RISK AND UNCERTAINTY ANALYSIS IN GOVERNMENT SAFETY DECISIONS

Elisabeth Paté-Cornell
Department of Management Science and Engineering
Stanford University

ABSTRACT:
Probabilistic risk analysis (PRA) can be an effective tool to assess risks and uncertainties and to set priorities among safety policy options. Based on systems analysis and on Bayesian probability, PRA has been applied to a wide range of cases, three of which are briefly presented here: the maintenance of the tiles of the space shuttle, the management of patient risk in anesthesia, and the choice of seismic provisions of building codes for the San Francisco Bay Area. In the quantification of a risk, a number of problems arise in the public sector where multiple stakeholders are involved. In this paper, I describe different approaches to the treatments of uncertainties in risk analysis, their implications for risk ranking, and the role of risk analysis results in the context of a safety decision process. I also discuss the implications of adopting conservative hypotheses before proceeding to what is, in essence, a conditional uncertainty analysis, and I explore some implications of different levels of "conservatism" for the ranking of risk mitigation measures.

Key Words: risk, uncertainty, probability, conservatism, ranking

1.0 INTRODUCTION

If done consistently and accurately, the quantification of risks (probability and consequences of different outcome scenarios associated with a hazard) allows ranking risk mitigation solutions and setting priorities among safety procedures (Paté-Cornell, 1998). Obviously, this quantification is not always necessary, nor is it the sole relevant input to a balanced decision process, but it is an important one in a world of limited resources when the best option is not obvious because tradeoffs have to be considered. Therefore, the way we describe the sensitivity of a risk to different factors is by computing the effect of the uncertainties about these factors on the uncertainties about human safety and systems' performance. A preliminary sensitivity analysis limited to extreme values is generally used to decide which factors matter in the decision and must thus be included in the analysis.

Unfortunately, risk analysis can seldom be performed on the basis of large statistical databases because they may not exist, and full information may not be available at the time when decisions need to be made. Under those circumstances, the best that can be done is to focus on an accurate representation of uncertainties, a problem for which the Bayesian framework can be most useful. An example of this kind of problem is the choice of policies designed to address the issues associated with global climate change (Paté-Cornell, 1996a). The Bayesian approach implies identifying and structuring a set of possible hypotheses, examining all existing evidence, updating the prior probability of these hypotheses given the evidence, and presenting the risk analysis results along with the quantification of uncertainties (e.g., Press, 1989; Apostolakis, 1990). This process often includes the use of expert opinions. In spite of problems of subjectivity, this use of expert judgment is simply unavoidable to compute probabilities under those circumstances (Morris, 1974; Winkler, 1974; Hora and Iman, 1989; Keeney and Winterfeldt, 1991; Kaplan, 1992; Budnitz et al., 1998).

In the absence of sufficient statistics and of a firm understanding of basic mechanisms, an alternative approach to risk management policies is to require a "zero risk," or to base a strategy on the "principle of precaution" without any attempt to quantify the probabilities and the consequences of the different scenarios. These concepts may be applicable either when the costs are low and the decision is fairly obvious or when the potential consequences are extraordinarily high. But when there are clear resource constraints to the implementation of such policies, better risk assessment methods are needed, even if they include an element of subjectivity.

The results of a risk analysis are generally meant to answer two kinds of questions: is a particular risk acceptable? and what measures can be adopted to maximize safety under resource constraints? The response to the first question cannot always be limited to a simple risk computation and the use of an acceptability threshold. Numerical results, including risk magnitudes and the corresponding uncertainties, generally provide one of the input into decision processes. But to be acceptable, this process must include other aspects of the situation such as the controllability and the voluntariness of the risk (e.g., Slovic et al., 1980; Slovic, 1987).

In a sound decision process, the ranking of options often requires comparisons of both costs and risk reduction benefits. Yet, the current legal framework in the U.S. may not allow explicit cost-benefit comparison, and since we are not infinitely rich, priorities will be set implicitly rather than explicitly. If risk ranking is recognized as a practical necessity and if resource limitations are acknowledged, the maximum overall safety is obtained by ranking the risks using the means of the risk results (i.e., expected value of losses). If expected values are not deemed appropriate, other utility functions can be used to reflect risk aversion. Furthermore, in addition to risk aversion, one may want to use other characteristics of the uncertainty structure to reflect ambiguity aversion (Davis and Paté-Cornell, 1994).

The treatment of uncertainty in risk computations is thus critical to what can be done with the results (Morgan and Henrion, 1990). Several levels of sophistication in the analysis of the uncertainties can be considered according to the circumstances (Paté-Cornell, 1996b). One can simply ask whether the risk exists or not, compute a "worst-case" result, assess a "plausible upper bound" of the risk, use a "best-estimate" approach, or proceed to a Bayesian estimation of the risk and compute loss distributions. In this Bayesian approach, two levels of complexity can be envisioned: a first-order analysis that results in mean estimates of future losses (or a mean "disutility" for these losses) based on "mean future frequencies" of critical events in the face of epistemic uncertainties, or a second-order analysis separating both types of uncertainty that results in a family of risk curves describing the effect of epistemic uncertainties on the overall results. Each type of analysis corresponds to a specific type of decision process and to the intended use of the risk figures in that decision process.

Sometimes, risk analyses have to be performed under "conservative" assumptions required by a particular agency with the laudable (but sometimes misguided) intention of providing better protection to U.S. citizens. Given these hypotheses, a full uncertainty analysis can then be required, leading to a family of risk curves that represent an overestimation of the risk. The problem lies with the deceptive appearance of an actual and accurate representation of all uncertainties when, in fact, conservatism has displaced all curves towards higher probabilities of high consequences. This happened, for example, in the analysis of the risks to human health that may be caused over the next 10,000 years by the Waste Isolation Pilot Plant (WIPP) in New Mexico (Helton et al., 1999). The problem, again, is that these results do not actually represent a proper quantification of uncertainties, but the cumulative effects of quantified uncertainties and conservative estimates of unspecified probability (Paté-Cornell, 1995, 1999). Therefore, in the end, it is impossible to tell how such results could be used to set priorities under any given risk- or ambiguity-aversion conditions.

The case of a few conservative assumptions followed by full uncertainty analysis is only one of many instances where "conservatism" can perturb the ranking of risk benefits and, therefore, where the risk analysis cannot meet its objectives. Ultimately, these objectives are to provide information that allows accounting not only for risk aversion on the part of the decision makers, but also aversion towards ambiguity, and eventually determines the right priorities in the framework defined by the preferences of an elected body. By default, one reasonable and simple version of these goals might be to support the optimal use of limited resources to provide the best protection to the maximum number of people.

This paper briefly presents three examples of probabilistic risk analyses that have been performed in the past in the Stanford Engineering Risk Research Group to reflect the sensitivity of risk results to various factors. It then examines the characteristics of an acceptable decision process for safety decisions in the public sector and the role of quantitative results in that process. This discussion is followed by a description of different approaches to uncertainty representations, their role in decision making under different types of preferences, and how several levels of sophistication in the mathematical treatment of uncertainties can be envisioned so that the risk analysis results are adequate to support these preferences. The paper then identifies the problems that arise with using conservative assumptions before proceeding to a conditional uncertainty analysis for other risk factors and using the results to set priorities in risk management. More generally, it discusses the effects of different levels of conservatism on the ranking of risk mitigation measures. It concludes on the necessity and the nature of consistency in risk analysis methods and results to be able to rank risks, set priorities, and generally support policy making for different kinds of criteria (meeting a threshold of tolerable risk, choice of the most effective risk mitigation options, etc.)

2.0 THREE EXAMPLES OF ENGINEERING RISK ANALYSIS: FINDING AND FIXING SYSTEM WEAKNESSES

The three engineering risk analyses studies that follow have been performed on the basis of probability and systems analysis. There were not enough statistical data at the global level of the whole system to base the results on observed frequencies. However, by decomposing the problem into subsystems and classes of accident scenarios and by accounting for anticipated changes in the systems' evolutions, we could use available information to obtain useful results. What we identified were the weakest points in the systems and the most economical way to fix them. We had some surprises in that often we could not have predicted the results given the number of parameters involved.

2.1 Seismic risk analysis for the San Francisco Bay Area

In that study, the problem was to assess the benefits of reinforcing buildings of different types of use (e.g., residential) and different types of structures (e.g., wood frame) to achieve higher standards than the seismic provisions of the building codes enforced at the time (Paté and Shah, 1980). The study was based on the superposition of two sets of maps. The first one characterized the seismic hazard. They showed zones of ground motions (e.g., peak ground accelerations) not to be exceeded per year with a given probability (e.g., 0.1, 0.2, etc.,) and also some of the existing regional features that could contribute to the damage, for example, the existence of dams or of liquefaction zones. The second set of maps represented, per area, the occupation of the ground in terms of building types (use and structure). The data included the value of the buildings in each category, the human occupancy at different times, and distributions of losses in each type of structure under different ground motion intensities.

For each seismic hazard map (and the associated probability), we computed the corresponding losses (human casualties, property damage, and subsequent economic losses) not to be exceeded annually with the same probability. We then computed the mean value of the losses with and without the proposed seismic code. We could then compare the costs and the expected benefits of the proposed measure. As it turned out, the most vulnerable buildings appeared to be the unreinforced commercial buildings along the edge of the Bay. Focusing resources in that area was showed to be a very good option.

2.2 The tiles of the space shuttle

The objectives of the shuttle study were first, to compute the probability of a shuttle accident caused by a failure of its black tiles; second, to identify the most risk-critical tiles; third, to study the maintenance process to find out what errors could contribute to the loss of tiles (and their probabilities); and finally, to recommend management measures that could reduce the failure risk most effectively (Paté-Cornell and Fischbeck, 1994).

This study, like the previous one, was based on two concepts: the susceptibility of the orbiter (i.e., the probability of losing a tile in the first place) and its vulnerability (i.e., the probability of losing the orbiter given that a tile had been lost). It was based on the first 32 missions, during which only two tiles had been lost without severe consequences. Two classes of "initiating events" were defined: a first tile is lost because (1) it is hit by a piece of debris or (2) it has not been properly glued to the felt pad that connects it to the aluminum surface of the orbiter. The orbiter surface was divided into 33 zones with roughly similar characteristics defined by four parameters: density of debris hits, aerodynamic forces (hence the probability of losing adjacent tiles given the initial tile loss), heat loads (hence the probability of a hole in the aluminum skin given a loss of tiles), and the criticality of the subsystems under the skin in different locations.

We found that the risk of losing a shuttle because of a tile failure was about 10% of the risk of a shuttle accident, but not quite as high as some astronauts feared at that time. We also found that about 85% of the risk was attributable to 15% of the tiles. We provided NASA with a map of the orbiter that showed the criticality of different tile zones so that priorities could be set in the last-minute inspection of tiles before flights. We also found out that errors during the maintenance of the tiles could be a risk factor, and we traced the probability of an error back to a number of management problems. For example, we found that there was an unusual turnover among the technicians because their wages were lower than those of electricians and machinists. We also found that time pressures were leading to short cuts, and we realized that some of them were caused by competition and poor communications among some of the contractors. Furthermore, we found that an important source of debris was in the insulation of the external tank, which was processed in another space center; therefore the link between the vulnerabilities of the two systems had not been properly addressed. We made a number of recommendations to NASA and showed, for instance, that the risk of a shuttle accident caused by the tiles could be reduced by about 70% at little cost.

2.3 Anesthesia patient risks in modern western hospitals

In that study, the challenge was to link the risk to healthy patients (e.g., during knee surgery) caused by anesthesia in modern hospitals and the management factors that affect the performance of anesthesiologists (Paté-Cornell et al., 1996). First, we divided the risk scenarios among different accident sequences caused by different types of initiating events (e.g., anesthetic drug overdose, or disconnection of the ventilation tube). We then modeled the dynamic evolution of two systems in parallel. The first “system” is constituted of the anesthesiologist and other operating room participants. The second one is the patient whose state evolves in parallel. We used a Markov model involving “super states” characterizing both at the same time. For example, the state of the patient following a tube disconnect can evolve from normal state to hypoxemia to cardiac arrest. Meanwhile, the problem must be observed, diagnosed, and fixed by the operating room crew to allow the patient to move towards recovery. We used an existing database from Australia to assess the probability of each accident initiator, and we used expert opinions combined with statistics to assess the probability of transition among states per time units of 10 seconds. The Markov model that we used assumed that this transition time is exponentially distributed which was not unreasonable. We found that the risk to the patient (in the order of 10^{-4} per operation) was about equally distributed between accidents starting with a ventilation problem and those caused by anesthetic drugs.

We then linked the parameters of the model (e.g., the probability of the initiating events and the time to detection and correction) to the state of the anesthesiologist in terms of alertness and competence. Finally, we considered a number of organizational factors that can affect the state of the practitioner, including the supervision of residents, the level of training and competence of practicing anesthesiologists, as well as screening for substance abuse among the practitioners. We found that the most effective safety measures were not the latter because substance abuse, although it had caused several visible and outrageous accidents, is relatively rare and difficult to detect among anesthetists. The most beneficial options were more mundane ones such as the periodic retraining (e.g., on a simulator) of people who may have worked in the operating room for years but not often enough to encounter rare events and remember how to handle them. In this case again, the surprise was that the most important causes of accident were not the ones that made the headlines (and motivated the study in the first place) such as substance abuse or a conflict in the operating room, but more common problems that were not receiving as much attention, perhaps because they were too widespread and people were used to them.

3.0 RISK ANALYSIS RESULTS AS INPUT TO AN ACCEPTABLE DECISION PROCESS

Contrary to popular belief, perception is not necessarily reality, especially when risks are involved. The magnitude of the risk and the uncertainties about it are also important input to an optimal allocation of resources (money, time and attention). As described elsewhere, an acceptable decision process involves at least the following elements (Paté, 1983):

- a sound legal basis with clear understanding of individual and societal risks, burden of proof, and treatment of economic effects (including when and how cost and benefit considerations are legally acceptable);
- a monitoring system that allows early detection of chronic problems, hot spots, repeated accidents, clear threats, etc.;
- an information system including a risk analysis with appropriate characterization and communication of uncertainties and assumptions;
- a communication system such that this information can circulate and be fully understood among concerned individuals and organizations;
- a sound criterion for selection of experts and a mechanism of aggregation of expert opinions that reflects the characteristics of the problem;
- a public review process in which the information used and the risk analysis method can be examined and criticized by members of the public, industry, interest groups, etc.;
- a clear but flexible set of decision criteria that reflect public preferences given the nature of the hazard, the state of the information, and the economic implications of considered regulations;
- an appropriate conflict resolution mechanism (mediation, arbitration, etc.); and
- a feedback mechanism such that data are gathered *post facto* and used in an appropriate and predictable way to measure the regulatory effects, including those that may have escaped initial policy analysis.

A sound legal basis is ideally a clear one that is explicit about the treatment of the costs of regulation. In the U.S., laws have generally evolved in the last 30 years from a “zero risk” concept (e.g., the Delaney amendment), to Best Acceptable (or Practicable) Technologies, to suggestions that cost and benefits be balanced in regulatory decisions, back to the notion that costs should be irrelevant. Swings of the pendulum in this respect have left to the regulatory agencies the task to make practical risk management decisions, sometimes without clear guidance. These include deciding what constitutes a tolerable individual risk, i.e., the *de minimis* level, and the cost-per-life-saved or per-added-life-year. These decisions sometimes sometimes have to be made as if the cost issue did not exist because the U.S. Supreme Court has said so. For individual risks, the *de minimis* threshold below which "*lex non curat*" (the law does not concern itself) often seems to be around one in a million per year (Paté-Cornell 1994). The cost-per-life-saved varies widely from tens of thousand dollars to several millions and at times billions (Teng et al., 1995). In any case, if costs are to be considered (and in practice, they often have to), the law should provide guidance on how this should be done. If costs are irrelevant, the law should specify the *de minimis* threshold. There should be, however, some flexibility in the decision criteria. For example, for industrial structures, required safety levels should be less stringent for existing facilities than for the new ones.

Monitoring mechanisms must be set to alert the public that a risk exists, for example, that tires of a particular model have created an unusual number of accidents or that there is an extraordinary concentration of leukemia among young children in a particular region. Furthermore, the data must be analyzed so as to minimize the probability of errors of both types (false positives and false negatives). It is important, in particular, to be able to recognize and expose false positives that may be the result of politically motivated claims by biased individuals, interest groups, and scientists of questionable credentials. Finally, after a regulation is implemented, and especially when the costs are high and when there are large uncertainties about both risks and benefits, the monitoring of results is essential to provide feedback about the effectiveness of the rules in place.

Most risks, for instance, those involving exposure to carcinogens, include epistemic uncertainties, i.e., incomplete fundamental knowledge. The multiple hypotheses thus have to be weighted in a Bayesian analysis, based on their prior probabilities and on the likelihood functions that characterize the probability of the evidence given each hypothesis (Press, 1989). This analysis often involves the use of expert opinion to assess both the priors and the likelihood functions. The choice of experts is therefore critical, first to ensure that they are qualified, and second that they provide a balanced representation of the spectrum of scientifically supported opinions. This choice may be left to peer groups such as the scientific societies of the different countries, or in the U.S., to the National Research Council. The problem is that the scientific community may be polarized between two groups with divergent opinions and interests. This is the case, for instance, of expert opinions regarding the risks posed by nuclear wastes, global climate change or genetically modified foods. In other instances, the majority of the scientists may be biased, for example, under the pressure of public opinion, or by research funding mechanisms that may favor a majority that may simply be wrong.

The aggregation of expert opinions and the mechanism that is used is also essential to the quality of the result. These mechanisms can be iterative, (e.g., the Delphi method), interactive (i.e., meeting of experts with the objective to identify and structure the hypotheses, and to link evidence and the various hypotheses), or analytical (e.g., a Bayesian integration of expert opinions based on the confidence of the decision maker in each source). It can also be a simple weighted average of expert-provided figures.

It is particularly important that the generation of probability distributions be as independent as possible of value judgments regarding the outcomes. Again, these value judgments have to be left to an elected body whose role is to inject in the process the political preferences of the citizens in addition to the results of best available science. What must be avoided are experts making value judgments according to their own interests and some interested party causing and exploiting misguided perceptions of the risk. Yet, it should be recognized that often, such value judgments are made at the onset of an analysis in order to ensure "conservatism". As discussed further, the problem with this practice (which I widespread in the U.S. as well as in Europe) is that it distorts the results in such a way that they are not comparable across the spectrum of hazards to which the public is exposed.

In any case, it is essential that public opinion be heard even if the popular views do not fit those of the scientists, and to make a clear difference between misconceptions of risk magnitude and differences of values. To eliminate public opinion pressures on matters of science (as opposed to value judgments), the option of a "science court" is a possibility. The advantage is that it requires that the opposing parties defend their viewpoints in the face of an adversarial position. The challenge is to avoid that the opposing sides truncate the evidence to support their interests and arguments in the debate thus leaving aside potentially critical facts that may not fit any of the considered hypotheses (Paté-Cornell, 1996b).

4.0 DIFFERENT APPROACHES TO THE TREATMENT OF UNCERTAINTIES IN RISK ANALYSIS

Uncertainties can generally be divided into two categories: epistemic and aleatory Epistemic uncertainties, as mentioned earlier, result from the lack of fundamental knowledge, while aleatory uncertainties reflect randomness in a well-defined statistical sample (i.e., one in which the probability of a particular factor is "firmly" known such as flipping a fair coin). Probabilistic analysis that allows quantification of risks can be based on the frequentist approach of classical statistics when uncertainty is essentially reduced to randomness. However, when the evidence regarding the fundamental phenomena is incomplete, uncertainty analysis calls for more sophisticated Bayesian methods (e.g., Apostolakis, 1990) that allow for displaying in the results the magnitude and the effects of epistemic uncertainties. The results of a Bayesian analysis can be either a simple mean probability (first-order analysis) or a mean future frequency when the problem can be characterized by a description of the future frequency of a specified event as a random variable. Full display of both types of uncertainties requires a family of risk curves that are shown both in Figure 1 and Figure 2 (second-order analysis). These curves can be very useful, but also sometimes, complicated to generate (e.g., by Monte Carlo simulation) and to interpret.

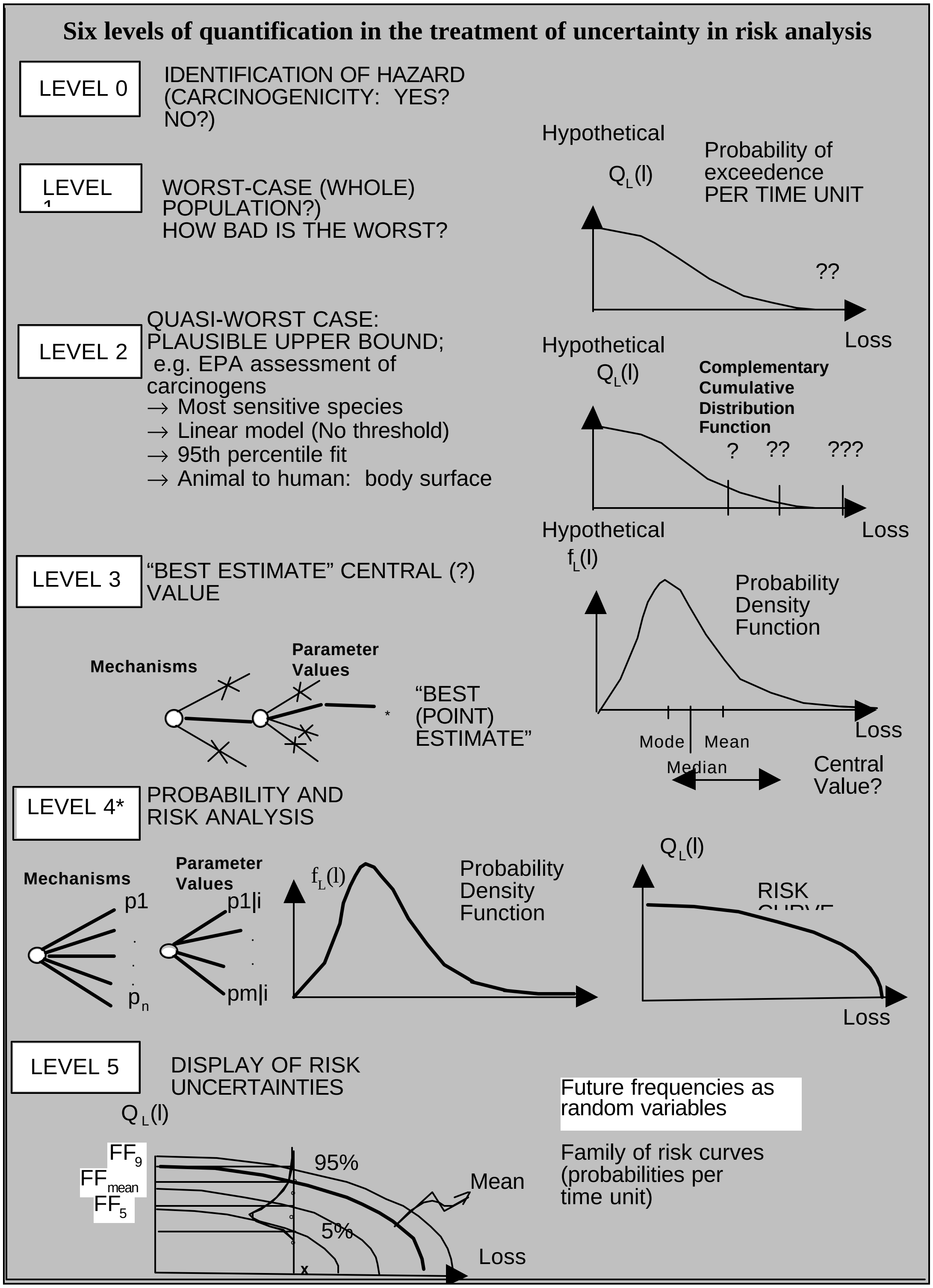

Figure 1. Six levels of complexity in the characterization of risks.
(*Source:* Paté-Cornell, 1996a)

Many problems, however, do not require such complex analyses. For instance, one can identify at least six levels of sophistication in the treatment of uncertainties, each adapted to a particular type of situation (Paté-Cornell, 1996a). As shown in Figure 1, these levels are the following:

- simple identification of a hazard (e.g., is a specified chemical a carcinogen?);
- worst-case approach (e.g., what is the maximum number of potential victims in a specified event?);
- quasi-worst cases and plausible upper bounds (e.g., what is the "maximal probable flood" or the "maximum credible earthquake" in an area?);
- "best estimates" (i.e., what is the "most credible" estimate of the probability of an accident or of losses in an accident in a chemical plant?);
- first-order probabilistic risk analysis based on mean probabilities or future frequencies of events (e.g., what is the probability of exceeding specified levels of losses in different degrees of failure of a particular dam?); and
- second-order probabilistic risk analysis based on full representation and separation of epistemic (fundamental) and aleatory uncertainties (randomness) (e.g., the full uncertainty analysis currently required in the nuclear power industry).

The first approach (hazard identification) is sufficient if the hazard is clearly defined and the solution is simple and inexpensive or if it is so poorly known with such potentially catastrophic results that the benefits of available solutions would dwarf the costs in any case. Other qualitative analyses include methods that lead to the display of risk matrices: on one axis, the probability (high, medium, or low), and on the other axis, the consequences (high, medium,or low) of undesirable events. These are perfectly sufficient when no numerical tradeoff is required. As shown later, this is seldom true.

The second approach (worst-case analysis) is often sufficient when the worst case is clear, and especially if there is a reasonable solution to address the worst case. Unfortunately, no matter how conservative the hypotheses regarding each parameter, one can often identify still worse and unlikely outcomes, calling for solutions that would be impractical. Hence the next approach ("plausible upper bounds" or quasi-worst cases) that represent, in effect, a truncation of the probability distribution of the potential loss distribution. The problem of policies based on such concepts is that one does not know what level of safety they provide and if they are the same everywhere. This raises issues of fairness in the protection of people in different locations.

The problem here is one of consistency and of simply knowing what this truncation implies; for example, "maximum probable floods" as well as "maximum credible earthquakes" may be neither (maximum nor credible) and may have very different probabilities in different parts of the country.

It would thus appear desirable to quantify the risks by estimates of event probabilities and consequences that are "somewhere" in the center of the loss distribution, hence the notion of "best-estimate" analyses. The problem in doing so is that such estimates are often based on the maximum likelihood hypothesis, then for this chosen hypothesis, on the most likely parameter values. Therefore, this approach may lead to an underestimation of the effect on the mean losses (and, in general, on the risk) of a particularly threatening hypothesis with low probability. For example, it may be that the most likely hypothesis of a remotely possible hazard is that nothing occurs; yet, there may be a small chance that it could have a drastic effect and high consequences. Therefore, the results that correspond to such "best estimates" may be anywhere on the (here implicit) density function of the potential losses and are problematic when used in policy decisions.

Probabilistic risk analysis based on Bayesian probability allows accounting for the identified hypotheses. This method, developed in large part in the nuclear power industry, requires structuring them into an exhaustive set of mutually exclusive elements (e.g., USNRC, 1975; Henley and Kumamoto, 1981; Garrick, 1984). Structuring the possible hypotheses may be difficult in cases where they are poorly defined. System complexity may make the exercise extremely difficult, for instance, when identifying the relevant factors is difficult in itself. In this case, however, one may need to simplify the hypotheses description and to group them into manageable sets. The principle that allows problem formulation is that the "scenarios" may not need to (and often cannot) be described in great details, but families or sets of scenarios can be useful to overcome the challenge of complexity. For instance, in the study of the tiles of the U.S. space shuttle previously described, each tile could not be practically examined individually, but they could be grouped into a reasonable number of zones with similar values of the relevant parameters.

The results can be, for instance, a ranking of elements based on their contribution to the failure risk of a specified system or a risk curve representing the complementary cumulative distribution of the losses per time unit in the operation of a particular facility. Each point of this curve (level 4* in Figure 1) represents the mean future frequency of exceeding the corresponding loss level on the basis of the epistemic uncertainties. Therefore, the effects of epistemic uncertainties are represented in aggregation with those of randomness, a characterization that does not allow representing, for example, the spread of results that corresponds to experts' disagreements about the interpretation of the existing evidence.

Representing the spectrum of possible hypotheses requires a higher level of analysis, separating the two types of uncertainties. The results are represented by a family of risk curves showing the effect of epistemic uncertainties on the probability of future loss levels (Figure 2). A cut of this family of curves by a line parallel to the vertical axis provides a discretization of the probability distribution of the probability (or future frequency) of exceeding the corresponding loss level (on the horizontal axis). Each curve, and therefore each intersection point with such a vertical line, represents a percentile ("fractile") of the distribution of that future frequency.

The spread of this family of curves thus represents the degree of epistemic uncertainty in the loss assessment. Note that the mean that has no reason to correspond to any specific fractile may be a different curve.

This last level of Bayesian analysis may be useful for at least two reasons: the event of interest can repeat itself, and the decision maker may want to account for the remaining epistemic uncertainties in his or her preferences (see for example, Fishburn, 1983).

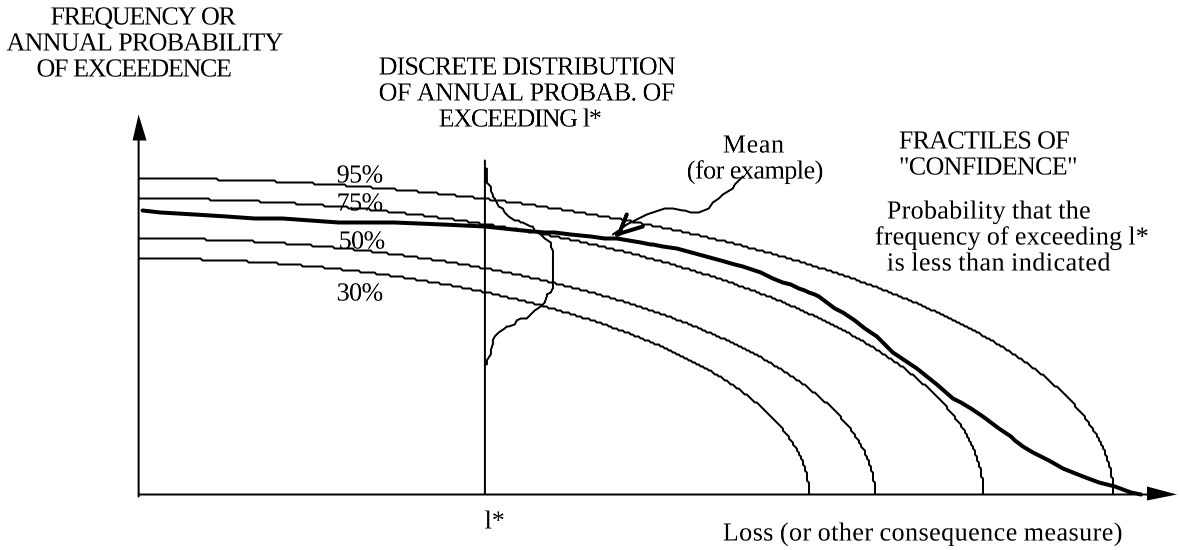

Figure 2. Display of the effect of epistemic uncertainties on the risk by a family of risk curves. (*Source:* Paté-Cornell, 1999)

Consider, for example, an event that can repeat itself (two independent realizations of the same event) and is so poorly known that its probability (or future frequency) can only be characterized by a uniform distribution between 0 and 1 (flat priors; "mean" probability: 1/2). The probability of two independent realizations of this event before any trial is 1/3 (the mean of the square of the U[0, 1]), whereas if the event has a "sharp" and well-established probability of 1/2, the probability of two independent realizations is 1/4 (the square of the mean of U[0, 1]). Alternatively, consider a poorly understood event E that, to the best of our knowledge is conditioned by two equally likely underlying hypotheses H1 and H2 (probabilities 1/2), with the likelihood of event E equal to 0.4 given Hypothesis 1, and 0.6 given Hypothesis 2. The probability of E under these conditions is 0.5, and the probability of two independent occurrences of E in two independent trials considering the possibility of the two different hypotheses is: 0.5 x 0.16 + 0.5 x 0.36 = 0.26 (mean of the square). Instead, it would be of 0.25 (square of the mean) if both randomness and epistemic uncertainty had been aggregated in p(E). Therefore, the separation of epistemic and aleatory uncertainties (randomness) matters in practice to accurately compute the probability of a repeated event before the fact.

It also matters if the decision maker wants to get a feeling for the spread of the probability distribution (either discrete or continuous) of an uncertain event and of its variation with the integration of new pieces of evidence in the risk computation (see for instance, Paté-Cornell and Fischbeck, 1995, for the analysis of the probability of a nuclear attack on the U.S. and of the corresponding uncertainties given the possibility of signal errors in the U.S. command and control system). One reason to consider the distribution of the probability of such a drastic event is that the decision maker may be "ambiguity-averse" in addition to being risk averse (see Appendix). This means that he or she will prefer a "sharp" probability of a loss to a dispersed one, even if the mean (future frequency or probability of the event) is the same. In other words, the decision maker may not be indifferent between two lotteries that result in the same distribution of the outcomes, but with different spreads of the probabilities of each of their realizations. This kind of preference does not satisfy the von Neuman axioms of rationality but may be considered as "rational" in a different axiomatic framework (e.g., Fishburn, 1983, Davis and Paté-Cornell, 1994).

The choice of a method of risk analysis (with or without quantification) thus depends on the desired type of consistency (if consistency matters at all) and also, on tradeoffs among different factors that may affect the risk or the decision in opposite ways. It also matters when costs have to be balanced against benefits, or when the individual risk has to be shown to be below a certain threshold of tolerance. The nature of the analytical method also affects the comparability of the risk results. Other types of consistency can be defined in legal terms, but the economic problem and the individual fairness problem may remain.

5.0 RISK COMPARISON AND THE RELEVANCE OF DIFFERENT RISK ANALYSIS METHODS

As discussed earlier in Section 3.0, the search for an acceptable level of risk should focus on an acceptable decision process. Clearly, the cost of saving a life is only one of many aspects of a risk mitigation policy. Yet, an investigation of risk tolerance in practice, across several fields and several countries, reveals an emerging pattern (Paté-Cornell, 1993, 1994). For example, if the risk inflicted on an individual by someone else is in the order of 1/100 (either per year or per operation), it is unacceptable and probably out of the range of cost-benefit analysis. Below this range of clearly unacceptable figures (e.g., below 10^{-4}), costs and benefits generally enter the picture one way or the other. This brings the important point that costs and benefits should matter, *if they are judged relevant*, only in a specific range of individual risks. This range lies below a figure that characterizes a risk that is "tolerable" compared to those that one has to face in life (e.g., 10^{-4}) and above a *de minimis* threshold where the risk is negligible (e.g., 10^{-7}). Figure 3 shows a possible structure of such criteria that seems to emerge in practice. But again, in addition to cost and benefit issues and as illustrated in the literature, risk acceptance also depends on, among other factors, the level of voluntariness, controllability, familiarity, and (epistemic) uncertainty.

Focusing now on the quantification of the risk, two estimations of different risks (whether individual or societal) based on different methods of treatment of uncertainties, such as those described in Section 4.0, *cannot* generally be directly compared. For example, a "plausible upper bound" estimate cannot, in general, be compared to a risk that has been actually and statistically observed, or to a risk estimation based on computed mean future frequencies. Indeed, in the case where the plausible upper bound is based on an accumulation of conservative assumptions, one can generally conclude that the former is smaller than the latter. Obviously, one problem is that when one uses a presumably plausible upper bound for the risk assessment to compute the cost-effectiveness of a safety measure, one obtains a plausible (?) lower bound of the actual cost per life saved or per added life year.

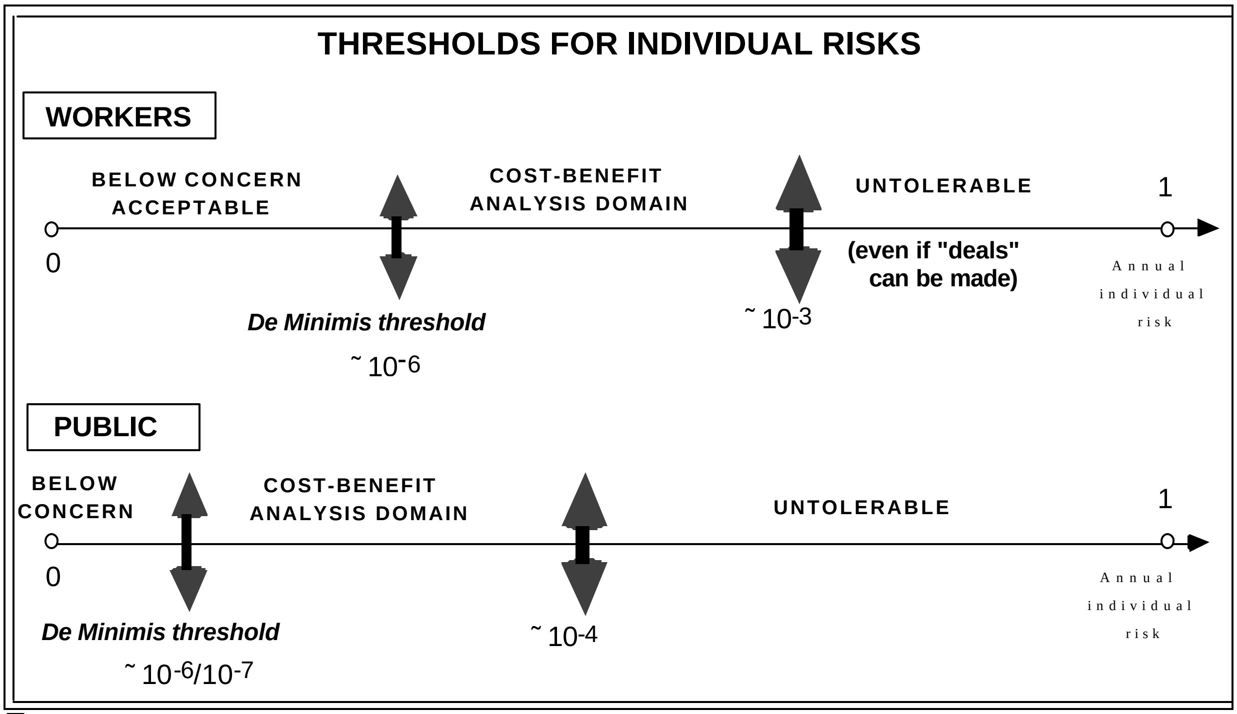

Figure 3. An emerging structure of risk acceptance criteria and safety goals.
(*Source:* Paté-Cornell, 1994)

For example, consider a hazard for which the plausible upper bound of the risk has been estimated at 10,000 potential victims based on linear extrapolation of exposure of sensitive mice to high doses of a toxic substance, but for which a mean based on a probabilistic analysis (including expert opinions about the existence and the magnitude of a threshold) is 25 per year. Consider now the desirability of a safety measure that can eliminate this risk at the cost of $10 billion per year. The cost-per-life-saved based on the plausible upper bound estimate is $1 million, which makes it generally acceptable by the current standards in the U.S.; but it is $400 million, which is hardly cost-effective in the current world given the other life saving opportunities that may exist. How this option compares with alternatives also meant to save human lives depends not only on the risk result but also on the method of computation.

Of course, these alternative opportunities vary from one country to another because, in order to set priorities among life saving measures, the current safety and health situation as well as the level of economic resources matter. This implies that health and safety regulations cannot be blindly transferred from one economy to the next without careful examination of that context.

In a study of 500 life saving interventions (Teng et al., 1995), it was shown that large discrepancies existed among the different criteria by which the standards had been set. Upon further inspection, it appears that different approaches to risk estimations had been used, including plausible upper bounds, risk computations based on mean future frequencies, and losses observed in the past. Such discrepancies among risk analysis methods make the results difficult to compare, except in cases where the lower bound of the cost of mitigation of one risk is clearly much higher than the corresponding expected value for another. Then, the difference between the two levels of cost-effectiveness is only increased by the overestimation of cost-effectiveness for the former and by the underestimation of the costs of protecting the public from it. Therefore, the ranking is correct.

6.0 THE PROBLEMS OF CONDITIONAL UNCERTAINTY ANALYSIS

Conditional uncertainty analysis is a more sophisticated form of "conservative" analysis, which can also lead to a misallocation of resources because it may inflate the results of a risk assessment. A conditional uncertainty analysis is an uncertainty analysis based on a few conservative estimates for some critical factors (therefore, there is no uncertainty analysis for those factors) followed by an in-depth uncertainty analysis for all other factors.

The problems of comparing a plausible upper bound and a mean future frequency are obvious: one is an overestimation or a high fractile of the probability distribution of an event frequency, the other is a mean and the net results cannot be directly compared for risk ranking. Less obvious is the effect of performing a conditional risk analysis given a number of conservative assumptions because, like in a proper Bayesian uncertainty analysis, the result is a family of risk curves that *appear* to represent all uncertainties. But because these "conservative" assumptions lead to an overestimation of each of the probabilities of exceeding specified levels of losses per year, this whole family of risk curves is displaced towards the upper right corner of the figure. Therefore, they are somehow misleading if they are not clearly labeled as conditional analyses and if they are interpreted as full representation of all the uncertainties involved.

For example, this is the case of the uncertainty analysis performed for the Waste Isolation Pilot Plant (WIPP) in New Mexico (Helton et al., 1999). In that study, a number of conservative hypotheses had been imposed by the U.S. Environmental Protection Agency (USEPA, 1996). A team of analysts based at Sandia National Lab did an in-depth uncertainty analysis conditional on these hypotheses. The results, however, do not represent what they seemed to in the sense that the n^{th} fractile as read on the risk curves probably corresponds to lower fractiles of the risks results (probability distribution of the future frequency of exceeding different levels of losses) that would have been obtained if a full uncertainty analysis had been performed for these very factors for which conservative estimates had been imposed. Figure 4 (Paté-Cornell, 1999) shows the effect of restricting the risk analysis to a single realization of one hypothesis concerning a particular, poorly known phenomenon.

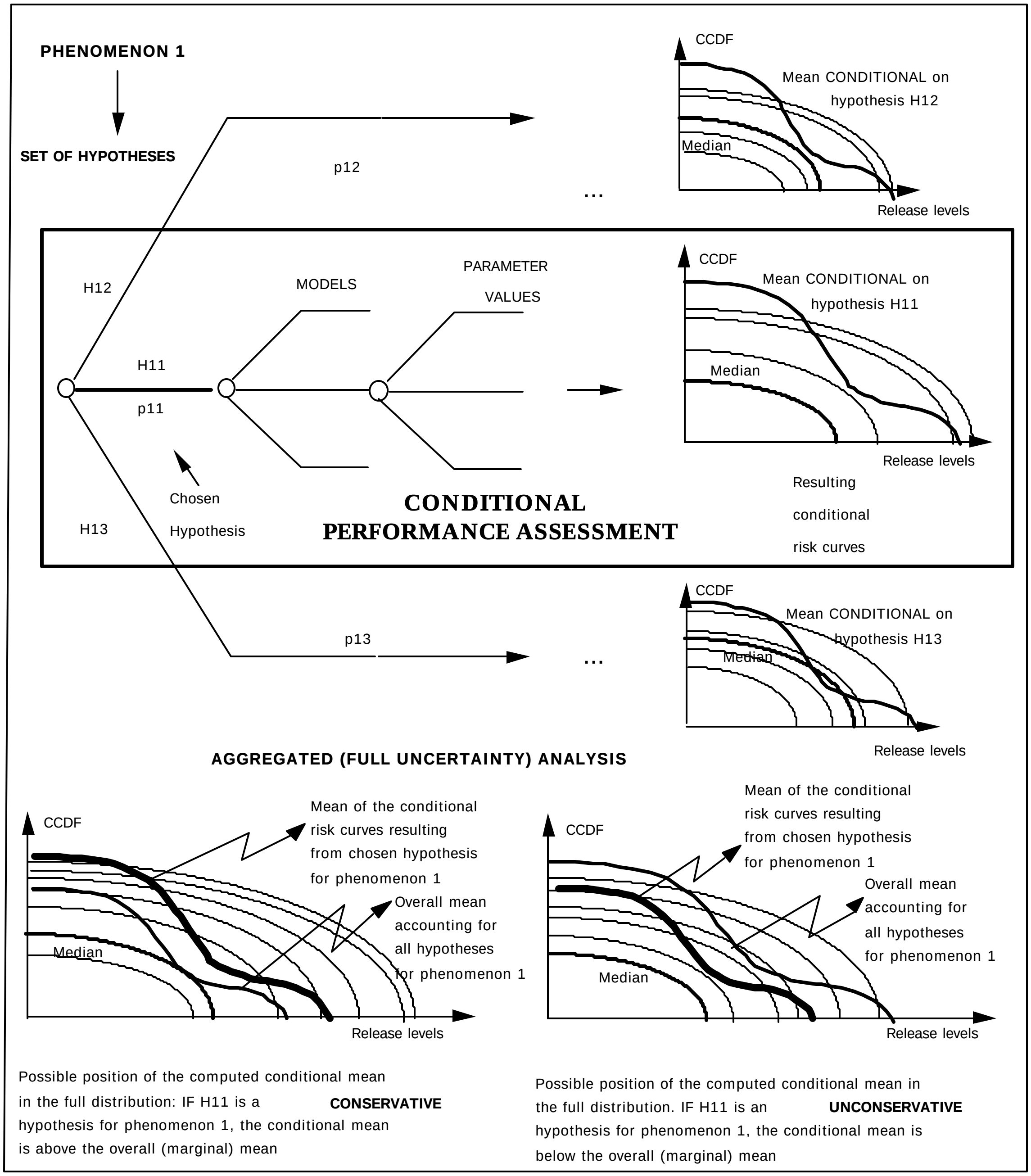

Figure 4. Effect of conservative estimates on the results of a conditional risk analysis. (*Source:* Paté-Cornell, 1999)

It shows that if the conditioning hypothesis is conservative, the risk curves are displaced towards higher values, but in the opposite case, the risk curves are displaced towards lower values.

Therefore, the risk curves provided by a conditional uncertainty analysis such as the one that was performed for WIPP have to be interpreted as conservative results (if indeed the hypotheses are conservative) instead of those of an actual full uncertainty analysis. Again, these results are difficult to interpret and to compare with those of other studies which may not have been based on the same hypotheses or on the same level of conservatism.

Note that the effects of conditional risk analysis can go either way and that optimistic results have also been imposed in the past on some studies for other political reasons. For example, the initial studies for the NASA space shuttle were performed under optimistic assumptions imposed by NASA for the probability of failure of the Solid Rocket Boosters (Paté-Cornell and Dillon, 2001).

In any case, politically driven assumptions (whether conservative or not) lead to results that are simply wrong. They are useful only if that does not matter. In any case, they provide a basis for comparison among alternatives only if one clearly dominates the other.

7.0 CONCLUSIONS

Risk analyses are generally performed for two reasons: to check that individual or societal risks are acceptable and to assess the benefits of various risk mitigation measures with the objective to optimize the allocation of scarce resources for the maximum safety benefits. The latter implies that the risk analysis results are comparable so that the order of priorities is right. Difficulties in assessing risks in the face of epistemic uncertainties are unquestionable. In that case, the use of expert opinions with the subjectivities that they imply is unavoidable. Furthermore, it is clear that the magnitude of the risk is only one of the elements of the problem. Other factors, including social characteristics such as voluntariness and controllability of the risk, also affect the rankings and preferences.

If it is decided that for legal and/or political reasons money is no object, then risk analysis is irrelevant and political forces alone will drive the decisions. The process may then lead to raising unnecessary fears and wasting scarce resources, or ignoring important problems. If priorities have to be set and if it has been decided that the risk magnitude matters, then the analysis has to be properly done. All assumptions (conservative or not) have to be described clearly and results can be compared only to the extent that they have been computed on comparable bases. Bayesian methods can then be useful to account for the possibility of all conceivable basic hypotheses (and their probability based on available evidence) when a health or safety policy has to be made before all uncertainties have been eliminated.

REFERENCES

Apostolakis, G. (1990); "The concept of probability in safety assessments of technological systems," *Science*, 250: 1359-1364.

Budnitz, R. J., G. Apostolakis, D. M. Boore, L. S. Cluff, K.J. Coppersmith, C. A. Cornell, and P. A. Morris, (1998); “Use of Technical Expert Panels: Applications to Probabilistic Seismic Hazard Analysis,” *Risk Analysis*, 18(4): 463-470.

Davis, D. (1990); “The effect of ambiguity on lottery preferences,” Ph.D. dissertation, Department of Industrial Engineering and Engineering Management, Stanford University, Stanford, CA.

Davis, D. and M. E. Paté-Cornell (1994); “A challenge to the compound lottery axiom: a two-stage normative structure and comparison to other theories,” *Theory and Decision*, 37: 267-309.

Ellsberg, D. (1961); “Risk, ambiguity and the Savage axioms,” *Quarterly Journal of Economics*, 75: 643-669.

Fischbeck, P. S. (1991); “Epistemic uncertainty in rational decision making,” Ph.D. dissertation, Department of Industrial Engineering and Engineering Management, Stanford University, Stanford, CA.

Fishburn, P. C. (1983); “Ellsberg revisited: a new look at comparative probability,” *Annual Statistics*, 11: 1047-1059.

Fishburn, P. C. (1991); “On the theory of ambiguity”, *International Journal of Information and management sciences*, 2(2): 1-16

Fishburn, P. C. (1993); “The axioms and algebra of ambiguity”, *Theory and Decision*, 34(2): 119-137.

Garrick, J. B. (1984); “Recent case studies and advancements in probabilistic risk assessment,” *Risk Analysis*, 4(4): 267-279.

Helton, J. C., D. R. Anderson, H. N. Jow, M. G. Marietta, G. Basabilvaso (1999); “Performance assessment in support of the 1996 compliance certification application for the waste isolation pilot plant,” *Risk Analysis* 19(5): 959-986.

Henley, E. J. and H. Kumamoto (1981); *Reliability engineering and system safety*, Englewood Cliffs, N.J.: Prentice Hall.

Hora, S. C. and R. L. Iman (1989); “Expert opinions in risk analysis: the NUREG-1150 methodology,” *Nuclear Science and Engineering*, 102: 323-331.

Kaplan, S. (1992); “Expert information vs expert opinion,” *Reliability Engineering and System Safety*, 35: 61-72.

Keeney, R. L. and D. von Winterfeldt (1991): “Eliciting probabilities from experts in complex technical problems,” *IEEE Transactions on Engineering Management*, 38: 191-201.

Morgan, G. and M. Henrion (1990):*Uncertainty,* New York, N.Y.: Cambridge University Press.

Morris, P. A. (1974); "Decision analysis expert use," *Management Science,* 20(9): 1233.

Paté, M. E. and H. C. Shah (1980); "Public policy issues: earthquake engineering," *Bulletin of the Seismological Society of America*, 70(5): 1955-1968.

Paté, M. E. (1983); "Acceptable decision processes and acceptable risks in public sector regulations," *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-13(3): 113-124.

Paté-Cornell, M. E. (1993); "Risk management for existing facilities: a global approach to numerical safety goals," *Applied Mechanics Reviews*, 46(5): 242-245.

Paté-Cornell, M. E. (1994); "Quantitative safety goals for risk management of industrial facilities," *Structural Safety*, 13(3): 145-157.

Paté-Cornell, M. E. and P. S. Fischbeck (1994); "Risk management for the tiles of the space shuttle," *Interfaces*, 24: 64-86.

Paté-Cornell, M. E. (1995); "Conservatism of the performance assessment and decision criteria for WIPP," Report to the Office of the Attorney General of New Mexico.

Paté-Cornell, M. E. and P. S. Fischbeck (1995); "Probabilistic interpretation of command and control signals: bayesian updating of the probability of nuclear attack," *Reliability Engineering and System Safety*, 47(1): 27-36.

Paté-Cornell, M. E. (1996a); "Uncertainties in global climate change estimates," *Climatic Change*, 33: 145-149.

Paté-Cornell, M. E. (1996b); "Uncertainties in risk analysis: six levels of treatment," *Reliability Engineering and System Safety*, Vol. 54, pp. 95-111.

Paté-Cornell, M. E., D. M. Murphy, L. M. Lakats and D. M. Gaba (1996); "Patient risk in anesthesia: probabilistic risk analysis, management effects and improvements," *Annals of Operations Research,* 67: 211-233.

Paté-Cornell, M. E. (1998); "Risk comparison: uncertainties and ranking," in *Probabilistic Safety Assessment and Management* (Proceedings of PSAM4) A. Mosleh and R.A. Bari Eds., 3: 1991-1996, New York N.Y.: Springer Pub.

Paté-Cornell, M. E. (1999); "Conditional uncertainty analysis and implications for decision making: the case of the waste isolation pilot plant," *Risk Analysis*, 19(5): 995-1002.

Paté-Cornell, M. E. and R. Dillon (2001) (in press): "Probabilistic risk analysis for the NASA space shuttle: a brief history and current work," *Reliability Engineering and System Safety*.

Press, S. J. (1989); *Bayesian statistics: principles, models, and applications*, New York, N.Y.: John Wiley & Sons.

Savage, L. J. (1954); The foundation sof statistics, New York, N.Y.: John Weiley & Sons.

Slovic P., B. Fischhoff, and S. Lichtenstein (1980); "Facts and fears: understanding perceived risks," in *Societal Risk Assessment*, R. Schwing Ed., New York, N.Y.: Plenum.

Slovic, P. (1987); "Perception of risk," *Science*, 236: 280-285.

Teng, T. O. et al. (1995); "Five-hundred life-saving interventions and their cost-effectiveness," *Risk Analysis*, 15(3): 369-390.

U.S. Environmental Protection Agency (USEPA) (1996); "40 CFR Part 194: criteria for the certification and re-certification of the waste isolation pilot plant's compliance with the 40 CFR part 191 disposal regulations; final rule," *Federal Register* 61: 5224-4245.

U.S. Nuclear Regulatory Commission (USNRC) (1975); *Reactor Safety Study: Assessment of Accident Risk in U.S. Commercial Nuclear Plants*, WASH-1400 (NUREG-75/014). Washington, D.C.

Von Neuman, J. and O. Morgenstern (1947); *Theory or games and economic behavior*, Second Edition, Princeton N.J.: Princeton University Press.

Winkler, R. L. (1974); "Combining probability distributions from dependent information sources," *Management Science*, 27(4): 479-488.

APPENDIX: AMBIGUITY AVERSION vs. RISK AVERSION

The von Neuman axioms of rationality (von Neuman and Morgenstern, 1947) have been designed for a single decision maker and expressed in several equivalent forms (e.g., Savage, 1954). Their implication is that the rational decision maker wants to choose the alternative that maximizes his or her expected utility.

One of these axioms, for example, Savage's "compounded lottery" axiom, states that the rational decision maker is indifferent between a sequence of lotteries and a single lottery representing their compounding using Bayesian rules. This implies, in particular, that the nature of the uncertainties involved in each of these sequential lotteries does not matter. Both epistemic uncertainty (about fundamental phenomena) and aleatory uncertainty (randomness) are characterized by Bayesian probability. The result of this compounding includes a single probability distribution for the outcomes and the corresponding probability distribution of the decision maker's utility for the outcomes. Therefore, according to this theory, the rational decision maker is indifferent between two lotteries that result in the *same distribution of outcomes*, regardless of the "softness" and of the "pedigree" of these distributions. For example, he or she must be indifferent between a "firm" lottery based on two sequential flippings of a fair coin, and a "softer" lottery that results from the compounding of "rain tomorrow" (to which the weatherman has attributed a probability 0.5) followed by a flipping of a fair coin. This is true regardless of the nature of the outcomes of these two compounded lotteries. Independently from this fundamental characteristic of rationality, the decision maker can be risk-averse, risk-prone or risk-indifferent. His or her risk attitude is characterized by the concavity (or convexity, or linearity) of the utility function. Again, the only things that matter at the time of a rational choice are the probability distributions of the outcomes and of their utilities after compounding of the probability distributions of all relevant factors.

One may challenge, however, this compounding axiom and argue that rational decision makers may feel differently about lotteries involving uncertainties of different nature (therefore, some "firmer" than others), even though these lotteries may be characterized by the *same distribution of outcomes and utilities* (Davis, 1990; Fischbeck, 1991). For example, a rational individual may have a higher "certain equivalent" (selling price or value) for a firm lottery based exclusively on aleatory uncertainties than for a softer one based on the compounding of epistemic uncertainty (lack of basic information) and aleatory uncertainty (randomness). This phenomenon is often referred to as the "Ellsberg paradox" (Ellsberg, 1961). One can reject this attitude as "irrational" or, on the contrary, find it quite reasonable and treat it systematically and consistently (e.g., Fishburn, 1991, 1993).

The axiomatic treatment of ambiguity aversion (or preference) requires modifying the compounding lottery axiom, for example, giving different "weights" to lotteries of different pedigrees. We have approached this problem by introducing a second utility function in the computation of the value of an alternative (e.g., Davis and Paté-Cornell, 1994). Instead of computing a simple expected utility, we have separated the two parts (epistemic and aleatory parts of the problem) and used two separate utility functions to represent the way some decision makers may want to "discount" softer lotteries compared to firmer ones. The solution that we proposed is for the rational (but ambiguity-averse) decision maker to maximize a nested function (Expected U_1[Expected U_2]) of these two utilities, one characterizing the risk attitude (U_2) and the other the attitude towards ambiguity (U_1). Note that according to the classic von Neuman or Savage axioms, the rational decision maker can be as risk-averse or risk-seeking as he or she wants, but is always ambiguity-neutral and therefore that his or her objctive is simply to maximize Expected U_2.

One practical illustration of this theory is the choice between two reinforcement systems of a nuclear reactor, one improving human performance (considered poorly known) and the other equipment performance (considered better known). In that case, the firmer lotteries involved made the latter more attractive to the risk-averse *and* ambiguity averse decision maker, even though, according to the classical axioms of rationality, the same degree of risk aversion (but ambiguity indifference) would have lead to the choice of the former (Fischbeck, 1991). Another illustration was a computatio of the risks of a nuclear attack on the U.S. territory given imperfect signals from sensors of the Command and Control system. We computed the probability distribution of the probability of attack itself (instead of its mean alone) to allow an ambiguity-averse decision maker to consider the softness of the information, in addition to the probabilities themselves, in any critical decision (Paté-Cornell and Fischbeck, 1995).