Att… ney General of New M… ico

PO Drawer 1508
Santa Fe, New Mexico 87504-1508

505/827-6000
Fax 505/827-5826

TOM UDALL
Attorney General

MANUEL TIJERINA
Deputy Attorney General

September 13, 1995

RECEIVED
SEP 19 1995

Environmental Protection Agency
Docket No. A-92-56, Air Docket
Room M-1500 (6102)
401 M Street, S.W.
Washington, D.C. 20460

To The Docket:

Enclosed is a report prepared for this office by Prof. Elisabeth Paté-Cornell, entitled Conservatism of the Performance Assessment and Decision Criteria for WIPP. Prof. Paté-Cornell is Professor of Industrial Engineering and Engineering Management at Stanford University and is currently President of the Society for Risk Analysis. She has written and lectured extensively on probabilistic risk assessment and has testified in Congress on proposed legislation on the subject.

The report constitutes further comments on the proposed compliance criteria, 40 CFR Part 194, and should be examined fully and carefully. Briefly, Prof. Paté-Cornell concludes as follows:

1. Generally, the 1992 performance assessment ("PA") in fact constitutes a conditional risk assessment, predicated on certain fundamental assumptions (made by EPA) as to the linkage between radionuclide releases and health effects and other assumptions (made by DOE or its contractor, Sandia) as to the probability and consequences of various release events. Whether the EPA assumptions or the DOE assumptions are conservative as judged by the outcome of a full probabilistic risk assessment of the WIPP repository is not known.

2. In inquiring whether the PA curve deemed determinative of compliance meets Prof. Paté-Cornell's standard--that high fractiles of the future frequency of exceedence of potential loss levels should be required to meet the performance criteria with a high level of confidence--it is important to know (a) what fractile of the current CCDF distribution the suggested curve--the mean--corresponds to, and (b) what fractile would the mean correspond to if some of the assumptions of the PA were instead treated probabilistically?

3. In pursuing the same inquiry, it is also important to know where the mean would fall if methods other than the use of expert opinion were used to obtain probability distributions for input variables.

4. Concerning the specific issue of the selection of experts for purposes of expert judgment elicitation, such persons should be required to meet a test of recognition by their peers in the scientific community.

5. Concerning elicitation of expert opinions on parameter values, the process must include the elements of (a) clarity of question, (b) identification of desired central value--probably the mean--and (c) the description of the thought process leading to the estimate.

6. Concerning the selection of variable parameters for PA, the test should be whether the variation of an input value across the possible range could change the final decision.

7. DOE should justify its decision to treat variables whose distribution is critical to the results through expert opinion rather than through experiments or measurement where feasible.

8. Concerning elicitation of expert opinions as to distributions of variable parameters, the process must include (a) construction of a probability distribution for a set of possible hypotheses, (b) identification of the appropriate distribution model for an identified model variable, and (c) given such model, identification of the distribution for the value of the variable.

9. Concerning aggregation of expert opinions of multiple experts, the process must include methods to reduce the range of disagreement, such as requirements that all experts (a) agree on the substance of the question, (b) consider and account for all available data, and (c) articulate the relationship between the data and their judgment as to the probability of the various models. Further, to aggregate different opinions, it is preferable to employ an interactive process wherein the experts (1) discuss the data, (2) explain their models, (3) discuss the probability of each of the models, (4) assess such probabilities, and (5) generate a composite distribution. Aggregation of multiple opinions must be performed systematically as to all expert elicitations. The task of quantifying the uncertainty of alternative assumptions cannot be ignored.

10. The rationality of the mean as a relevant characteristic of a probability distribution does not apply to collective decisions (such as governmental decisions), in which the administrator is concerned not only with the probability distribution of the levels of release but also with the health and safety of the most exposed members of the public, which involves

the choice of a threshold based on prudence. The mean may or may not reflect that threshold, depending on the fractile it represents and the practicality of so demonstrating.

11. A full uncertainty analysis includes (1) structuring alternative hypotheses into realizations so that probability distributions can be assigned to them, (2) aggregation of expert probabilities for each set of assumptions, (3) identification of the models and parameter values (with probabilistic treatment) which correspond to each hypothesis, including interdependencies, (4) propagation of uncertainties for each fundamental hypothesis, and (5) aggregating the results of conditional analyses according to the probabilities of the underlying hypotheses.

12. The full uncertainty analysis of WIPP has not been done and would be extremely difficult. In this situation, it is sensible to apply a test of reasonable expectations to the results of a conditional risk analysis based on fixed hypotheses, provided (1) that the hypotheses are globally conservative and (2) that the mean curves generally correspond to high fractiles of the CCDF families. In such situation the combination of hypotheses and means may provide "reasonable assurance." It must be demonstrated that the global model (health effects plus PA assumptions) is conservative and that a full uncertainty analysis achieves "reasonable assurance." It is appropriate for EPA to find a "reasonable expectation" only if its assumptions as to health effects (including its cancer risk model) provide the additional level of safety consistent with the NRC language of "reasonable assurance."

13. Such demonstration involves identifying the major hypotheses from EPA and DOE and assessing, by analysis of their probabilities and outcomes, their effect on the placement of the current mean curves.

14. EPA cannot simply frame a conditional risk analysis based on certain assumptions and then claim without checking that the conditional means resulting from this analysis necessarily support "reasonable expectation" of human safety. The effects of the hypotheses as to health effects and release models on the mean curves must be assessed. EPA must show that the combination of "reasonable expectation" for the PA and conservatism (if it is so) of the health effect model provides "reasonable assurance" of actual safety.

15. DOE, for its part, must identify the major hypotheses in its PA and show the effects of those hypotheses on the family of release curves. As an example, one can take the five or six most important assumptions of the PA (such as the hypotheses about the frequency, means, and effects of drilling; borehole diameters; groundwater flow model; solubility model; engineered barrier model), generate a set of reasonable alternatives, and show that

the mean curves generated with proper probabilistic analysis of the alternatives show compliance and do not move the mean curves toward lower fractiles of the CCDF families.

16. Depending on how far the current means are (assuming full probabilistic treatment of hypotheses) from a reassuring (but not sacred) 95% fractile, it may be appropriate to ask for additional analysis or a change in risk management strategy.

17. The test of 95% confidence to account for sampling error should be sufficient.

18. It is essential to deal with correlations among variable parameters.

* * *

We have undertaken to draft proposed regulatory language for §§194.26 and 194.34 following the analyses by Prof. Paté-Cornell, and it is attached to our comments, filed today. We request that the Agency consider and adopt the proposed regulatory language.

Very truly yours,

Lindsay A. Lovejoy Jr. (by AS)

LINDSAY A. LOVEJOY, JR.
Assistant Attorney General

LAL:mh

Dr. M. ELISABETH PATÉ-CORNELL

110 Coquito Way,

Portola Valley, CA 94028

Tel: (415) 723-3823
(415) 854-8052
Fax: (415) 725-8799
(415)854-8053
e-mail: ng.mep@forsythe.stanford.edu

Portola Valley, July 5th, 1995

Mr Lindsay Lovejoy
Assistant Attorney General
Office of the Attorney General of New Mexico
P.O. Drawer 1508
Bataan Memorial Building
Santa Fe, New Mexico 87504

Dear Lindsay:

You will find enclosed here my final report to the Attorney General of New Mexico entitled: "Conservatism of the performance assessment and decision criteria for WIPP". I enjoyed my interaction with you in this work and I hope that we will have the opportunity to continue.

Sincerely yours,

M. Elisabeth Paté-Cornell

CONSERVATISM
OF THE PERFORMANCE ASSESSMENT AND
DECISION CRITERIA FOR WIPP

Report to Office of the Attorney General of New Mexico

by

Elisabeth Paté-Cornell

July, 1995

CONSERVATISM OF THE PERFORMANCE ASSESSMENT AND DECISION CRITERIA FOR WIPP

by

M. Elisabeth Paté-Cornell

July, 1995

---

1. BACKGROUND AND PROBLEMS

WIPP has been basically constructed and is scheduled to start operating in 1998. At this time, the remaining operational decisions concern the potential need for engineered barriers, the management of the facility in the future, and the timing of the start of operations. EPA has been required by Congress to certify that WIPP will comply with Federal regulations for the storage of high-level wastes. A "performance assessment"(PA) has been done by Sandia National Laboratories for the Department of Energy which is in charge of the design, construction, and operation of the facility.

This performance analysis is in essence, a *conditional* probabilistic analysis based on *mixed methods* involving both a set of fundamental assumptions provided by EPA and a probabilistic release analysis conditional on these assumptions. (There is no probabilistic risk analysis *per se* because the consequences have been determined by EPA through a single-estimate method, presumably using a conservative model.)

1. <u>The set of fundamental assumptions</u> that have been adopted by the EPA concern mostly the linkage between health effects and radionucleide releases. The figures presented in Table 1, Appendix A of 40 CFR 191 are based on a number of hypotheses that I could not all identify. They involve, for instance, the assumption that the different isotopes are released to a large stream of water. EPA's assumptions also affect the framing of the risk analysis problem. For example, EPA has set the requirements "in terms of cumulative

releases of radionucleides at the accessible environment, either at the ground surface or anywhere at depth, 5 kms horizontally from emplaced wastes, over 10,000 years" (Lee). These hypotheses can generally be assumed to be conservative with respect to health effects, but it may not be the case, and their effects on the overall result has to be checked.

2. Sandia's performance assessment is a conditional analysis of the radionucleide release given the hypotheses and constraints set by EPA. It includes an uncertainty analysis within this framework. This analysis is restricted to uncertainties associated with the distributions of the variables of the conditional release model (such as λ, the mean number of human intrusions in 10,000 years per km^2). This uncertainty analysis does not involve the fundamental assumptions originally set by EPA in the containment requirements: these are taken for granted. Therefore, it does not reflect the uncertainties about the outcome of interest: the health effects of the potential release. A second set of assumptions were made by Sandia in the performance analysis. For instance, some of these distribution models were fixed, such as a Poisson model for human intrusions and a uniform distribution for its mean λ. The value of the parameter(s) of these distribution were either based on past data or on expert opinions (here λ=U[0,30]). The propagation of these uncertainties through the analysis has been performed using simulation (Monte Carlo or Latine Hypercube sampling methods) to obtain a description of the uncertainties about the release levels given the uncertainties about the inputs of the analysis (e.g., solubility factors).

The results of the performance assessment are thus families of risk curves that represent, for each release level, a discretization of the conditional probability of exceeding this value in a specified time window (10,000 years in most cases) given the analytical hypotheses specified by EPA. Note again that Sandia does not address directly the uncertainties about the health effects.

I do not know, given the way the performance analysis was done, whether these conditional results and their implications for health effects are conservative or not. In other terms, if instead of using the EPA assumptions plus additional assumptions of their own about the shape of the variable distributions, Sandia had done a complete uncertainty analysis (i.e., had assessed probabilities for these assumptions), would the curves obtained by this full uncertainty analysis about the release be above or under the current conditional curves? (I recognize, of course, that the uncertainty analysis has to stop somewhere). I can only *presume* that, in the EPA's generic studies that led to the release criteria, the accumulation of hypotheses that are generally intended to be conservative in the first place, lead to conservative results in specific analyses such as that of WIPP.

The question is thus whether the EPA hypotheses are in fact conservative with respect to the WIPP site. One of them, as mentioned above, concerns the release of radionucleides to the environment through a large stream, part of which will provide drinking water to the population. Whether this large stream assumption is conservative or not given that WIPP is in the desert, I do not know. Also, the argument was made that the assessment of cancer risk that led to this table was based on japanese epidemiologic data and that they have been found to be unconservative in later studies (EPA, background info, 1993, p 6-5).

It is important to note that the results of this kind of conditional PRA are not directly comparable to the results that one would have obtained if the EPA and Sandia's assumptions had been incorporated and weighted along with alternative assumptions in a fully probabilistic risk analysis of the health effects. Restricting the scope to release levels alone and to the hypotheses that led to Sandia's current results, Figure 1 shows a schematic representation of the full uncertainty analysis of release (for one single Hypothesis 1), and the restriction of Sandia's analysis to one particular realization of Hypothesis 1.

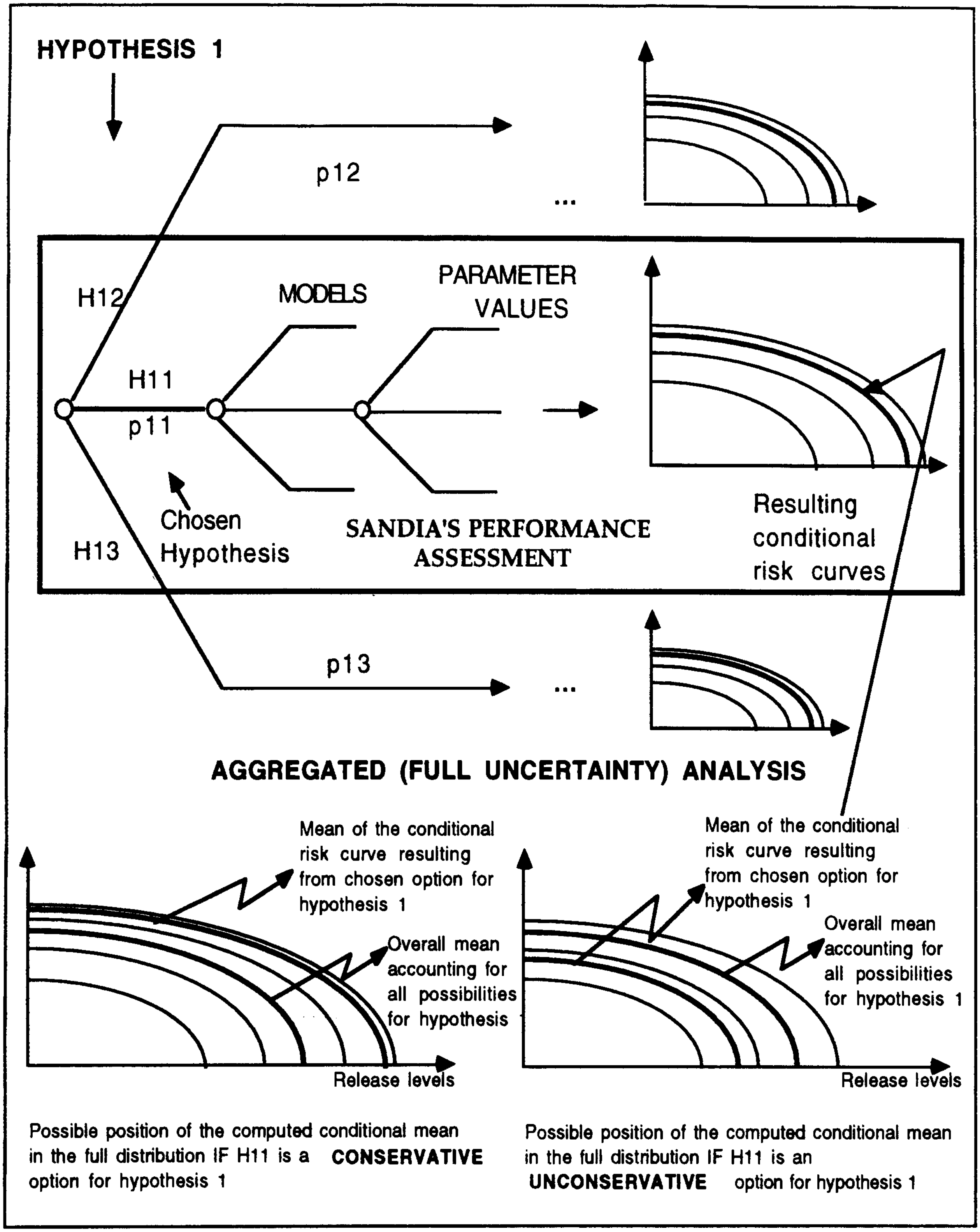

Figure 1: Conditional risk curves (CCDFs), and position of the conditional mean in a full uncertainty analysis.

This would have required using probabilities for the different possible realizations of each hypothesis instead of adopting what is probably (but not necessarily) conservative assumptions. Clearly, a full analysis of this type is complex. It would also involve much larger uncertainties because it would require assessing the probabilities of additional poorly known phenomena. Yet, it may be possible once the conditional analysis is done for the chosen case, to assess (even coarsely) the probable effects of alternative key assumptions on the final results. Therefore, the conservatism of the final safety levels achieved under the proposed criteria (e.g., specifying that the risk curve corresponding to the *mean* must meet the EPA release requirement) has to be examined in the light of the conservatism of the EPA (and later, Sandia's) assumptions.

EPA's compliance criteria involve several components: the assumptions behind the containment criteria, the criteria themselves, the CCDF characteristics (fractiles or moments) to be used to show compliance (second order), and the confidence level that the compliance criteria are met (third order). There are infinite combinations of such choices that lead to the same level of safety. EPA first made its own (single point, presumably conservative) analysis of the link between health effects and release levels. Then, they chose the mean CCDF for its robustness in the face of large uncertainties for the 10,000-year horizon, a set of analytical assumptions for the release model, and a high level of confidence for the mean. The first question is whether the choice of this combination is both prudent and practical. The second question is whether it provides a sufficient level of safety in terms of health effects.

In a 1986 paper, in the Journal of Nuclear Engineering and Design, I wrote an article entitled "Probability and Uncertainty in Nuclear Safety Decisions". In that article, I argue that both qualitative and quantitative safety goals are useful tools, and that *high fractiles* of the future frequency of exceedence of potential loss levels should be required to meet the performance criteria with a *high level of confidence* (hence the accumulation of two layers of conservatism). I assumed in that discussion that the PRA results came out of a full uncertainty analysis *on the outcomes* (here: health effects). I did not specify that a designated fractile should systematically be used (95% is one possibility). I believe that such a goal should remain flexible, depending on the case, alternative risk management options, and the difficulty of showing compliance (uncertainties, time horizon, etc.).

Therefore, I think that you want to know which one(s) of the *real* risk curves (that would result from a full uncertainty analysis as shown in Figure 1) are tangent to the compliance criteria curve, and possibly, what level of health safety do they represent. Two questions thus arise:

° What fractiles of the current distributions (conditional on Sandia *and* EPA hypotheses) do means correspond to in the *current* analysis? The means are shown in Sandia's results but, in each case, they do not correspond to one single CCDF curve on the whole range of release levels. Although the mean (predictably) appears to be in the high fractiles, I cannot tell which ones.

° Where would these means be in the full (marginal) distributions if some of the assumptions of the release model were treated probabilistically? In other words, how do the current assumptions affect the position of the current mean in the family of CCDFs?

The mean has several advantages in many PRA cases. First, it is compatible with economic efficiency criteria. Second, in the face of large uncertainties, the mean is relatively robust compared to specified fractiles (for example, it can be estimated with smaller sample size). Third, in PRAs as performed so far for nuclear power plants, the means are often among the high fractiles of the risk curves (e.g., 70%, 80% or 90%). This is true because many of the distributions that represent uncertainties in the results are skewed right. The position of the mean does reflect the level of uncertainty. Typically, the distance between the mean and the median is one measure of the uncertainty: the higher the level of uncertainty, the higher the fractile corresponding to the mean. Therefore, altogether, for studies involving a very high level of uncertainty, the advantage of the mean is that it is a robust estimate that generally corresponds to high fractiles.

In a recent paper entitled "A Perspective on the 1992 Performance Assessment for the Waste Isolation Pilot Plant" (EEG, 1995), William Lee argues that the analysis has been incomplete. I concur with him, to the degree that the Sandia conditional performance analysis does not allow me to estimate the conservatism of some of the basic hypotheses in the WIPP case. Eventually, the issues are: (1) To what extent will additional uncertainty analysis change the PA results? (2) Is it likely to make a difference in policy decisions given the release criteria as set? And (3) what combination of change in performance criteria and performance analysis would result in a change of risk mitigation measures? Lee also argues that expert opinions play a critical role in the PA results, that they may not have been encoded with sufficient care, and that they may not be appropriate given that experimental data could be reasonably obtained instead (e.g., for solubility). I tend to agree with Lee on this last point. Some parameters can be better estimated. The decision to gather more data depends on the "value of information" and on the difference that experimentation would make in the final decision. I believe, however, that the use of panels to assess what might happen in the distant future is unavoidable and appropriate in a probabilistic framework. Clearly, the results are subjective probabilities. There is nothing wrong with that if the encoding is well done: they are the only ones available for this kind of exercise.

At this point, the concern of the Office of the Attorney General of New Mexico is that the combination of the release criteria as set by EPA, the (conditional) performance analysis as done by Sandia for the DoE, and the compliance requirements proposed by EPA may not provide sufficient conservatism to ensure the long-term safety of the citizens of the state of New Mexico. The central issues are thus:

i. What was the level of conservatism used by EPA in its model linking health effects and release levels?

ii. What is the actual conservatism of Sandia's PA (conditional risk analysis) given the combination of EPA hypotheses and Sandia's choice of distribution models and parameter values for the input variables? In particular: What are the potential problems and possible effects on the PA results of the procedures that were used to obtain probability distributions for the input variables including the choice of distribution models and parameter values, based on expert opinions?

iii. If the proposed compliance criteria are adopted and the case is judged based on the combination of the existing analyses and these criteria, what would be the actual level of safety and with what level of confidence?

iv. Does the Office of the Attorney General of New Mexico want to require (or can require) that EPA issue fractile-based compliance criteria *on the release level* as a general numerical standard?

v. Does the Office of the Attorney General of New Mexico want to require (or can require) that EPA provide a full probabilistic version of its model of the link between radionucleide release and health effects so that one can perform a probabilistic risk analysis (complete with uncertainty analysis) of *the final health effects*?

[Since a large part of the problem relies on the treatment of uncertainties in risk analysis in general and in the Sandia study in particular, you will find in appendix of this report a discussion of this problem based on a report that I recently wrote for the Electric Power Research Institute].

Probabilities are understood in two different ways by different people. For classical statisticians, probability means frequency in very large samples. For the Bayesians, probability is a degree of belief and is updated in a systematic way given each new piece of information.

Uncertainties themselves are also of two different types. The first type is randomness in samples (or aleatory uncertainty to which you refer as stochastic). It can be treated by statistical methods and the frequentists' definition of probability. The second type reflects the limits of fundamental knowledge and can be called epistemic uncertainty (you refer to it as subjective). It cannot be addressed by the frequentist approach to probability. For this second case, one needs Bayesian probability and expert opinions, with the understanding that there are numerous problems associated (1) with the encoding and the validity of this type of information and (2) with the aggregation of expert opinions.

Because of the unavoidable subjectivity of Bayesian probability and expert opinions, some government agencies (such as the EPA) have used, since the late seventies, "plausible upper bounds" of the risks, for instance, for dose-response relationships for carcinogens. These plausible upper bounds are single numbers meant to provide conservative estimates based on an accumulation of worst-case assumptions. This approach, however, has led in the past to regulations that the present Congress found unacceptably costly.

Currently discussed (or recently voted) legislations such as HR 1022 require a "soft" cost-benefit analysis approach to regulation, and therefore, an estimate of central values and a description of uncertainties in addition to plausible upper bounds. EPA, as well as many other governement agencies, is in a state of transition in its approach to risk assessment as they are trying to adapt their methods to this new sensitivity to consistency in rule making. There are several problems with a full probabilistic quantification of health risks. First, the methods of risk analysis are not yet fully developed, especially in environmental and health risk assessments. This is true, for instance, for cancer risk assessment, for which there is no full probabilistic analysis method. This is why I think that it would be very difficult for EPA, at this time, to provide a probabilistic version of the model that they used to set the release criteria for WIPP, especially over 10,000 years. This model would be necessary for Sandia to do a full probabilistic risk analysis of health effects of radionucleide release at WIPP which would be dominated by cancer risks. Second, there is not enough consensus in the scientific community to base risk acceptance and degree of confidence on analytical results alone. Furthermore, the acceptability of a particular risk level depends on many more factors than its computed magnitude.

As I shall discuss further, the issue of aggregation of expert opinions is still unresolved, i.e., there is no consensus about how to do it. Therefore, the agencies (such as the EPA) tend to focus their initial efforts about uncertainty analysis on the development of methods for the quantification of *randomness* in parameter values. As far as fundamental (epistemic) uncertainties, they are not generally ready to incorporate them in a probabilistic risk analysis and still tend to base their risk assessment on specified hypotheses that are generally conservative. Therefore, for the moment, the results tend to be *conditional risk analyses* of the type performed by Sandia for WIPP and not the type of full PRA that is state-of-the-art in the nuclear power industry.

As a result, the curves that are produced in that way are difficult to interpret. This point is at the core of the problem that you have with the WIPP analysis. Yet, Sandia had no choice: the hypotheses had been set for them by the EPA in the generic studies that led to the release criteria. The methods are still in flux and criteria that one may want to adopt for complete probabilistic risk analysis in which *all* uncertainties have been quantified are not the same as criteria appropriate for mixed methods. In any case, showing that expert data have been gathered in a way that is as objective as possible, then properly aggregated is going to be both difficult and necessary.

2. QUESTIONS FROM THE OFFICE OF THE ATTORNEY GENERAL OF NEW MEXICO

It is against this background that I will address the five questions posed to me by the Office of the Attorney General of New Mexico.

Question 1:

How should expert judgment elicitations be conducted?

I interpret this question as: elicitation of "best estimates" and elicitation of probability distributions for either a spectrum of hypotheses or for the numerical value of an uncertain parameter.

Question 2 is made of two independent parts:

How should the judgments of multiple experts be combined?

How should the results (and the uncertainties) be incorporated in the regulatory agency's decision making? (I will answer this second question as part of question five).

Question 3: How should variable parameters be selected?

I interpret this question as: in which case should a parameter value be represented by a "best estimate" (i.e., some central value of the distribution to be determined), and when should the uncertainties about a parameter be represented by a full probability distribution?

Question 4: How should probability distribution functions be developed?

Question 5: When a family of risk curves has been generated (by propagating uncertainties about models and parameter values through the risk analysis model), how should compliance with the containment requirements be determined? i.e., what fractile or other characteristic of the CCDF family should be required to meet the criteria and with what level of confidence?

I shall address these questions *from the point of the view of the risk analyst*. I will leave to the Office of the Attorney General of New Mexico the task to translate the scientific answer into regulatory language, which is out of my domain of expertise. It is also important to note that some issues are objective (the discussion of the soundness and the practicality of a methodology), while others are purely subjective and reflect a desired degree of *prudence*. There is no scientific basis for the latter and one can only approach it from the point of view of consistency and practicality.

Q1. HOW SHOULD EXPERT JUDGMENT ELICITATION BE CONDUCTED?
I will separate the question into two parts: choice of the experts, and elicitation of individual expert opinions. Note that the encoding of a distribution for a single expert is fairly standard. The difficulty is in the aggregation of the opinions of several experts to obtain a composite distribution.

Q1.1 Choice of experts and definition of expertise:

The choice of the experts should be limited to people who have demonstrated scientific competence in the field, and have been recognized by a substantial fraction of the corresponding peer scientific group as part of the scientific expert community for this particular domain. The notion of expertise includes knowledge and understanding of the generally admited theories and of the available base of evidence, and capability to reason about the different hypotheses *given the evidence* (i.e., mastering the scientific method of reasoning about existing data). The demonstration of such competence may have been achieved in different ways: publication in the refereed literature, reasoned support of one or several hypotheses, and contribution to research, development, or practice in the field of interest.

In particular, the sole role of advocate, on political grounds alone, of one view or another is not sufficient to constitute expertise. A scientific understanding of the current evidence base and of the spectrum of possible hypotheses is an essential component of the definition of expertise. Obviously, this definition is not black and white. There is a spectrum of expertise levels based on experience and the ability of an individual to reason scientifically from the evidence base. Therefore, there remains an unavoidable subjective element in the definition of the degree of expertise.

Other issues, such as "no conflict of interest", seem to have been adequatly addressed by EPA and in this respect, EPA language (40 CFR 194.26) generally appears reasonable.

Q1.2 Elicitation of expert opinions for a single point estimate

Clarity test

First of all, questions to the experts must be phrased in such a way that a hypothetical individual who would know the variables with certainty could immediately answer with a single number. This requires that the input variable is clearly defined and that there is no ambiguity such that given perfect information, different values could be given in good faith (in the literature, this is called "the clarity test" Ref. Ron Howard).

Best estimate

Second, if the objective is to elicit a "best estimate", one must understand the thought process by which the expert is going to come up with this figure. Suppose that there are several possible models for this best estimate and several parameter values for each of these models. A simple way for the expert to find a best estimate is to take the most likely model, and for this model the maximum likelihood estimate of the parameter value(s). Note that if the expert does that, the result is unlikely to be equal (or even close) to the mean, and one cannot use this figure as such. Indeed in some cases, e.g., for a remote risk (low probability, high consequences), the most likely mechanism may well be "nothing happens", which does not require any further treatment of parameter values. This process may thus yield an "unconservative" answer which could be inappropriate because of the possibility of severe consequences.

Central values

Central values of the distribution are generally what one wants from the expert.

° The mode (the maximum of a probability density function) is not very helpful because it cannot be easily combined in a risk analysis with other variables (treated either deterministically or by a probability distribution). In other terms, one does not know what the results mean at the end of computations where distributions, means and modes have been mixed.

° The median is more helpful because the experts can think about it relatively easily (variable X is as likely to be larger as smaller than the revealed value). It is not easy, however, to include it in the analysis (i.e., to combine it logically with other variables) except for lognormal distributions of which it is a natural characteristic.

° The mean is the most robust of the central values. But for an expert to come up with a mean sometimes requires a more sophisticated thought process: what are the different possible underlying models for that variable, what is the spectrum of parameter values for each model, and given these, what is the mean that one gets after combining models and parameter values. For skewed distributions, the mean may be driven by extreme values and correspond to high fractiles; in that case, it may not be easy for the expert to assess it directly without analytical support.

Note that for variables for which there is little uncertainty, the mean, the mode, and the median are close enough and the distinction does not matter much.

Q1.3 Elicitation of expert opinions for a distribution

Encoding the probabilities of fundamental hypotheses

Assume first that the issue is to assign a probability distribution to a spectrum of possible hypotheses. The first step is to structure them as a set of exhaustive, mutually exclusive possibilities. The second step is to get the expert to elicit a probability distribution for this structured set of hypotheses. At that stage, most experts require first some training and explanation about probabilities, and for what the figures actually mean. To do the actual encoding, one of the best tools is a "wheel of chance". The expert is asked to divide the wheel into "pie portions" whose relative angles represent the relative probabilities of the different hypotheses. Therefore, when the wheel is spinned, for any hypothetical "lottery", the expert is indifferent between playing the lottery with the wheel, and with the true nature of the phenomenon of interest (as if it were to be revealed). The result thus represents the expert's degree of belief in each hypothesis.

This method is adequate for relatively large probabilities. Very small ones must generally be either decomposed into a sequence of conditional variables whose probablities can be more easily assessed, or based on revealed models.

Encoding a distribution for a variable of the model

To encode the distribution for a model variable X, one generally needs: (1) to identify the appropriate distribution model for X (e.g., normal), and (2) given this model, to encode a distribution for the value of its parameter(s) (e.g., the mean and standard deviation of X). The probability distribution for a parameter value can be obtained in two ways: a non-parametric approach based on the wheel of chance described above (e.g., interval by interval), or by a specified probability distribution (e.g., Normal) for which the expert assesses secondary parameter values (e.g., mean and standard deviation for the mean of X).

It is clear that this process of embedded uncertainty analyses has to stop somewhere. A general rule is to stop when additional information is unlikely to influence the final choice given the decision criteria.

Q2. AGGREGATION OF EXPERT OPINIONS

How should the judgments of multiple experts be combined?

One of the greatest challenges of risk analysis is the treatment of expert opinions when they disagree. Note again, that there is no standard, widely accepted procedure to do it at this time. First, one must understand why the experts disagree. One can then proceed to obtain a family of risk curves that represent, for each value on the consequence axis, a composite distribution reflecting the spectrum of opinions.

Q2.1. Sources of disagreement

They can include semantic misunderstandings, differences in experience and evidence base, fundamentally different mental models to treat the evidence base, and disagreement about parameter values (Ref. Bonduelle). Note that, of course, some of the experts may also want to influence the decision to fit their own value system, and may for instance, choose to ignore part of the evidence in their assessment of probabilities.

° Semantic disagreement is often overlooked. Therefore, one should first check that the variables are precisely defined and understood in the same way by all the experts.

° Bases of evidence can differ entirely from expert to expert. First, different experts may have observed the same phenomenon but in different settings. In addition, someone who has seen only "real-world" data (e.g., epidemiological data) may have gathered information that differ significantly from laboratory results. This is why, in the processing of real-world data, all relevant confounding factors must be taken into account. For laboratory experiments, it is their adaptability to the case *in situ* that has to be questioned. In all instances, the experts should not be allowed to arbitrarily truncate the evidence base to fit their views of what should be done.

° Disagreements about models and parameter values are the most difficult to resolve. The first thing to do is to examine the relationship between the probability of the different models and the complete set of data and evidence (e.g., by Bayesian methods). The second is to decide what approach is required by the level of complexity and the importance of the variable in the final decision.

Q2.2 Different approaches to aggregation of expert opinions

There are three classical approaches to this problem.

° The iterative approach: for example, the Delphi technique, in which the experts are required to elicit independently their probabilistic opinions. These opinions are gathered and sent back to the experts who then have the opportunity to revise their assessments in the light of the colleagues'estimates. The process generally converges quickly, but perhaps towards the wrong figures and for the wrong reasons. One of the major problems is that the experts do not have the opportunity to argue about their models, to exchange their evidence bases, and to discuss the probability of each theory given the evidence.

° The analytical approach:

An example of this type of approach is the Bayesian treatment of the opinion of each expert by a "super expert" (presumably the decision maker). The super expert is supposed to compute the probability of different values conditional on the opinion of each expert treated as different pieces of evidence, with possible dependences (Ref. Morris, Winkler). The problem with this approach is the role of the "super expert" who acts as an aggregator, adding one more layer of subjectivity to the process. Besides, it is often politically difficult to attribute different likelihood functions to the opinions of different experts. A simplified version of this procedure is to simply weight the opinions of the different experts, often with equal weights as if they were independent. Unfortunately, in such a case, the result is a direct product of the choice of the group of experts, without a real chance for them to interact and debate the problem. Hence my preference for the third approach:

° <u>Interactive procedures</u>

In an interactive procedure, the experts meet (1) to share the evidence and discuss the existing data, (2) to explain their models and their reasonings of how they conceived the model given the data, (3) to structure the set of models so that they can begin to talk about the probability of each of them, (4) to assess (individually) the probabilities of the different models, and (5) to participate actively and directly in a debate leading to the generation of the composite distribution.

It is important to note that *there is currently no standard procedure for the aggregation of expert opinions*, and that this exercise will remain subjective in nature. I believe that the key to success (matching the evidence and the distributions, and respect of the internal consistency of the probabilistic logic) is to focus on the probabilities of the models and assumptions as opposed to weighting the experts. Having said that, I have to recognize that the two are frequently linked and that the problem often involves personalities and conflicts as well as a scientific issues.

One promising such procedure has been designed and implemented by the Seismic Hazard Assessment Committee (SHAC) chaired by Robert Budnitz. The work of this committee is now in the publication process. Basically, the committee asked the experts to play successive roles in the aggregation process (from proponent of their own model, to technical integrator of the spectrum of opinions). The result of this work is similar, in its form, to the family of risk curves that Sandia has obtained for potential release levels at WIPP. For a given site, the SHAC committee modeled first the different sources of seismic activity, then the propagation of energy from the source to the site. They obtained a family of risk curves representing a discretization of the frequency of exceedence of different peak ground acceleration levels at the choosen site (Ref. SHAC). Note that this analysis integrates uncertainties about both the source model and the attenuation model (as opposed to the use of conservative hypotheses). These curves are similar to the CCDFs generated by Sandia for WIPP.

How the results (and the uncertainties) should be incorporated in the regulator's decision making is discussed in details in my answer to the fifth question. This answer is based on the assuption that the aggregation of expert opinions will be done systematically for all fundamental assumptions, and that the resulting distributions will be integrated in the risk analysis. Otherwise (e.g., if the disagreement is simply represented by a set of consequence distribution, one per expert), I do not know how to recommend to a decision maker to systematically treat a collection of results, or the results of a conditional risk analysis based on unweighted assumptions. It becomes a matter of faith in the conservatism of the assumptions.

Q3. SELECTION OF INPUT VARIABLES THAT REQUIRE PROBABILISTIC TREATMENT: SENSITIVITY ANALYSIS

It is not necessary in many risk analysis problems to put a probability distribution on all variables. In the decision analysis cycle (Ref. Howard), the first step is to develop models by a deterministic analysis of the link between the consequences and the input variables. Second, a sensitivity analysis for each variable reveals whether or not the variation of an input value across the possible range can change by itself the final decision. Third, the probabilistic analysis is performed: for the variables that do not require full treatment of uncertainty, the mean value is encoded and included in the model. For the variables that do require a probability distribution, this distribution is encoded as described above. The uncertainties are then "propagated" through the analysis by different methods (closed-form solutions, relevant moments, logic/event trees, or simulation, for example, using Monte Carlo or Latin Hypercube sampling).

Q4. DEVELOPMENT OF PROBABILITY DISTRIBUTIONS

Incorporating *all* uncertainties is a risk analysis is indeed a challenge. Therefore, it is important to proceed first to the sensitivity analysis discussed above so as not to lose sight of the ultimate goal (to support a specific decision).

The development of probability distributions is currently a hot topic within the EPA and the environmental/health risk analysis community. [Note, however, that for many years, it has been done systematically for industrial facilities such as nuclear power plants]. Because of the controversial nature of the treatment of epistemic uncertainties by Bayesian probabilities, the solution is often to do only what I consider a partial uncertainty analysis, focusing on randomness in statistical samples and on distributions for the variables explicitly included in the model. The default solution is thus to focus on randomness and on *some* epistemic uncertainties.

There is seldom any attempt to quantify systematically the epistemic uncertainties (about partially known fundamental phenomena) because it requires quantifying explicitly the probabilities of alternative assumptions and, in order to do that, proceeding to an aggregation of expert opinions. For example, in a recent expert-based study of global climate change, Granger Morgan chose to simply present the range of results for each of the different experts without any attempt to come up with a composite distribution. I personally believe that one cannot escape this full uncertainty analysis (i.e., to include the probabilities of alternative hypotheses). Otherwise, the problem is exactly the one that you are facing with WIPP: how to judge of the degree of conservatism of a conditional risk analysis without looking at the conservatism of the hypotheses.

The structure of a full uncertainty analysis is thus the following:

1. Structuring of the different hypotheses into sets of alternative realizations so that probability distributions can be

attributed to these sets of assumptions.

2. Encoding and aggregation of expert probabilities for each set of assumptions.

3. For each fundamental hypothesis, identification of the subsequent models and parameter values (probabilistic treatment). Conditional risk analyses of the type performed by Sandia, but one for each possibility (e.g., each H1i in Figure 1) in a complete set of assumptions, including a measure of possible dependencies through conditional probabilities.

4. Propagation of all relevant uncertainties for each hypothesis (the results are the sets of risk curves shown in Figure 1 for each realization of a given hypothesis).

5. Summing of the results of the conditional analyses weighted by the probabilities of the fundamental underlying assumptions (one then obtains an overall set of risk curves like those presented at the bottom of Figure 1).

(Alternatively, the overall set of risk curves can be obtained directly through the use of a logic tree).

Again, there are different methods for the propagation of uncertainties through each model: closed-form solutions (which is sometimes possible, for example, to treat lognormal distributions and products of variables), computation of the relevant moments, use of logic (event) trees that layout all possible combinations of hypotheses, models, and parameter values, or full simulation (by various methods including Monte Carlo and Latin Hypercube sampling)

Q5. COMPLIANCE CRITERIA GIVEN A FAMILY OF RISK CURVES

How this full uncertainty analysis is used by the decision maker (DM) is a function of his or her own preferences (including risk attitude). Therefore, it is by nature subjective. The consistency of the process, however, can be treated somehow objectively.

For individual decisions, these preferences are represented by a utility function that allows representing risk aversion by putting higher weights (than linear functions would) on the possibility of higher losses. Note that by virtue of the axioms of rationality for individual decisions, it is the *mean future frequency* that is the relevant characteristic of the probability distribution for the future frequency of the potential loss levels (in the WIPP case: the release level as an intermediate descriptor, but more importantly, the helth effects).

This rationality paradigm does not apply to collective decisions, except if one assumes that one elected decision maker (administrator) has been given complete power to make these decisions according to his or her utility function (which, presumably, would have to be revealed if it were to be used in an analytical model). This is impractical because it does not fit our political process and because there are many attributes to each decision that would require some adaptation of any revealed preferences.

The administrator is not only concerned about the probability distribution of the levels of release and about the economic costs of release (for which mean future frequencies would theoretically suffice), but also about the health and safety of the most exposed individuals in the public. The choice of a threshold and the way one demonstrates that it has not been exceeded should reflect directly a concern for prudence. The mean may or may not do that depending on the fractile(s) that it represents in the family of risk curves, and the practicality of demonstrating by analytical means that the goal has been achieved.

I would like, at this point, to go back to what I wrote in my 1986 paper:

> " The next question is to ensure that the goals have been satisfied with "reasonable certainty". A common procedure is to use "conservative estimates" at every step which means to overestimate the probabilities of initiating events, failures, accidents, etc. The overestimation of the

> final result, however, is impossible to assess. It is a wrong approach that may lead to absurd figures and quite possibly to suboptimal decisions, thus defeating the pupose of conservatism itself. This is why the analysis of uncertainties and their explicit treatment in the final decision are critical.
>
> Once this analysis has been done, safety decisions must be made to ensure that with a high probability (e.g., 0.95) the plant is in compliance with a the maximum acceptable individual risk constraint and with the maximum allowable frequency of failure. There is no compelling theoretical reason to use one fractile or a mean value rather than another criterion. In a framework involving numerical safety goals, this certainty level must be specified by the U.S. NRC along with the safety goal"

The example that I was using was safety of nuclear reactors for which the time horizon is relatively short and the uncertainties can be approached systematically. Therefore, the Probabilistic Risk Analyses that are performed for these plants do not involve the types of uncertainties faced with WIPP. Hence the possibility of "reasonable certainty" (which the USNRC calls "reasonable assurance"). In the case of WIPP, part of the analysis (the EPA linkage of release and health effects) is non-probabilistic and presumably, based on conservative modeling. Therefore, given the time frame and the level of uncertainties (e.g., about the future of civilizations in the next 10,000 years), the chosen approach has been different: to start with a set of preliminary results and framing hypotheses, then do a conditional performance analysis based on a mixed method (probabilistic and pre-set health effects estimates). First, one cannot judge directly which fractile(s) the mean curves of the future release levels would actually represent if Sandia had included in the analysis (1) the presumably conservative hypotheses that EPA had specified (complete with alternative assumptions and their probabilities), and (2) the uncertainties attached to the hypotheses that they generated themselves. Second, one cannot derive from this analysis a probabilistic distribution for the health effects. The problem is that a full risk analysis of this type would be extremely difficult given the state of the art, and that the uncertainties over the next 10,000 years would be so large that the results may not be very informative.

In this highly uncertain, long-term case, I believe that the approach based on some fixed hypotheses, *then* on "reasonable expectations" for the conditional risk results is generally sensible provided (1) that the hypotheses are globally conservative (health effects given release as well as assumptions in the release computation) and (2) that the mean curves for the release of the different radionucleides generally correspond to high fractiles of the risk curve families (CCDFs). If that is the case, the combination of *hypotheses* and *means* may indeed provide the level of "reasonable assurance" that you wish and that is consistent with the USNRC requirements for much shorter life facilities. To check that the overall analysis is "globally conservative" you need to verify that the global model (Health Effects + Performance Assessment) yields conservative results and in particular that the hypothetical health risk results that would have come out of a fully integrated analysis meet the level of "reasonable assurance" that you want to see. This requires that the combination of the health effect model and the Sandia hypotheses provides a higher level of safety than the one demonstrated by the position of the PA mean curves in the PA alone.

Therefore, you may want to examine the effects of hypotheses on the position of the *current means* in the family of CCDFs (fractiles) for release accounting for the EPA/DoE hypotheses as shown in Figure 1. Of course, you do not want to ask Sandia to redo the whole uncertainty analysis, but to give you a feeling for the final degree of conservatism of the release results after this accumulation of assumptions. This involves listing the main hypotheses (both from EPA and from the Sandia PA) and assessing (even coarsely) their cumulative effects on the position of conditional (current) means in the CCDFs families. If it is the case that the EPA/DoE assumptions are generally *conservative*, it is likely that what are now mean curves in the current conditional performance analysis (Sandia's PA) would correspond to higher fractiles of distributions that would account probabilistically for all hypotheses (Figure 1, bottom left). If the set of assumptions turns out to be altogether *unconservative*, introducing alternative assumptions will tend to make the current means go down in the families of risk curves towards lower fractiles (Figure 1, bottom right).

When you receive this information about the probabilities and the effects of alternatives to the main hypotheses on the position of the mean curves, you want to examine whether the *final* levels of fractiles that would correspond to the *current means* meet the level of conservatism that you want. You may also want to go one step further and look closely at the health effects themselves and at the conservatism of the EPA model of cancer risk. I do not believe that at that stage it would be realistic to require EPA to proceed to a full probabilistic risk assessment (they do not have the methods as far as I know). Yet, you can argue that their "reasonable expectations" are reasonable only if their hypotheses and health effects model provide the additional level of safety that is consistent with the NRC language of "reasonable assurance". In other terms, first their current means for the release of the different radionucleides have to provide at least as much safety as the overall "expected value" of the release that one would from a probabilistic analysis of the hypotheses. Second, the EPA health effect model should provide an additional layer of safety that convinces you that you are indeed in the high fractiles of a hypothetical full risk analysis.

Should you push EPA to specify a fractile level applicable across the board to all cases? I don't believe so, simply because each problem has to be replaced in its context (uncertainties, existence of alternatives, economic and political context, etc.). I believe, however, that examining carefully the range of fractiles corresponding to the mean in the consequence distribution is a reasonable way to address the question of uncertainties. In the WIPP case, the choice of the mean conditional on a set of hypotheses was based on the long-term nature of the project, the fact that the computation of the mean is more robust than that of specified fractiles, and that the means (given the uncertainties) are likely to be among the high-fractiles anyway. And in any case, requiring the EPA to make a general statement about a "high level of confidence" in the final health effects analysis including all uncertainties would be helpful.

Regulatory language

I think that you can require that EPA be more rigorous in its implementation of the "reasonable expectation" language. They cannot just set hypotheses and models (as those leading to Figure 1, Appendix A of 40CFR191), frame the conditional risk analysis for the applicant, then claim without checking that the conditional means (even with infinite sampling size) resulting from this analysis necessarily support "reasonable expectation" of human safety. Whereas it may be unreasonable (and perhaps, even hazardous given how uncertain the results would be) to leave the choice of hypotheses and model framing to the applicant, it is not unreasonable to require that the effects of these hypotheses on the mean curves be assessed (i.e., simply to check how they displace the mean curve: up or down). In the WIPP case, I would focus on the hypotheses of the intrusion model (frequency, means and effects of drilling) which are the most likely to significantly affect the release results. I would also examine very closely the EPA health effect model.

I would want EPA to show that, in the end, the combination of "reasonable expectation" for the performance assessment and of the conservatism (if it is the case) of the health effect model that they have used to set the release criteria provides "reasonable assurance" of actual safety (i.e., for the ultimate health effects). Because EPA has done the health effects modeling, they are in a good position to show the conservatism of their own results and of the final health risks when these results are combined with those of the performance assessment.

Therefore, you want to require EPA:
(1) to fully reveal the models that they have used to come up with the release standards,
(2) to list all the major assumptions that they have made (those that are likely to affect the risk analysis results), (3) then, to ask the applicants to show that the combination of these models, hypotheses and their own performance analysis supports the requirement that the current conditional mean is indeed "above" the marginal (overall) mean, and that altogether, the assumptions are in fact "conservative".

By comparison, the uncertainties that result from the sampling are probably (1) cheap to reduce and (2) not very significant compared to effects of the basic hypotheses. Therefore, you may choose either to accept their 95% confidence language, or to require a third level of confidence in the analysis. I do not think that it will make much difference.

3. ADDITIONAL COMMENTS ON ISSUES RAISED IN YOUR LETTER

<u>3.1 Level of confidence in the fractiles (or mean) given the sampling size</u>

This issue is easy to resolve because it is cheap to require additional computer runs if you do not think that the level of confidence achieved is what you want. Of course, the tail of the distribution will not be often reached in the simulation by definition of high consequence/low probability modeling. You may want to press EPA to specify the confidence level in this process (third order treatment of uncertainty, i.e., one level further than what I describe as Level 5 in Figure 2 of the Appendix). But you have to realize that the results will be somehow artificial given the variety of the sources of uncertainties. So, I would not focus so much on the uncertainties due to sampling size because they are probably "in the noise", as I would on the uncertainties about the fundamental hypotheses.

3.2 Encoding of expert opinions

I agree that you may want EPA to specify better their encoding procedures. Anyway, in the case of WIPP, you want to find out how Sandia exactly did it (especially for parameter values).

3.3 Use of the mean

I generally agree with EPA that the mean *does* convey "a sense of the whole ensemble of the CCDF's generated". It represents an aggregated description of the risk by a single probability distribution (Level 4 of Figure 2) without displaying the higher level of uncertainties (Level 5 in Figure 2). I do not believe, as you do, that the applicant can vary the number of realizations and dilute at will the effects of *any particular CCDF*. What is true, however, is that with a small number of realizations (in the simulation) one may not reach the tail of the distribution. You want Sandia to specify case by case what level of assurance the mean represents (it varies, of course, along the release axis).

3.4 Additional comments

a. Specific guidance for the form of probability distribution functions seems to me impractical.

b. Need to deal with correlations: I agree, this is essential.

c. Appropriateness of the mean: in the case of WIPP, I think that the coupling of EPA assumptions (*if* they are globally conservative) and mean release level (which is likely to be among the high fractiles given the uncertainties) should provide the level of safety that you want. This is what you want Sandia to demonstrate.

d. Calling explicitly for a 95% fractile with 99% confidence would require a full probabilistic treatment of all EPA/DoE hypotheses regarding the release, introducing still more uncertainties in the analysis and probably producing highly questionable results. [I would not suggest this kind of fractile *on top* of the EPA hypotheses.] Again, I would start by checking what the current mean represents (roughly) in the full picture. To call for the 95% fractile of the *real risk curves (i.e., the health effects)*, would require a whole new risk analysis including both the release model and the health effect model. It is obviously not the direction that was chosen *a priori*.

f. Of course, the process of sampling of 50 parameters, even with an infinite sampling size would dilute the effects of the extremes. It is the nature of probability: the extremes are much less likely than the central range of the distribution. But you want sufficient sample size to have confidence that you have given the extremes their proper weight.

h. Reducing uncertainties can be done in many different ways. Increasing the sampling size of course is one of them; but again, these uncertainties are probably minor compared with the uncertainties involved in the fundamental assumptions.

j. No, it is not easy to identify the various percentiles of crossed curves. Indeed, any mean curve will represent different percentiles in different release ranges.

4. CONCLUSIONS:

I believe that the case of WIPP as it stands now raises issues that are different from those that I addressed in my 1986 article regarding nuclear power plants. But the fundamental concern is the same: reaching an acceptable level of safety with reasonable certainty (or assurance). In the 1986 article, I proposed to do it using high fractiles of the risk curves (which is often where the means are anyway) based on full PRAs including the treatment of all identified and relevant uncertainties (as determined by sensitivity analysis). For WIPP, we do not have risk curves (in the sense of full probability distributions for the consequences,i.e., the health effects). Because of the 10,000-year time horizon, the uncertainties in the case of WIPP are such that this kind of analysis may be a futile exercise. Instead, EPA has chosen to make some assumptions in its performance criteria and to require a conditional performance analysis given these assumptions. Then, EPA specified the use of the conditional means as the basis for the compliance criteria.

In order to verify that the conditional means (conditional on specified health models and hypotheses) provide indeed "reasonable expectation" of safety once the effects of the hypotheses on expected values are carefully considered, you want to ask Sandia to provide additional information about what these conditional means really represent for future release and what they imply for human safety. In particular, you want to question assumptions regarding engineered barriers and the hypotheses that have been made to support the currently planned storage system. This is where you may be able to show that some of the assumptions are unconservative and that the real mean curves are below the conditional ones. Therefore, you may be able to conclude that the current analysis based on conditional means does not meet, on the whole, the "reasonable expectation" standard. I would not focus much on the effect of the sampling size (although it probably does not cost much) because increasing it may not provide large variations of the position of the mean in the overall CCDF family. The hypotheses about the frequency, the means and the effects of drilling are more likely to provide significant variations.

To summarize my conclusions:

4.1 I do not know where the current means stand in terms of fractiles on the distribution of release curves presented by Sandia.
=> You may consider asking Sandia to specify which fractiles are involved in the mean release curves that are presented in their final PA report (these fractiles will vary along the release axis; but Sandia may be able to bracket them).

4.2 I cannot judge the degree of conservatism of the Performance Assessment results because I do not know the effects of the EPA and DoE hypotheses on the release curves.
=> Ask Sandia to list the major hypotheses that have been taken for granted in their PA and to give you an idea (if not a full analysis) of the effects of these hypotheses on the results (i.e., the family of release curves). For example take the five or six most important assumptions of the PA (e.g., the Poisson model of human intrusions,

the diameter of the bore holes, the water flow model, the solubility factor of the main nucleides, etc). Ask Sandia to generate a set of reasonable alternatives to these hypothese and to show you that the mean curves that would be generated with proper probabilistic analysis of these alternative assumptions actually meet the criteria (and that they do not pull the means towards lower fractiles of the risk curve families).

4.3 The expert opinion procedures of encoding could be made more rigorous.

=> You may want to ask Sandia to identify the variables whose distributions are critical for the results (could make WIPP violate the performance criteria), to justify their decision to treat them through expert opinions (as opposed to experiments or measurements when feasible), to better justify their findings by describing exactly how they have encoded and aggregated expert opinions, or to redo the encoding and aggregation of these judgments if you conclude that some of the variables have nor been properly treated.

4.4 The uncertainties about WIPP are such that full probabilistic treatment of all assumptions is likely to introduce large additional uncertainties in the results if they were to be systematically treated through probabilities.

=> You may want to find out what is the level of release risk obtained given the combination of EPA and DoE assumptions and the results of the corresponding conditional risk analysis, judge whether it is reasonable, and if it is not, ask EPA to reveal how it is going to inject additional levels of prudence in its decision. Depending on how far the current means are (assuming full probabilistic treatment of hypotheses) from a reassuring (but not sacred) 95% fractile, you may want to ask for additional analysis or for a change of risk management strategy.

4.5 If you really want to estimate the long-term health risks associated with the possibility of release, you need a probabilistic version of the EPA health effect model and a true risk analysis involving both release and health effects. I doubt that this is feasible. But:

=> You may want to ask EPA to better justify what they have done to obtain Table 1 of 40CFR191 and DoE to show that the overall risk results (their model plus the PA) provide "reasonable assurance" of safety.

5. APPENDIX

[What follows on this topic is based on a report that I recently wrote for the Electric Power Research Institute].

<u>Six levels of treatment of uncertainties in risk analysis</u>:

The form under which one would like uncertainty analysis to be done depends in large part on the use that one intends to make of the results, i.e., what criteria will apply in the decision making. All decisions do not need full treatment of uncertainties. Different degrees of sophistication in the assessment of the risks can be envisioned depending on the management rule that one intends to apply. Six different levels in the treatment of uncertainty (see Figure 2) can be identified.

Level 0 simply involves the detection of a potential hazard without attempt to assess the risk in any way. It is sufficient, in theory, to support strict zero-risk policies, or to make risk management decisions when the costs are low.

Level 1 is the "worst-case" approach. It does not involve any notion of probability. It is based on the accumulation of worst-case assumptions and yields, in theory, the maximum loss level. In practice, however, whatever the worst-case scenario that has been constructed, it is often possible to imagine still more unlikely circumstances that could worsen the result. It is therefore necessary to truncate the loss distribution.

probability. [Looks conservative. To be checked]
p.114: the criterion itself: "demonstrate that there is at least a 95% level of statistical confidence that the mean of the population of CCDFs meets the requirements of section 13(a) of 40CFR19" [The mean is the most robust measure under the circumstances (smaller sample size required) and it may already be in the 80 to 95% fractile].

3. Background information: EPA 402-R-95002

p.3-7: Disposal systems shall be designed to provide reasonable expectation based upon performance assessments that cumulative releases of radionucleides to the accessible environment for 10,000 years after disposal from all significant processes and events that may affect the disposal system shall:

(1) Have a likelihood of less than one chance in ten of exceeding the quantitties calculated according to table 1 (Appendix A) and

(2) Have a likelihood of less than one chance in 1,000 of exceeding ten times the quantities calculated according to Table 1.

Table 1 defines a set of permissble releases ("normalized release" for each isotope). [The question is; what were all the hypotheses underlying Table 1].

4. Compliance criteria: March 21, 1995

p.55: results of performance assessments.

5. EEG Comments. April 28, 1995.

p.5: the WIPP site does not meet the there stated criteria of 40 CFR §91.149 (because it is in a resource rich area)=> unconservative assumption. On the other hand, (p.6)EPA claims that the hypotehse are favorable because of the favorable characteristics of the WIPP (located in the desert). [net result??]

p.11: Engineered barriers:

Argument for engineered barriers: unconservative assumptions regarding human intrusion in a resource rich environment. Also: benefits will be small because it would only delay the arrival of actinides in the environment.