into a serious issue (unless the probabilities are very small) because probability numbers are not cardinal in the sense that twice the probability *0.5* is *1.00*. Using hazard rates can solve the problem via the transformation $r(pr) = -ln(1-pr)$ in a some situations common to risk assessment. In passing, we note that probabilities can be used with partial knowledge, as upper and lower probabilities (Kofler and Meges, 1976; Pratt, Raiffa, and Schlaifer, 1995; Howson and Urbach, 1993; Hampel, 2001).

Elsewhere, the phrase *well established* appears. Then, another phrase appears: *uncertainty is substantial*. Does this suggest that a weighted average is appropriate or that the most probable model should be used? Suppose that the single-hit and the multi-hit are equally probable (Laplace's criterion): weighing each by *50%* does not really solve the problem. While this Solomonic judgment appear plausible, it can also yield nonsensical results because it combines two different approaches (even though each of them may be based on a common theoretical basis) by splitting the difference. These raise questions such as: Are Bayesian methods (Lindley, 1984; Gelman, Carlin, Stern, and Rubin, 1995), acceptable by the courts? How is the public going to assess expert opinions in prior distributions and who pays for the assistance need to do so?

The OMB states (OMB, 2006):

> Statistical uncertainty sometimes referred to as data uncertainty or parameter uncertainty occurs when some data exist on the value of an input, but the value of the input is not known with certainty. If a sample of data

exists on an input, the degree of statistical uncertainty in the input value is influenced by the size of the sample and other factors. Risk assessors should document and disclose the nature and degree of statistical uncertainty.

Suppose that the sample size cannot be augmented either because of cost, or for some other practical reason, then what? Should the analyst stop and live with possibly very large variability such that the measure of central tendency is smaller that the standard deviation (or standard error of the estimate)? Should she change from using classical statistical analysis and to Bayesian methods? Should she use bootstraps? Should she impute missing data; and, if so what would be the preferred method that meets the standards set by the OMB? In short: Where is the standard so that society can assess, according to the statutes that govern the OMB, the validity of the risk analyses? It is important to recognize that the combination of different estimates, either experts' or samples or both, requires complex methods and inevitably results in a judgmental choice of one method over the other. If there is a stopping rule, What is it?

We have already discussed Bayesian updating, (but other methods such as envelope, mixtures, logarithms pooling or other can also be used, depending on the circumstances). What matters, is that different properties are maintained. For example, Ferson et al., (2003) discuss some of the desirable properties that include:

1. Generality,
2. Closure,
3. Idempotence,
4. Commutability,
5. Continuity,
6. Associativity,
7. Symmetry,
8. Quasi-associativity,
9. Enclosure-preserving
10. Intersection-preserving,
11. Insensitivity to vacuoussness, and
12. Narrowness.

Although the discussion of these properties is beyond this work, we suggest that the OMB's guidelines might benefit from considering at least some of them.

OMB's *Standard for Characterizing Results*

The OMB (2006) also states that:

Results based on different effects observed and/or different studies should be presented to convey how the choice of effect and/or study influences the assessment. Authors of the assessment have a special obligation to evaluate and discuss alternative theories, data, studies and assessments that suggest different or contrary results than are contained in the risk assessment. When relying on data from one study over others, the agency should discuss the scientific justification for its choice.

This is an important idea, although it lacks quantitative content on how to assess different studies and -- more importantly -- how to choose from those studies the one or ones to be used in a risk assessment. The *special obligation*, moreover, is part of the scientific

process but its level of inclusiveness needs additional explanation and quantification. Had the OMB supplied that quantification, it could have provided a very useful contribution. The essence of the problem is that the desiderata stated are well understood. What is needed is actual numerical and qualitative guidance that, unfortunately, is not forthcoming. Sound solutions to the standard for the characterization of results have been discussed in the earlier sections of this paper.

OMB's *Standard for Characterizing Variability*

The standard states that (OMB, 2006):

> A risk is variable when there are known differences in risk for different individuals, subpopulations, or ecosystems. In some cases variability in risk is described with a distribution. Where feasible, characterization of variability should include sufficient description of the variability distribution so that the lower and upper percentiles and the median, mean, and mode are apparent. This section should also disclose and evaluate the most influential contributors to variation in risk. This characterization should reflect the different affected populations (e.g., children or the elderly), time scales, geography, and other parameters relevant to the needs and objectives of the risk assessment. If highly exposed or sensitive subpopulations are highlighted, the assessment should also highlight the general population to portray the range of variability.

This standard -- applicable to sampled data by its wording -- does not provide sufficient guidance for those assessors who must confronts heterogeneous information (possibly represented by multiple modes in the empirical distribution), outliers, and missing or censored data.

Recalling that the OMB's is providing a standard of performance, more guidance is needed to provide an agency with the minimum set of analyses to be performed. As we think it appropriate for the several other standards in this *Bulletin*, the OMB should provide minimum guidance to inform federal agencies. Clearly, the OMB cannot provide all of the methods and techniques, but it should provide the minimum set in a way that integrates across the disciplines used in the assessment of risk. In this sense, the OMB does not have to discuss molecular biology or epidemiology or even how to obtain solutions using Monte Carlo or bootstraps, but it should inform of the range of acceptable methods. That is, from the content of the standards, what is needed is the minimal set of statistical and probabilistic methods that apply to the entire spectrum of disciplines that provide samples or *results* generated by sampling methods used in risk assessment. If dealing with risk is synonymous with dealing with sampled values, it seems obvious that, if the OMB can repeat basic statistics, it should also be able to add some value where it is needed most, while providing for enough uniformity of application truly to provide a standard for characterizing data variability. The heterogeneity argument applies with force to the children and elderly, temporal and spatial distributions of effects and exposures.

OMB's *Standard for Characterizing Human Health Effects*

The standard that follows is unobjectionable as to its contents:

Where human health effects are a concern, determination of which effects are adverse shall be specifically identified and justified based on the best available scientific information generally accepted in the relevant clinical and toxicological communities. In chemical risk assessment, for example, measuring the concentration of a chemical metabolite in a target tissue of the body is not a demonstration of an adverse effect, though it may be a valid indicator of chemical exposure. Even the measurement of a biological event in the human body resulting from exposure to a specific chemical may not be a demonstration of an adverse effect. Adversity typically implies some functional impairment or pathologic lesion that affects the performance of the whole organism or reduces an organism's ability to withstand or respond to additional environmental challenges.

…

Although the language in this section explicitly addresses human health endpoints, for other endpoints, such as ecological health, it is expected that the agency would rely upon information from a relevant group of experts, such as ecologists or habitat biologists, when making determinations regarding adversity of effects.

OMB's *Standard for Discussing Scientific Limitations*

This standard is unobjectionable as to its contents:

Influential risk assessments should, to the extent possible, provide a discussion regarding the nature, difficulty, feasibility, cost and time associated with undertaking research to resolve a report's key scientific limitations and uncertainties.

OMB's *Standard for Addressing Significant Comments*

This standard is unobjectionable as to its contents:

An agency is expected to consider all of the significant comments received on a draft influential risk assessment report. Scientific comments shall be presumed to be significant. In order to ensure that agency staff is rigorous in considering each significant comment, it is typically useful to prepare a "response-to-comment" document, to be issued with, or as part of, he final assessment report, to summarize the significant comments and the agency's responses to those comments. Agency responses may range from revisions to the draft report or an acknowledgement that the agency has taken a different position than the one suggested by the commenter. Where agencies take different positions than commenters, the agency response to comments should provide an explicit rationale for why the agency has not adopted the position suggested by the commenter (e.g., why the agency position is preferable or defensible).

There is little need specifically to comment on the OMB's Sections below. A weakness is that these sections, in particular §VI, should also include the means to update information. For example, it could provide that the minimum acceptable standard for updating information is through Bayesian updating rules or other rules that are more consistent with partial information. Specifically, Bayesian updating rules can account for the value of information (e.g., perfect, sample, and other information-related knowledge) so that the value of any *additional inquiry* (below, §VI) can be formally and replicably assessed. This is not to say that policy choices must be determined by quantitative methods (assuming that probabilistic methods are merely quantitative); rather, guidance assessed by compliance with standards must contain *sufficient* information to permit comparisons in sound, formal and replicable means. The *necessary* aspects of any policy decisions include heuristic choices based on considerations that may not be quantifiable.

A fortiori, the OMB should make sure that when sound methods can be used to inform a decision and select from alternative choices, this task is competently guided. Procedural (in the sense of administrative law) consistency is accounted for by §§ VII and VIII, below.

Section VI: Updates

Influential risk assessments should provide information or analysis, within the intended scope of the assessment, which assists policy makers in determining whether more data needs to be gathered or whether the assessment can be based on the data and assumptions currently available. Since risk assessment is typically an iterative process, with risk estimates subject to refinement as additional data are gathered, it is useful for assessments to disclose how fast the relevant database and assumptions are evolving and how likely it is that the database and assumptions will be significantly different within several months or years. While risk assessments should offer insight into what additional scientific understanding might be achieved through additional data collection and/or analysis, the decisions about whether to invest in additional inquiry, whether to take interim protective steps while additional inquiry is underway, or whether to act promptly without additional inquiry are policy decisions that are beyond the scope of the risk assessment report.

The statement is unobjectionable.

Each agency should, taking into account the resources available, priorities, and the importance of the document, consider revising its influential risk assessments as relevant and scientifically plausible information becomes available. Each agency should (1) have procedures in place that would ensure it is aware of new, relevant information that might alter a previously conducted influential risk assessment, and (2) have procedures in place to ensure that this new, relevant information is considered in the context of a decision to revise its previously conducted assessment. In addition, as relevant and scientifically plausible information becomes available, each agency shall consider updating or replacing its assumptions to reflect new data or scientific understandings.

We provide no comments on §§ VII and VIII, below.

Section VII: Certification

For each risk assessment subject to this Bulletin, the agency shall include a certification, as part of the risk assessment document, explaining that the agency has complied with the requirements of this Bulletin and the applicable Information Quality Guidelines, except as provided in Section VIII.

Section VIII: Deferral and Waiver

The agency head may waive or defer some or all of the requirements of this Bulletin where warranted by compelling rationale. In each such instance, the agency shall include a statement in the risk assessment document that the agency is exercising a deferral or waiver as well as a brief explanation for the deferral or waiver. …

CONCLUSIONS

We posed 3 questions. The short answers are:

1) Can risk assessment-based judgments and methods correctly – i.e., formally, soundly, and replicably -- predict the probability of adverse outcomes to inform risk managers of the soundness of the options developed by risk assessors?

YES: but the OMB has to strengthen its proposed standards as contained in the proposed Bulletin.

2) Can the combination of choice of data and methods used in regulatory risk assessments be shown to be the best possible, given the state of information and knowledge, with what level of probative value, and under what circumstances?

YES: the choice of the literature used in justifying choice of data, models, and so on does guide, but its use in risk assessment should include statistical analysis (according to meta-analysis and other pooling approaches) to assess the bias introduced by exclusion of studies that show null or negative effects. The state-of-the-art of statistical and probabilistic model currently can go well beyond what federal agencies currently do (as it is implicit in the reasons fro the proposed Bulletin).

3) Because risk assessment results combine qualitative choices of data and techniques to assess this data within the scientific method and with scientific judgments: Are those regulatory results equitable on those who are at risk but may not benefit form the activity to be regulated?

Unclear: we have provided some understanding of the issues that affect the equitable distribution of risk, ex ante to selecting and implementing a choice to reduce exposure to one or more hazards, and thus reduce risks. The full answer can only be given when risk-cost-benefit assessments are conducted.

The OMB should revise its proposal by strengthening its *Bulletin* without embarking on a handbook for risk assessment. We have provided comments on each of the relevant standards proposed to help public decision-making. We think that, while this agency has attempted a very difficult task, it has fallen short of its objectives because it has focused on what is known but has omitted to address the difficulties with risk assessments conducted to meet regulatory standards or even guidelines.

We suggest that the OMB may become more ampliative in its several proposed standards to provide the minimum thresholds and leave the technical aspects to the agencies. The public cannot be expected to be able to validate risk assessments conducted by those agencies because the public does not have the funds or the expertise that is available to the agencies. If the remedy is to leave serious disagreements to litigation, clearly the OMB misses the opportunity to give bite to its standards.

REFERENCES & BIBLIOGRAPHY

Anand P., (2002) Decision-Making when Science Is Ambiguous. *Science*, 295:1839.

Andrews J.D., Moss T.R., (1993) *Reliability and Risk Assessment*, Longman, Harlow.

Apostolakis G.E., (1974) *Mathematical Methods of Probabilistic Safety Analysis*. UCLA-ENG-7464.

Applebaum D., (1996) *Probability and Information: An integrated approach*. Cambridge University Press, Cambridge.

Bedford T., Cooke R., (2001) *Probabilistic Risk Analysis: Foundations and methods*. Cambridge University Press, Cambridge.

Calabrese E.J., Baldwin L.A., (2001) Special Issue on the Scientific Foundations of Hormesis. *Toxicology*, 31: 351-695.

Cox, L.A. jr, (2002) *Risk Analysis: Foundations, models and methods*. Kluwer, Boston.

Crump K.S., (1984) Issues Related to Carcinogenic Risk Assessment from Animal Bioassay Data, in P.F. Ricci, L.A. Sagan, C.G. Whipple (Eds), *Technological Risk Assessment*. NATO ASI Series, No. 81, M. Nijhoff Publishers, The Hague.

Day N.E., (1985) Statistical Considerations, Wald N.J, Doll R.(Eds), *Interpreting Negative Epidemiologic Evidence for Carcinogenicity*. IARC Scientific Pub. No. 65, Lyon.

De Finetti B., (1970) *Theory of Probability*. Vol. 1. Wiley, Chichester.

Ferson, S, Kreinovith, V., Ginzburg, L., Meyers, D. S., and sentz, K., Constructing Probability Bozes and Dempster-Shafer Structure, Sandia, SAND2002-4015, Sandia national Laboratory, Albuquerque, NM, 2002.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B., (1995) *Bayesian Data Analysis*. Chapman & Hall, London.

Granger-Morgan M., Henrion M., (1990) *Uncertainty: A guide to dealing with uncertainty in quantitative risks and policy analysis*. Cambridge University Press, Cambridge.

Grimmett G.R., Stirzaker D.R., (1982) *Probability and Random Processes*. Oxford University Press, Oxford.

Hampel F, (2001) An Outline of Unifying Statistical Theory. *II Int. Symposium on Imprecise Probabilities and Their Applications*, Ithaca, NY.

Howard R.A., (1988) Uncertainty about Probability: A decision analysis perspective. *Risk Analysis*, 8:91-96.

Howson C., Urbach P., (1993) *Scientific Reasoning: The Bayesian approach*. Open Court, Chicago.

Jeffrey R.C., (1983) *The Logic of Decision*. University of Chicago Press, Chicago.

Jensen F.V., (1996) *An Introduction to Bayesian Networks*. Springer-Verlag, NY.

Klir G.J., Folger T.A., (1988) *Fuzzy Sets, Uncertainty, and Information*. Prentice Hall, Englewood Cliffs.

Kofler E., Meges G., (1976) *Decision under Partial Information*, Springer, Berlin.

Linhart H., Zucchini W., (1986) *Model Selection*. Wiley, NY.

Lindley D.V., (1984) *Bayesian Statistics: A review*. SIAM, Philadelphia.

Luce R.D., Where Does Subjective Utility Theory Fail Prescriptively? *Journal of Risk and Uncertainty*, 5:5 (1992).

Madigan D., Raftery A., (1994) Model Selection and Accounting for Uncertainty in Graphical Models

Pearl J., (2000) *Causality: Models, reasoning and inference*. Cambridge University Press, NY.

Pratt J.W., Raiffa H., Schlaifer R., (1995) *Introduction to Statistical Decision Theory*. MIT Press, Cambridge.

Office of Management and Budget, OMB, *Proposed Risk Assessment Bulletin*, 2006, Wash DC

Ramsey F.P., (1990) *Philosophical Papers*. Mellor D.H. (Ed.), Cambridge University Press, Cambridge.

Ricci P.F., Molton L.S., (1981) Risk and Benefit in Environmental Law. *Science* 214:1096-2001.

Ricci P.F., Sagan L.A., Whipple C., (1984) *Technological Risk Assessment*. NATO Advanced Study Institute No. 41, Suijthoff, The Netherlands.

Ricci, P.F*, Environmental Health and Risk Assessment and Management: Principles and practices*, 2006, Springer, Dordrecht, NL.

Ricci, P.F., Cox,Jr., L. A., and T. R. MacDonald, Hormesis and Precaution: The twain shall meet*, Hormesis in Precautionary Regulatory Culture*, *BELLE Newsletter*, 14:24 (2006).

Ricci, P. F., Rice, D., Ziagos, J., Cox Jr., L. A., Precaution, Uncertainty and causation in environmental decisions, *Env. International*, 103:1-19 (2003).

Ricci, PF, Gray NJ, 1998.Toxic Torts and Causation: Towards an Equitable Solution in Australian Law (Part I). *University of New South Wales Law Journal*, 21: 787-209

Ibid., 1999. Toxic Torts and Causation: Towards an Equitable Solution in Australian Law (Part II*). University of New South Wales Law Journal,* 22: 155-175

Kahnemann D., and Tversky, A., (1979), *Intuitive predictions: biases and corrective procedures*, TIMS studies in Management Sciences, 12:313-327.

Richards, M. J., Smith, J. L., and Kritzer, H. M., Does *Chevron* Matter? *Law and Policy* (2006)

von Plato J., (1994) *Creating Modern Probability*. Cambridge University Press, Cambridge.