CERTIFIED

ACNW-0031
PDR-5-30-91

DATE ISSUED: April 16, 1991

MINUTES OF THE WORKING GROUP ON THE USE OF EXPERT JUDGMENT
OF THE ADVISORY COMMITTEE ON NUCLEAR WASTE
JANUARY 25, 1991, BETHESDA, MARYLAND

The Working Group on the Use of Expert Judgment in Performance Assessments for Repositories of the Advisory Committee on Nuclear Waste was convened by Working Group Chairman Paul W. Pomeroy at 8:30 a.m. on Friday, January 25, 1991, at 7920 Norfolk Avenue, Bethesda, Maryland.

[N.B.: For a list of attendees, see Appendix A. ACNW members, Drs. William J. Hinze, Dade W. Moeller, Paul W. Pomeroy and Martin J. Steindler were present. ACNW Consultant, Dr. David Okrent was also present.]

The Chairman stated that the agenda of the meeting had been published in the Federal Register [55 FR 50901, Tuesday, January 25, 1991]. He also stated that the meeting was being held in conformance with the Federal Advisory Committee Act (Public Law 92-463). He also noted that a transcript of the meeting was being made. [Copies are available in the NRC Public Document Room in the Gelman Building, 2120 L Street, N. W., Washington , D.C.]

[N.B.: Copies of the transcript taken at this meeting are also available from the Ann Riley & Associates, Ltd., 1612 K Street, N.W., Washington, D.C. 20006]

DESIGNATED ORIGINAL

Certified By EMB

[Ms. Charlotte Abrams was the Designated Federal Official for the meeting]

Speakers at the meeting included: M. Blanchard, Department of Energy (DOE) Yucca Mountain Project Office (YMPO); S. Copland and P. Brooks, NRC, Office of Nuclear Material Safety and Safeguards (NMSS); S. Frishman, State of Nevada; W. North (Member) and L. Reiter (Staff), Nuclear Waste Technical Review Board (NWTRB); C. Stepp, Electric Power Research Institute (EPRI); A. Cornell, Stanford University; and E. Bonano, Sandia National Laboratories.

The ACNW Working Group on the Use of Expert Judgment was convened to discuss the proper role of expert judgment in site characterization, performance assessment, and the licensing of waste repositories. The Chairman pointed out that the ACNW's goal was to delineate the areas of agreement and disagreement within the community with regard to the use of expert judgment in order to advise the Commission of potential problems and their solutions.

The Chairman went on to summarize a series of questions, principally suggested by Dr. Okrent, which were asked of the participants. These were:

1. What are the issues, events and processes in which empirical data are likely not to be adequate to serve as a

principal basis for judgment?

2. For such issues, what approach do you think is practical and tenable; i.e., expert opinion, boundary calculations, or some other device?

3. What issues, if any, seem to be beyond quantitative grasp?

4. If expert opinion is the recommended approach for an issue, please discuss briefly:

- how to select experts,
- how to aggregate multiple expert opinion, and
- how to deal with bipolar expert opinion?

5. What constraints, if any, should there be on the use of expert judgment?

6. How can we ensure that the use of expert judgment is clearly identified at each point of the process?

The Chairman noted that the EPA high-level radioactive waste (HLW) standard recommended that a combination of quantitative data and expert judgment be used to provide reasonable assurance of compliance.

The Chairman introduced the first speaker, Mr. Maxwell Blanchard of DOE/YMPO.

Presentation of M. Blanchard, Yucca Mountain Project Office

Mr. Blanchard discussed two aspects of expert judgment to support DOE managerial and technical decisions affecting the suitability of the HLW disposal site and to support the process of obtaining an NRC license. He identified the less formal application of using technical and peer review, as "implicit" expert judgment. The more formal application using decision analysis was described as "explicit."

He explained that the DOE philosophy in using expert judgment is generally in agreement with that described in NUREG/CR-5411 [Elicitation and Use of Expert Judgment in Performance Assessment]. Expert judgment is used in screening events and scenarios, characterizing site features such as geology, tectonics, etc..., as well as in evaluating models for assessing conditions of the site, and even in data collection.

He identified the role of expert judgment in three specific areas:

- In the case where data are unavailable or highly uncertain, expert judgment is used to provide

probabilities of events and even parameter values.

- Expert judgment is used to build confidence in the validity of the interpretations of processes based on uncertain data.
- Expert judgment adds confidence with respect to model validity.

Key questions which keep coming up in the use of expert judgment include: (1) What elicitation methods should be used, and (2) how can the conclusions be validated? The first question has to do with a choice between the implicit versus the explicit. The second has to do with credibility, reviewability and verifiability of the judgment and how it was accomplished.

Mr. Blanchard then discussed the past uses of expert judgment with respect to the Office of Civilian Radioactive Waste Management (OCRWM). The 1986 DOE Environmental Assessment used expert judgment to establish the presence and the absence of favorable and potentially adverse conditions, as well as to present disqualifying and qualifying conditions per 10 CFR Part 960. The lack of quantitative criteria in Part 960, and the low-probability nature of the events necessitated the reliance on expert judgment. Multi-attribute utility analysis reduced five potential HLW repository sites to three, and an explicit expert judgment elicitation prioritized the three candidate sites.

Similarly in the Site Characterization Plan (SCP) and numerous proposed Study Plans, expert judgment -- in the form of implicit peer reviews and more formal explicit elicitations -- played a key role in identifying site data needs, as well as forming strategies and confidence evaluation.

It should be noted that DOE considers peer review and formal expert elicitations to be the same thing, differing only by their degree of formality. Ongoing OCRWM activities where expert judgment plays a role include the exploratory shaft facility (EFS) alternatives; the Calico Hills risk/benefit study, where multi-attribute utility analysis is being used to identify potential adverse impacts of site characterization; and the determination of the relative advantage of further testing.

Dr. Hinze inquired about the criteria used in deciding whether to use explicit or implicit approaches. Mr. Blanchard responded that the sensitivity of the subject and magnitude (in cost or complexity) of the problem were the major criteria that would lead to a formal process. An example of this was the exploratory shaft construction method. In contrast, if the subject is restricted to a specific technical area, then implicit approaches are preferable.

In response to Dr. Hinze's question on selecting experts, Mr. Blanchard proffered relying on nationally renowned individuals, with a significant publication history. Dr. Hinze pursued his

question in light of the dynamics of using experts on expert judgment per se. Mr. Blanchard responded that this was not consistently done by DOE.

Mr. Blanchard went on to indicate that expert judgment was used in the testing strategy; not only in what kinds of tests, but also the timing of the testing. The focus of the DOE was to optimize the use of available funds for testing, by trying to predict the value of the information expected from the tests. If the expected information would only lead to a better characterization of uncertainty, DOE would assign it some value towards priority in testing.

Dr. Okrent pointed out the pitfall of using experts to find answers that don't really exist. Even if the answers exist, how or where the experts go about finding them raises further questions. Mr. Blanchard indicated that the vagueness and qualitative nature of 10 CFR 960 made interpretational judgments suitable. In effect, one may not need a full answer, but enough of a qualitative understanding of an important factor to dismiss the suitability of a site.

Mr. Blanchard moved on to some lessons learned from DOE's experiences with formal expert judgment. Expert judgment provided a means of combining qualitative and quantitative information; he cited probabilities for large magnitude earthquakes as an example.

Another lesson was in getting an insight into asking the right question. Tailoring the expert panel to the problem is another concern; how specific must the experts' knowledge be to resolve the problem? And finally, the time factor to properly execute a formal elicitation can be significant, especially in educating or training the experts to the problem at hand.

Dr. Steindler asked whether highly specialized talent is always a prerequisite. Dr. Steindler was interested in whether generalists, or just specialists, should be included in expert panels. Mr. Blanchard indicated that a decision analyst should be included but having generalists also would be useful.

In closing, Mr. Blanchard noted that expert judgment will be necessary to quantify a significant number of processes and events; e.g., effects of coupled processes such as tectonics and hydrology. But even using bounding calculations requires at least the use of implicit judgment to provide the degree of confidence or to assure that the result will be conservative. And finally, expert judgment is needed by both DOE and NRC to interpret ambiguities in the regulations, in a formal and explicit manner to provide documentation for the course of action taken to achieve a compliance decision.

Dr. Moeller asked whether NUREG/CR-5411 serves as the primary DOE reference for using expert judgment, and Mr. Blanchard said it was.

Dr. Moeller asked for the demarcation where peer review ends and expert judgment begins. Mr. Blanchard responded that peer review was implicit expert judgment. Explicit expert judgment consists of a peer review coupled with the addition of a decision analyst to facilitate the elicitation process.

Dr. Bonano (Sandia) distinguished expert judgment as a means to obtain a solution; peer review critiqued that solution. In response to Dr. Okrent, Dr. Bonano characterized the main weakness of NUREG/CR-5411 in that it was not based on an actual expert judgment elicitation study. A further flaw indicated by Dr. Bonano is in dealing with the issue of how to ask the right questions; failure to do so could lead to very biased results. Dr. Okrent pointed out that the explicit expert judgment process has not yet been validated. No impartial experimentation has been performed to characterize the degree of reliability. Dr. Bonano agreed, but pointed out efforts were under way to deal with the use of expert judgment especially in the area of combining judgments of different experts.

In a final exchange, Mr. Blanchard told Dr. Okrent that DOE doesn't have a delineation of exactly which places in the performance assessment expert judgment would be used or of what degree of uncertainty would result from such use. Although Mr. Blanchard was not current on the degree of use of expert judgment in the performance assessment to generate the complementary cumulative

distribution function (CCDF) curve, he indicated that the performance assessment side of the YMPO was using implicit expert judgment and expects that decision analysis will be used eventually in the CCDF efforts.

Dr. Hinze and Mr. Blanchard engaged in a discussion about using expert judgment in such a manner that evaluating insufficient data or information could result in a significant impact on the decision process. This is a concern, since it appears that DOE will rely heavily on expert judgment to arrive at conclusions, which would not otherwise be achievable.

Presentation by NRC Staff

Mr. Seth Coplan (NRC/NMSS staff) discussed the historical use of expert judgment in NRC. He pointed out that the formal elicitation process was a new process. Since DOE began using expert judgment in an extensive manner in the HLW area, NRC developed guidance for reviewing areas where such elicitation was used. Sandia prepared the NUREG/CR-5411 guidance document, as a result of its experience with the NUREG-1150 reactor risk reference document.

Ms. Pauline Brooks (NRC/NMSS staff) discussed NRC's review of the DOE Site Characterization Plan, in light of the use of expert judgment. The review criteria were directed to ensuring that the use of expert judgment placed high priority on obtaining objective

data. These criteria include:

1. Expert judgment was to complement other sources of information (data), not to substitute for them.
2. The issues, planned to be addressed by formal expert judgment, had to be identified and described.
3. In the licensing phase, the results of formal expert judgment will be reviewed in a substantive manner.

Ms. Brooks addressed the steps involved in the elicitation. These involve:

-Identifying issues and information needs.

-Selecting and training experts.

-Conducting the elicitation sessions.

-Combining and communicating expert judgments.

The first step is extremely useful to later reviewers; in effect, it serves as documentation of what was done. Since this is an explicit process, normative experts (decision analysts and experts in the use of expert judgment) are involved in the selection and training. The second and third steps consist of a process, whereby the technical experts are sensitized to bias, and they are aided in articulating their technical judgment in a useful manner; e.g., a quantitative value. This assistance can involve decomposition of the problem into smaller more malleable components; e.g., forming a fault tree. In all of these steps, the process is documented, so that any or all of the elicitation steps can be

revisited and understood.

Dr. Steindler questioned the value of eliminating biases. Dr. North (NWTRB Member) interjected that biases can exist at two levels: the conscious level which is a deliberate shading of expectations to make an outcome look favorable; the less conscious level is more complicated. Dr. North pointed out that a simple example of this unconscious bias is asking someone to estimate the error probability of an estimate he makes. In most cases, the individual underestimates the error in his estimates, in some cases significantly. (This phenomenon is called overconfidence).

Dr. North characterized the expert elicitation process as an art form as much as it is a science. As much as possible, the goal of elicitation is to transform simple intuition into a process that is explicit, cleansed of psychological biases, and communicated in a fashion that can be brought into the decision process. But this process cannot be perfect.

Dr. Bonano offered a breakdown of biases:

Motivational biases - these are those Dr. North characterized as conscious.

Overconfidence - an individual underestimates the error distribution of a judgment he makes.

Anchoring biases - an individual has a preconceived notion and filters out any information which conflicts with that notion.

Mr. Coplan resumed the NRC staff portion of the presentation by describing past experiences in dealing with expert judgment. He began with the NRC staff's role in assisting the Commission's licensing decision regarding the HLW repository. In doing this, the NRC staff obtains significant quantities of material in support of a licensing decision. He cited DOE, DOE contractors, the Center (Center for Nuclear Waste Regulatory Analyses), and NRC consultants among the many sources of input to such a determination.

Mr. Coplan indicated that in the cases of conflicting conclusions in these sources of information, the staff inspects the quality and the completeness of the information, as well as the reasonableness of their assumptions and the soundness of the reasoning. This can lead to a hybrid or composite of the various data or information sets available. Then the applicant is requested to address any inconsistencies.

In the end, according to Mr. Coplan, the staff produces a Safety Evaluation Report (SER), which promotes a conservative position minimizing the likelihood of underestimating adverse consequences. The key issue in HLW is to comply with the reasonable assurance

that waste can be disposed without undue risks to the public health and safety.

The next step in the process following the NRC staff's conclusions involves the licensing boards. Mr. Coplan went on to describe the adjudicatory hearing process involving DOE as the applicant, the NRC staff, the State, and other affected parties. In his discussion, Mr. Coplan brought up the concept of "layman's proof," which is a very clear, reasonable, logical, well-laid out argument that can be used to convince an intelligent layman. The board resolves conflicting testimony as mentioned above, by considering the quality and the completeness of the information, as well as the reasonableness of the assumptions and the soundness of the reasoning of those presenting testimony. The board in effect provides another level of independent review over and above, yet following the same criteria used by, the NRC staff in documenting their SER.

Mr. Coplan acknowledged that expert judgment will be relied upon in the justification by DOE that the repository complies with 10 CFR Part 60. However, he shares Dr. Okrent's concern about the lack of validation. There is an intrinsic value in such a validation in that it can provide a clearer definition of the problem, and it is a useful way of sampling the state of knowledge in an area. The NRC staff would critically scrutinize those areas where elicitation was used; i.e, does the elicitation provide a

truly representative sample of the state of knowledge?

Mr. Coplan indicated that expert judgment on a lower scale could be more readily reviewed than expert judgment on a higher scale; e.g., an expert judgment on absorption data for a transport calculation as opposed to an expert judgment on the quantity of plutonium released over 10,000 years. Furthermore, the NRC staff would be skeptical on the use of aggregating algorithms, rather than evaluating each opinion on its own merits.

Dr. Okrent and Dr. Pomeroy brought up the question of how much expert judgment is too much? Mr. Coplan indicated that he couldn't answer that question in quantitative terms; i.e, 10% expert judgment is acceptable or more acceptable than 90% expert judgment. As far as the HLW repository licensing process is concerned, the use of expert judgment is so pervasive that it would be difficult to separate expert judgment from hard science. The concern is more in terms of the amount of uncertainty associated with expert judgment as opposed to that associated with real data or analyses. When that uncertainty is too large, then it compromises the process. The iterative process of performance assessment may help to reveal where the cutoff might be.

Mr. Coplan responded to a question from Dr. Pomeroy on whether there might be sufficient layman's proof to license the repository. He indicated that he thought there would be, but that it was going

to be a difficult process.

An ensuing discussion between Mr. Wolf (OGC), Mr. Abramson (RES), Dr. Okrent and Dr. Bonano followed in a free-format exchange. Mr. Wolf indicated that Congress had not charged the participating Federal Agencies to do an impossible task. He indicated that with reasonable and conservative means, one could come to a supportable licensing decision in terms of a layman's proof. It might not be the decision that DOE wants, but NRC can come to a licensing decision.

Dr. Okrent didn't see how a finding of reasonable assurance could be made in certain areas such as future climate changes, especially to the point of complying with the EPA's Standards. Mr. Abramson pointed out that a key effort should be to obtain an honest expression of uncertainty. The uncertainty may be a key piece of information to be used in an assessment. In fact, a significant purpose of using expert judgment is to reflect the uncertainty in the technical community on the issues at hand. It should be factored into the licensing decision, as well as into the results of the elicitation itself.

Presentation by S. Frishman, State of Nevada Dept. of Nuclear Waste

Mr. Frishman remarked that he generally agreed with Mr. Coplan's approach. In focusing on the important issues and events where

empirical data may be inadequate to the task, the list of such items ranges from the extremely simple to the very difficult issues of trying to balance out elements of performance. Mr. Frishman questioned the movement toward performance assessment, when the program should still be in the data collection phase. He characterized Part 60 as consisting of two main parts: one is description, the other assessment. He referred to a 1957 National Academy of Science (NAS) Report, which he believes to provide an important lesson to apply to the present situation; i.e., the question should be: is it possible to dispose of waste at a given site, not how can we dispose of waste at that particular site?

He observed that the transition to assessment has been forced. In this assessment phase, he characterized two aspects of performance: the geologic barrier component and the engineered component. He pointed out that the effectiveness of the engineered and natural barriers is discussed in Subpart E (of Part 60). It is in this aspect where all the iterations, and hence the heaviest emphasis of expert judgment, come into play.

He also cautioned that expert judgment should only be used as a last resort and encouraged the continued reliance on data acquisition. He cited examples of misuse and filtering of information, which are forms of expert bias, and which led to distortion of information and systematic failures; e.g., the Challenger disaster. He specifically cited the present effort in

Nevada as a case whence there is a systematic distortion in the data itself. This distortion was compounded with expert interpretation of biased data.

He outlined what he thought should be done in the program. He stated that the analysis of the geological data needs to be continued. He believes that the engineering and underground construction are straight-forward, and that the meshing of these two should not be tremendously difficult; except that this meshing must result in optimum isolation, which is a new twist. This meshing is in the form of the implicit expert judgment mentioned earlier by Mr. Blanchard. But, Mr. Frishman did not believe that the project had achieved this stage. He faulted the alternative exploratory shaft analysis for prematurely entering into this explicit arena. Again he expressed the opinion that, when one has to come to terms with the effectiveness of the barriers in complying with the regulation; at that point judgment enters but not earlier.

Mr. Frishman returned to the initial questions on the role of expert judgment:

- Bounding calculations are the essence of site characterization.
- Selection of experts: the key factor is independence.
- Bipolar opinion: mathematical treatment produces results that do not reflect either of the poles. Instead,

analogues should be sought, which might resolve the bipolarity.

Mr. Frishman cautioned the trend being taken in the exploratory shaft alternative study and the Calico Hills risk-benefit analyses. He warned that although this approach has the most appealing structure and the best documentation, it might be completely wrong. He proposed a penalty system of increasing (doubling) the error term in those cases where expert judgment, instead of bounding conditions or more data acquisition, is used. Ultimately, too much expert judgment will result in a set of inputs for the CCDF, which produces unacceptable results. This is a disincentive to going the "easy" route; i.e., using expert judgment in places where data collection or further analysis may be difficult but not impossible.

At this point, Dr. Steindler and Mr. Frishman discussed the relationship between the reliability of expert judgment and the time period for its extrapolation. Mr. Frishman brought up the analogue of the oldest surviving structure, with its corresponding level of technology compared to a technologically advanced bridge structure. Mr. Frishman observed that the program for a HLW repository is in its infancy, yet expectations are being required for 10,000 years of future performance.

Dr. Hinze suggested that one of the advantages of expert judgment is to assemble a panel of experts who could assist in deciding if

the previously acquired data are of sufficient quality, quantity, heredity, integrity, etc.... He went on to ask Mr. Frishman whether it was too late in the program to do this? Mr. Frishman agreed that would be a proper function for expert judgment, and that, regrettably, the process had advanced too far to revisit the data adequacy question.

Mr. Galpin (EPA) asked Mr. Frishman whether, in generic terms, a geologic repository could ever be licensed; i.e., that inherent in the process there was a feasible regulatory framework? Mr. Frishman indicated that he thought there was. He went on to discuss what he thought could be the real stumbling block in licensing: the "what if" syndrome. Since we can't guarantee proof of isolation, we try to convince the boards, public, and so forth, that all possible contingencies have been covered. Inherent in this position is the assurance that all contingencies have either been solved or do not present significant concern.

Dr. Okrent disagreed that the "what if" syndrome was the operating mechanism either for licensing nuclear power plants or for HLW repositories. Risk analyses accompany licensing decisions with not insignificant probabilities of large releases. There do remain open issues; no claim of knowledge of all contingencies is made. Dr. Okrent pointed out that there is no acceptable way of evaluating low-frequency, high-damage floods in this country; yet dams are constructed. To expect data for everything and no

reliance on expert judgment is unrealistic.

Mr. Blanchard pointed out that the DOE would like to do more data collection, but one of the largest obstacles is legal, and it prevents DOE from such collections. In fact, the State of Nevada has contributed to this dilemma.

Dr. Bonano pointed out that the objective in using expert judgment is not to obtain correct judgments, but rather to guarantee that they were obtained in a structured manner, which can be peer reviewed and critiqued. Correct judgments cannot be further guaranteed by using a straight forward QA methodology than by using expert judgment. Mr. Frishman remarked that this was precisely the problem with the program.

Dr. Cornell pointed out that the goal of elicitation is not only to try to approximate the larger level of consensus by carefully using a smaller sample to approximate the average consensus, but also to get a grip on the breadth (variation) of the larger level of judgment. The objective is not the truth, but rather prevalent opinion of what the truth is.

Mr. Frishman pointed out an observation from a 1957 NAS report. In effect, the problem was that hydrologists and geologists needed to be educated in radioactive waste and its disposal problems. Too often nuclear engineers become instant geologists and hydrologists.

He claims that this is still a problem 33 years later.

Dr. Steindler asked which was of more concern: the extrapolation of expert judgment or the quality and applicability of the judgment? Mr. Frishman pointed out that it was the applicability. Mr. Frishman observed that, when geologists predict for less than 100 years or for greater than a million years, the comfort index is acceptable. It is the middle area where the geologist has a difficult time.

Dr. Stepp (EPRI) indicated that extrapolation in geology was more credible. The fundamental property in geology that the past is the key to the future has lead to our understanding of deep time as a continuity [continuum]. Dr. Stepp said that this understanding is embedded in Part 60. He stated that our understanding of processes in geologic time frames should stand us well in 10,000-year extrapolations.

Presentation by Dr. Warner North, Affiliated with the Nuclear Waste Technical Review Board.

Dr. North indicated that he would present some of his past experiences in decision making and the use of expert judgment. He pointed out that, as in the case of some of the other participants in the workshop, he felt uncomfortable about the use of expert judgment when data are available.

In many of the instances of expert judgment, we are not dealing with science. He used the analogue of a doctor practicing medicine, where truth isn't the object; but healing or improving the patient's condition is.

The first example addressed by Dr. North was the government program to seed hurricanes. The idea was to place certain chemicals in the hurricane to lessen its destructiveness by lowering the speed of the maximum winds sustained around its eye. Due to its complexity, the problem is broken down (decomposed) into pieces.

The first piece is wind speed data over time. The three possibilities are that seeding would make things worse, better or has no determinable effect. Using statistics on wind data, a model was developed to result in three probabilities for the three possible outcomes. This approach was severely criticized, because it gave no reasonable assurance determination. The main problem was that this model development did not include an iterative process for interaction. This interaction needed to be on a larger scale than just dealing with a few experts.

The next example was to use expert judgment to assess the possibility of contaminating Mars by landing the Viking probe. Specifically the interest was from a probabilistic regulatory constraint viewpoint: 0.001 chance that Mars would be contaminated

in the next 50 years. Here, the first step was to interview a large number of scientists on how even to undertake this type of assessment. As in the case above, the problem was broken down in a sequence of events from microbes being present on the spacecraft through all the mechanisms including their reproduction in the microenvironment of Mars. Factors that entered the Bayesian thought process included the UV radiation flux through the Martian atmosphere -- could it kill the microbes or would they be shielded? Sensitivity analyses indicated that the variability of any factor was not enough to violate the above constraint.

The next example dealt with the Middle East's stability and the flow of oil to the U. S. This was one of the first uses of the influence diagram: a map showing the linkages of events composing a scenario. The diagram was linked to U.S. and Soviet military presence, but also factored in Israeli-Arab conflicts and political instabilities of various states. Even though the information was largely non-quantitative, the influence diagram enabled the analyst to calculate and assign probabilities to the various interactions, and hence to the millions of different scenarios resulting from the influence diagram. Even though much of the analysis has been borne out by recent events, there was a great deal of difficulty in communicating those results not having to do with Soviet and U. S. interaction.

Dr. North pointed out that there is great value in looking at past

examples of successes and failures of expert judgment as well as looking ahead for shorter periods of time to assess the credibility of trying to predict the same kinds of things 10,000 years into the future.

The next example involved the Tokamak Fusion Test Reactor. The expert judgment was directed to performance of the Tokamak, when compared to a magnetic mirror type, as well at to others still being developed. Again, experts were used to decompose the assessment into manageable pieces. An influence diagram was developed to assess the performance. The influence diagram laid out the components in order to address the performance. The conclusion was that the Tokamak was unlikely to give an economically competitive system. However, Dr. North observed that this study did not go far enough into investigating the possibilities of alternatives to the Tokamak.

Next Dr. North discussed a state study on controls for sulphur oxide emissions in Wisconsin. The additional controls would be extremely costly (100s of millions of dollars). Time and budget constraints were set, and experts were sought in the areas of atmospheric science, as well as on impacts of acid deposition on lakes, soils and forests. The iterative process included a public meeting, where a computer was available to run through scenarios suggested by the attendees. Not only was information elicited from the public, but the public was able to see how the conclusions were

made, and why. This open architecture for the elicitation was a distinguishing feature from the typical Delphi process.

Dr. Hinze questioned the potential for bias, since a number of the experts were from the state, and could have hidden agendas. Dr. North remarked that no apparent bias was observed, except for a preference for statistical versus mechanistic views. This only affected the response time, not the conclusions, of the process.

Dr. North moved on to the next example: low-level lead exposure and its effect on childrens' IQ. EPA used decision analysis to assess probabilities of decrement of childrens' IQ versus blood levels of lead. Five of the panel of six concluded that there was decrement as a function of blood lead level. Dr. North criticized this elicitation, since there did not appear to be a documentation by each expert of the analysis, rationale and assumptions used.

Dr. North summarized that there are more than 20 years of experience in decision analysis and expert judgment, much of which is available. He was glad that NRC and DOE are trying to learn from this body of knowledge.

Dr. North noted a trade off with regard to the formality of the expert judgment process. If it becomes too formal too soon, you chance losing the valuable interaction between the assessor and the expert. Before addressing probabilities, the assessor(s) and the

expert(s) should discuss the subject, what kind of data and models should be used, and so forth. From this, the assessor often discovers information that was not previously considered.

Dr. North indicated that the decomposition of the problem into smaller components and the models selected was as much a feature of expert judgment as was the assignment of uncertainty or probabilities.

Finally, Dr. North recommended that expert judgment should be opened to a wider view of the process in order to facilitate credibility. Whether this is done by using an outside peer review or by open meetings may depend on how sensitive the information is. But the process should be reviewable to ensure that the prevailing opinions and minority viewpoints have been captured in the process.

Dr. Steindler asked Dr. North to expand on the documentation of rational as a necessary ingredient. Dr. North pointed out that this is more a question of accountability, rather than quality of the decision. The decision must be evaluated in light of its basis. In the public arena, this is usually the central issue. Frequently, expert judgment processes have failed because of a lack of peer review or oversight -- something slips through the process.

Dr. Bonano brought up the iterative nature of the expert judgment process. He indicated that a major result of using expert judgment

is the conceptualization of what is known and what isn't. An explicit articulation of this leads to what kind of information need to be collected. He also noted that there is a trade-off between bias introduced by certain experts and the loss of the prevailing state of knowledge by disqualifying those individuals. The decision analyst can try to minimize the bias impact through the training/educational phase of the elicitation. Dr. North indicated that there is precedent for expert disqualification, where doubts exist regarding the expert's judgment.

Dr. Stepp asked for an expansion on the disqualification of an expert; specifically, what criteria would one use? Dr. North indicated that the other experts serve as a check on outliers. In many cases, the outlier seems to be in conflict with the available information and to be deviating from the training/educational criteria. The outlier's opinion represented a departmental objective or some personal agenda. Some straightforward tools can help minimize this; e.g., at the beginning of each meeting everyone pledges allegiance to the objectives of the elicitation, as opposed to other objectives. Those who couldn't pledge allegiance can be disqualified.

Dr. Pomeroy noted that in the previous examples, the time frames for the period of performance were reasonably short; he asked Dr. North about the cases where long periods of performance were involved; e.g., human intrusion over 10,000 years. Dr. North

pointed out that WIPP had used expert judgment elicitation regarding this issue, and that it should have been identified much earlier in the process.

Dr. Okrent cautioned that prevalent opinion changes. So basing a decision on prevalent opinion now, could be overturned in the future.

Presentation by Dr. C. Stepp, Electric Power Research Institute (EPRI).

Dr. Stepp discussed the problem of transference, the tendency of shifting the emphasis from the real issues to other areas more familiar to the participant(s). He thought this was happening in the HLW program. He also wanted to make a distinction between judgments and interpretations, specifically in the degree of support and the documentation. He indicated that judgments have little support and would be more of a gross bottom line. Interpretations are decomposable into manageable stages, whose uncertainties can be documented. He indicated that he would focus on expert interpretation.

In the area of expert selection, he indicated that the range of disciplines and knowledge needed would require more that one expert and dictated a team approach. Moreover, the team approach provided some relief to the single expert bias problem.

He moved on to the question of quantification of uncertainty, which is composed of three types:

- uncertainty of our knowledge of processes, models, etc....
- uncertainty of the data and information available to evaluate these models and processes.
- unquantified uncertainty -- uneven distribution of both knowledge and data.

In response to Dr. Okrent's question on defensibility of quantifiable uncertainties for all repository issues, Dr. Stepp replied that it could be done for the key interpretations.

Dr. Stepp next discussed the bias of knowledge, which was addressed earlier, and the bias of information. This second type exists when different information is made available to different team members. A structured feedback mechanism is needed to show individuals how their interpretations impacted the end results. This would allow them to rethink and adjust their judgment in the iterative process.

As repeated before, the basis for the interpretation needs to be transparent; i.e., documentation of the hypotheses, processes, models and specific criteria used at every stage. This feeds into the trackability of the elicitation, not only of the basis for the

interpretation, but also of the elicitation process, itself. This is especially crucial for regulatory actions. Perception of credibility is also necessary, and this is achieved by using accepted methods having credible scientific support. Having a documented data base on which to make the interpretation is also important for trackability and credibility.

Dr. Stepp went on to aggregation of multiple interpretations, the purpose of which is to arrive at a single set of probabilities reflecting the multiple interpretations. He indicated that the problem should be decomposed into manageable elements, as a first step. This is followed by a definition of the necessary data sets use for the interpretation. In a specific EPRI case study, the next step was to convene workshops. The first workshop was a state of knowledge workshop. All the information was compiled and sent to participants prior to the workshop. Key researchers in the subject were also invited to present the current state-of-the-art knowledge to the participants.

Based on this uniform set of information, six teams of interpreters were then instructed to take this common base of information and to proceed with their interpretation. In the next workshop these teams were requested to present to their peers their interpretations and their basis for them. No consensus was required at that time. This process was repeated for each stage of the overall interpretation.

Dr. Pomeroy brought up the case of significant divergence in the estimate in one particular team; i.e., zero versus 100 percent. He asked Dr. Stepp whether that was desirable. Dr. Stepp indicated that due to the construction of the teams -- no one individual member had a sufficient range of disciplined knowledge to fully address all of the data sets -- some level of pooling of information was necessary -- thereby encouraging some level of consensus, but yet missing some of the inherent uncertainty.

Dr. Steindler asked whether different partitions of the problem could lead to different conclusions. Dr. Stepp indicated that, if the decomposition is properly done, there could be different sets of intermediate interpretations, but the final judgment should be the same. Mr. Galpin asked whether this approach didn't propagate uncertainties, to which Dr. Stepp agreed.

Returning to the aggregation step, Dr. Stepp noted that the teams were passive in that they were not asked to develop the state of knowledge independently. But, having this information, they were asked to perform interpretations for their particular stage of the process. There was a core team of nationally prominent scientists who had assembled the state of knowledge and were instrumental in devising the central core processes of the elicitation. Both normative (experts on expert judgment) and substantive experts composed the core team. This core team met with each of the teams

after the first round of interpretations for feedback purposes.

The next step was for the teams to revise their interpretations, which then became the final input into the overall process. Using logic trees the uncertainties were propagated to the end results. Each path through the logic tree represented a possible scenario.

Dr. Stepp indicated that this master logic tree approach is very applicable to the HLW repository. He also stated that the structure was flexible and could be as simple or as complex as necessary. The only missing node is the human intrusion situation; this omission was due to time constraints. He did note that the structure could drive site characterization and licensing and avoid transference. Further advantages include the ability to trace back the uncertainties to each of the interpretations (nodes in tree) and to compare them to the overall uncertainty. This becomes a strong basis for setting priorities both in site characterization and licensing issues.

Dr. Pomeroy indicated that the past EPRI case history was small when compared to the HLW repository case. Dr. Stepp agreed that there was an order of magnitude increase in complexity, but pointed out that decomposing the problem into logic nodes, the use of workshops for each of the nodes (which could be done in parallel), and the maintenance of a continuity of transfer of information among them, kept the problem manageable.

Dr. Steppe indicated that the approach used by EPRI had been scrutinized.

Firstly, a panel of nationally prominent experts reviewed the entire process. Among other things, two workshops were conducted by EPRI using that peer review panel. Based on their comments, the methodology was revised; e.g., the aggregation approach. Outside parties, including the NRC and the USGS, reviewed the approach, and the NRC ultimately accepted the approach for application to the definition of the seismic hazard for nuclear plants. A parallel effort by Lawrence Livermore National Laboratories (LLNL), under contract to NRC, resulted in an SER that defined total uncertainty to be that which incorporated both the EPRI and LLNL models' uncertainty.

Dr. Okrent remarked that logic trees are useful to organize the thinking process and to help towards the issue of completeness. When it came to a specific site action, however, information was lacking to answer the necessary questions. Dr. Stepp indicated that when that happens, it defines an area where the uncertainty towards the end result needs to be determined; i.e., if an area doesn't have enough information available <u>and</u> such knowledge of its uncertainty would significantly affect the end result, then this area should be investigated (testing).

Presentation by Dr. Allin Cornell, Stanford University.

Dr. Cornell pointed out that in structural safety there is the same dearth of information and knowledge of processes as in the case of Yucca Mountain, but things still are constructed. The main focus for structural uncertainty is on loads. So earthquakes are the main safety problem, but seismologist and geoscientists were not providing applicable information on probabilities of earthquakes. So stochastic models of earthquake occurrence and seismic hazards began to be developed. He indicated his interest in expert judgment estimation of parameters used in these models specifically with respect to uncertainties. This process evolved into experts providing alternative models, multiple values of parameters, and statements regarding their uncertainty. He thought that there might be some value in clarifying some questions on the model-to-model, and parameter-to-parameter variabilities, on stochastic models and on definitions of probabilities.

Dr. Cornell thought that the EPRI seismic hazard experience addressed all the necessary facets of the HLW repository problem. He listed the CCDF output, uncertainty of modelling the physical process (such as earthquakes), uncertainty of parameters (dimension of faults), and the aleatory (stochastic) component (randomness of size or time of future earthquakes). He also pointed out that the EPRI seismic risk study received very positive support by normative experts in its use of expert judgment.

Dr. Cornell briefly recounted the developments, since the 1960s of the use of expert opinion, specifically in the seismic risk arena. He outlined the process of integrating the effects of the site location relative to possible earthquake faults or sources. This was followed by the development of the recurrence parameters for each source, for which there is statistical evidence. From the curve produced, there is a truncation point -- the maximum magnitude -- which is reliant on expert judgment. Further integration of the ground motion attenuation leads eventually to a CCDF of probability (of maximum magnitude of seismic event) over acceleration.

In the 1970s, several applications of this approach led to different results at the same sites. This inconsistency led to the use of multiple models, different estimates of the upper bound magnitude, and so forth. Weights were then assigned to the various models, parameter value choices, and so forth, indicative of the experts' basis for using them. This assignment was done by the analyst, who was not necessarily a normative expert.

In the late 1970s, the NRC began to use expert judgment; e.g., the systematic evaluation program (SEP). Many nuclear power plants, as well as dams and other structures, were being evaluated in this expert judgment seismic risk manner. The LLNL eastern U.S. seismic hazard study and the WASH-1400 study on reactor safety were

produced. The need to display the uncertainties associated with these probabilities was demanded. The rise of alternative models reflected the applicant/licensee versus regulator debate over parameter inputs and how the seismic assessments should be done.

Resolution of the debates saw the use of frequency weight assignments; i.e., number of experts in literature favoring a specific choice divided by the total number of experts with any opinion. The scientific literature was used to estimate prevailing opinion.

Dr. Cornell next discussed the NRC/LLNL project. It was decided to preserve the anonymity of experts on the panel. Dr. Reiter indicated that there was a belief that a more candid expression of uncertainty would be gained in using this approach. The elicitation was performed by using questionnaires, which each individual answered separately (not in group session). In aggregating the results, Dr. Cornell indicated that each expert was given equal weight. There was a delayed feedback loop. This approach was criticized for lack of expert accountability -- there was no documentation of the basis for their assessments.

Dr. Reiter corrected Dr. Cornell's understanding that the LLNL study used more than limited percentiles in the process; a continuous band of probabilities was integrated into the risk curve. Another clarification of the deliberation process was that

each expert was the final judge of whether to override the data. Dr. Reiter indicated that the formal process could be overridden even in the EPRI seismic risk study; in some cases this did happen.

Dr. Cornell proceeded to describe the last decade as moving strongly to aleatory modeling. He mentioned examples such as the USGS Charleston and Cascadia Subduction Zone studies, which led to questioning of past uses of deterministic and probabilistic model applications.

The experience of the past studies was beneficial to the second NRC/LLNL study. More care was taken, and the quality improved. However, he related a case of bipolar opinion where one expert put all of his allocated weight to just one relatively conservative model; this has yet to be resolved. Dr. Cornell also cited the EPRI study described above by Dr. Stepp. He pointed out that the defensibility of the uncertainties assigned to probabilities was underscored by the requirement that the experts had to defend their rationale in front of eminently-qualified peers.

The aggregation problem seems to be one of how to assign weights to the various opinions. The options seem to be equal weights or else use a mechanical aggregation procedure. The normative experts provide the calibration for doing this; Dr. North's presentation addressed this situation. But in the case of the seismic risk, this calibration is not a viable option (how well can an expert

estimate a .0001 probability?). The mechanical aggregation was performed by assigning higher weights to those experts closer to the norm of consensus. This approaches professional consensus rather than the truth, but that may be as much as one can expect.

In agreement with the DOE approach, Dr. Cornell indicated that there was value in performing the probabilistic assessment early in order to determine what the data needs are.

Dr. Cornell had a few overall comments. He felt that normative experts were needed, because of the lack of extensive experience with probability and stochastic modeling on the part of the substantive experts. Likewise, time must be allocated for the normative experts to come up to speed in the substantive fields of knowledge. Another problem is separating the aleatory uncertainty from the epistemic uncertainty (lack of or limited information). Frequently such uncertainties are improperly blended, so that certain geologic properties are treated as random, when they aren't; e.g., spatial variability of conductivity.

His conclusions included:

> There is no substitute for experience -- for the substantive experts in doing expert judgments and for normative experts in becoming familiar with the subject area.

Uncertainty should be expressed as probabilities -- both the uncertainty of the individual and expert-to-expert variability.

The formal expert judgment process takes resources and time.

The experts must be accountable -- of the rationale/ defensibility of their choices must be documented for scrutiny by their peers.

Aggregation of opinion is tricky -- equal weighting may be the only feasible way to handle it.

Presentation by Dr. Leon Reiter, NWTRB

Dr. Reiter made a disclaimer regarding his presentation as representing either NRC or NWTRB; he provided his personal opinions. He described the reluctance by the NRC staff to use multiple experts in the late 1970s. The NRC staff cited the example of Nicolaus Copernicus, whose geocentric theory was rejected by the "experts" of the early 1550's.

Dr. Reiter cast the real question to be, not whether expert judgment should be used, but rather how is it used? He agreed with Dr. Bonano's earlier point that expert judgment has, is and will always be used, regardless of whether anyone acknowledges it as

subjective opinion. He indicated that his presentation would address four main issues:

- o implicit versus explicit use of expert judgment
- o informal versus formal use of expert judgment
- o deterministic versus probabilistic models
- o use of single versus multiple experts

Dr. Reiter went on to discuss the lack of understanding in the scientific community regarding earthquake predictions: why, where and when? He noted the harm that some of the predictive theories has caused in the area of public panic, lost resources, and so forth. The situation improves somewhat in the western U.S., where the "whys" and "wheres" are better understood, but the "when" still remains a mystery.

Dr. Reiter addressed the distinction between explicit and implicit expert judgment somewhat differently than did the previous speakers. He defined implicit expert judgment as the hidden use of expert judgment in making a decision; i.e., in effect, interpretations are made without openly indicating the rationale of the decision process. In more concrete terms, if an earthquake engineer, a seismologist and a geologist were each responsible for parameter estimations, which in turn were fed into a earthquake model, the more prominent scientist would alter his/her parameters to offset the other experts, in order to produce an end result to his/her liking.

Dr. Reiter went on to discuss informal expert judgment. He indicated that even in cases of very explicit and open use of expert judgment, the basis for that judgment is not laid out; sometimes the basis is semi-formal. His example was the Diablo Canyon Plant licensing case, in which Dr. N. Newmark, an eminent earthquake engineer, played a significant role. Dr. Newmark's original opinion was based on an analogue of large ships' sensitivity to small waves, which was challenged. As each of his explanations were challenged, he modified or changed the explanations to address the challenges. His intuition didn't change so much as his basis for it.

Dr. Reiter observed that even when formal judgments clearly describe and document the basis for the judgment, as well as the associated uncertainty, the decision makers are uncomfortable with that uncertainty. They tend to prefer the best estimate as a decision tool, and they consider the uncertainty as compromising the estimate's usefulness.

Dr. Reiter moved on to the distinction between deterministic and probabilistic approaches. He characterized the NRC reactor licensing approach of discrete single-valued events or models, as an example of a deterministic approach. He alluded to Dr. Cornell's presentation of going from the maximum earthquake to the ground motion value in a series of dependent steps. Dr. Reiter

took issue with the idea that no matter how you decompose the problem, you should expect the same answer. How you draw the curve to fit the analysis is subject to interpretation, which results in different conclusions.

On the issue of multiple versus single experts, Dr. Reiter indicated that use of one expert immediately reduces the variability of the judgment, which in turn reduces one aspect of the uncertainty. Occasionally a single, but very prominent, expert can suggest interpretations, which are at odds with the scientific community. But in most cases, an expert's views can by challenged by equally-credentialed experts. In these cases, multiple experts provide a solution to the challenge problem.

Dr. Reiter recalled the LLNL probabilistic seismic hazard analysis, where one of the experts was an outlier. For ground motion inputs, the outlier expert came up with consistently higher values than the other four experts. The staff tried a number of approaches in dealing with this incongruity. The expert was not disqualified, because the aggregated results of not using him were no better than those using him. In the end, the staff decided to use the total aggregation as one study and the partial aggregation as a second study. Only conclusions that could be supported by both "studies" would be used. This may or may not be a feasible alternative depending on the regulatory context.

In some cases the differences are in the fundamental methods used to assess performance. The cited example involved the LLNL and EPRI seismic risk studies, which used different methods, different models, and so forth. Again, the robust approach of discarding conclusions not supported by both studies was selected by the staff. He did note that the revised LLNL approach used a number of the EPRI aspects, which could lead to some convergence.

Dr. Reiter's conclusion was to recommend using explicit, formal multiple expert elicitation. Although this approach presents a greater degree of difficulty, it provides the best treatment of uncertainty. The uncertainties should be addressed early, not at the end of the project. He cautioned against the tendency of allowing the tail to wag the dog, in that outliers can sometimes drive an assessment instead of the majority of the experts.

In response to the initial questions posed at the beginning of the workshop, Dr. Reiter had the following remarks:

- You should select experts, who reflect the range of responsible views.
- Aggregation of multiple expert opinion may or may not be sensitive to the means by which it is done.
- In case of bipolar opinions, bounding calculations or even using central estimates may be useful; more detailed or refined estimates present very difficult problems.
- If you can obtain the needed information without using

experts and within the uncertainty constraints, don't use expert judgment.

- To ensure that the use of expert judgment is transparent at each stage of the process, one should rely on documentation.

Presentation by Dr. Evaristo (Tito) Bonano, Sandia National Laboratories (Sandia).

Dr. Bonano referred to the elicitation report (NUREG/CR-5411) that had been cited earlier. This report was not based on a specific elicitation of experts, but was a collaboration of ideas by performance assessors and normative experts. The focus of the report was on how to start the process and to generate further effort in expert judgment, rather than to provide solutions. He focused on two major areas:

- Areas in performance assessment, where the use of expert judgment should be formalized.
- Procedures for reviewing the use of expert judgment.

He agreed with Dr. Reiter that the main questions are when to formalize the use of expert judgment and to what degree should it be formalized. Problems of significant technical, environmental and socioeconomic impact, such as the HLW repository, should employ

a formal expert judgment approach. He went on to say that formal expert judgment should be a well-documented, systematic process that should be peer reviewed. This process should allow for traceability of judgments and should be able to withstand the scrutiny of a formal peer review. In the example of the EPRI seismic hazard performance assessment, he characterized the instances of the use of expert judgment as occurring in two major areas: site implementation and site investigations.

He agreed with Dr. Reiter in that there is a great deal of expert judgment factored into data collection. Much of the "hard science" parameters rely on expert judgment. Setting priorities for data collection is done by expert judgment. Dr. Bonano described the Sandia approach of iteration, called the preliminary performance assessment, which consists of executing the performance assessment to conduct a sensitivity analysis of what the major sources of uncertainty are. This is used as a basis for setting priorities for data collection. Once these priorities are determined, the level of resource allocation is established.

Once the preliminary performance assessment is completed, the performance assessment analysis is used to develop and to screen scenarios, as was done in the WIPP performance assessment and in the Swedish Project 90 effort. They are used to quantify uncertainty in (developing probability distribution functions) key parameters, as well as providing the ranges of values for these

parameters. They would also be used in selecting methods for propagation of uncertainty in performance assessment.

Dr. Bonano returned to the four areas in performance assessment where expert judgment should be formalized: Scenario development and screening; model development; parameter estimation; and information gathering. Experts have been involved in all of these areas. Dr. Bonano noted the following points:

- Estimation of the probability of occurrence of the scenario is critical, and will need to be formally and thoroughly documented.
- There will be debates over data selection, especially when site-specific data aren't available. The alternative options must be thoroughly documented and will probably be critically scrutinized.
- The conceptual models need to address all the possible alternatives, so that the necessary correct data and information are collected.
- Confidence-building -- sometimes called validation and verification -- of models and other computational tools require an experiment, which establishes the adequacy of the model for performance assessment.
- Some of the problem with model validation is determining whether the model is correctly representing reality or whether the model is merely adequate for its intended use

-- this is a subtle distinction.

- The criteria selected for model performance must reflect the purpose of the model use; the wrong criteria could significantly complicate any validation exercise.

Dr. Steindler asked whether it was self-evident that expert judgment is a legitimate tool for the exercise of validation. Dr. Bonano responded that there wasn't much of an alternative at this time. Specifically, Dr. Steindler asked how one validates a model of geology performance over the next 10,000 years? Dr. Bonano went back to his point of the purpose of the model. To validate it against reality will require a complexity that usually demands more data. Then one must determine the chances of collecting such data, because without this data the model would be useless. However, if the model is a decision aiding tool, which helps in selecting a course of action, the complexity may not be as confining. An example would be a rough computer code that evaluates strategies; although the absolute numbers may not be precise, the relative values may be sufficient to permit the selection of a preferable strategy.

Dr. North pointed out that in validation is frequently not possible. He gave an example of a military strategy, which could only be validated by having a military conflict. He returned to an earlier statement that "...a warm feeling in the gut..." of someone who understands the issues may be all that can be achieved.

In the area of parameter estimation, expert judgments are still used, but to a lesser degree than in other aspects of performance assessment. Expert judgment plays a significant role in identifying the important parameters. In the case where this influences testing priorities, a degree of formality is needed; e.g., quantification of uncertainty in parameter values.

At this point, Dr. Bonano discussed information gathering. Expert judgment plays a major role in the identification of newly required information. As part of this, experts are looking at the feasibility of obtaining the additional, new information; also they address what impact not obtaining the data will have on the performance assessment.

Statement by Ms. Jean Younker, representing the DOE YMPO

Ms. Younker thanked the Committee on behalf of Mr. Blanchard and the DOE for the opportunity to participate. She characterized the overall meeting as being more optimistic than she expected, in the uses of expert judgment.

The Chairman asked if there were any other observations from participants. Mr. Galpin (EPA) indicated that the WIPP effort was a good and relevant source of information on past or present uses

of expert judgment in waste management performance assessment. He noted that the WIPP was further along than any other effort in that regard.

The Chairman indicated that there were a few requests to make statements from the public; he first introduced Dr. Rex Brown.

Summary of Statement by Dr. Rex Brown, Decision Science Consortium

Dr. Brown made the following observations:

> In dealing with expert judgment in areas of greater uncertainty, such as human intrusion, one should determine whether the available information can be refined to be less uncertain. In cases of unintentional intrusion, you can improve confidence by using resource mining information and data. However, for malevolent intrusion, it is pointless to bother with expert panels. In trying to reduce uncertainty by using expert judgment, it might be worthwhile to examine whether the effort would make any difference in the ultimate licensing decision.
>
> Aside from whether one can quantify malevolent human intrusion, there might be design features, which would be handled differently with the possibility of such intrusion.

Restricting expert judgment to that which can be accomplished formally may well limit information available from informal expert judgment; this could be a serious loss.

Some attention should be paid to the knowledge that can be extracted from, as well as the knowledge that is inserted into, a formal elicitation.

One needs to distinguish the bias an expert may have towards some interpretation, which could be based on useful information, versus the bias in extracting information from the panel. The first "bias" may be useful information, the second is a distortion of the information in the process of eliciting it.

Anonymous versus shared elicitation strategies each have pluses and minuses. Anonymity can avoid political or dominating personality problems, which can occur when the panel interacts directly. However, in an open elicitation process each expert has access to the information and knowledge of the others. This could help in achieving a consensus.

Mr. Jerry King, SAIC

Mr. King had a few observations for the participants; he cautioned that he didn't necessarily agree with them. These had to do with the fact that in practical terms there are no experts on the potential for human intrusion 1,000 years from now. He voiced the position that in cases of competing models, design bases should be set via deterministic analysis using the worst-case hazard model. He also noted that prevalent scientific opinion, which was fully documented, disaggregated and displayed with its bases, formed the best possible information for decision making, since truth was not achievable.

Round-Table Discussion

Dr. Pomeroy suggested starting with those areas where there might potentially be the greatest disagreement among the experts. A second question to be addressed is whether a credible performance assessment could be performed that would satisfy the EPA requirements, given our state of knowledge and tools.

S. Coplan started the discussion with the belief that credible performance assessments could be done, but he had some reservations. He cited the human activity question in terms of

being factored into the CCDF. He used human activity to broadly include the unpredictability of future human actions beyond just intrusion. He didn't think that expert judgment could help with the way the EPA Standards are presently fashioned. Secondly, he had reservations in using unvalidated models, since these are a form of expert judgment without empirical support. He didn't think that confidence could really be built from expert judgment panels. Mr. Coplan indicated that except for the human activity, he was optimistic about estimating the probabilities for the scenarios. He believes that one can make estimations regarding natural processes, because the data, assumptions and reasoning for using the models can be rationally justified.

Dr. N. Eisenberg (NRC/NMSS) noted that whether one can make a finding to license a repository depends on its character. If the repository is robust in terms of its design -- if changes in the natural system or human behavior have little effect on the repository's performance -- then one should be able to arrive at a licensing decision with little reliance on expert judgment. Secondly, whether a repository can be licensed is a function of being able to predict its long-term performance, rather than the character of the standards.

Dr. Bonano clarified his comments about using expert judgment in model validation for purposes of choosing the criteria against which the model would be tested. As an example, a model for heat

transfer was used on data collected from a geothermal reservoir. For predicting average heat fluxes, the model agreed with the data; in the case of predicting actual temperature distribution, it failed. The validation criteria must be a function of the scope of the intended use or purpose; the experts must formalize their judgment criteria with that in mind.

S. Frishman highlighted the design effectiveness analysis as an area where the greatest reliance is made on expert judgment for large consequence decisions. He pointed out that whatever decisions are made on the design's effectiveness, the host rock is altered to some extent (drilling and construction) both in space and time. Whatever is characterized is not going to be the final repository; the final repository is no longer the site that was characterized. This is especially troubling with respect to the credit taken for engineering. He thought the less reliance on engineering and design, the better.

Mr. Frishman cautioned that there was a significant difference between DOE's understanding of expert judgment and that of other participants, specifically with respect to peer review. He thought that expert judgment and peer review are sequentially related, not synonymous. He also pointed out that some of the probability estimation had lost sight of the chronological time frame in which we find ourselves. For example, for something to have a probability of occurrence of one in a billion is impossible,

because that expectation would involve the age of the earth; i.e., if it hasn't happened, it probably won't.

Dr. Cornell suggests that in dealing with competing models, formal quantitative weighting (through formal expert judgment) should be considered for selecting alternative models. Dr. Cornell indicated that this would be better than just using the worst consequence (most conservative) model. In the area of aggregation of multiple expert judgment, the weighting factor corresponding to each expert's opinion (equally weighted or not) corresponds to a probability, which in turn can be interpreted as a credibility factor for each expert. In order to make decisions, the use of weighting strategies -- besides simply assigning equal weights -- needs to be seriously considered.

Dr. Cornell indicated that the real issue regarding the use of CCDFs is the credibility of the uncertainty assessments placed by experts on alternative models and parameter values (specifically to the extent that the experts would be willing to stand up and defend their judgments in front of their peers). The trouble comes up when you disaggregate the problem, if possible, down to a finer structure, e.g., branches of a logic tree, and you face a single dominant component that obscures a safety issue in one or more components. Even so, this disaggregation of the problem(s) into elemental pieces, and being able to reassemble them, is a crucial capability in using expert judgment in performance assessments.

In the end, should one end up with a very low probability event with everything hinging on it, perhaps one should reconsider going forward with the course of action.

L. Abramson noted that the problem of weighting the experts may not be such a difficult issue. He indicated that in his experience different weighting schemes didn't affect the end result to any great degree. Frequently the strategy of weighting was dictated by convenience or political expediency, and relied on equal weighting to achieve those ends. He recalled that different weighting schemes were tried in the EPRI study, and the conclusions were not significantly different. He thought that the weighting strategies were not a profitable expenditure of effort.

Dr. Cornell rebutted by raising the disqualification situation; he thought that attributing a low weight to an option, which deserved a low weight, was an effective way of dealing with an outlier situation without completely eliminating the option.

Dr. Bonano noted that early use of expert judgment in evaluating the whole process is essential to avoid having to respond during licensing hearing, to questions of validity of the process. Going through the steps of the performance assessment can lead to identification of the need for further data collection and other analyses and should be done, even if only in a generic manner. During the licensing process, the focus should be on whether the

judgments are correct, rather than on the validity of the elicitation process, itself. Mr. Galpin asked whether this generic evaluation should still be pursued, if it required a rulemaking ahead of time. Dr. Bonano stated that it should.

Dr. Steindler raised the question of situations, topics or areas where expert judgment is not a legitimate tool. He posed an example of a truly random process. He observed that there are many examples of people using the wrong empirical tools (ammeters versus voltmeters) in a situation; how does one deal with the question of the validity of the reliance on expert judgment in an application?

Mr. Coplan recalled that early on in the program, he had thought that an expert judgment procedure should have been established by rule; this expert judgment regulation would bind the NRC to an elicitation process. However, OGC informed him that this would effectively replace the Commission's legally required deliberation and decision by the judgment of an expert panel.

Mr. Coplan noted that the day's discussion had been directed to the expert elicitation in isolation from the context of the HLW repository program. There hadn't been much discussion of the DOE license application with its myriad of technical issues. He indicated that the situation would likely consist of a scenario with a probability of, say 0.004, as well as a description of the

elicitation rationale. He recalled Dr. Okrent's earlier observation that such an elicitation and use of expert judgment will be met with questions of validation of the application and of the elicitation itself. He indicated that the NRC staff would need to see some rationale for giving credence to the use of expert judgment.

Mr. Galpin understood Mr. Coplan's concern regarding the credibility of and confidence for the use of expert judgment. However, he indicated that there were going to be areas, where expert judgment would be needed to some extent, for example in the human activities/intrusion area. He pointed out that the repository licensability could be addressed otherwise. If the experts could not reach a consensus, this could be an indication that the repository should not be sited at that location.

Dr. Stepp remarked that expert judgment fits into the overall reasonable assurance process, which is the basis for making licensing board decisions. The NRC and the DOE should give this process considerable thought for that reason. He observed that this working group meeting had concentrated more on quantifying the uncertainty on interpretations, rather than on elicitation of expert opinion. The elicitation process corresponds to a lesser level of assurance than does the estimation of uncertainty of a judgment. He appealed to the participants to be sensitive to this distinction. He concluded with the observation that the formal

elicitation process may be the best option for resolving those kinds of problems that our technologies and empirical science cannot adequately address.

Dr. Reiter suggested that human intrusion should be handled separately with respect to performance assessment of a repository; it should not be handled in the same context as other geological phenomena. He cautioned against being too specific; this could restrict flexibility in the application of expert judgment. The guiding light should be to protect public health and safety. He recommended an honest and open use of outside experts, in order to avoid self-fulfilling prophecies. On the other hand, he cautioned that at some levels the formalism could obscure the goal, while implying an unwarranted level of accuracy.

Mr. Abramson indicated that exercising caution to avoid using expert judgment, unless absolutely necessary, was unwarranted; the problem was more of the case of people not using expert judgment because it was difficult and arduous in its demands on time, resources and patience.

The Chairman thanked the participants and encouraged them to provide the Committee with suggestions for future consideration about the expert judgment process. The meeting was adjourned.

PAGES 61-66 DELETED DUE TO PREDECISIONAL
INFORMATION - FOIA EX(5)